Reading in a Foreign Language October 2016, Volume 28, No. 2
ISSN 1539-0578 pp. 161–182

The complexity of second language reading: Investigating the L1-L2 relationship

Lisbeth M. Brevik
University of Oslo
Norway

Rolf Vegar Olsen
University of Oslo
Norway

Glenn Ole Hellekjær
University of Oslo
Norway

Abstract

This article contributes to the field of reading assessment in English as a second language (L2). Few reading studies have been carried out at the upper secondary school level, and the present study provides insight into upper secondary school students' L2 reading proficiency. It examines whether such proficiency can be explained by reading proficiency in Norwegian as their first language (L1). The analysis uses data from two national reading tests, comprising a large sample of 16-year-old students (*N*=10,331), and it is the first time reading across these languages has been investigated at this level. The results show a significant and meaningful relationship between students' reading proficiency in the two languages. The results also reveal marked reading differences in reading proficiency in the two languages among poor readers.

Keywords: assessment, crosslinguistic reading, reading in a second language, reading comprehension, reading tests

In "Reading and Linguistic Learning: Crosslinguistic Constraints on Second Language Reading Development," Koda (2007) explains that, unlike in the first language (L1), second language (L2) reading involves two languages (p. 16). Indeed, research indicates a structural relation between L1 and L2 reading comprehension (Bernhardt, 2011; Brantmeier, Sullivan, & Strube, 2014; Grabe, 2009; Jeon & Yamashita, 2014). Koda (2007) argues that a primary focus within L2 reading research should therefore be to get a clearer understanding of *how* reading in the L1 and the L2 interact in L2 reading. As Alderson, Haapakangas, Huhta, Nieminen, and Ullakonoja (2015) point out, assessing the ability to read in the L1 is a complex process, and assessing the ability to read in an L2 "is even more complicated because it involves not only the ability to read but also the knowledge of and the ability to use the second or foreign language" (p. 68).

In the present study, we have analysed reading comprehension among adolescents in the first year of Norwegian upper secondary school (16-years-old students) to learn whether there is a relationship between their reading comprehension in English as an L2 and in Norwegian as the L1, and to examine whether background variables such as gender and study programme can explain variations in their reading comprehension. This first large-scale assessment of reading across these languages at this level uses quantitative data from 10,331 upper secondary students' scores from two nationally distributed reading tests: a paper-based test in L1 and a digital test in L2. We have analysed the entire sample, paying particular attention to the readers in the lowest quintile.

Reading Comprehension in the L1 and the L2

A commonly used definition of reading comprehension is "the process of simultaneously extracting and constructing meaning through interaction and involvement with written language" (RAND Reading Study Group [RAND], 2002, p. 11). This definition is in line with the constructs of the two tests in the present study (Norwegian Directorate for Education and Training [UDIR], 2010a, 2010b). It also aligns with the more recent PISA definition, which adds engagement as an integral part of reading by establishing that "reading literacy is understanding, using, reflecting on and engaging with written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society" (OECD, 2010, p. 23). The latter definition was influenced by contemporary and current theories of reading, which emphasize reading's interactive nature, models of comprehension, and theories of performance in solving reading tasks (OECD, 2013, p. 4). Thus, "reading literacy" seems to denote "reading comprehension," and it is the latter term we will use in this article.

Reading comprehension is a cognitive as well as a social process that involves extracting and constructing meaning (Bernhardt, 2011; Duke, Pearson, Strachan, & Billman, 2011; Koda, 2007, 2010). As Alderson et al. (2015) point out, "it is relatively uncontroversial to say that reading consists of at least two sorts of processes, commonly called low-level and higher-level processes" (p. 75). Current models of reading describe it as an interactive process between bottom-up and top-down processing (Alderson, 2000; Bråten, 2007; Grabe, 2009; Koda, 2005). The low-level, bottom-up process involves recognizing the written words in the text along with relevant grammatical information, which in turn hinges upon automatic word recognition (decoding words and relating print to sound) (Droop & Verhoeven, 2003; Jeon & Yamashita, 2014). This process provides the basis for top-down, higher-level processing, i.e., the creation of meaning in an interactive process between the information in the text being read, the reader's knowledge of language and content, and the reader's processing skills and strategies (Alderson, 2000; Bernhardt, 2011; Grabe, 2009).

With good readers, the word recognition process proceeds effortlessly and rapidly. This finding means vocabulary knowledge is essential for good reading comprehension (Alderson, 2000; Alderson et al., 2015; Grabe, 2009; Jeon & Yamashita, 2014; Koda, 2005; National Reading Panel, 2000). Furthermore, when good readers encounter problems, such as unfamiliar words or concepts, "they deal with inconsistencies or gaps as needed" when trying to determine the meanings in the text (Duke et al., 2011, p. 56). It also involves other cognitive processes,

metacognitive monitoring in particular, and the use of content knowledge to repair comprehension (Alderson, 2000; Brevik, 2014; Duke et al., 2011; RAND, 2002). In fact, monitoring is one of the main factors distinguishing good readers from poor (Alderson, 2000; Bråten, 2007). Alderson (2000) argues that good readers "tend to use meaning-based cues to evaluate whether they have understood what they read whereas poor readers tend to use or over-rely on word-level cues, and to focus on intrasentential rather than intersentential consistency" (p. 41).

Reading comprehension also involves the use of skills and strategies. While the use of skills is automatic, strategy use is under the conscious control of the reader (Afflerbach, Pearson, & Paris, 2008; Grabe, 2009; McNamara, 2011). Examples of reading strategies would be re-reading to sort out a discrepancy in meaning (Block & Duffy, 2008; Brevik, 2014), using context to sort out the meaning of unknown words (Brevik, 2015; Duke et al., 2011; Grabe, 2009), or, alternatively, ignoring these if possible. Another example would be adjusting how one reads to suit the reading purpose, such as skimming to understand main points in a text or scanning to find particular details (Brevik, 2014, 2015; Grabe, 2009). Reading for a specific purpose might also mean engaging in careful reading at the local level in order to understand the syntactic structure of a sentence or clause, or careful reading at the global level to understand the main ideas of a text (Brevik, 2014; Duke et al., 2011). In fact, the ability to adjust one's reading to a specific purpose is a key reading requirement in the Norwegian English syllabus (Norwegian Ministry of Education and Research [KD], 2006, 2013).

The Relationship between L1 and L2 Reading

As mentioned, an important difference between L1 and L2 reading is that readers approach L2 reading with a dual-language system (Koda, 2005, 2007). This distinction echoes Cummins's (2000) argument that "academic proficiency transfers across languages such that students who have developed literacy in their L1 will tend to make stronger progress in acquiring literacy in their second language" (p. 173). Cummins proposed his Linguistic Threshold Hypothesis (1979) in the context of attempts to improve the educational chances of bilingual children, and argued that this transfer depends upon language proficiency. If a reader's L2 proficiency falls below a certain level, the transfer of these skills and strategies from the L1 to the L2 is prevented, even if the student is a good reader in the L1. In contrast, in the Threshold Hypothesis (TH) of Alderson (1984), L1 refers to a native language that is the official school language, while the L2 refers to any non-native language. Thus, the TH relevant in this Norwegian study seems more in line with the TH of Alderson than that of Cummins. As Alderson (2000) notes, "this linguistic threshold is not absolute but must vary from task to task: the more demanding the task, the higher the linguistic threshold" (p. 39). Another uncertainty in this hypothesis is that it assumes adequate levels of L1 proficiency and knowledge, which is by no means a certainty.

An advantage of a dual-language system can be found in the compensatory hypothesis, which claims that deficiencies at one level can be compensated drawing on other levels (Stanovich, 1980). Based on Stanovich's (1980) model, Bernhardt's (2011) compensatory model of L2 reading claims that reading variables interact and that a weakness in one area might be compensated for by knowledge from another. She also attempts to quantify the importance of

"L1 literacy" (e.g., vocabulary, text structure), "L2 language knowledge" (e.g., cognates, L1-L2 linguistic distance), and an "unexplained variance" (e.g., comprehension strategies, engagement, domain knowledge). Specifically, she argues that L1 literacy accounts for up to 20% of a reader's L2 reading comprehension, that L2 language knowledge accounts for up to 30%, and that unexplained variance accounts for the last 50% of the variance.

Several studies support Bernhardt's (2011) model, but with great variation in the levels of explained variance between the L1 and L2 (Alderson, 1984; Bernhardt & Kamil, 1995; Brantmeier, Bishop, Yu, & Anderson, 2012; Jeon & Yamashita, 2014; Lee & Schallert, 1997; Yamashita, 2002). For example, for reading comprehension in Spanish L2, Bernhardt and Kamil (1995) found that English L1 literacy explained 10%–16% and Spanish L2 language knowledge 30%–38% of the variance (p. 25). In contrast, Lee and Schallert (1997) found Korean L1 literacy to explain only 3% and L2 language knowledge 57% of English L2 reading comprehension, while Yamashita (2002) found an explained variance of 40% across Japanese L1 and English L2, with L2 language knowledge explaining more than L1 literacy (Grabe, 2009, p. 147–148). Similarly, Brantmeier et al. (2012) found that L2 language knowledge explained more of the participants' English L2 reading comprehension than did their Chinese L1 literacy. These studies indicate that the explained variance between reading comprehension in L1 and L2 may vary with the linguistic distance between the two languages (Jeon & Yamashita, 2014), which echoes Koda's (2007) comments on the explanatory power of linguistic distance. As Norwegian and English are both Germanic languages, they are closer linguistically than the languages in the studies referenced above (Grabe, 2009; Koda, 2005), which means that Norwegian as an L1 may explain more of the variance in L2 English reading comprehension. In line with Bernhardt (2011) and Koda (2007), we hold that L2 reading research needs to develop a clearer understanding of how reading in the L1 and the L2 interact in L2 reading comprehension.

The Norwegian Context

For Norwegian students, elementary school (Years 1–4), middle school (Years 5–7), and lower secondary school (Years 8–10) are mandatory. They can then move on to three years of upper secondary school (Years 11–13), which are voluntary, and where the students choose between general and vocational educational programmes. English is a compulsory common core subject taught from Year 1 (6 years) to at least Year 11 (16 years) (KD, 2006, 2013). While it is taught in Year 11 in general programmes, the same course is taught in the vocational programmes across Years 11 and 12. Further, English is offered as an elective subject in Years 12 and 13 of the general programmes. The level of English proficiency has long been fairly high (Bonnet, 2004; Ibsen, 2002). Recent research shows that L1 and L2 reading skills have improved markedly among Norwegian secondary school students (Hellekjær & Hopfenbeck, 2012; Ibsen, 2002; OECD, 2013; Olsen, Hopfenbeck, Lillejord, & Roe, 2012; Roe, 2013). In a 2000 European reading assessment in English as L2 in eight countries, Norway came in second (Bonnet, 2004; Ibsen, 2002). Regarding gender differences in English as L2, the European test showed "a large significant difference for Finland and Norway in favour of girls" (Ibsen, 2002, pp. 144–145). While this gender gap is consistent with findings in L1 reading comprehension in the PISA test in Norway (Frønes, Narvhus, & Aasebø, 2013), recent Norwegian national tests in English L2 for students in Years 5 and 8 show little difference between boys and girls (UDIR, 2013).

In 2012, at the time when the student data in the present study were collected, 58% of the students in upper secondary school attended general programmes, with the remaining 42% in vocational programmes (UDIR, 2013). The school results between students in these programmes reveal major differences. On average, students in the general programmes perform better in common core subjects, such as Norwegian and English, than the students in vocational programmes (UDIR, 2013). However, while these results are based on overall achievement and examination grades in the subjects, there are no available data on these students' reading proficiency in L1 or L2.

Students in Norway participate in national L1 and L2 reading tests annually, at the beginning of Years 5, 8, 9, and 11. UDIR administers these tests, and the upper secondary tests (Year 11) are mapping tests "used to enable early intervention for students with learning difficulties by identifying the 20% with lowest skills (intervention benchmark)" (Tveit, 2014, p. 224). However, while a few studies have examined L2 reading in Norwegian upper secondary school, no research has systematically compared reading in Norwegian L1 and English L2, neither for students in general, or for poor readers. Furthermore, no previous studies have made use of the upper secondary level reading tests, as we have done in this study.

The overall question for our study is therefore: *How do Norwegian upper secondary students read across Norwegian as the L1 and English as the L2?* In order to investigate this question, we explore three specific research questions:

1. To what extent is a poor reader in English L2 also a poor reader in Norwegian L1?
2. How do gender and study programme relate to the students' L1 and L2 reading scores?
3. To what extent is there a statistical relationship between students' L2 reading scores and the variables L1 reading, gender, and study programme?

In the present study, *poor readers* are defined as those who perform among the 20% lowest performers in the L1 and the L2, respectively (UDIR, 2010a, 2010b). The following section presents the data and methodology in further detail.

Data and Methods

This study is based on secondary data from the two previously mentioned national reading tests conducted at the beginning of upper secondary school (Year 11); a paper-based test in Norwegian L1 and a digital one in English L2. While the L1 test was mandatory for all students at this level, the L2 test was voluntary for each school, which means that if the school enrolled, all students at the school participated. Since its inception in 2010, the student population in Year 11 has increased: 76,028 in 2010, 76,659 in 2011, and 78,012 in 2012 (UDIR, 2011, 2012d, 2013). The number of participants in the optional L2 test has increased from 22% in 2010 (N=16,381) to 42% in 2011 (N=31,942) and 45% in 2012 (N=34,882) (UDIR, 2012b).

The tests are based on the competence aims in the criterion-based national curriculum (KD, 2006, 2013) that are to be achieved at the end of lower secondary school (Year 10). These tests are designed to provide teachers with indicators on individual students' reading performance early in the school year by identifying the 20% weakest performers and the areas in which the students have particular strengths and weaknesses. This information provides a guide for the students' development in L1 and L2 reading comprehension.

Overlapping test constructs

As mentioned, both reading tests are based on overlapping construct descriptions from UDIR, which in turn are based on curricular guidelines. These guidelines state that the students in upper secondary school are to have developed reading skills that enable them to read increasingly more complex texts in all subjects, in the L1 as well as the L2. In practice, the students should be able to find, interpret, and make inferences based on information in various text types and formats (KD, 2006, 2013; UDIR, 2012c). Each test included a set of items that together measured the students' language and text comprehension in L1 and L2 respectively. The test frameworks described the constructs they are to assess, with language constructs corresponding to the decoding aspects of reading, while the reading comprehension constructs draw upon the PISA and RAND frameworks for reading (OECD, 2010; RAND, 2002).

As Brantmeier (2004) points out, "though interactive models of L2 reading emphasize different components involved in the process, all models include and underscore the importance of comprehension" (p. 52). Moreover, Alderson et al.'s (2015) elaboration on the aspects of reading comprehension echoes the test construct for these two reading assessments by noting that "understanding text involves drawing inferences, making subjective interpretations, as well as recognizing explicit statements" (Alderson et al., 2015, p. 69). Table 1 presents an overview of the reading constructs and their operationalization for the two tests (UDIR, 2010a, 2010b).

Table 1. *Test construct for the L1 and L2 reading assessments*

	Description	*Norwegian L1 test (paper)*	*English L2 test (digital)*
Language (vocabulary & grammar)	Tasks require the reader to recognize words.	Separate words in word chains (max 75 points)	Add missing words in sentences (max 5 points)
Reading comprehension (RC)	Tasks require the reader to (a) find explicitly stated information in the text, (b) to understand main points in the text, and (c) reflect and make inferences based on information in the text.	Multiple choice (max 34 points)	Multiple choice, click word, move paragraph (max 23 points)
Texts in the RC part	Fact and fiction	Two long texts (1300-1700 words)	11 shorter texts (40-300 words)
Intervention benchmark	The lowest 20% reading achievement, set the first time the tests were conducted (2010)	Language: 41 points RC: 20 points	L2 language & RC: 11 points
Total points	For each reading test	Max: 109 points	Max: 28 points

As displayed, although the two tests are based on overlapping constructs, they have three main differences apart from languages; (a) the test formats (paper vs. digital), (b) the text length (long vs. short), and (c) the task format. Regarding task format, both tests have multiple choice items, with additional ones in the L2 compatible with the digital test format (click word, move paragraph). These assessment tasks are largely in line with formats used in recent L2 reading assessments (e.g., Brantmeier, 2004).

The intervention benchmarks identifying the lowest performers were set in 2010, based on representative samples in each test (Heber, Mossige, & Kittel, 2010; UDIR, 2012b, 2014). However, the benchmarks should not be considered absolute; for example, a student performing immediately above the benchmark might need support, while a student performing below the benchmark might not (Heber et al., 2010; UDIR, 2012b, 2014). Furthermore, the tests by design have ceiling effects in order to maximize the information about the poor readers. As a result, the tests produce fewer details about the average and good readers. Nevertheless, the actual tests are not notably skewed (see Table 6), allowing for reasonable separation also for students with higher scores. Furthermore, the large and representative samples involved allow for fairly robust and reliable inferential statistics, including population means.

Participants in the present study

UDIR granted us permission to collect the L1 data from upper secondary schools on a national basis. This procedure was complicated but necessary, since no central register for the paper-based L1 test exists. We contacted all public upper secondary schools. To avoid selection bias, privately owned schools were excluded as neither test is mandatory for them. They make up only a small percentage of upper secondary students (7% from 2007 to 2012). Since the L2 assessment was administered electronically, we had access to all schools and students participating in this voluntary test. Regarding the L2 data, there is no reason to expect a selection bias in participating schools; the results have been consistent since 2010, although the participating schools have not been identical every year (UDIR, 2012b).

Table 2 provides details about the sample. After merging the two datasets and including only schools and students participating in both assessments, the final sample for the present study (L1-L2) includes 10,331 students from 87 public schools.

Table 2. *The L1-L2 sample for the present study: 87 schools with a total of 10,331 students*

	L1 (Norwegian) population				*L2 (English) population*			*L1-L2 sample*
	Invited (public)	No reply	Refused	Accepted	Provided data	Excluded (private)	Included (public)	Participated in both tests
Schools	346	113	42	194	167	21	223	87
					25,962 (36%)	1,153 (1.6%)	33,729 (47%)	10,331 (14%)

Note. Student percentage is based on the 2012 reference population, which comprises 72,551 students (total population of 78,012 minus 5,461 at private schools) (UDIR, 2013)

To control for how representative the 87 participating schools were, we compared the L1 and L2 participants across geographical regions and the L1-L2 sample with the total L2 population. First, by dividing Norway into five regions, we found a strong similarity in the distribution of the L1 test population (*N*=25,962) and the L2 population (*N*=34,882) across the regions. However, when comparing the L1-L2 sample (*N*=10,331) to the L1 and L2 populations, we discovered differences in two regions. To the best of our knowledge, these differences did not relate to any systematic bias; rather, they indicated that in one region most schools provided L1 data *and* participated in the voluntary L2 assessment, while the opposite was the case in the other region, namely that fewer schools provided data for both tests.

Second, we compared L2 test performance for students in the L1-L2 sample and the L2 population regarding gender, mean scores, standard deviation (SD), and z-scores. The consistency in patterns indicated in Table 3 suggests that the L1-L2 sample is representative of the L2 population tested.

Table 3. *Descriptive information for L2 means and standard deviations for raw scores, with the gender distribution and effect size for the L1-L2 sample and the L2 population*

	L1-L2 sample (N=10,331)		*L2 population (N=34,882)*	
	Boys	Girls	Boys	Girls
Percentage	52	48	51	49
L2 mean (max: 28)	18.3	19.4	18.5	19.7
SD	7.6	7.2	7.7	7.2
Z-scores L2	-0.07	0.08	-0.08	0.08

Note. L1=First language (Norwegian). L2=Second language (English). L1-L2=Across first and second languages. SD=Standard deviation

Based on the comparison in Table 3 and the geographical distribution, we therefore contend that the L1-L2 sample provides a reasonably representative sample.

Data collection

We received the L1 data as Excel files from the individual schools, including separate sum scores for language tasks and text reading tasks (see Table 1), along with background information (county, school, student ID, study programme). UDIR delivered the L2 data as a single digital file, including scores for each item and additional background information (gender). We transferred the L1 and L2 data to the statistical software SPSS (Statistical Package for the Social Sciences), and merged the two SPSS files using student ID as the key variable across the datasets.

Data analysis

Table 4 provides an overview of the data analysis.

Table 4. *Steps of analysis*

	Aim	*Analysis*	*Research question*
Step 1	To identify crosslinguistic characteristics for the poor L2 readers	Frequency analysis and cross tabulation	RQ1: To what extent is a poor reader in English L2 also a poor reader in Norwegian L1?
Step 2	To relate gender and study programme to L1 and L2 test performance	Effect size analysis (Cohen's *d*)	RQ2: How do gender and study programme relate to the students' reading scores?
Step 3	To build a model for the relationship between L1 and L2.	Correlation, reliability, and regression analyses	RQ3: To what extent is there a statistical relationship between students'L2 reading scores and the variables L1 reading, gender, and study programme?

Note. L1-L2=Across first (Norwegian) and second (English) languages

Step 1: Identifying crosslinguistic characteristics for the poor L2 readers. In simple terms, compensatory reading theory (Bernhardt, 2011) claims that reading comprehension in L2 draws on reading comprehension in L1. Although it is reasonable to expect that a poor reader in the L1 is also a poor reader in the L2, the relationship is not necessarily completely linear, as some might be better in one language than in the other. We classified the students into quintiles according to their scores on the L1 and L2 tests, which enabled us to identify the poor readers who read below the intervention benchmark. Since the original test measures focused on identifying the lowest quintile of readers in both languages, it is reasonable to assume that the classification precision is highest in the lower end of both scales. By cross tabulating the L1 and L2 quintiles, we can identify how the poor L2 readers perform across the two reading tests.

Step 2: Relating gender and study programme to L1 and L2 test performance. We created z-scores for L1 and L2 reading proficiency and performed effect size analysis (Cohen's *d*) to see how gender and study programmes were related to the students' reading proficiency in each language. This step was motivated by findings in the first L1 test in 2010, where there were significant differences between study programmes both in the L1 language and L1 text reading measures in favour of students in general programmes (Heber et al., 2010). In the L1 language measure, the students are asked to separate words in several word chains consisting of five words each, where the space between the words have been deleted. Thus, L1 language measures recall of words in a separate section of the test, which is quite different from the L1 reading comprehension items that measure the students' understanding of two long texts. Related to Bernhardt's (2011) compensatory model, gender and study programme might be part of the unexplained variance which, according to Bernhardt, may account for up to 50% of L2 reading comprehension.

Step 3: Building the regression model. In order to explain the relationship between students' reading comprehension in the L1 and the L2, we developed a regression model. Since the sample consisted of students clustered in schools, a multilevel regression model using SPSS mixed (with

restricted maximum likelihood estimation) was performed (Heck, Thomas, & Tabata, 2010). No school level variables were used and only the within school component of the analysis is reported. Having data from a large sample of students allowed us to conduct this analysis not only for the poor readers, but for all students. We were duly aware of the ceiling effect, as will be further discussed.

Validity is calculated as internal and external correlations within and across the two tests. First, we found high internal correlations between the overall L1 test scores (L1 reading proficiency) and the lower order constructs (L1 language r=.90, L1 reading comprehension r=.71), and a moderate correlation between L1 language and L1 reading comprehension (r=.44). For the L2, we found high internal correlations between the overall L2 test scores (L2 reading proficiency) and the lower order constructs (L2 language r=.89, L2 reading comprehension r=.97), as well as between L2 language and L2 reading comprehension (r=.70). Moreover, we discovered a moderate external correlation between the overall L1 and L2 reading proficiency scores (r=.55). Reliability estimates (Cronbach's α) for the tests were high both for L1 reading comprehension (α=.88) and L2 reading proficiency (α=.93), the latter being a consistent finding since 2010 (Heber et al., 2010; UDIR, 2012b). Based on the validity and reliability analyses, we used the following variables in a multiple regression model:

1. *L2 reading proficiency*: The overall test score for L2 language and L2 reading comprehension.
2. *L1 reading proficiency*: The overall test score for L1 language and L1 reading comprehension.
3. *L1 reading comprehension*: The text component (see Table 1).
4. *L1 language*: The language component (see Table 1).
5. *Gender*: Dummy variable coded 0 for boys and 1 for girls.
6. *Study programme*: Dummy variable to separate between vocational programmes (0) and general programmes (1).

In the regression models, we used L2 reading proficiency as the dependent variable, while the independent variables or predictors were the overall L1 reading proficiency and the components L1 reading comprehension and L1 language. We introduced gender and study programme to control for potential confounding of the findings. In addition, we tested for non-linearity by including the square of L1 reading comprehension, and given the results from steps 1 and 2 presented above, terms representing interaction effects between L1 gender and study program, respectively were included.

Findings

To what extent is a poor reader in English L2 also a poor reader in Norwegian L1?

The student scores were divided into quintiles, or groups of 20%, 40%, 60%, 80%, or 100% based on their score out of the total score in each language. Table 5 shows that 2,123 students performed below the intervention benchmark (20th quintile) in the L1, and 2,208 students in the L2. These students are therefore labelled poor readers. Interestingly, only half of these performed in the lowest quintile in both languages (*n*=1,192).

Table 5. *Cross tabulation across reading in L1 and L2, identifying the poor readers who perform in the lowest quintile in one language*

	The poor L1 and L2 readers' proficiency in the other language					*Total (poor readers)*
	20th	40th	60th	80th	100th	
Poor L1 readers (20th quintile)	*n*=1,192 (54%-56%)	*n*=468 (22%)		*n*=463 (22%)		*n*=2,123 (100%)
Poor L2 readers (20th quintile)		*n*=511 (23%)		*n*=505 (23%)		*n*=2,208 (100%)

Note. The percentages are calculated as parts of the total number of poor readers in L1 and L2, respectively

Since these tests have a ceiling effect and therefore do not provide as good information about average and good readers, we grouped those who read in the 20th quintile in one language and in the 60th to 100th quintiles in the other. We investigated the patterns among these students, who read markedly differently in the L1 and the L2.

First, among all the poor L1 readers (*n*=2,123), most (79%) of the students were in vocational programmes (52% boys and 27% girls), with only 20% in general programmes (12% boys and 8% girls)[1]. However, the pattern among the group of students, who were poor readers in the L1 while being markedly better readers in the L2 (*n*=463, 22%), is quite different. These students included a larger number of boys (66%) who were equally distributed across the study programmes.

Second, among all the poor L2 readers (*n*=2,208), the majority (85%) were in vocational programmes (50% boys and 35% girls), with 14% in general programmes (7% boys and 7% girls). This pattern is rather similar to the pattern among the poor L1 readers. However, in this group, who are poor readers in the L2 and good readers in the L1 (*n*=505, 23%), the clear majority is girls in vocational studies (78%). We also found this fairly complex relationship between L1 and L2 reading proficiency in the sample as a whole, which will be described below.

[1] For the remaining 1%, study programme is unknown.

How do gender and study programme relate to the students' reading scores?

First, Table 6 shows an almost equal number of boys (52%) and girls (48%), and students in vocational (52%) and general (47%) programmes. Next, using raw scores, the analysis showed that the girls (L2: 19.4 points, L1: 79.7 points) read better than the boys (L2: 18.3 points, L1: 72.9 points), and that the students in general studies (L2: 22.2 points, L1: 82.6 points) read better than the vocational students (L2: 15.8 points, L1: 70.2 points). The scores also indicate that the tests are skewed towards higher scores (56%–79% in L2; 64%–76% in L1), which is as expected for this type of mapping tests that are designed to have a ceiling effect.

Table 6. *Descriptive Information for L2 and L1 Mean and Standard Deviations for Raw Scores, with the Distribution and Effect Size (Cohen's d) of Gender and Study Programme for the L1-L2 Sample*

L1-L2 sample (N=10,331)	*Gender*		*Study programme*	
	Boys (*n*=5,398)	Girls (*n*=4,943)	Vocational (*n*=5,345)	General (*n*=4,900)
Percentage	52	48	52	47*
L1 mean (SD)	72.9 (18.6)	79.7 (17.7)	70.2 (19.2)	82.6 (15.2)
L2 mean (SD)	18.3 (7.6)	19.4 (7.2)	15.8 (7.6)	22.2 (5.5)
Cohen's *d* in L1	0.37		0.7	
Cohen's *d* in L2	0.15		0.9	

Note. All effect sizes are statistically significant with p<0.01. *Study programme is unknown for the remaining 1%

Table 6 further gives the standardized total scores divided by the students' gender and study programme. First, the L1 z-scores revealed a gender effect size (Cohen's *d*) of approximately 0.37 in favour of the girls. It is interesting to note that the gender effect was far less for L2 reading (approximately 0.15), with girls still being the more proficient readers. It is also worth noting that the standard deviations for both genders were fairly equal in both assessments. Second, when divided between study programmes, the analysis showed a different pattern and much larger effects reflecting that the general programmes comprise a more homogeneous group of fairly proficient readers than do the vocational programmes. For L1, the effect size was close to 0.7, and for L2 it approached 0.9. This finding showed that the difference in test scores between the groups of students in the two study programmes was markedly higher in the L2 than in the L1.

Before analysing the statistical relationship between the entire sample and the poor L2 readers, we want to mention that many of the main characteristics of the univariate description of the test scores in L1 and L2 reading proficiency are in line with what is usually observed for reading in the L1 and the L2 in Norway (Heber et al., 2010; Ibsen, 2002; Roe, 2013; UDIR, 2012a). Still, Table 6 shows an interesting pattern; namely, the gender effect was relatively smaller for the L2 than the L1, while the study programme effect was relatively larger for the L2 than the L1. Moreover, the gender effect size was relatively smaller than the effect size for study programme.

To what extent is there a statistical relationship between students' L2 reading scores and the variables L1 reading, gender, and study programme?

Based on the studies mentioned in the review section, it is reasonable to expect a strong positive relationship between L1 and L2 reading proficiency (Bernhardt, 2011; Bernhardt & Kamil, 1995; Brantmeier et al., 2012; Grabe, 2009). We applied a multilevel multiple regression analysis to examine this relationship.

The regression analysis used L2 reading proficiency as the dependent variable, and measures of L1 reading proficiency, gender, and study programme as independent predictors. Table 7 summarizes the results from six regression models (A to F). We show the simple bivariate effects of gender, study programme, and L1 reading on L2 reading proficiency separately first (Models A to C). Model D and E are multiple linear regression models where Model D includes the two measures of L1 reading comprehension and L1 language as predictors, while Model E also controls for study programme (with vocational programmes coded as 0 and general programmes coded as 1) and gender (coding boys as 0 and girls as 1). In Model F, two product terms were used to model the interaction between the overall reading performance in L1 and the gender and study programs, respectively.

Table 7. *Results from a multilevel regression model predicting L2 reading comprehension*

Variables entered in the models		*B*	*SE*	*t*	*p*	R^2
Model A	Gender	.07	.02	3.7	.00	.00
Model B	Study programme	.79	.02	37.2	.00	.11
Model C	L1 Reading proficiency (overall L1 score)	.51	.01	61.6	.00	.27
Model D	L1 Reading comprehension	.58	.01	70.7	.00	.41
	L1 Language	.09	.01	10.9	.00	
Model E	Intercept	-.18	.02	-7.6	.00	.43
	Gender	-.02	.01	-1.7	.10	
	Study programme	.37	.02	20.6	.00	
	L1 Reading comprehension	.54	.01	65.1	.00	
	L1 Language	.07	.01	8.7	.00	
Model F	Intercept	-.19	.02	-8.1	.00	.43
	Gender	-.02	.01	-1.6	.10	
	Study programme	.38	.02	20.6	.00	
	L1 Reading comprehension	.57	.01	58.1	.00	
	L1 Language	.12	.01	10.1	.00	
	Gender * L1 Reading proficiency	-.03	.01	-2.3	.02	
	Study programme * L1 Reading proficiency	-.08	.02	-4.5	.00	
	Squared L1 Reading proficiency	.02	.01	4.0	.00	

Note. L1= First language (Norwegian). L2=Second language (English). R^2 expresses amount of within-school variance accounted for as compared to the empty model (the model where only the intercept is allowed to vary between schools). The Bs are standardized regression coefficients. Gender is a dummy with girls=1. Study programme is a dummy with general programmes=1

First, a so-called empty model was estimated. This is a multilevel regression model which is equivalent to a one-way ANOVA where the sole purpose is to decompose the total variance into

two components, one component representing the differences between schools, and one representing the variability of students' performance within the schools. 14% of the total variance is accounted for by the differences between schools. Although this is a rather low proportion, it is at a level where ordinary regression would likely lead to attenuated effect sizes and an underestimation of measurement error.

Model A demonstrates that gender by itself is only weakly related to students' performance in L2. *Model B* demonstrates that study programme has a fairly strong relationship with L2, accounting for 11% of the within school variance in L2 reading proficiency. On average, a student in a general study program scores nearly 0.8 standard deviations higher than a student in a vocational program. *Model C* shows that the bivariate relationship between students' overall L1 proficiency and L2 reading comprehension is relatively strong and accounts for 27% of the variance in their L2 reading proficiency.

In *Model D*, L1 reading proficiency is decomposed into reading comprehension and language. The effect of L1 is primarily related to L1 reading comprehension with a regression coefficient of 0.58. In addition, students' ability to recognize words in the L1 language measure has a small, but unique and statistically significant effect on their L2 reading. When taken together, these two predictors account for 41% of the variance in the students' overall L2 reading proficiency. The intercept is not reported for Models A-D since it is not significant.

Model E, which includes study programme and gender as control variables, does not change this picture substantially. It is interesting to note that the small gender effect observed in Model A disappears when controlling for L1 reading proficiency, L1 language, and study programme. Although reduced from an effect of approximately 0.8 to 0.4, study programme still has a unique effect on L2 reading even when controlling for the students' L1 text reading, L1 language, and gender. The intercept is negative and statistically significant. The value of -0.18 represents the predicted standardized score in L2 for a boy in a vocational programme who has average scores on both the L1 components.

Model F is included to study the potential effects of the interaction between gender and study programme for students which are otherwise equal. In addition, by including the squared L1 term this model tests for potential non-linearity in the relationship between L1 and L2 reading proficiency. These additional terms do not increase the amount of variance accounted for as compared to Model E, and the effects are small. The interaction between study programme and L1 reading proficiency is most pronounced. The most straightforward interpretation of this interaction is that the effect associated with higher L1 reading scores is slightly less for students in the general (academic) studies than for students in the vocational studies.

As stated above, the purpose of this paper is not to study the differences between schools. The purpose of using a multilevel modelling approach, instead of an ordinary multiple regression model, was to improve the estimation of the effects at the student level in a sample where the students are clustered within schools. In our sample, the between-school effect is rather low, accounting for only 14% of the variance. However, even small between-school effects may lead to deflated standard errors and biased estimates of the regression coefficients (Kreft & De Leeuw, 1998). For the analysis presented here, when comparing with the results from an

ordinary regression model, where all the students are assumed to be independent units, we observe that in particular the within-school effect of the students' study programme is strongly reduced. This is not surprising, given that the proportion of students in the different study programmes varies a great deal between schools. Looking at the between-school component of the solution, the effect of students' study program alone accounts for more than 50% of the variation in the average scores of the schools.

In sum, our regression models suggest a positive relationship of up to 11% between L2 reading proficiency and study programme (Model B), which might account for parts of Bernhardt's (2011) unexplained variance. Moreover, the regression models confirm a strong positive relationship of up to 41% between L1 and L2 reading proficiency (Models C and D). Seen together, the combination of gender, study programme, L1 reading comprehension, and L1 language (with or without crossed terms) accounts for up to 43% of L2 reading proficiency (Model E and F).

Discussion

Reading in an L2 "share[s] many features with the same tasks in the first language" (Bunch, Walqui, & Pearson, 2014, p. 539), while also being more complex than reading in the L1 (Koda, 2007). Not surprisingly, the findings presented support the close relationship and shared characteristics of reading in the L1 and L2. Thus, in order to better inform instructional decisions for poor readers, it is also important to include assessments of students' L1 transfer. Koda (2007) identifies the importance of research that not only aims to identify the statistical relationship between reading in L1 and L2, but also to study *how* reading in the L1 and the L2 interact in L2 reading comprehension. Although the design of the presented study did not allow for the identification of such qualitative relationships in the reading transfer between languages, the findings clearly highlight that the relationship between the L1 and the L2 in the reading process is a complex one. One example of this complexity is how study programme moderates the relationship between reading proficiency in the two languages. Clearly, the relationship between L1 and L2 reading proficiency may qualitatively be very different for poor readers, depending on their study programme. We will in the following discuss three aspects of this crosslinguistic relationship.

The relevance of identifying poor readers across the L1 and the L2

The literature demonstrates a consensus among researchers about the utility of identifying readers as either good or poor (e.g., Alderson, 2000; Bråten, Amundsen, & Samuelstuen, 2010; Duke et al., 2011; Grabe, 2009). Duke et al. (2011) argue that "we must understand how skilled comprehenders construct meaning, so we can help students learn to construct meaning in the same way" (p. 52). Our findings expand on this dichotomous notion of good and poor readers by identifying how some of the poor readers in either the L1 or the L2 appear to be markedly better readers in the other language (Table 5). This finding challenges Bernhardt's (2011) notion that a poor reader in one language is most likely a poor reader in the other.

However, these results are in line with Alderson's (1984, 2000) Threshold Hypothesis, suggesting that, even though some of the struggling L2 readers are more proficient in the L1, their L2 reading proficiency might be too low to profit from L1 transfer (Alderson, 2000; Bernhardt & Kamil, 1995; Koda, 2007). If so, these poor L2 readers would need to improve their proficiency in the L2 before they can profit from L1 transfer, which has implications for L2 instruction. It is also of interest that these readers are mostly girls, suggesting that, even among poor readers, girls outperform boys in L1 reading, which in turn echoes the PISA reading results (OECD, 2010, 2013).

One quite unexpected finding was that in the other group among the poor readers, those who read markedly better in the L2 than in the L1, boys are in the majority, outnumbering as well as outperforming the girls. We recommend studying this group in further depth in a future study, for example considering whether some influence from the youth culture among boys leads to their reading English in their spare time.

Furthermore, since vocational students in the present study represent the majority of the poor L1 and L2 readers, the implication of such information is important for English L2 instruction and policy-level decisions in light of the OECD report *Education at a Glance* (2014). The OECD report stated that only 40% of the students in Norway who entered a vocational programme graduated within the stipulated time. In comparison, among the 26 participating countries with available data, 64% of students in vocational programmes and 76% of students in general programmes graduated within the stipulated time (p. 63). The question is to what extent poor reading proficiency contributes to this situation.

Implications of identifying the relationship between L2 reading and study programme

It was not unexpected that L2 reading proficiency varied markedly according to study programme. This finding is in line with previous research in L1 reading among the same reference population (Heber et al., 2010), and also reflects the students' overall grades in the subjects Norwegian L1 and English L2. However, this is the first time the relationship between reading in L2 and study programme has been systematically analysed in Norwegian upper secondary school. We found that study programme is particularly relevant as a background variable for L2 reading proficiency, since study programme is dependent upon the students' own choices when moving on from lower to upper secondary school (Years 10–11).

Our analysis showed that, on average, general studies students were better readers in both languages than vocational students, and the difference was larger for the L2 than for the L1 (Table 6). On the one hand, this finding reflects the selection process from lower to upper secondary school in Norway (UDIR, 2013). On the other, if this difference were due to selection factors only, we would expect that controlling for the students' L1 reading proficiency would heavily reduce the difference between the L2 reading scores in the two study programmes. It turned out that inclusion of measures of L1 reading comprehension and L1 language did reduce the impact of study programme significantly. Indeed, in our regression model, a student in the general studies is predicted to score 0.38 standard deviation units higher in the L2 reading assessment than a similar student in a vocational programme. However, as expressed by the interaction term between study programme and L1 score, the effect of higher L1 reading score is

less for a student in a general (academic) programme. This could be an artefact of the somewhat skewed distributions in the tests analysed, leading to more precise and reliable measures for students at lower levels. Given that more students in general programmes perform at higher levels for both languages, it is to be expected that the estimates of the effects for these students to some degree is attenuated by relatively lower reliabilities. However, this result is also consistent with the threshold hypothesis (Alderson, 1984, 2000). According to this hypothesis, a student needs to achieve a certain level of reading performance in L1 to be able to read with understanding in L2. It is therefore to be expected that a unit increase from very low levels of L1 reading performance is associated with a larger effect than a one-unit increase higher on the scale.

The value of identifying the relationship between L1 and L2 reading proficiency

The present study contributes to the existing research on aspects of the relationship between reading in the L1 and the L2. While Bernhardt's model (2011) indicates that L1 literacy accounts for up to 20% of L2 literacy, we have found an explained variance of 27% to 41% of L1 on L2 reading, depending on the specifications of the model. Our findings revealed that, for all the students, L1 reading was the strongest predictor of their L2 reading proficiency. The explained variance may be higher in our study than in Bernhardt's model (2011) and other studies (e.g., Bernhardt & Kamil, 1995; Brantmeier et al., 2012; Grabe, 2009) because of the linguistic distance between the languages involved (e.g., Koda, 2007). After all, Norwegian and English are Germanic languages, and far closer linguistically than are Spanish, Korean, Japanese, and Chinese, which were involved in the reviewed studies (Bernhardt, 2011; Grabe, 2009).

Strengths and limitations

To sum up, Table 8 provides an overview of strengths and limitations of the present study.

Table 8. *Strengths (+) and limitations (-) in the research reported in this article*

	Strengths (+) / limitations (-)	*Consequences*
+	Unique design	The merging of the L1 and L2 test results enabled a comparison of reading proficiency across the two languages for the first time at this level.
+	Large sample of students (N=10,331)	The results might be applicable to the general upper secondary reference population at this level.
+	Geographically distributed across the country	This has positive influence on representativity.
-	The L1-L2 sample included only 14% of the student population at this level, and they were not randomly selected.	This adds uncertainty to the generalizability of the data. However, the sample is fairly large and representative.
-	Existing tests (secondary data)	Unable to influence test construct. No information on omitted data, such as socio-economic status (SES), and L2 language knowledge related to Bernhardt's (2011) compensatory model of second-language reading.

We contend that the main strength is the design, which has enabled us to compare reading proficiency across the L1 and the L2, for a large sample of upper secondary school students that are geographically distributed across Norway. The main limitations are that the sample includes only 14% of the student population at this level, and is based on secondary data.

Conclusion and Avenues for Further Research

In this article, we provide new information about the relationship between reading in English L2 and Norwegian L1. Our study includes an examination of the effects of gender and study programme, as well as the ways L1 reading and study programme differentially relate to overall L2 reading scores for poor readers. Although we identified poor readers, only about half of them were poor readers in both the L1 and the L2. We have argued that, in order to better inform our knowledge about students' reading in L2, measures of L1 reading are needed to identify converging and diverging aspects of reading in different languages. Such measures would aid and improve decisions regarding what kind of support groups of students might need to further develop their L1 and L2 reading comprehension. Large-scale national assessments may therefore become important tools for supporting teachers in this process. In this light, our findings highlight the importance of such tools for monitoring the progress of poor readers in both languages, and in particular in vocational study programmes.

In this study, merging the information from the two assessments was made possible by teachers who were willing to provide the researchers with data. The teachers provided us with test results that were not readily available through systematic and automatic procedures. We would therefore suggest that large-scale national assessments could profit from installing a logistic routine where students' results on the two assessments are effectively merged and reported back to students, teachers, and schools in a coordinated fashion.

We believe that our findings contribute to the field of L2 reading research through the study's comparison of the ways upper secondary students read across L1 and L2. Our regression analysis suggests that this relationship is a strongly positive one for all the readers in this study, where the variance in L2 reading proficiency was related to a combination of L1 reading, gender, and study programme. Together, these variables account for up to 43% of overall L2 reading scores. We recommend that future studies investigate whether the high statistical relationship of 11% between study programme and L2 reading in our study partially explains unexplained variance, as described in Bernhardt's model (2011).

While the observed relationships were consistent with and expanded prior findings on reading across the L1 and the L2 (e.g., Bernhardt, 2011), this is the first study comparing Norwegian as L1 and English as L2. Based on our findings, a follow-up study investigating the importance of the language distance might be of interest. Likewise, building upon the present study with results from the examined reading tests in a longitudinal perspective, analysing data from 2010 onwards, would provide insight into whether the patterns found in the present study are confirmed over time.

Hopefully, our finding can benefit researchers, practitioners, and policymakers, not least since these findings support Koda's (2007) claim that reading in an L2 is a complex phenomenon involving two languages. They also show the importance of taking the crosslinguistic aspect of reading in an L2 into consideration in further research, particularly related to school reading instruction and testing.

References

Afflerbach, P., Pearson, D., & Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher*, *61*, 364–373. DOI: http://10.1598/RT.61.5.1

Alderson, C. (1984). Reading in a foreign language: A reading problem or a language problem? In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a Foreign Language* (pp. 1–27). London: Longman.

Alderson, C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Alderson, C., Haapakangas, E. L., Huhta, A., Nieminen, L., & Ullakonoja, R. (2015). *The diagnosis of reading in a second or foreign language.* New York: Taylor & Francis.

Bernhardt, E. B. (2011). *Understanding advanced second-language reading*. New York: Routledge.

Bernhardt, E. B., & Kamil, M. L. (1995). Interpreting relationships between L1 and L2 reading: Consolidating the linguistic threshold and the linguistic interdependence hypotheses. *Applied Linguistics*, *16*, 15–34. DOI: http://10.1093/applin/16.1.15

Block, C. C., & Duffy, G. G. (2008). Research on teaching comprehension. Where we've been and where we're going. In C. C. Block & S. R. Parris (Eds.), *Comprehension instruction. Research-based best practices* (2nd ed.) (pp. 19–37). New York: The Guilford Press.

Bonnet, G. (2004). The assessment of pupils' skills in English in eight European countries 2002. The European network of policy makers for the evaluation of educational systems. Retrieved from: https://www.eva.dk/projekter/2002/evaluering-af-faget-engelsk-i-grundskolen/projektprodukter/assessmentofenglish.pdf

Brantmeier, C. (2004). Statistical procedures for research on L2 reading comprehension: An examination of ANOVA and regression models. *Reading in a Foreign Language*, *16*, 51–69.

Brantmeier, C., Bishop, T. V., Yu, X., & Anderson, B. (2012). Readings on L2 reading: Publications in other venues 2011-2012. *Reading in a Foreign Language*, *24*, 256–272.

Brantmeier, C., Sullivan, J. A. H., & Strube, M. (2014). Toward independent L2 readers: Effects of text adjuncts, subject knowledge, L1 reading, and L2 proficiency. *Reading in a Foreign Language*, *26*, 34–53.

Brevik, L. M. (2014). Making implicit practice explicit: How do upper secondary teachers describe their reading comprehension strategies instruction? *International Journal of Educational Research*, *67*, 52–66. DOI: http://10.1016/i.ijer.2014.05.002

Brevik, L. M. (2015). Strategies and shoes: Can we ever have enough? Teaching and using reading comprehension strategies in general and vocational programmes. *Scandinavian Journal of Educational Research*. DOI: http://dx.doi.org/10.1080/00313831.2015.1075310

Bråten, I. (2007). Leseforståelse – komponenter, vansker og tiltak. In I. Bråten (Ed.),

Leseforståelse. Lesing i kunnskapssamfunnet - teori og praksis (pp. 45–81). Oslo, Norway: CappelenAkademisk.

Bråten, I., Amundsen, A., & Samuelstuen, M. S. (2010). Poor readers – good learners: A study of dyslexic readers learning with and without text. *Reading and Writing Quarterly: Overcoming Learning Difficulties*, *26*, 166–187. DOI: http://10.1080/10573560903123684

Bunch, G. C., Walqui, A., & Pearson, D. P. (2014). Complex text and new common standards in the United States: Pedagogical implications for English learners. *TESOL Quarterly*, *48*, 533–559. DOI: http://10.1002./tesq.175

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, *49*, 222–251. DOI: http://10.3102/00346543049002222

Cummins, J. (2000). *Language, power and pedagogy*: *Bilingual children in the crossfire*. Clevedon, UK: Multilingual Matters.

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading Research Quarterly*, *38*, 78–103. DOI: http://10.1598/RRQ.38.1.4

Duke, N. K., Pearson, P. D., Strachan, S. L., & Billman, A. K. (2011). Essential elements of fostering and teaching reading comprehension. In S. J. Samuels & A. E. Farstrup (Eds.), *What research has to say about reading instruction* (4th ed.) (pp. 51–93). Newark: International Reading Association.

Frønes, T. S., Narvhus, E. K., & Aasebø, M. C. (2013). Nordic results from the PISA digital reading assessment. *Nordic Journal of Digital Literacy*, *8*(01-02), 13–31. Retrieved from: https://www.idunn.no/file/pdf/61221164/nordic_results_from_the_pisa_digital_reading_assessment.pdf

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge: Cambridge University Press.

Heber, E., Mossige, M., & Kittel, A. (2010). *Kartleggingsprøver i lesing for Vg1. Resultater fra standardiseringsundersøkelsen september 2010* [Mapping tests in reading for upper secondary school year 1. Results from the standardization test September 2010]. Internal report to the Norwegian Directorate for Education and Research. University of Stavanger: Norwegian Reading Centre.

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2010). *Multilevel and longitudinal modelling with IBM SPSS*. New York/London: Routledge.

Hellekjær, G. O., & Hopfenbeck, T. N. (2012). *CLIL og lesing. En sammenligning av Vg3-elevers leseferdigheter og lesestrategibruk i 2002 og 2011* [CLIL and reading. A comparison of students' reading skills and reading strategy use in upper secondary school, Year 13]. In B. W. Svenhard (Ed), *CLIL: Kombinert fag- og engelskopplæring i videregående skole. Fokus på Språk*, *28*, 84–124. Retrieved from: http://www.fremmedspraksenteret.no/neted/modules/archive/front/file.php?data=4661d2a1e517eb4f134b3938b25ce12b

Ibsen, E. B. (2002). *English in Norway*. Retrieved from: http://www.eva.dk/projekter/2002/evaluering-af-faget-engelsk-i-grundskolen/projektprodukter/assessmentofenglish.pdf

Koda, K. (2005). *Insights into second language reading*. Cambridge, England: Cambridge University Press.

Koda, K. (2007). Reading and linguistic learning: Crosslinguistic constraints on second language reading development. *Language Learning*, *57*, 1–44. DOI: http://10.1111/0023-8333.101997010-i1

Koda, K. (2010). The role of reading in fostering transcultural competence. *Reading in a Foreign Language*, *22*, 5–10.

Kreft, I., & De Leeuw, J. (1998). *Introducing multilevel models*. Thousand Oaks: Sage.

Lee, J. W., & Schallert, D. L. (1997). The relative contribution of L2 language proficiency and L1 reading ability to L2 reading performance: A test of the threshold hypothesis in an EFL context. *TESOL Quarterly*, *31*, 713–739. DOI: http://10.2307/3587757

McNamara, D. S. (2011). Measuring deep, reflective comprehension and learning strategies: Challenges and success. *Metacognition and Learning*, *6*, 195–203. DOI: http://dx.doi.org/10.1007/s11409-011-9082-8

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific literature on reading and its implications for reading instruction: Reports of the sub-groups.* Bethesda, MD: National Institute of Child Health and Human Development.

Norwegian Directorate for Education and Training [UDIR]. (2010a). *Kartleggingsprøve i engelsk for vg1. Veiledning* [Mapping tests in English for upper secondary school year 1. A guide]. Oslo: Author.

Norwegian Directorate for Education and Training [UDIR]. (2010b). *Kartleggingsprøve i lesing. Lærerveiledning for vg1* [Mapping tests in reading. A teacher's guide for upper secondary school year 1]. Oslo: Author.

Norwegian Directorate for Education and Training [UDIR]. (2011). *Utdanningsspeilet 2011* [The Education Mirror 2011]. Oslo: Author. Retrieved from: http://www.udir.no/globalassets/upload/rapporter/education_mirror_2011/the-education-mirror-2011.pdf

Norwegian Directorate for Education and Training [UDIR]. (2012a). *Analyse av nasjonale prøver i lesing 2012* [Analyses of national tests in reading 2012]. Oslo: Author

Norwegian Directorate for Education and Training [UDIR]. (2012b). *Kartleggingsprøve i engelsk VG1 – 2012* [Mapping test in English year 1 - 2012]. Internal report from the test developers. Oslo: Author.

Norwegian Directorate for Education and Training [UDIR]. (2012c). *Framework for Basic Skills*. Oslo: Author. Retrieved from: http://www.udir.no/Udir/PrintPageAsPdfService.ashx?pid=54996&epslanguage=no

Norwegian Directorate for Education and Training [UDIR]. (2012d). *Utdanningsspeilet 2012* [The Education Mirror 2012]. Oslo: Author. Retrieved from: http://www.udir.no/globalassets/upload/rapporter/2012/us2012_eng_nettversjon.pdf

Norwegian Directorate for Education and Training [UDIR]. (2013). *Utdanningsspeilet 2013* [The Education Mirror 2013]. Oslo: Author. Retrieved from: http://www.udir.no/globalassets/upload/rapporter/theeducationmirror_2013.pdf

Norwegian Ministry of Education and Research [KD]. (2006, 2013). *Læreplan for grunnskolen og videregående skole* [Curriculum for elementary and secondary school]. Oslo: Author. Retrieved from: http://www.udir.no/kl06/eng1-03/Hele/?lplang=eng#

OECD. (2010). *PISA 2009 results: What students know and can do – Student performance in reading, mathematics and science* (Vol. I). DOI: http://dx.doi.org/10.1787/9789264091450-en

OECD. (2013). *PISA 2012 results: Ready to learn – students' engagement, drive and self-beliefs*

(Vol. III). Paris: OECD Publishing. DOI: http://dx.doi.org/10.1787/9789264201170-en
OECD. (2014). *Education at a glance 2014: OECD indicators*. OECD Publishing.
Olsen, R. V., Hopfenbeck, T. N., Lillejord, S., & Roe, A. (2012). *Elevenes læringssituasjon etter innføringen av ny reform. Synteserapport fra evalueringen av Kunnskapsløftet*. [*Students' learning situation after the new reform. Report from the evaluation of the Knowledge Promotion*]. Acta Didactica Oslo. Oslo: Department of Teacher Education and School Research. Retrieved from: http://www.udir.no/Upload/Rapporter/2013/Synteserapport%20RVO%20endelig%20jan%202013.pdf
RAND. (2002). *Reading for understanding. Toward an R&D program in reading comprehension.* RAND Reading Study Group. USA: Rand Education. Retrieved from: http://www.rand.org/content/dam/rand/pubs/monograph_reports/2005/MR1465.pdf
Roe, A. (2013). Lesing [Reading]. In M. Kjærnsli & R.V. Olsen (Eds.), *Fortsatt en vei å gå. Norske elevers kompetanse i matematikk, naturfag og lesing i PISA 2012* (1st ed.) (pp. 177–200). Oslo: University Press.
Stanovich, K. E. (1980). Towards an interactive compensatory model of individual differences in the development in reading fluency. *Reading Research Quarterly*, *16*, 32–71. DOI: http://10.2307/747348
Tveit, S. (2014). Educational assessment in Norway. *Assessment in Education: Principles, Policy & Practice*, *21*, 221–237. DOI: http://10.1080/0969594X.2013.830079
Yamashita, J. (2002). Mutual compensation between L1 reading ability and L2 language proficiency in L2 reading comprehension. *Journal of Research in Reading*, *25*, 81–95. DOI: http://10.1111/1467-9817.00160

About the Authors

Lisbeth M. Brevik is an Associate Professor of English Didactics in the Department of Teacher Education and School Research at the University of Oslo, Norway. Her research interests relate to reading comprehension instruction, strategies, assessment, and classroom observation in English as a second language. E-mail: l.m.brevik@ils.uio.no

Rolf Vegar Olsen is a Professor at the Centre for Educational Measurement at the University of Oslo (CEMO), Norway. His research interests relate to the measurement and international large-scale assessment of students' skills and competencies. E-mail: r.v.olsen@cemo.uio.no

Glenn Ole Hellekjær is a Professor of Teaching English as a Foreign Language in the Department of Teacher Education and School Research at the University of Oslo, Norway. His research interests include academic reading in English, language needs analysis, and English teacher education. E-mail: g.o.hellekjar@ils.uio.no