

Glenn Ole Hellekjær

Associate professor, Department of Teacher Education and School Development,
University of Oslo.

A Case for Improved Reading Instruction for Academic English Reading Proficiency

Abstract

This article presents a study of the academic reading proficiency in English of 217 senior level Norwegian upper secondary school students who upon graduation are considered qualified for higher education. Testing with an International English Language Testing System (IELTS) Academic Reading Module revealed that two thirds of the 178 respondents with ordinary EFL courses did not achieve the equivalent of the IELTS Band 6 score minimum that is usually required for admission to British and Australian universities. In comparison, two thirds of a sample of 39 respondents with a single, sheltered Content and Language Integrated Learning (CLIL) subject achieved a Band 6 score or better. Closer analysis indicates that the poor test scores can be attributed to weaknesses in current English as a Foreign Language (EFL) instruction where reading is neglected, where students do not learn to adjust how they read to reading purpose, and where they do not learn how to handle unfamiliar words to avoid disrupting the reading process. The article ends with suggestions on how to improve EFL instruction, in Norway and elsewhere.

Introduction

In a small language community such as Norway, a good command of English is of vital importance in almost all domains, from the purely personal to the domains of public administration, business and higher education. To give an example, a recent study showed that Norwegian exporters used English for 95% of their international communication and that many of the firms lacked staff with adequate levels of proficiency (Hellekjær 2007a). English is just as important in higher education because a small language community means that English texts and textbooks must be used extensively, in particular at advanced levels and in specialized courses (Dahl 1998; Hatlevik & Norgård 2001). This makes the ability to read and learn from English texts an essential academic skill, even though lectures and examinations are usually in Norwegian.

In Norway it has been taken for granted that upper secondary level instruction in English as a Foreign Language (hereafter referred to as EFL) develops the proficiency needed for higher education. Indeed, this might even seem reasonable since English is taught for all ten years of elementary school and from one to three years of upper secondary. Norwegian students also receive extensive media exposure from subtitled English films and TV programs. Furthermore, Norwegian 16-year-olds did well in a comparative, eight-country European study of English proficiency (Bonnet 2004).

However, this assumption of adequacy has been challenged in two doctoral studies. Lehmann (1999), and Hellekjær (2005) found serious weaknesses in university and college students' writing and academic reading proficiency respectively. Both studies are highly critical of the effectiveness of Norwegian EFL instruction at the upper secondary level with regard to preparing for higher education. This article is based on data from one of the surveys in Hellekjær's (2005) study of the academic reading proficiency in English of 217 Norwegian, senior level, upper secondary school students (Hellekjær 2005). The respondents underwent tests based on an Academic Reading Module from the International English Language Testing System (hereafter referred to as IELTS) from the University of Cambridge Local Examinations Syndicate (hereafter referred to as UCLES) (see www.ielts.org). The main goal of the 2005 study is to ascertain to what extent Norwegian upper secondary students have attained the levels of academic English reading proficiency needed for higher education. Second, it attempts to identify key variables that co-vary with test scores. Third, it attempts to identify deficiencies in the respondents' reading proficiency that might be attributed to EFL instruction. Finally, it uses the findings to suggest changes in EFL teaching in Norway, changes that might be relevant in other countries as well.

In the following, I will start with a short description of the skill that is tested, academic English reading proficiency, also known as a *construct definition*.

Defining the construct: academic reading in EFL

In *Assessing Reading* Alderson (2000 p. 118) defines a construct as a “*psychological concept*” derived “*from a theory of the ability to be tested*” that can be used for testing purposes. It is

a definition which focuses on an aspect of the ability that is of particular relevance to our testing purpose, or it may be a definition that we adopt wholesale from previous research or practice.

(Alderson 2000 p. 119)

In the following I will briefly present a construct definition for reading.

For the most part, current models describe reading as an interactive, but first and foremost a lower-level (bottom-up) process that also draws upon higher-levels (top-down) processes (see for instance Alderson 2000; Bråten 2007; Grabe 1999; Koda 2004). The core, the bottom-up process involves recognizing the written words in the text along with relevant grammatical information. This process of automatic word recognition, in turn, forms the basis for higher-level processing, i.e. the creation of meaning in an interactive process between the information in the text being read and the reader's knowledge of language, content, and processing capabilities.

With a fluent reader, the process of word recognition proceeds effortlessly and rapidly in the working memory. In case of difficulty, for instance an unfamiliar word, the process may slow down, or even stop up while the reader attempts to use “*other knowledge sources, regardless of their level in the processing hierarchy*” to deduce meaning (Stanovich 1980 p. 3). However, due to the limited processing capacity of the working memory, this reduces reading speed and fluency (Rayner & Pollatsek 1989; Stanovich 1980). Such a slowdown highlights the importance of a large sight vocabulary for fluent reading, an area where one finds the main difference between first and foreign language reading. In fact, Grabe (1988 p. 63) argues that the lack of

a massive receptive vocabulary that is rapidly, accurately and automatically processed [. . .] may be the greatest single impediment to the fluent reading by ESL students.

Alderson puts this as follows:

Measures of a readers’ vocabulary knowledge routinely correlate highly with measures of reading comprehension, and are often, indeed, the single best predictor of text comprehension.

(Alderson 2000 p. 35)

The importance of vocabulary notwithstanding, fluent reading in an academic context also requires the ability “*to integrate text and background information appropriately and efficiently*” (Grabe & Stoller 2002 p. 28). This involves using background knowledge, that is to say content knowledge on the one hand, and knowledge of the language and text types on the other. It also involves other cognitive processes, where I will focus on meta-cognitive monitoring and reading strategies.

To start with background knowledge, research has shown that such knowledge not only influences what a reader remembers from a text, but also his or her understanding of the content (Alderson 2000 p. 35). Indeed, knowledge of the world in general, and knowledge of the topic in question can, within limits, support understanding when it is necessary to compensate for language difficulties (Stanovich 1980). However, in her 1996 study, Clapham (1996 p. 197), found that

language proficiency appeared to have a much stronger effect on students’ scores than did background knowledge. However, the comparative importance of the variables seemed to depend on the specificity of the tests.

In other words, poor language proficiency prevented her respondents from compensating for their lack of understanding by using a top-down strategy such as drawing on subject matter knowledge to guess the meaning of unknown

words and phrases, or, if the specific topic is unfamiliar, to build up understanding from the text using a bottom-up strategy. In contrast, the linguistically proficient readers in her sample could “*compensate for a certain lack of background knowledge by making full use of their language resources*” (Clapham 1996 p. 196).

Furthermore, when faced by an apparent inconsistency in a text, or in their understanding of the content, readers can make use of metacognitive monitoring. This is the ability to monitor understanding across the text and use linguistic and/or content knowledge to repair comprehension (Alderson 2000 p. 43). In fact, this is one of the main factors distinguishing good from poor readers (Alderson 2000, Bråten 2007; Bråten & Olaussen 1997). Alderson argues that good readers are:

. . . more sensitive to inconsistencies in the texts. . . and 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.

(Alderson 2000 p. 41)

This focus on word-level cues and on intrasentential relations may explain the tendency of poor readers to focus on, and be hindered by word problems, which is particularly relevant for L2 reading. The good readers, on the other hand, seem better able to decide when to ignore unfamiliar words, which, as mentioned, is of particular importance when reading in a foreign language.

Last in this brief overview is skill and strategy use. A strategy can be described as a set of abilities that are under the conscious control of the reader, as compared to a skill where the use is automatic. Examples would be re-reading to sort out a discrepancy in meaning, guessing in order to sort out the meaning of unknown words, or, alternatively, ignoring these if possible. Another would be adjusting how one reads to reading purpose, such as using skim reading to get the main points of the text, search reading to find particular information, or scanning through a text to find a particular name or phrase. It might also mean engaging in careful reading at the local level to understand the syntactic structure of a sentence or clause, or careful reading at the global level for comprehension of the main ideas of a text. Depending upon the reader’s proficiency some of these decisions will be made consciously, others automatically (see for instance Urquhart & Weir 1998).

In fact, skill and strategy use is considered one of the weak areas in L1 reading instruction in Norway. Bråten cites several studies claiming that far too little emphasis is put on teaching Nordic students how to *read to learn*, which would entail instruction in reading as well as in learning strategies (Bråten 2007; Bråten & Olaussen 1997; Ro 2006). Instead, there is an almost exclusive emphasis on what Urquhart & Weir (1998 p. 103) call “*careful reading at the*

global level for comprehension of the main ideas of a text.” They attribute this to careful reading being “*favoured by many educationalists . . . to the exclusion of all other types*” (Urquhart & Weir 1998 p. 103). Urquhart & Weir claim this focus on one type of reading is problematic because it prevents students from learning to adjust how they read to reading purpose.

To sum up, reading proficiency can be described as more than just the ability to decode the written words in the text, but as the active creation of meaning in an interactive process between information in the text on the one hand, and the reader’s knowledge on the other.

This study presents academic reading in English as a foreign language, involving older learners who have already learnt to read in their L1 and who can call upon their L1 skills when reading in a foreign language (Koda 2004). This ability depends upon their proficiency in the foreign language in question, also known as the Linguistic Threshold Level. Students whose L2 proficiency falls below a certain level, despite their being fluent readers in the L1, prove unable to transfer these strategies and skills to the L2 (Alderson 2000; Bernhardt & Kamil 1995; Carrell 1991; Laufer 1997). Furthermore, Alderson (2000 p. 39) points out that “*this linguistic threshold is not absolute but must vary from task to task: the more demanding the task, the higher the linguistic threshold*” Alderson also states that “*second-language knowledge is more important than first-language abilities*” (Alderson 2000 p. 39). Thus, the IELTS test scores reported in this study reflect the respondents’ English proficiency as well as their reading skills and strategy use.

Method, sample and procedure

The quantitative study reported here uses a quasi-experimental, one-group, post-test design (Shadish, Cook, & Campbell 2002 pp. 106-107). Its lack of a pre-test and of control groups means that it is of limited utility for the identification of causal relations. This goes to say that not all instances of covariation found reflect cause and effect; some probably do and can lead to hypotheses about causal relations. What it does mean is that this design does not allow for the satisfactory “*identification and study of plausible threats to internal validity*” needed to identify such causal relations by excluding other explanations (Shadish, Cook, & Campbell, 2002 p. 105). Therefore, the statistical analysis of the data will for the most part concentrate on presenting mean scores, score and respondent distributions, and covariations between dependent and independent variables. I will, however, in some cases go beyond the limitations of the design to suggest causal relations that should be investigated further.

With regard to the sample, at the outset of the study in spring 2002 I did not have the time and opportunity to contact a representative selection of upper secondary schools. Instead I contacted ten schools with college preparatory branches where I knew teachers who might be willing to help out. Three also had sheltered, single subject Content and Language Integrated Learning (here-

after referred to as CLIL) classes that I wanted to include for comparison with EFL instruction (for more about CLIL in Norway, see Svenhardt, Servant, Hellekjær & Bøhn 2007). Since three of the schools contacted declined to participate, I ended up with a sample from seven schools in different parts of Norway comprising 217 senior level, upper secondary students, all around 18 to 19 years old. Of these students 178 had had EFL instruction only, while 39 had had a single subject, sheltered CLIL course in either Physics or History in addition to their EFL courses. In the following analysis these will be treated as separate samples, the EFL and CLIL sub-samples.

The upper secondary students in question specialized in either the natural sciences, social studies, or in modern languages. Since English is a popular elective subject, a high proportion, 56% of the 178 students in the EFL sub-sample had opted for the Advanced English Course. This is an elective course with five lessons per week in the second and third year following a compulsory first year, five lessons per week Foundation Course. Among the remaining respondents, 45 out of 178 (25%) had the first year course only while a further 30 out of 178 (17%) had a three or five-lessons-per-week course in their second year. Although a somewhat lower proportion of the CLIL students, 41% compared to 56%, had the Advanced English Course, the sub-samples are otherwise roughly comparable apart from the single CLIL subject.

From my knowledge of the schools in question I would argue that the convenience sample I ended up with is skewed in favor of better than average students. The main reason is that at least four of the seven schools had particularly good reputations, and consequently attracted good students. Furthermore, no Supplementary Course classes, in which students from vocational backgrounds qualify for higher education by completing the first year English Foundation Course were included. In other words, it is therefore reasonable to assume that the resulting test scores are probably higher than would have been the case with a representative sample from a reference population comprising all Norwegian students qualifying for higher education.

The tests and questionnaires were sent out by mail to the participating schools. A teacher was present during the two consecutive lessons, of about 45 minutes each, used for the test. During the survey, the respondents filled in a questionnaire with 74 items ranging from indicators of independent variables expected to covary with reading comprehension scores to those of variables providing information about student background and EFL instruction. As mentioned, the respondents also completed an IELTS Reading for Academic Purposes Module, a specimen test used with UCLES permission and claimed to be identical in difficulty to comparable tests for admission purposes (UCLES 2001a; 2001b). Australian, British, Canadian, and New Zealand institutions use the IELTS test to assess the English proficiency of students for whom English is a foreign language. The results of the tests are converted to composite band scores, Bands 1 to 9, where Band 1 is the lowest and Band 9 the highest.

Although some institutions accept students with scores at the Band 5 level and others require Band 7, most institutions consider Band 6 the minimum score for admission purposes.

The test itself comprises three texts of about 950 words on different topics. Test items vary from matching paragraphs and headings, agreeing or disagreeing with given statements in relation to the text in question, to stating that there is no relevant information in the text on a particular subject or filling in tables or graphs using information in the texts. Respondents are also required to make use of background knowledge, both of content and text type, to interpret and understand the texts. Some items drew upon the ability to make inferences needed for correct understanding, to read strategically, and to use metacognitive processing – the ability to monitor and realign comprehension while reading. In addition, the one-hour time limit allotted for the answering of 38 items gives an indirect measure of reading speed by forcing the respondent to read fairly quickly and vary how they read according to their purpose, such as scanning to find a key date or reading carefully to find a piece of information.

For the specimen test used in this study the conversion tables used to calculate Band scores were not available. The results are therefore tallied as correct or incorrect answers, and based upon the guidelines for the test. A score achieving 25 of 38 points is designated the rough equivalent of Band 6. The test scores were also combined into additive indices to serve as dependent variables for statistical analysis.

About 10 questionnaires where the respondents had not even started on the IELTS test were rejected. The data was processed by means of the statistical processing program Statistical Package for the Social Sciences (SPSS).

Results and analysis

In the following presentation of my findings and analysis I will start by presenting the IELTS scores of the EFL and CLIL sub-samples. I will then continue the analysis by examining data that can explain these scores.

I mentioned above that the IELTS Academic Reading Module used for this survey comprises 38 items drawn from three texts, out of which 25 correct answers are considered the rough equivalent of Band 6. In Table 1 below the mean test scores for the EFL and CLIL sub-samples are presented.

Table 1. IELTS Academic Reading Module scores for the EFL and CLIL sub-samples. The maximum possible score is 38.

Sample	Mean IELTS score (\bar{X})	Standard Deviation (SD)	Respondents (N)
EFL sub-sample	21	9.0	178
CLIL sub-sample	28	7.9	39

As can be seen, the mean score of $\bar{X}=21$ for the EFL sub-sample is not only four points below the Band 6 level; it is almost seven points below that of the CLIL sub-sample. In Figure 1 below the group means and confidence intervals for the EFL and the CLIL sub-samples are displayed.

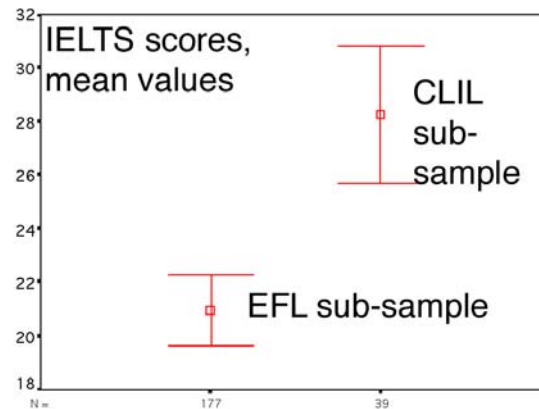


Figure 1. Confidence intervals for the EFL and CLIL sub-samples with 178 and 39 respondents with IELTS scores as the dependent variable. The difference between group means is statistically significant at the 95% level of certainty. Maximum IELTS score is 38.

The gap between the mean scores for the two sub-samples displayed in Figure 1 goes to show that the difference between the two sub-samples is statistically significant at the 95% level.

Mean scores are one thing, how the scores are distributed is another. The distribution of scores for the EFL sub-sample is displayed in Figure 2 below.

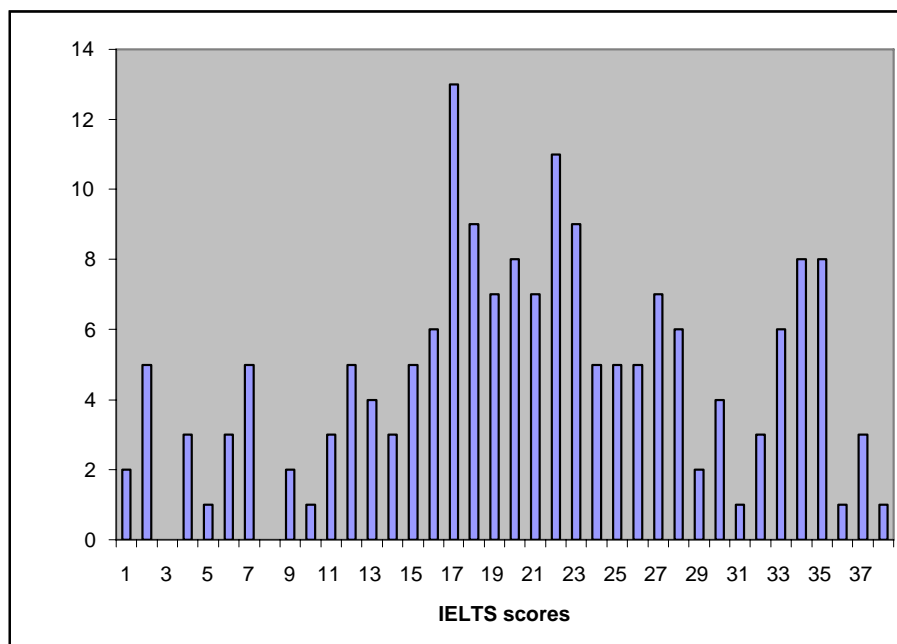


Figure 2. IELTS Academic Reading Module scores for the EFL sub-sample. The maximum score is 38, $\bar{X}=21$, $SD=9.0$. $N=178$.

The distribution of the IELTS Academic reading scores in Figure 2, with two-thirds of the EFL sub-sample scoring less than 25 points, stands in strong contrast to the expectations that upper secondary EFL instruction effectively prepares for higher education, not to mention the reputed Norwegian high proficiency in English. Of course, one explanation might be that the IELTS test format is an unfamiliar one for most Norwegian students. Many would undoubtedly have scored higher if they had had more experience with this kind of test. On the other hand, the respondents were allowed 90 minutes to answer, considerably more than the hour those taking the IELTS Academic Reading Module are usually allowed to use. This extra time should, at least partially, have compensated for an unfamiliar test format. Consequently, the fact that as many as 66% of the 178 respondents score below the Band 6 level is disquieting. The scores give an even more disturbing indication of fundamental weaknesses in Norwegian EFL instruction when compared to the scores of the CLIL sub-sample displayed in Figure 3 below. We should keep in mind that the CLIL students had no more experience with the IELTS test format than did the EFL sub-sample.

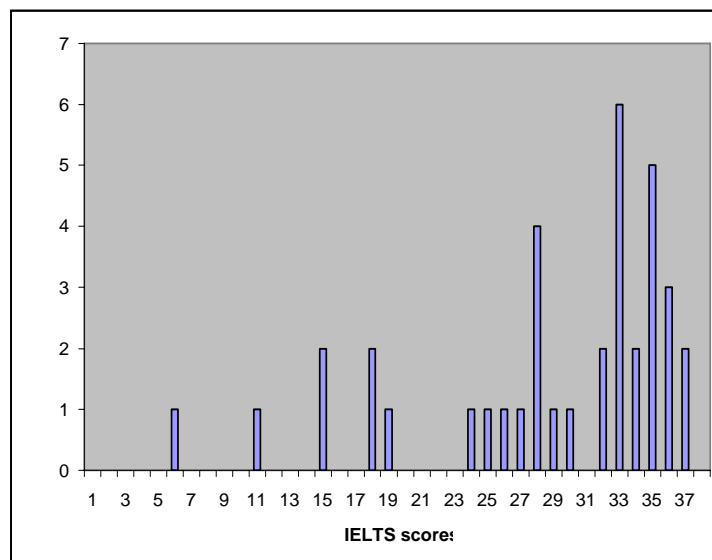


Figure 3. IELTS Academic Reading Module scores for the CLIL sub-sample. The maximum score is 38, $\bar{X} = 28$, $SD = 7.9$. $N = 39$.

As for the CLIL sub-sample, Figure 3 shows that only 26% of the respondents score less than 25 points, compared to 66% in the EFL sub-sample. Likewise, while 74% of the CLIL sample score 25 points or above, only one third of the EFL sample do so. In fact, the IELTS scores for this sub-sample, $\bar{X} = 28$ ($SD = 8.0$), are roughly comparable to those of experienced Norwegian university level students tested in a different study who had a mean score of $\bar{X} = 30$ ($SD = 8.1$) (Hellekjær 2005). This difference might in part be explained through selection factors, since CLIL students must be volunteers (Svenhardt & al. 2007). However, closer examination of the test papers of the EFL sub-sample showed that the low scores were largely due to unanswered items, not errors. It seemed

that a great many of the respondents with low scores only managed from 12 to about 17 correct answers in the time available and left the remaining items unanswered, which indicates that they read and worked very slowly. In contrast, the respondents in the CLIL sub-sample were able to answer many more of the items, which matches my experience of upper secondary students in CLIL courses. At the outset they read slowly and carefully for detailed understanding in the way typical of textbook reading in EFL instruction, but quickly had to change pace in order to keep up (Hellekjær 1996). That is to say, the CLIL students must quickly learn to read for general meaning and tolerate some uncertainty with regard to unfamiliar words in order to manage the course. I will return to this issue below.

Upper secondary EFL instruction and reading proficiency

A possible explanation of the poor IELTS scores of the EFL sub-sample is that many of the respondents with low scores have had fewer hours of EFL instruction than those who did well. This can be tested by seeing whether completing the elective Advanced English Course correlates positively with the IELTS test scores, since this course comprises five lessons per week over two years and represents a considerable expenditure of time and effort. Alternatively, if there is no positive correlation, problem may be systemic and due to deficient reading instruction. The latter conclusion was corroborated when a dummy variable for completing the Advanced Elective English Course showed no significant correlation with the IELTS scores for the EFL sub-sample, $r = .01$, ($p = .85$, $N = 178$).

Of course, negative selection might be a possible explanation. Usually the academically strong upper secondary students choose courses in the Natural Sciences instead of languages (Ibsen & Lie 1990). In fact, when other third year, advanced elective courses the students had completed were correlated with IELTS scores, Mathematics had a positive and statistically significant correlation, $r = .25$, $p < .01$, $N = 178$, as did Physics $r = .23$, $p < .05$, $N = 178$. For the other subjects the results were low and not significant. These correlations indicate that selection factors might, at least partially, contribute to the lack of positive correlation between the Advanced English Course and the test scores. On the other hand, if the considerable expenditure of time and effort the Advanced English Course represents is taken into consideration, selection factors should not be an acceptable explanation.

Exposure to English

One of the variables that was expected to covary positively with the test scores was exposure to English through reading or the media. The questionnaire therefore included a number of items about such exposure to English. One item asked the students how many English novels they had read. Likewise, they were asked how often they read English books and magazines. There was a similar

question about their reading of English on the Internet, on watching English language films without subtitles and the number of English books in their homes. The correlations between these independent variables and the IELTS scores for the EFL sub-sample are displayed in Table 2.

Table 2. Correlations for English media consumption with IELTS scores for the EFL sub-sample. N=178.

	Eng books read, number	Eng books, reading frequency	Eng periodicals, reading frequency	Internet reading, frequency	English films/videos, frequency	Eng books at home, number
r	,21	,17	,15	,21	-,04	,13
p	,00	,00	,02	,00	,55	,08

Although the correlations displayed in Table 2 are fairly low, they still go to show that the number of books read and reading on the Internet correlate positively with the IELTS scores, while the correlation with watching English films/videos is low and not significant. Furthermore, multiple regression analysis with IELTS scores as a dependent variable and the items for number of books read, frequency of reading English in magazines, and on the Internet as independent variables give an explained variance of only $R^2 = .08$ for the EFL sample (that is to say how much of the variation is accounted for by these variables). The low correlations and the low explained variance for reading were somewhat unexpected. As can be seen in Table 3 this is probably due to the low number of books the students in the EFL sub-sample have read.

Table 3. The number of English books read by the EFL sub-sample, N=178.

Number of books read	None	1-5	6-10	11-15	16-20	21-50	51 or more
Respondents	7	91	37	15	10	15	2

The numbers in Table 3 suggest that most of the students seem to have read only the absolute minimum required by the current Norwegian EFL syllabus, where reading is given low priority (Hellekjær 2005). Consequently, the lack of variation with regard to reading can largely explain the low correlations. It is, nevertheless, interesting to note that the mean IELTS score for the 27 respondents who had read the most, 16 books or more, was fairly high, $\bar{X} = 24$ ($SD = 9.4$) compared to $\bar{X} = 20$ ($SD = 8.9$) for the remaining 150 who had read less. Other Scandinavian studies have also shown that extracurricular reading gives an advantage (see for instance Hellekjær 2005; Sylvén 2004).

Last, the low correlation between the frequency of watching English videos and films and IELTS scores that emerges from Table 2 shows that this input should not be overrated with regard to language development for other skills than listening. A possible explanation is that most English language films and videos available in Norway are subtitled in Norwegian. However, it would

hardly be possible to pick up many of the low-frequency words necessary for fluent academic reading from films and videos with predominantly oral, everyday language. It is also possible that the relatively high consumption of films and videos some respondents admit to having comes at the expense of activities such as homework or reading, a fact which could also explain the negative correlations.

How students handle unfamiliar words

As mentioned, it is important for the students' reading proficiency how they handle unknown words when they read in a foreign language, and this was expected to be an important variable in the survey. Respondents were therefore asked to indicate on a seven point Likert scales from 1 (never) to 7 (frequently) how often they resorted to dictionaries, deduced or guessed word meaning from context or from their knowledge of the subject, asked a teacher, parents or fellow students, ignored the word and kept on reading, or simply gave up entirely. The following correlations are from the EFL sub-sample.

Table 4. Ways of coping with unfamiliar words correlated with the IELTS test scores in the EFL sample. N=178.

	Dictionary use	Guess from subject knowledge	Guess from context	Ask teacher	Ask parent	Ask fellow students	Continue reading	Give up reading
r	-.00	-,03	,10	-,20	-,24	-,31	-,11	-,28
p	,99	,67	,17	,01	,00	,00	,15	,00

As is displayed in Table 4, most of these ways of handling unfamiliar vocabulary correlate negatively with the IELTS scores. This is also indicated by multiple regression analysis with these items as an independent and the IELTS scores as the dependent variable, which gives an explained variance of $R^2 = .22$. One possibility is that language problems force the respondents to look up unfamiliar words. Alternatively, it may be that the respondents feel obliged to look up unfamiliar words, and the more they do and interrupt the reading process, the lower they score on the IELTS test. Likewise, the low but positive correlations for guessing from context can be because they have reached a level of language proficiency that allows them to use this strategy. It might also mean that they have learnt to tolerate some uncertainty and vagueness of meaning. Whatever the reasons, it is interesting to note that all activities that disrupt the reading process in order to ascertain the meaning of unfamiliar words have negative correlations.

Discussion

The findings presented above show that the assumption that Norwegian upper secondary EFL instruction adequately prepares for academic reading in higher

education is highly questionable. It is unacceptable that 66% of the EFL sub-sample, including respondents from seven different schools in a country with a reputation for high levels of English proficiency do not attain the equivalent of a Band 6 score on the IELTS Academic Reading Module. It is just as unacceptable that additional EFL instruction which results in a completed Advanced English Course, does not improve IELTS scores while a less time consuming, sheltered CLIL class does. The mean IELTS score for the CLIL sub-sample was $\bar{X}=28$ compared to $\bar{X}=21$ for EFL sub-sample.

The study also shows that respondents in the EFL sub-sample have read very little beyond the bare minimum required by the current EFL syllabus. Closer analysis of the IELTS test papers also indicates that all too many used a counterproductive strategy of careful reading for detail which is typical of textbook reading in EFL instruction. Third, the reading problems seem to be further exacerbated by poor word handling strategies. All too few seem willing or able to guess the meaning of unfamiliar words from context or they simply ignore them and continue reading. For some though, it cannot be ruled out that many simply lack the vocabulary needed for fluent reading. In sum, for this sample the findings are a clear indication of quality problems in Norwegian EFL instruction with regard to the development of vocabulary and reading proficiency.

Validity

Before continuing, it first is necessary to address the issue of the validity of these findings. One crucial issue is external validity, i.e. whether the findings are valid for Norwegian EFL instruction in general. The other is the construct validity of the IELTS Academic Reading Module, i.e. whether the test scores provide an accurate picture of the respondents' reading proficiency. Last I will touch upon the issue of the predicative validity of the IELTS test.

External validity depends on how representative the sample, in particular the EFL sub-sample, is of Norwegian senior upper secondary level students in general. The 178 respondents come from seven different schools in as many parts of the country and the selection process focused first and foremost on finding willing participants. Since, as mentioned above, there is reason to believe that the participating schools and classes had better than average pupils, it means that the scores from the resulting convenience sample are most probably skewed positively. That is to say that the IELTS test scores of the EFL sub-sample are probably higher than would have been the case with a representative sample. However, since these "higher" scores are still disturbingly low, I would argue that my findings present a useful picture of the quality of Norwegian EFL instruction with regard to academic reading proficiency. It could also be mentioned that another survey with respondents in higher education confirms that the type of academic English reading problems found persist at the college and university levels (Hellekjær 2005).

The next issue is the construct validity of the IELTS test used, i.e. whether the test scores give a valid “picture” of academic English reading proficiency. Clapham (1996) has outlined many of the considerations and discussions underlying the development of the IELTS tests (see also Fulcher 1999). In the following I will limit the discussion to content validity, one of six aspects of construct validity in Messick’s unified framework for validation, namely content, substantive, structural, transfer, external and consequential validity. These aspects “*function as general validity criteria and standards for all educational and psychological measurement*” (Messick 1996 p. 248).

Content validity implies that the knowledge, the skills, and the other factors assessed are relevant to the construct domain in question. One aspect of this is the extent to which the tasks selected “*sample domain processes in terms of their functional importance*” (Messick 1996 p. 249). Examples here would be the IELTS test’s selection of text topics and whether the test items sample student reading in a manner that reflects “real world” academic reading. I would argue that the three texts on different topics used in the test are at a level of difficulty comparable to or just below that of undergraduate level English textbooks. Furthermore, the items tap reading proficiency in a way that reflects academic reading. The time limit in particular makes it possible to get an impression of the respondents’ reading speed and fluency as well as their ability to adjust how they read to reading purpose, since failure to do so would result in many unanswered items.

Next, a number of studies of predicative validity, i.e. whether the IELTS test scores give an accurate prediction of academic success, have been carried out (see Feast 2002; Fulcher 1999; Lee & Greene 2007). The consensus is that high test scores, be they from IELTS or other EFL/ESL placement tests, are

. . . a necessary, albeit insufficient, prerequisite for graduate school success (Sharon, 1972). However, there is little agreement about the exact nature of the relationship between English-language proficiency and successful academic performance.

(Lee & Greene 2007 p. 366)

In other words, although high IELTS scores do not guarantee success as a student, they are, nevertheless, a prerequisite. What there does seem to be some uncertainty about in these studies is whether Band 6.0 on the IELTS test, which is used as a cut-off point in this study, is an adequate minimum requirement for admission purposes (Feast 2002). If that is not the case, it makes the findings of the present study even more worrisome.

To conclude on the issue of validity, although these findings should, of course, be checked by follow-up studies, I would argue that they give a sufficiently useful picture of Norwegian upper secondary students reading

proficiency to argue for comprehensive changes in Norwegian EFL instruction. It also suggests avenues of further research.

Implications for Norwegian EFL teaching

The main conclusion that can be drawn from this study is that the current Norwegian complacency about the quality of upper secondary EFL instruction as preparation for higher education, or for occupational purposes, is unmerited (see also Lehmann 1999). For institutions of higher education it means that these institutions need to offer their own English courses to prepare their students for their future careers (see Lehmann 1999; Hellekjær 2007a). For upper secondary EFL instruction, or to be more exact, for Norwegian EFL instruction in general, it means serious changes in teaching practices and learning objectives as well as in examinations and testing.

As for EFL instruction at the lower as well as upper secondary school level, it is, above all, necessary to give far higher priority to reading outside the perennial EFL textbooks and to instruction in reading and learning strategies (Hellekjær 2007b). Second, I would argue that these changes must be supported by the systematic testing of academic English reading proficiency at the upper secondary school level by means of specially designed reading tests. Otherwise, the washback effect from the more traditional oral and written examinations will counteract any change of emphasis in favor of reading in EFL instruction and preserve the status quo. Last, but not least, the systematic use of CLIL instruction for all upper secondary students should be considered, for at least two reasons.

First, this study has shown that a sheltered CLIL class in just a single subject can be quite effective in developing reading proficiency (see also Hellekjær 1996). Second, it has shown that traditional EFL instruction is not effective in this respect. Therefore, unless traditional EFL instruction is improved in both quality and outcome, serious consideration should be given to replacing advanced English courses with CLIL instruction. An alternative solution would be to combine systematically EFL classes with one of more CLIL classes, also known as adjunct-instruction (Brinton, Snow, & Wesche 1989).

In conclusion

This study has shown the need for further research on the quality and outcomes of Norwegian EL instruction. First, there is a need to test a larger and more representative sample from upper-secondary school using the IELTS test or by means of a comparable reading test. Next, there is a need for causal studies examining the effects of CLIL instruction and extensive reading on academic English reading proficiency. Last, but by no means least, there is the need to find out more about EFL students' reading and word handling strategies at different levels of the educational system.

The last point is whether the findings presented in this article should be of interest outside the borders of Norway. I would argue that, to the extent EFL instruction elsewhere is heavily textbook dependent, to the extent there is little emphasis on extensive reading, to the extent vocabulary development is not focused upon, and to the extent systematic instruction in reading and learning strategies is neglected, the findings of this study should be of great interest for other countries as well. In fact, Norwegian 16-year-olds did quite well in a comparative study of English proficiency in eight European countries (Bonnet 2004). This could also imply that the level of academic English reading proficiency displayed by the 217 Norwegian respondents in the present study is actually better than is the case elsewhere.

References

- Alderson, C. J. 2000. *Assessing Reading*. Cambridge: Cambridge University Press.
- 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(1), 15-34.
- Bonnet, G. (Ed.). 2004. *The assessment of pupil's skills in English in eight European countries 2002: The European network of policy makers for the evaluation of educational systems*.
- Bråten, I. 2007. Leseforståelse - komponenter, vansker og tiltak. In *Leseforståelse* (pp. 45-81). Oslo: Cappelen.
- Bråten, I., & Olaussen, B. S. 1997. Lærings- og studiestrategier i høgre utdanning. *Norsk Pedagogisk Tidsskrift*, 5, 306 - 318.
- Brinton, D. M., Snow, M. A., & Wesche, M. B. 1989. *Content-Based Second Language Instruction*. Boston: Heinle and Heinle.
- Carrell, P. L. 1991. Second Language Reading: Reading Ability or Language Proficiency? *Applied Linguistics*, 12(2), 159-179.
- Clapham, C. 1996. *The Development of IELTS: A Study of the Effect of Background Knowledge on Reading Comprehension*. Cambridge: Cambridge University Press.
- Dahl, J. 1998. *Pensumlitteratur i høyere utdanning - en undersøkelse av forholdet mellom norsk og engelskpråklig pensumlitteratur i fire grunnfag* (No. 14). Oslo: Norsk institutt for studier av forskning og utdanning.
- Feast, V. 2002. The Impact of IELTS Scores on Performance at University. *International Education Journal*, 3(4), 70-85.
- Fjeldbraaten, A.-L. 1999. Undervisning i lærings- og studiestrategier i sammenheng med allmennlærerutdanningens pedagogikkundervisning. In I. Bråten & B. S. Olaussen (Eds.), *Strategisk læring* (pp. 122-138). Oslo: Cappelen.
- Fulcher, G. 1999. Assessment in English for Academic Purposes: Putting Content Validity in Its Place. *Applied Linguistics*, 20(2), 221-236.
- Grabe, W. 1988. Reassessing the word interactive. In P. L. Carrell, J. Devine & D. E. Eskey (Eds.), *Interactive Approaches to Second Language Reading* (pp. 56-70). Cambridge: Cambridge University press.

- Grabe, W. 1999. Developments in reading research and their implications for computer adaptive-reading assessment. In M. Chalhoub-Deville (Ed.), *Issues in computer-adaptive testing of reading proficiency* (Vol. 10, pp. 11-47). Cambridge: Cambridge University Press.
- Grabe, W., & Stoller, F. L. 2002. *Teaching and Researching Reading*. London: Longman.
- Hatlevik, I. K. R., & Norgård, J. D. 2001. *Myter og fakta om språk* (No. 5). Oslo: NIFU . Norsk institutt for studier av forskning og utdanning.
- Hellekjær, G. O. 1996. Easy does it: Introducing Pupils to Bilingual Instruction. *Språk og Språkundervisning* (3), 9-14.
- Hellekjær, G. O. 2005. *The Acid Test: Does Upper Secondary EFL Instruction Effectively Prepare Norwegian Students for the Reading of English Textbooks at Colleges and Universities*. University of Oslo, Oslo.
- Hellekjær, G. O. 2007a. *Fremmedspråk i norsk næringsliv - engelsk er ikke nok!* (Vol. 3). Halden: Fremmedspråksenteret.
- Hellekjær, G. O. 2007b. Reading: From a Forgotten to a Basic Skill. *Språk og Språkundervisning* (2), 23-29.
- Ibsen, E., & Lie, S. 1990. *Språk og struktur: En undersøkelse om valg av fremmedspråk*. Oslo: Senter for lærerutdanning og skoletjeneste, University of Oslo.
- Koda, K. 2004. *Insights into Second Language Reading*. Cambridge: Cambridge University Press.
- Laufer, B. 1997. The lexical plight in second language reading. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition* (pp. 20-34). Cambridge: Cambridge University Press.
- Lee, Y.-J., & Greene, J. 2007. The Predictive Validity of an ESL Placement Test: A Mixed Methods Approach. *Journal of Mixed Methods Research*, 1(4), 366-389.
- Lehmann, T. M. 1999. *Literacy and the Tertiary Student: Why has the Communicative Approach Failed?* Unpublished Dr. Philos., Department of English, University of Bergen.
- Messick, S. 1996. Validity and washback in language testing. *Language Testing*, 13, 241-256.
- Rayner, K., & Pollatsek, A. 1989. *The Psychology of Reading*. Englewood Cliffs: Prentice-Hall.
- Ro, A. 2006. Leseopplæring og lesestrategier. In E. Elstad & A. Turmoe (Eds.), *Læringstrategier*, (pp. 67-92). Oslo: Universitetsforlaget.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- Stanovich, K. E. 1980. Towards an interactive compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16, 32-71.
- Svenhardt, B. W., Servant, K., Hellekjær, G. O., & Bøhn, H. 2007. Norway. In A. Maljers, D. Marsh & D. Wolff (Eds.), *Windows on CLIL* (pp. 139-146). The Hague: European Platform for Dutch Education.
- Sylvén, L. K. 2004. *Teaching in English or English Teaching?* Ph. D, University of Gothenburg, Gothenburg.
- UCLES. 2001a. *IELTS Specimen Materials Handbook*. Cambridge: UCLES, The British Council, IPD Education Australia.
- UCLES. 2001b. *IELTS Specimen Materials Test Texts*. Cambridge: UCLES, The British Council, IPD Education Australia.
- Urquhart, S., & Weir, C. 1998. *Reading in a Second Language: Process, Product and Practice*. London: Longman.