*Glenn Ole Hellekjær**

Lecture Comprehension in English-Medium Higher Education

Abstract

In European higher education the growing number of English-Medium (EM) courses, i.e. non-language subjects taught through English, has led to discussion about, and research on, whether the use of a foreign language for instruction has a negative impact on teaching and learning. The present quantitative study investigates this issue by comparing student lecture comprehension in English and the first language (L1) at three Norwegian and two German institutions of higher education, with a sample comprising 364 Norwegian and 47 German student respondents. It compares self-assessment scores for lecture comprehension in English and the L1. Analysis shows that while the difference between English and L1 scores was not substantial, a considerable number of students still had difficulties understanding the English-Medium lectures. Among the main problems, which in fact were similar in English and the L1, were difficulties distinguishing the meaning of words, unfamiliar vocabulary, and difficulties taking notes while listening to lectures. The study argues the need to improve the quality of lecturing in English and L1 as well as the lecturers' and students' English proficiency.

1. Introduction

The 1999 Bologna Declaration on the reform and convergence of European higher education led to a dramatic increase in the number of English-Medium (EM) courses and programs at European colleges and universities (Maiworm/Wächter 2008). English-Medium instruction is when non-language courses in for instance medicine, physics or political science are taught in English, to students for whom it is a foreign language. As often as not, it is also taught by a lecturer who does not have English as a first language (L1). The main driving forces behind this expansion have been the need to offer courses to international students, co-operation schemes between universities, and the hiring of foreign lecturers. They may also come about due to "English from below" initiatives. Ljosland (2008) for instance, presents a Norwegian case study where a Department of Engineering voluntarily changed to teaching in English because they felt theirs was an international subject, because they hoped the change would attract international students, and above all, because they felt that it would be important for study quality.

However, Ljosland's (2008) and other Nordic studies have shown that the quality of EM courses has suffered from their being set up without due consideration for the practical and pedagogical implications of teaching and learning in a foreign language, English (Hellekjær 2007, 2010, Hellekjær/Westergaard 2003, Halvorsen/Faye 2006, Ljosland 2008, Räsänen 2000, Tella et al. 1999). One reason might be that it is taken for granted that lecturers will not have appreciable difficulties teaching in a foreign language. Another assumption is that Norwegian, Danish, Swedish or Finnish students have few difficulties understanding lectures, taking part in discussions and seminars, and not to mention writing papers and taking examinations in English.

Sometimes even the students share the assumption that changing the language of instruction to English is entirely unproblematic. In a qualitative study of 23 Physics students conducted by Airey/Linder (2006: 555), it was found that "the students initially report no difference in their experience of learning of physics when taught in Swedish or English." This was, however, contra-

*   Glenn Ole Hellekjær
    University of Oslo
    Department of Teacher Education and School Research
    PO. Box 1099 Blindern
    N-0317 Oslo
    g.o.hellekjar@ils.uio.no

dicted by videos of the students during lectures, and "the students' own accounts of their learning experience during stimulated recall. These indicate a number of problems related to learning in English rather than Swedish" (Airey/Linder 2006: 555). One example is the students' difficulties taking notes during lectures, another their reluctance or inability to ask and answer questions in English. Airey and Linder also found that students developed compensatory strategies. One was "that a number of students, though silent in the lecture, came forward at the end of each session to ask questions" (Airey 2009: 79). Another was that they engaged in additional preparatory reading and relied on follow-up reading and discussions to ensure comprehension.

With regard to teaching, the assumption that lecturers with another L1 are able to switch to teaching in English without undue difficulty has also been contradicted. Hellekjær/Westergaard (2003) found that the number of linguistically capable lecturers was limited, and that too rapid an expansion in the number of EM courses invariably led to student complaints and other quality issues. Two Dutch Ph.D. theses, Klaassen (2001) and Vincke (1995), and a smaller study by Hellekjær (2007) examined this issue in further detail. They found that EM lecturers felt constrained when teaching in English, that is to say that they found that they lacked the ability to vary their language or explain in different ways or from different perspectives. Many also experienced language difficulties in less formal situations such as group discussions and conversations. Furthermore, even highly proficient lecturers managed to cover less ground in English lectures than they were able to in their L1 (Vinke 1995). Moreover, it was found that the overall quality of the lecture with regard to structure, pacing, the use of visual aids, meta-discursive comments and taking the audience into consideration, could be more important for comprehension than the lecturers' or students' language proficiency. Indeed, Klaassen (2001) found that such "effective lecturing behavior" could have a greater effect on how Dutch students experienced lectures than did the language used.

To sum up, a number of studies have found that many students in EM courses have difficulties understanding and learning from lectures in English. These difficulties may be due to language difficulties among students and lecturers on the one hand, and to the quality of the lectures on the other. However, given the implicit assumptions about students' and lecturers' English proficiency, effectively arguing for change will require hard data. Providing such data is the goal of the present quantitative study of EM lecture comprehension comprising 391 respondents from three Norwegian institutions of higher education (Hellekjær 2010) and 47 from two German universities.

The main aim of the study is therefore to examine whether, and to what extent, EM students actually experience lecture comprehension problems in English compared to in their L1. The second is to identify the main variables that covary with lecture comprehension, and use the findings to suggest what can be done to improve EM lecture comprehension.

1.1. Lecture comprehension in English

In the following I will give a brief description of the phenomenon studied, i.e. academic English listening comprehension, also known as a construct definition (Alderson 2000: 118). When doing so I draw mainly upon Vandergrift (2007) and Buck (2001).

Listening comprehension is an interactive process combining bottom-up and top-down processing. In bottom-up processing listeners build meaning by combining information from phoneme, word and discourse-level features. Vocabulary knowledge and word-segmentation skills – the ability to find word boundaries in spoken discourse – are particularly important at this level. In the top-down processing the listener uses context and/or general or prior knowledge of the topic to build, check or repair understanding. An example would be using knowledge of the topic in question to infer the meaning of unfamiliar words. This is also an example of the interaction between processing levels, of how the lack of comprehension at the word/bottom-up level can be compensated for by using top-down "compensatory mechanisms – contextual, visual or paralinguistic information, world knowledge, cultural information and common sense" (Vandergrift 2007: 193).

The ability to compensate, however, is limited by the transient nature of the working memory. This means that interruptions longer than about 30 seconds to ponder the meaning of unfamiliar words or expressions can lead to the listener losing track of what is being talked about (Buck 2001, Rayner/Pollatsek 1989).

To return to the interaction between processing levels during the listening process, the bottom-up and top-down processes function in parallel, with the focus depending on the "purpose for listening, learner characteristics such as the level of language proficiency, and the context of the listening event" (Vandergrift 2007: 193). Variations in second language (L2) listening proficiency will therefore depend on language proficiency on the one hand, and strategy use on the other. Less proficient listeners, for instance, often focus on the word level cues to build understanding, and some even attempt to translate what they hear to their first language. In contrast, more skilled and/or more linguistically proficient listeners will rely more on top-down compensatory strategies to infer what is not immediately understood. In addition, they will to a greater extent use cognitive and meta-cognitive strategies such as comprehension monitoring and elaboration to repair or enhance comprehension.

This brings us to the question of what to focus on when assessing listening proficiency, linguistic knowledge or strategy use. Buck (2001) argues that although comprehension depends on the combination of strategic knowledge and language proficiency, variation in performance most often correlates with linguistic competence. He therefore argues that it "makes more sense to put the emphasis on testing language competence rather than strategic competence" (Buck 2001: 105). Buck (2001: 114) also proposes a default listening construct that can be adapted to different contexts. It describes listening as the ability to:

> 1) process extended samples of realistic spoken language, automatically and in real time; 2) understand the linguistic information that is unequivocally included in the text; and, 3) make whatever inferences are unambiguously implicated by the content of the passage

With this construct definition as a point of departure, the questionnaire in the present study uses a combination of items to tap into low-level linguistic processing on the one hand, and the outcomes of this and higher-level processes such as content understanding on the other. The context is academic lectures, and the respondents' actual experience is with real-time lectures in English and the L1.

2.   Method

The present quantitative study uses a quasi-experimental, one-group, post-test research design, in this case a single questionnaire (see Shadish et al. 2002: 106-107). This design precludes making hypotheses about causal relations, although I might suggest such relations in the analysis and discussion.

The survey in Norway (Hellekjær 2010) and Germany took place in the 2008 spring and fall semesters. In Norway, websites were used to identify undergraduate level EM courses at different faculties at the University of Oslo, BI Norwegian School of Management, and Oslo University College. While the original plan was to limit the survey to the University of Oslo only, the Norwegian School of Management was included because it had an EM three-year Bachelor of Business Administration, and Oslo University College an EM engineering program. The German universities volunteered to take part in the survey. It turned out that the available EM courses were Master's level only, and that the number of possible respondents was quite limited at both universities, which resulted in a smaller sample than anticipated. The two German universities have asked to remain anonymous.

The procedure was to contact the lecturers by phone or e-mail to ask for permission to survey their students. To ensure a reasonably high reply rate, paper questionnaires were handed out during lectures. Table 1 provides an overview of the sample.

Table 1. Overview of the sample according to institution, faculty and level of study

Faculty and institutions	Undergraduate level	Graduate level	Total
Faculty of Mathematics and Natural Sciences, University of Oslo	123		123 (28%)
Faculty of Humanities, University of Oslo	70	32	102 (23%)
Faculty of Medicine, University of Oslo	57		57 (13%)
Faculty of Social Sciences, University of Oslo	49		49 (11%)
Bachelor of Business Administration, BI Norwegian School of Management	47		47 (11%)
Faculty of Engineering, Oslo University College		13	13 (3%)
German university 1		24	24 (5%)
German university 2 (Faculty of Engineering)		23	23 (5%)
Total	346 (79%)	92 (21%)	438 (100%)

In the statistical analysis I treat the 47 German respondents and 391 Norwegians as separate samples, and percentages are from the sample in question. As can be seen, the majority of the Norwegian respondents (79%) are undergraduate level students and the remaining (21%) graduate level. The German respondents are all graduate level. In Norway 227 (57.5%) were female and 166 (42.5%) male, in Germany there were 29 (62%) females, 17 (36%) males, and 1 missing answer. Next, there were 267 (68%) respondents in Norway who had Norwegian as their L1, 27 (7%) who had English, and 97 (25%) had another L1. In Germany 36 (77%) had German as their L1, 11 (23%) had another language than English as their L1, and none had English as their L1. In the analysis the 27 respondents in the Norwegian subsample who have English as their L1 are excluded, which in practice gives a Norwegian subsample of 364. There were, as mentioned, no L1 English speakers in the German subsample, so all are included in the analysis. This gives 97 (27%) with a different L1 in the Norwegian sample, in the German 11 (23%).

Neither the German nor the Norwegian samples meet the requirements of a representative sample, but these are generally very difficult to attain in education-based research. The Norwegian sample, however, can be described as a purposive sample of typical instances, which entails some limits with regard to external validity (Ary et al. 1996; Shadish et al. 2002). I would, nevertheless, argue that an effective sample of 364 respondents from 14 different courses at three Norwegian institutions of higher education provides reasonably useful data about EM students in Norway, where the population is about 4.8 million. The German sample, however, comprising 47 respondents from three courses at two univerisities in a country with about 82 million inhabitants is on the small side, and the data must therefore be interpreted with caution.

2.1. The questionnaire

The questionnaire (see Appendix) was in English due to the many exchange students and comprises 61 multiple choice and two open-ended items. It was designed to take about ten to 15 minutes to fill in at the end of the lectures so that lecturers would allow it to be handed out during lectures, and the need for speed led to the use of self-assessment items to measure comprehension instead of a listening test. There were also items about background variables such as education, prior experience with EM courses, motivation for EM courses, and a number of self-assessment

items. The latter measure lecture comprehension by tapping into different aspects of listening comprehension during lectures. These were on vocabulary, clarity of pronunciation/word segmentation, speaking speed, ability to follow the lecturer's line of thought, the speed of the presentation of information, difficulty in taking notes and finally, content understanding. There were identical items for L1 and EM instruction, all of them using four point Likert scales with 1 indicating a high level of difficulty, 4 no difficulty. The seven items selected for EM instruction are presented below. The wording for the equivalent L1 items is almost identical. It should be mentioned that in the instructions the respondents were asked to answer on the basis of their overall experience with EM and L1 lectures.

48. Indicate on the scale **to what extent do you find words and expressions in the English language** lectures **unfamiliar**.

49. Indicate on the scale to what extent **words and expressions are clearly pronounced and understandable in the English language lectures**.

50. Indicate on the scale to what extent you experience **that the lecturer in English language lectures speaks too fast.**

52. Indicate on the scale to what extent you can **follow the lecturer's line of thought during English** lectures.

53. Indicate on the scale **to what extent you understand the content of the English lectures**.

54. Indicate on the scale to what extent **the information in the English lectures is presented so quickly** that it hinders your understanding.

57. Indicate on the scale **how difficult you find taking notes** during English lectures.

Finally, there were a number of items designed to elicit whether students compensated for the use of English through preparatory reading or asking questions for clarification, as reported by Airey/ Linder (2006). Statistical analysis was carried out using the Statistical Package for the Social Sciences (SPSS). This comprises calculating percentages, mean scores and standard deviations, correlations, Cronbach's alpha coefficients, and factorial analysis.

2.2. Dependent variables

As mentioned, the questionnaire comprised a number of items designed to tap into the listening comprehension construct, with identical items for the L1 and English to allow for comparison between languages.

Testing which items measure, or do not measure, the same underlying trait can be done using Cronbach's alpha or factorial analysis. In the present study I used factorial analysis (principal axis factoring) to identify and select the items that loaded on the same (latent) variables, in this case the language proficiency aspect of lecture comprehension in the L1 and English respectively. The English versions of the seven items that were selected are listed above.

Factorial analysis shows that for the Norwegian subsample the seven English items that loaded on the same variable can explain 52% of the total variance in lecture comprehension scores. Furthermore, the Cronbach's alpha coefficient for the seven items was a fairly high $\alpha= .84$, which also confirms that the items tap into the same variable. For the L1, the explained variance was a lower 43%, while the alpha coefficient was $\alpha=.76$. For the German subsample the numbers for English are the same as in the Norwegian subsample, 52% and $\alpha= .84$. For the L1, however, both are markedly higher, 61% and $\alpha= .89$ respectively, meaning that these items have a higher loading on the same latent variable. Closer analysis shows that this difference might be due to the items in the Norwegian subsample loading on a second, but marginal latent variable. For the German sample, however, the loading fell just below the limit of one set for additional latent variables, giving a higher explained variance.

Since analysis confirmed that each of these seven items measures a different facet of the same trait, they could be merged, without weighting, into additive indices and used as measures of lecture comprehension. In the following these are designated *L1Index* for the L1, and *EngIndex* for English. Using additive indices simplifies analysis by making it possible to use one instead of several items as indicators of the same underlying trait. It also helps reduce the effects of possible measurement errors and improves both validity and reliability (Hellevik 1999: 303-310). In the following analysis these indices are used as dependent variables for listening comprehension in the L1 and in English. The items are also examined separately.

Airey and Linder's (2006) study provides an example of student self-assessment at a very general level being in error. However, the present study uses a different form of measurement; a set of items tapping into different aspects of lecture comprehension with focus on specific tasks, about which Bachman (1990: 148) claims that:

> self-relating questions that ask test takers to judge how difficult various aspects of language use are for them appear to be better indicators of specific language abilities than are questions that ask how well they can use various aspects of language.

In other words, it is possible that Airey and Linder's (2006) findings can be explained by how they asked students about the use of English, because other research indicates that self-assessment in general gives reasonably valid information in low stakes situations such as in this study (Bachman 1990, Oscarson 1997). Furthermore, in their validation studies of the self-assessment of language proficiency, Marian et al. (2007) and Ross (1998) also found self-assessment to be reliable predicators of listening proficiency. There is therefore good reason to expect the scores from the self-assessment items and indices used in the present study to provide a useful and valid comparison of student lecture comprehension difficulties in English and the L1, that is to say for the samples in the present study. A firmer conclusion, however, would require a separate validation study in which self-assessment item scores used are correlated against a relevant listening test.

3. Results and analysis

The tabels and figures used below, now including German data, are modified from Hellekjær (2010).

3.1. Lecture comprehension

To start with the indices, on a scale from one indicating high levels of difficulty and four no difficulties, the analysis found a clear, but not dramatic difference between the mean scores. For the Norwegian sample the L1 index mean score was 3.4 (SD=. 6) and a lower 3.1 (SD=. 5) for EngIndex. For the German sample, however, these were a lower 3.0 (SD=. 7) for L1Index and 2.8 (SD=. 6) for EngIndex.

For the Norwegian sample this indicates that most of the respondents do not have serious difficulties with EM instruction. For the German sample, however, the scores indicate higher levels of lecture comprehension difficulty, in the L1 as well as in English. This is despite all of these being master level students, compared to the majority of undergraduate level students in the Norwegian sample. The distribution of the scores is presented in Figures 1 and 2 below.
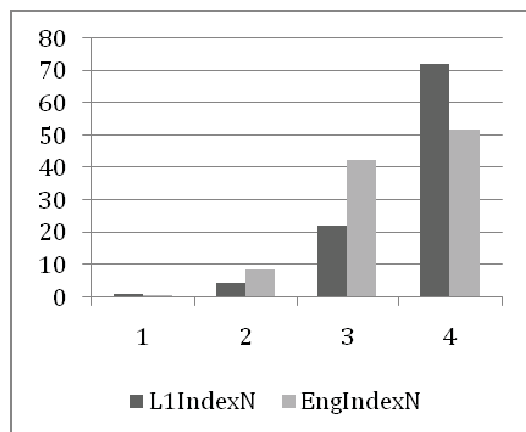
Figure 1. The Norwegian sample. Distribution of the lecture comprehension scores for the L1 (L1Index) and English (EngIndex) in percent. The y-axis represents the students involved in the study, and the x-axis represents the 4 points on the Likert scale. N= 364
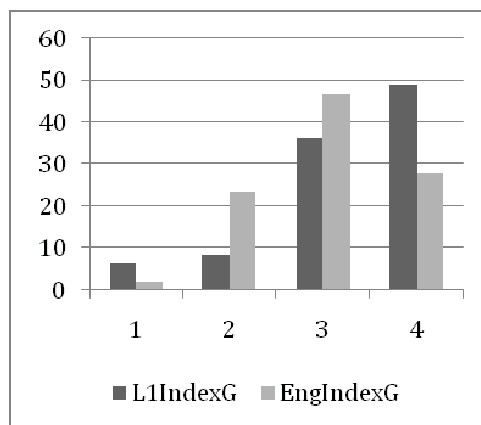
Figure 2. The German sample. Distribution of the lecture comprehension scores for the L1 (L1IndexG) and English (EngIndexG) in percent. The y-axis represents the students involved in the study, and the x-axis represents the 4 points on the Likert scale. N= 47

The score distribution for the Norwegian sample shows that for the L1 the distribution is skewed in favor of high scores, although 24% have some difficulties as indicated by a score of 3 or below. For English, however, scores indicate a higher level of difficulty, with 42% of the respondents scoring three or below. In the German sample, however, the lecture comprehension scores are lower overall. As many as 72% have a score of 3 or lower for English, compared to 44% with the L1. It should be kept in mind that the L1 lectures can either be in Norwegian, German, or in the L1s of the many exchange students among the respondents.

The next issue is what aspects of the lectures the respondents find most difficult. In Table 2 the mean scores and standard deviations for the seven items included in the indices are displayed separately.

Table 2. A comparison of the mean scores and standard deviation for the main items tapping into lecture comprehension in the L1 and English for the German (N=47) and Norwegian (N=364) samples

Items and item numbers	Norwegian sample N = 364				German sample N = 47			
	L1		English		L1		English	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
To what extent do you find words and expressions unfamiliar? (items 35, 48)	3.3	.8	3.0	.6	3.2	.8	2.7	.6
To what extent are words and expressions clearly pronounced and understandable? (items 36, 49)	3.4	.7	2.9	.7	3.1	.8	2.8	.6
To what extent does the lecturer speak too fast? (items 37, 50)	3.6	.7	3.4	.7	3.4	.7	3.1	.7
To what extent can you follow the lecturer's line of thought? (items 39, 52)	3.4	.6	3.1	.7	3.1	.8	2.9	.7
To what extent do you understand the content of the lectures? (items 40, 53)	3.5	.6	3.3	.6	3.2	.6	2.9	.6
To what extent is the information in the lectures presented so quickly that it hinders your understanding? (items 41, 54)	3.4	.7	3.3	.7	2.9	.7	2.8	.8
How difficult do you find taking notes during English lectures? (items 44, 57)	3.4	.8	3.1	.8	3.0	.7	2.7	.7

As can be seen, there are lower scores for English for all seven items in both samples. In Norway the largest difference between English and LI lectures is for words and expressions not being intelligible, which might be due to unclear pronunciation and/or word segmentation problems. In Germany, however, unfamiliar words and expressions are the main difficulty, followed by difficulties with unclear pronunciation. In both samples these are followed by problems following the lecturers' line of thought, and by difficulties taking notes. Last, higher scores and smaller differences between languages indicate that speaking speed, how fast information is presented, as well as content understanding, are less problematic areas in the Norwegian sample, but more so in the German sample.

It is interesting to note that student difficulties with L1 lectures in both countries are much the same as with English, although to a somewhat lesser extent. It is also probable that the respondents' difficulties with subject specific terminology and concepts in their L1 as well as English illustrate that learning these goes together with the learning of the subjects.

Another indication of lecture comprehension difficulties is to which extent the respondents feel they need to ask questions about language or content (Airey/Linder 2006, Hellekjær 2007). Items about the need to ask questions were therefore included in the questionnaire, and Table 3 displays the correlations between L1Index and EngIndex and the items about the need to ask questions about language and content.

Table 3. Comparison of the correlations between L1Index, EngIndex and items tapping into the need to ask questions about language or content in the Norwegian and German samples

Items	Norwegian sample N= 364		German sample N=47	
	L1Index	EngIndex	L1Index	EngIndex
How often do you want to ask about unfamiliar words and expressions during lectures?	-.38**	-.47**	-.62**	-.51**
How often do you want to ask about unclear content during lectures?	-.48**	-.46**	-.58**	-.47**

** Correlation is significant at the .01 level (2-tailed).

The table displays fairly strong, negative correlations (Pearson's r) between the items about the need to ask questions with L1Index and EngIndex respectively. These negative correlations show that the lower the respondent's lecture comprehension scores on L1Index and EngIndex, the greater their need to ask questions about language or content as indicated by their answers to the items about their felt need to ask questions about language or content. What is particularly interesting is that there are fairly high negative correlations for the L1 as well as for English. Indeed, in the German sample the correlations for the L1 are actually the highest. These correlations confirm the need to pay greater attention to language issues, more specifically key concepts and subject specific vocabulary and terms, in L1 as well as in EM lectures. They also support Airey and Linder's (2006) claims about the importance of making room for clarification questions in EM instruction, and show that this would be useful and important in L1 lectures as well. It also shows that when studying the role of English in EM context, it is important to also consider (compare) the L1 context, as is done in the present study. This is because if one only investigates L2 English lecture comprehension and assumes that lecture comprehension in the L1 is more or less perfect, this could give quite inaccurate conclusions.

An additional sign of language problems, or general difficulties with EM instruction, is how laborious EM instruction is compared to in the L1. This is an issue that involves more than lecture

comprehension, and should be of interest for course design as well as for teaching. The respondents' answers are presented in Table 4 below.

Table 4. How much work is attending a course in English compared to in the L1?

How much work do you find the EM course compared to in the L1 ?	Just like in the L1	2	3	Much more work than in the L1	N
Norwegian sample	35%	24%	27%	12%	352
German sample	17%	30%	30%	12%	46

As can be seen, 63% of the Norwegian respondents find EM courses more laborious than they do courses in their L1, compared to 72% in Germany. Furthermore, in both countries this item correlates negatively with EngIndex as a dependent variable with r= -.4 (p<.01, N=352) in Norway and a somewhat higher r= -.47 (p<.01, N=46) in Germany. In other words, these correlations show that the lower their EngIndex scores are, the more laborious respondents find EM instruction.

Yet another indication of language difficulties is the extent to which the respondents are dependent on visual aids to support comprehension. Table 5 below displays the answers to the items about the importance of visual slides for lecture comprehension in the L1 and in English.

Table 5. How important are the lecturer's transparencies/PowerPoint slides or other visual aids for your understanding of the lectures?

Sample	Lecture language	Not important for understanding	2	3	Very important for understanding
Norwegian sample N=364	L1 lectures	18%	32%	32%	18%
	English lectures	13%	25%	30%	31%
German sample N=47	L1 lectures	9%	18%	54%	18%
	English lectures	7%	16%	36%	42%

It should be kept in mind that some of the variation could depend on the discipline and topic in question, for instance if a complex system is explained in a graph, the visuals may be more important than the language. However, these answers show that irrespective of language, the majority of the respondents find visual aids important, some even very important, for lecture comprehension. Furthermore, when the importance of visual aids for understanding in the L1 was correlated with L1Index, this gave a negative correlation of r= -.26 (p<.01, N= 343), compared to r= -.32 (p<.01, N= 355) with EngIndex. For the German sample the correlation with L1 index is r= -.32 (p<.01, N=47) and with EngIndex r= -.58 (p<.01, N=47). In other words the greater the difficulties the students have understanding the lectures, the more they depend on visual aids to support comprehension. Furthermore, this is the case in L1 as well as in EM lectures.

To sum up, the analysis of different aspects of lecture comprehension difficulties so far shows that many Norwegian and most of the German respondents in the present study experience difficulties with EM lectures, and that these are largely due to language difficulties. It also shows that many of the problems evident in EM lectures, such as difficulties with words and expressions and dependence on notes for comprehension, are problematic in the L1 lectures as well. Indeed, the similarity of problems across languages is one of the more interesting findings in the present

study. In the following analysis I will examine a number of background variables to see whether these can explain the findings presented above, among these the differences between Norwegian and German scores.

3.2. Background variables

In a previous study based upon interviews of Norwegian EM lecturers (Hellekjær 2007), I found that lecturers did not believe that the Norwegian students had serious difficulties due to language difficulties. Instead, they were far more skeptical about the language proficiency of the exchange students – even though these are screened using TOEFL or IELTS admission tests. In this study I therefore used the independent sample T-test to examine whether there were differences in the L1Index and EngIndex scores between students with German or Norwegian as their L1 and those with a different L1, mostly exchange students. In the Norwegian sample 97 (27%) had a different L1, in the German 11 (23%). The T-test compares the difference in the scores between groups, and checks whether these are statistically significant. Keep in mind that the low number of German respondents means a lower probability of significant results. The comparisons are displayed in Table 6.

Table 6. Comparison of T-test scores of L1Index and EngIndex according to the respondents' L1

Indices and sample	L1	N	Mean scores	Std. Deviation
L1Index German	German	36	3.2*	.46
	Other	11	2.5*	.92
EngIndex German	German	36	2.9	.54
	Other	11	2.5	.64
L1Index Norwegian	Norwegian	267	3.5*	.42
	Other	97	3.1*	.86
EngIndex Norwegian	Norwegian	267	3.1*	.47
	Other	97	3.2*	.65

* The difference is significant at the 0.05 level

The first point that can be made in this comparison is that in the Norwegian sample there is almost no difference between students with Norwegian as their L1 and those with a different L1 with regard to lecture comprehension in English. This shows that the Norwegian lecturers' impression of the exchange students' poor English proficiency is a questionable one. In the German sample, however, there is a clear difference for English, in this case favoring those with German as their L1. For lecture comprehension in the L1, however, in both samples the respondents with a different L1 than Norwegian or German have lower scores for lecture comprehension in the L1 – although the low number of respondents in the German sample should be kept in mind. An alternative explanation is cultural differences in self-assessment scores. Blue (1994: 15) mentions a previous study (Blue 1988) showing that "nationality can be an important factor in self-assessment, with some nationalities having a tendency to overestimate their level and others tending towards underestimation." In other words, any conclusions about the lower scores for the German respondents, and not to mention international students, should be made with great caution pending a validation study in which national differences are also investigated as well as an international study with a larger sample.

3.2.1. Motivation

Two of the factors that were expected to correlate positively with higher levels of lecture comprehension were exposure to English, for instance through reading or the media, and student motivation with regard to the utility of English. Examples of the latter would be that they believed that English skills would be useful in future careers, or if they wished to work in an English speaking country. Table 7 presents the correlations between items for motivation and EngIndex.

Table 7. Correlations between motivational factors and EM lecture comprehension scores (EngIndex) for the Norwegian and German samples

Questions	Correlation, Norwegian sample	Correlation, German sample
How useful do you believe knowing English will be in your future career?	.23**	.05
How interested are you in working outside your own country in your future career?	.16**	.24
Are you interested in working in a job where English is your working language?	.17**	.36*
Do you think knowing English will be important for new jobs?	.12*	-.04
Is the extra work involved in taking an EM course worthwhile?	.21**	.06

** Correlation is significant at the 0.01 level
* Correlation is significant at the 0.05 level

The fairly low, but positive correlations between items for motivation and EngIndex might for the Norwegian sample be explained by the fact that the most internationally motivated respondents are also those who are most proficient in English. The other point is that their belief in the utility of English for future careers or interest in working abroad might stimulate their efforts to master EM lectures. The latter might well be indicated by the positive correlation for the extra work involved in taking an EM course being considered worthwhile. While the lack of significant correlations for the German sample can be due to the low sample size, it is, nevertheless, interesting to note that these students find English less important for future careers than do their Norwegian counterparts. The exception is their interest in working abroad or in a job where English is a working language. It would seem reasonable to assume that this reflects that English is less important in the larger German language community than it is in the Norwegian.

3.2.2. Exposure to English

Another explanation for the difference between the Norwegian and German samples with regard to EM lecture comprehension scores might be differences in the respondents' exposure to and use of English. Indeed, in Norway extensive exposure to English through the media is considered an important explanation for Norwegians' supposedly high levels of proficiency (Bonnet 2004/Ibsen 2004). A number of items on the exposure to and use of English were therefore included in the questionnaire. In Table 8 below the amount of books the respondents read are presented.

Table 8. How many English books do you read per year?

	None	1-3	4-6	7-12	13 or more
Norwegian sample (N=364)	7%	28%	28%	22%	15%
German sample (N=47)	34%	47%	15%	2%	2%

As can be seen the German respondents are less avid readers of English than their Norwegian counterparts. In Table 9 below the frequency of other forms of exposure to, or use of English is displayed.

Table 9. Exposure to and use of English, in percent. N= Norwegian sample, G= German sample

Questions	Never		Some-times		Monthly		Weekly		Daily	
	N	G	N	G	N	G	N	G	N	G
How often do you read English periodicals, magazines or newspapers?	11	15	39	70	17	4	19	6	14	4
How often do you read English on the Internet?	3	2	13	40	5	13	24	23	56	21
How often do you watch English language movies, videos, or TV programs?	2	11	19	40	22	21	27	21	30	6
How often do you speak English?	2	4	19	36	22	17	27	21	30	21
How often do you write English?	5	11	25	47	12	11	29	19	29	13

As with the reading of novels where the respondents in the Norwegian sample read more than those in the German, Table 9 shows that the respondents at the German universities are less exposed to, and use English less frequently than do their counterparts in Norway. Table 10 provides the correlations (Pearson's r) between these items and EngIndex as a dependent variable.

Table 10. Correlations between items for the exposure to and the use of English, and EM lecture comprehension scores (EngIndex)

Questions	Norwegian sample	German sample
How many English books do you read per year?	.26**	.33*
How often do you read English periodicals, magazines or newspapers?	.26**	.49**
How often do you read English on the Internet?	.21**	.37*
How often do you watch English language movies, videos, or TV programs?	.08	.42**
How often do you speak English?	.26**	.53**
How often do you write English?	.26**	.54**

** Correlation is significant at the 0.01 level

* Correlation is significant at the 0.05 level

These correlations show that the active use of English, by reading, speaking or writing in English, for almost all the items correlate positively with lecture comprehension scores in both samples. That is to say, the greater the exposure to English, the higher the EngIndex score. The higher correlations for the German sample can be due to greater variation in the respondents' exposure to English than is the case among Norwegian students, as can be seen in Tables 8 and 9. The latter could be indicated by the lack of any correlation for media exposure to English in the data from the Norwegian sample, while the correlations for this are markedly higher for the German sample. It is also possible that the lower correlations for the Norwegian sample can, at least in part, be due to a higher level of English proficiency than in Germany (Bonnet 2004).

4.   Discussion

In the following, I begin with a summary of the findings, continue with a brief discussion of their validity, and then end with a discussion of their implications with regard to improving EM as well as L1 lecture comprehension.

The main finding of the present study is that many Norwegian students, despite a reputation for high levels of English proficiency (Bonnet 2004), have difficulties with EM lecture comprehension, while German students seem to have even greater difficulties. As illustrated by Figures 1 and 2, 42% of the Norwegian sample and 72% of the German apparently have scores indicating that they experience lecture comprehension difficulties. For EM instruction, the most frequent source of difficulty appears to involve unclear pronunciation/word segmentation or unfamiliar vocabulary. This may be due to the respondents' listening proficiency and/or to the lecturers' pronunciation and lecturing skills. Other areas of difficulty are following the lecturer's line of thought, and difficulties taking notes. Another issue was that many respondents find EM courses more laborious than in the L1.

Next, the other key finding in the present study is the similarity of the problems found in EM and L1 lecture comprehension. Indeed, the low mean scores for L1 lecture comprehension displayed in Table 6, with the exception of those with Norwegian as their L1, are disquieting. This is further supported by the apparent need many have to ask clarification questions about language and content in the L1 lectures as well as in English. In addition, many students were highly dependent on visual aids for comprehension in the L1 as well as in English. All in all, these point out the need to use effective lecturing behavior (Klaassen 2001), that is to pay serious attention to lecture quality in the L1 as well as in English.

4.1. Validity

Before continuing, the validity of the findings in this study needs to be addressed. To start with the construct validity of self-assessment items, a number of studies indicate that these can be reliable predicators of listening proficiency. An optimal solution, however, would require a validation study where self-assessment scores are correlated against listening test scores. This would also provide information about their construct validity, that is to say to which extent the scores in EngIndex and L1Index actually reflect lecture comprehension and not other variables (see for instance Messick 1996). It could also be used to determine at which level in the self-assessment scores the respondents are starting to have serious problems. In the meantime, however, I would argue that the use of a set of items tapping into different aspects of lecture comprehension with focus on specific tasks (see Bachman 1990: 148) and the possibility to compare lecture comprehension scores between lectures in English and L1 provide a useful picture of which areas the respondents experience lecture comprehension difficulties, although interpreting the degree of the difficulties merits greater caution.

The next issue concerns external validity; to what degree the findings are generalizable beyond the sample studied, which to a large extent depends on how the sample is selected. As mentioned, practical constraints made it difficult to select random, representative sample from the relevant

reference populations, in this case all students receiving EM instruction at Norwegian institutions of higher education and in Germany. Nevertheless, I would argue that a purposive sample comprising 364 respondents from 14 courses and three different institutions should be adequate to provide useful insight into the extent and kind of difficulties many Norwegian students have, or do not have, with EM instruction. For the German sample, however, the data must be interpreted with even greater caution pending follow-up studies, in particular a validation study of the validity of the self-assessment scores. It is, nevertheless, interesting to note that the Norwegian and German respondents experience largely the same problems.

4.2. Improving lecture quality

As noted above, the perhaps most interesting finding of this study is that while Norwegian as well as German respondents experience much the same difficulties with EM instruction, many had exactly the same problems in the L1. In fact, this goes to confirm Airey's (2009: 84) claim that "changing the lecturing language [may simply accentuate] communication problems that are already present in first language lectures". This also supports Klaassen (2001), who found that effective lecturing behavior could have a greater effect on how Dutch students experienced lectures than did the language used. In other words, the perhaps most important implication of the present study is the need to use effective lecturing behavior in all lectures, in the L1 as well as in English (see Airey 2009, Airey/Linder 2006, Klaassen 2001, Vinke et al. 1998). In the following, however, I will focus on how to improve EM instruction based on the findings presented above.

The perhaps most important source of lecture comprehension difficulties found in the present study was due to unclear pronunciation. On the one hand this could mean working with the lecturers' pronunciation, stress and word segmentation, on the other it might mean screening EM lecturers with regard to language quality (see Klaassen, this issue). It also means that screening students with regard to English proficiency is imperative.

Yet another issue affecting comprehension is unfamiliar vocabulary, which seems to be a problem in the L1 as well as in English. This is probably due to the fact that university studies involve the socialization of students into domain specific academic genres and registers along with the learning of specialized vocabularies, independent of language. One way of working with this issue would be for lecturers to devote some time going through key terms and concepts as a pre-lecturing exercise, or taking time to explain these during or after the lecture. Another way would be the use of exercises in which the students get to use the terms and concepts in relevant contexts. It is also important, as Airey/Linder (2006) also suggest, to create extra space for clarification questions in connection with lectures, or in follow-up groups. Indeed, this is supported by the data indicating that respondents with low listening proficiency scores more often feel the need to ask questions about language and content correlates (see Table 3). Another option would be to encourage students to work together in preparation for, and as a follow-up after lectures, as students did in Airey/Linder's (2006) study.

Yet another difficulty the respondents experienced was following the lecturers' line of thought. An important way of aiding students' ability to do so is making sure that lectures are clearly and predictably structured along with the use of what is known as interactive discourse structuring. This is "the use of metadiscursive comments [signposting] such as 'First, let's look at' or 'what I will do now is' – [to] facilitate lecture comprehension, particularly for L2 listeners" (Vandergrift 2007: 202). Training in the structuring of lectures and in the use of signposting during lectures to help guide students through the lectures should be made part of in-service, and not to mention all pre-service courses.

Part of such courses should also include learning to make high-quality visual aids, since data on the EM and L1 lectures in the present study show that students rely heavily on these to support their understanding (see Table 5). Airey/Linder (2006) also found that students had problems simultaneously taking notes and listening to the lecturer. This also suggests the need to make

lecture notes available before the lectures as many Norwegian lecturers actually do (Hellekjær 2007). In fact, when during the survey I listened to a lecture at a well-established and successful EM program included in this study, I noticed that lecture notes for the entire semester had been made available to the students in a booklet. I also observed many students taking additional notes on these during lectures.

A final point worth mentioning is looking at the design of and goals of EM courses. One issue that should be addressed is that lecturers, as often as not, are unable to lecture as quickly in English as in their L1 (Vinke 1995). Does this mean a clearer prioritizing of topics, or additional lecturing time? Another issue is the need to make room for student questions, either by organizing follow-up seminars or group discussions, or for instance using text messages to ask questions with large groups of students (see Wilkinson, this issue). Working actively to improve students' language proficiency should also be considered. An example would be requiring students to write papers or give presentations in English, and assessing these for language quality as well as content (Hellekjær/Wilkinson 2003). This will, however, require the support of a language specialist, to teach language, sort out language problems, and to assist students in producing domain specific texts and presentations. Furthermore, such an effort to integrate content and language aims in teaching will require making room for cooperation between the subject matter lecturers and the language specialists (Jacobs 2006).

5.   Conclusion

The present study is of EM and L1 lecture comprehension among Norwegian, German, and international students. Its main findings are that there are lecture comprehension difficulties due to the use of English in such programs, that many of the same problems are evident in L1 lectures as well, and that investigating EM lecture comprehension under the assumption that comprehension in the L1 is more or less perfect, will probably lead to inaccurate conclusions. Indeed, it seems that changing the language of instruction to English merely exacerbates difficulties already present in the L1 lectures.

Given the fairly limited sample, and the use of self-assessment to measure lecture comprehension, the findings of the present study must be interpreted with some caution. Nevertheless, it has identified a number of avenues for further research. It also has clear implications for lecturing in English as L2 as well as in the L1.

With regard to further research, there is clearly a need for a large-scale, follow-up study with sufficient domestic and international student respondents to allow for the detailed analysis of each group. Although a representative sample might be difficult to attain, the goal should be to include respondents from different institutions of higher education with EM courses, from undergraduate and graduate levels, as well as from different types of courses ranging from single, English-on-demand courses to EM programs.

Next, while the use of self-assessment might well be unavoidable, for instance those used in the present study, these need to be validated against a relevant and valid test of English, and possibly of L1 lecture comprehension.

Finally, although the follow-up survey as well as the validation study can be done in a single country, the present study indicates that including other countries, for instance in Europe, would provide additional information.

Limited validity notwithstanding, the present study has a number of implications with regard to ensuring the quality of EM as well as L1 instruction. To start with the former, the main implication is that language proficiency is important for lecture comprehension, and that this needs to be taken into account when designing and teaching EM courses. This means, on the one hand, screening students to ensure they have the language proficiency necessary to benefit from EM instruction, and, on the other, making certain that lecturers have the language proficiency. It will also mean providing the necessary English courses for staff and students.

However, the present study has also shown that students can have difficulties understanding key terms and concepts in the L1 as well as in English as an L2. Indeed, they seem to have the same problems understanding content or following the lecturer's line of thought whether English or their L1 is being used, and seem just as reliant on visual aids to support comprehension. This means that care must be taken to help students understand key terms and concepts independent of lecturing language, to systematically support lectures with well-designed visual aids, to ensure that lectures are well structured, and that lecturers guide students though the lectures by using metadiscursive comments.

In conclusion, it would seem that effective lecturing behavior is just as necessary in L1 as well as in EM lectures.

References

Airey, John 2009: *Science, Language and Literacy. Case Studies of Learning in Swedish University Physics* (Vol. 81). Uppsala: Acta Universitatis Upsaliensis.

Airey, John/Linder, Cedric 2006: Language and the experience of learning university physics in Sweden. *European Journal of Physics* 27(3), 553-560.

Alderson, Charles James 2000: *Assessing Reading*. Cambridge: Cambridge University Press.

Ary, Donald/Jacobs, Lucy Cheser/Razavieh, Asghar (1996). *Introduction to research in education*. Fort Worth: Harcourt Brace College.

Bachman, Lyle 1990: *Fundamental Considerations in Language Testing* (2nd ed.). Oxford: Oxford University Press.

Bonnet, Gérard (ed.) 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.

Blue, George M. 1988: Self-assessment: the limits of learner independence. In A, B./Grundy, P. (eds.), *Individualization and Autonomy in Language Learning*. Oxford and London: Modern English Publications/British Council.

Blue, George M. 1994: Self-Assessment of Foreign Language Skills: Does It Work? *CLE Working Papers* 131, 18-35.

Buck, Gary 2001: *Assessing listening*. Cambridge: Cambridge University Press.

Halvorsen, Tor/Faye, Reidun 2006: *Internasjonalisering. Evaluering av NIFU STEP: Kvalitetsreformen, Delrapport 8*. Oslo-Bergen: Norges forskningsråd, Rokkansenteret.

Hellekjær, Glenn Ole (forthcoming) 2010: Language matters: Assessing lecture comprehension in Norwegian English-medium higher education. In Smit, Ute/Nikula, Tarja/Dalton-Puffer, Christiane (eds.), *Language in CLIL Classrooms.* John Benjamins.

Hellekjær, Glenn Ole 2007: The Implementation of Undergraduate Level English Medium Programs in Norway: An Explorative Case Study. In Wilkinson, R./Zegers, V. (eds.), *Researching Content and Language Integration in Higher Education.* Nijmegen, Maastricht: Valkhof Pers & Maastricht University.

Hellekjær, Glenn Ole/Westergaard, Marit R. 2003: An exploratory survey of content learning though English at Nordic universities. In van Leeuwen, Charles/Wilkinson, Robert (eds.), *Multilingual Approaches in University Education.* Maastricht: Universiteit Maastricht, 65-80.

Hellekjær, Glenn Ole/Wilkinson, Robert 2003: Trends in content learning through English at universities: a critical reflection. In van Leeuwen, Charles/Wilkinson, Robert (eds.), *Multilingual Approaches in University Education.* Maastricht: University of Maastricht, 81-102.

Hellevik, Ottar 1999: *Forskningsmetode i sosiologi og statsvitenskap* (6 ed.). Oslo: Universitetsforlaget.

Ibsen, Elisabeth B. 2004: *Engelsk i Europa – 2002* (Vol. 2). Oslo: University of Oslo.

Jacobs, Cecilia 2006: Integrated assessment pracices – when language and content lecturers collaborate. In Wilkinson, Robert/Zegers, Vera/van Leeuwen, Charles (eds.), *Bridging the Assesment Gap in English-Medium Higher Education*. Bochum: AKS Verlag. 141-160.

Klaassen, Renate 2001: *The International University Curriculum: Challenges in English-Medium Engineering Education*. Delft: Delft University of Technology.

Ljosland, Ragnhild 2008: *Lingua franca, prestisjespråk og forestilt fellesskap: om engelsk som akademisk språk i Norge: et kasusstudium i bred kontekst.* Norges teknisk-naturvitenskapelige universitet, Det historisk-filosofiske fakultet, Institutt for nordistikk og litteraturvitenskap, Trondheim.

Maiworm, Friedrich/Wächter, Bernd 2008: *English-taught programmes in European higher education: the picture in 2007*. Bonn: Lemmens.