

University of Halmstad

English Department

English 61-90

Teaching, Testing and Assessment; an Interrelation of English as a Foreign Language

- The Swedish National Test of English

Helen Adolfsson

Bachelor Thesis

Ms. Veronica Brock

14 August 2012

Author: Helen Adolfsson

Year: Spring Term 2012

Title: Teaching, Testing and Assessment an Interrelation of English as a Foreign Language

- The Swedish National Test of English

University: Högskolan Halmstad

Supervisor: Veronica Brock

Abstract:

The aim of this study was to investigate the interrelation between teaching, testing and assessment of English as a foreign language in the Swedish school system. The focus is on the Swedish National test of English, and teaching and assessment of English in Swedish schools. A comparison with the Cambridge testing system is part of the study which has given rise to questions regarding the validity and reliability of the Swedish National test of English.

The study was carried out through qualitative questionnaires and interviews aimed at 5 English teachers working within the Swedish School system. The findings of the primary data suggest that the test is, with the exception of its marking procedures, regarded highly by the interviewees. The results of the interviews and questionnaires also suggest that the common factor between teaching, testing and assessment of EFL (English as a Foreign Language) in the Swedish school system appears to be the syllabus and, thus, indirectly, the Common European Framework of Reference (CEFR).

Key words: teaching, testing, assessment, validity, reliability, the Swedish National test of English, the Cambridge ESOL Examinations, Common European Framework of Reference, English as a foreign language.

Table of Contents

1. Introduction and Research Questions	6
2. Literature Review	8
2.1 Teaching	8
2.1.1 Teaching and Learning English as a Foreign Language	8
2.1.2 Approaches and Theories of Language Acquisition	9
2.1.3 Models and Styles of Teaching	11
2.1.4 What Constitutes Language Proficiency?	12
2.2 Assessment	14
2.2.1 Self Assessment	14
2.2.2 The Portfolio Model	14
2.2.3 Foreign Language Testing	15
2.3 Testing	16
2.3.1 What is Tested?	16
2.3.2 How Are Language Tests Constructed?	17
2.3.3 Validity	19
2.3.4 Reliability	20
2.3.4.1 Inter-rater Reliability and Marking Procedures	21
2.3.5 Score Meaning	23
2.3.6 Test Items	24
2.3.7 Purpose	25
2.3.8 Fairness	25
2.3.9 The Washback Effect	26
3. The Study	28
3.1 Structure and Conduct – Primary Data Gathering	29

3.1.1	Subjects	29
3.1.2	Structure and Conduct – Questionnaire	31
3.1.3	Structure and Conduct – Interviews	32
3.1.4	Limitations and Analysis of Primary Data Collection	33
3.2	The Results - A Summary of Main Item Responses	35
3.3	Discussion	37
3.3.1	Research Question I – What do teachers working within the Swedish school system teach when they teach English, and how do they teach?	37
3.3.1.1	A) - How do they assess their students' proficiency?	39
3.3.2	Research Question II - What are the teachers' personal opinions of the NTE?	40
3.3.2.1	A) - How do they prepare their students for the NTE?	41
3.3.2.2	B) - Is the NTE regarded as a reliable and valid grading aid by the teachers who use it, and what status does it hold in contrast to the Cambridge testing system?	42
3.3.3	Research Question III - What does the interrelation of teaching, testing and assessing EFL in the Swedish school system appear to be?	43
	Conclusion	46
	Works Cited	48
	Appendices	52
I.	Questionnaire	52
II.	Interviews	58
III.	Web – Cambridge, Demaret and Skolverket	66

Tables

Table 1. Teaching Styles	11
Table 2. Comparing and Contrasting Marking Procedures	22
Table 3. Subject Profiles	29
Table 4. Summary of Main Item Responses	35
Table 5. The Common Reference Levels: global scale	45
Table 6. Questionnaire Responses	54
Table 7. Interview Responses - Oral	58
Table 8. Interview Responses - Written	62

1. Introduction

This study sets out to investigate the relationship between teaching testing and assessment in the Swedish school system. The focus is on testing, in particular the Swedish National test of English (henceforth NTE), a nationally standardised proficiency test. Interviews with five English teachers working within the Swedish school system were conducted in an attempt to acquire their personal opinions and experiences of the NTE. The test is compared and contrasted with the internationally recognised testing system of the Cambridge ESOL (English for Speakers of Other Languages) Examinations, and the related aspects of purpose, validity and reliability are also highlighted. These aspects are considered with regards to the examination of the NTE and the relation to both the interviewees and the Cambridge testing system.

Teachers play a central role in the three areas of foreign language learning mentioned above, and their thoughts and experiences have consequently been of value to this research. There are many approaches to teaching a second language with possibly just as many ways of assessing a student's level of proficiency. Testing is central to both of these elements of foreign language learning, in the sense that a test often reflects what has been taught, and is also used as a form of assessment. The different circumstances under which the teachers interviewed work provided opportunities for the comparison and contrast of the data. This in turn revealed the common factor between teaching, testing and assessment in this particular context. Furthermore, this study intends to investigate if the national test influences teaching, i.e. if there are any suggestions of a washback effect.

All tests should assess the area of language proficiency which they are intended to test, and for a test to be beneficial, it has to be consistent and the results should also carry meaning enabling a judgment to be made. Extensive research has been made on how a test can be made valid; some aspects of this research will be discussed in relation to both the NTE as well as to chosen Cambridge ESOL Examinations.

This essay begins with an introduction to EFL (English as a Foreign Language) teaching, its related theories and styles, and an examination of what is meant by language proficiency. It is then followed by an exploration of assessment. There is a deeper investigation of the concept of language testing with a closer look at both the NTE and the Cambridge testing system. The second half of the essay is devoted to a practical study which tries to establish the status of the NTE both from a teacher's point of view and in comparison to the Cambridge testing system. Additionally, the purpose, reliability and validity of the NTE is questioned, and an attempt to establish if there are any notions of a washback effect is also made.

Research questions

- I. What do teachers working within the Swedish school system teach when they teach English, and how do they teach?
 - A) How do they assess their students' proficiency in English?
- II. What are the teachers' personal opinions of the NTE?
 - A) How do they prepare their students for it?
 - B) Is the NTE regarded as a reliable and valid grading aid by the teachers who use it, and what status does it hold in contrast to the Cambridge testing system?
- III. What does the interrelation of teaching, testing and assessing EFL in the Swedish school system appear to be?

2. Literature Review

2.1 Teaching

2.1.1 Teaching and Learning English as a Foreign Language

When it comes to teaching a foreign language, a great deal of models and types are available. However, which ones are used often depends on what view one has on language acquisition and related factors such as end goals, age of the learners, and learning and teaching contexts, individual differences, environment, social situation etcetera. There are many different goals and purposes for learning a language. English is considered by many as a global language, in the sense that it is widely spoken in many parts of the world, in diverse contexts. In *Second Language Learning and Language Teaching*, Cook devotes a whole chapter to the different goals of language teaching and learning (194-213).

In particular he mentions three main goals for teaching: "Transitional language teaching" (206), teaching for "Language maintenance and bilingual language teaching" (207) all of which are directed towards people who, in this case, are in an English speaking society but with a different first language (L1), such as immigrants or exchange students. Cook expresses the three goals of teaching mentioned above, as "Central goals . . . within a society"(212). He further states career, higher education, travel as well as language as an aid in understanding research and information as "International goals" (207, 212). Lastly, he mentions the motivational and attitude related "individual goals" (209-10, 212) such as simply understanding the language or other cultures, which could be part of a curriculum as an academic subject.

In Swedish schools, English as a subject is introduced to students at an early age and it is a compulsory subject throughout the compulsory school up until the first level of English in upper secondary school (Skolverket F). The introduction of English in the curriculum for the compulsory school insists that languages are one of our best assets, and that knowing more than one language

will improve our lives both professionally as well as personally. English can be found in many diverse social situations and it “surrounds us in our daily lives. . .” (Curriculum 32). This diversity is incorporated into the curriculum and teaching English “should aim at helping the pupils to develop knowledge of the English language and of the areas and contexts where English is used”(Curriculum 32). It is further stated that one of the main purposes of teaching is to encourage students to use the language with confidence (Kommentarmaterial 8). It can thus be argued that the Swedish curricula relates to both the international and individual goals mentioned above.

2.1.2 Approaches and Theories of Language Acquisition

There are many different theories of language acquisition, which in turn have given rise to different approaches to language teaching. Below is a brief introduction to three major theories/approaches.

- Behaviourism sees language as a behaviour which need to be taught. This theory was popular in the 1940s and 1950s. The most well known theorist connected to behaviourism was B.F Skinner. Behaviourism “explain(ed) language learning in terms of imitation, practice, reinforcement, and habit formation” (Lightbown and Spada, 34). This theory sees language learning as something external, it does not come from within but rather through the environment around us (Cook, 220-1).
- The Innatist perspective, on the other hand, is very much about the internal and the mind. The best known theorist in this area is Noam Chomsky, who proposed that we are all born with an innate language device which he calls 'Universal Grammar' (UG) (Lightbown and Spada, 15; Cook, 215). The Critical Period Hypothesis (CPH) is also part of this approach, which suggest that humans as well as animals are genetically programmed to acquire

certain skills at certain times, over certain periods in our development. Furthermore, it is the Innatist's belief that we only acquire a language during the critical period (approximately between birth to puberty), after that it has to be learnt. It is also suggested that without exposure to language during the CPH a language will never be developed properly (Lightbown and Spada 17-19). Thus, the innatist approach sees language as a biological function, the same as walking, and it does not need to be taught but it will come automatically with exposure to it. It can be argued that both the Common European Framework of Reference (henceforth CEFR, see 2.1.4) and the language syllabus for the Swedish school system are partially based on research supporting the innatist approach. In the comment material to the syllabus it is stated that knowledge of the language's structure should be combined with an effort to understand, communicate and to express oneself in the language rather than to focus on separate building blocks, communication being the key factor, which in turn falls in line with the CEFR (Kommentarmaterial 6).

- There is a third approach also worth mentioning in this context. Both Jean Piaget and Lev Vygotsky were well known theorists within, what Lightbown and Spada entitle, "Interactionist/developmental perspectives" (19). This approach "focus(ed) on the interplay between the innate learning ability of children and the environment in which they develop" (Lightbown and Spada 19). Language learning and acquisition could be seen from the "interactionist/developmental" approach as a mix between the behaviourist approach and the innatist approach. Most importantly Piaget and Vygotsky believe that a language contains everything a child needs to be able to learn it (Lightbown and Spada 19). The child needs to use the language, practice and interact with the environment in which they live to be able to make use of the innate language device they are believed to be born with.

2.1.3 Models and Styles of Teaching

This study focuses on five different styles of teaching; the academic style, the audio-lingual style, the communicative style, the task-based style as well as the traditional EFL style. Table 1 below contains short descriptions of the five diverse styles in combination with their connection to the previously mentioned approaches to, and theories of, language acquisition.

Table 1. *Teaching Styles*

Teaching Styles		
A	The Academic Style	<p>* Focus lies on the written language with a majority of grammatical instructions and tasks as well as translation tasks. Repetitive.</p> <p>** This style is based on the early method used when teaching classical languages (Lightbown and Spada 138). Arguably this style could be connected to the behaviourist approach.</p>
B	The Audio-lingual Style	<p>* Contains three steps;</p> <ol style="list-style-type: none"> 1) Dialogue: The students listen, repeat and act the dialogue out in diverse ways. 2) Structure drill: The students repeatedly practice grammatical structures. 3) The students create new dialogues by using the information acquired in step 1 and 2. <p>** According to Lightbown and Spada this style is based on the behaviourist approach in combination with contrastive analysis (138).</p>
C	The Communicative Style	<p>* Communication and interaction is both the goal and means of teaching via information gap exercises, role play and different tasks.</p> <p>** When referring to Cook's description of the “interaction approach” (224-8) and “sociocultural theory” (228-30) this style resembles some of the theories of the "Interactionist/developmental perspectives" (Lightbown and Spada 19). It could be further argued that this style is connected to Chomsky's UG theory.</p>
D	The Task-based Learning Style	<p>* Contains meaning-based tasks with distinct results. The teacher decides on task, the students complete them (often in pairs), they report back and then discuss the results.</p> <p>** This style could be connected to, what Cook terms, the “processing model” (218-22) which in turn, approaches language acquisition the same way as the behaviourist theories.</p>

E	The Traditional EFL Style	<p>* Contains three steps;</p> <ol style="list-style-type: none"> 1) The teacher introduces new vocabulary and grammatical structures. 2) Dialogue: The students listen, repeat and ask questions. 3) Controlled role play with a template. <p>** The EFL style resembles points made in the argument of the behaviourist approach to language acquisition.</p>
---	---------------------------	--

* This section was used in the construction of the questionnaire and interviews used in the practical study (see 3.1.2, Appendix I and II).

** Additional information.

2.1.4 What Constitutes Language Proficiency?

To become fully proficient in a language usually means that learners need to be able to read, listen, write and speak (known as the four *main skills*). In turn, this means that they need to master the related *language systems*: vocabulary, grammar, discourse, etcetera. From my own observations, it would seem that the areas of language proficiency which are given the most attention are usually reading, listening, writing and speaking. Vocabulary, grammar and pronunciation are often, but not always, seen as integral parts of the four main skills. The micro skills of interaction, comprehensibility and production are also seen as parts of language proficiency. Through extensive research, the Council of Europe has drafted the Common European Framework of Reference (CEFR). The CEFR is a six level scale which helps define different levels of language proficiency (Cambridge E, Common European Framework of References for Languages). In general,

“The CEFR is a document which describes in a comprehensive manner

i) the competences necessary for communication,

ii) the related knowledge and skills and

iii) the situations and domains of communication.

. . . It facilitates a clear definition of teaching and learning objectives and methods and provides the necessary tools for assessment of proficiency.”

Common European Framework of References for Languages

The Swedish syllabus for English is based on criteria stated in the CEFR (Kommentarmaterial 6-7). Table 5 (section 3.1.1) contains the CEFR descriptions of proficiency levels; a good general view of what is considered proficiency in relation to different levels. The Swedish syllabus of English also highlights the importance of language in general, but it also contains some central aspects of learning English. According to *Kommentarmaterial till Kursplanen i Engelska* (as interpreted by the author of this essay) the three main parts of learning English are "Content of Communication", "Listening and Reading - Reception" and "Speaking, Writing and Converse - Production and Interaction" (11). As can be seen, the syllabus covers the four main skills in combination with the micro skills of comprehension, production and interaction. These three groups contain subdivisions with themes relating to society, culture, literature and so on. "Strategies" (Kommentarmaterial 15) are also seen as an element of learning English in Swedish schools, explained as conscious or unconscious strategies which students use to overcome difficulties and obstacles in their learning process. The strategies the students use are also seen as a means of continued progress for the students after having completed their English studies.

2.2 Assessment

In this essay "assessment" is defined as the process of evaluating students' knowledge and proficiency in English. In many cases, "testing" is part of the assessment process, and there are a number of different ways to test different elements of language knowledge. Tests and quizzes can thus cover a large part of the assessment. However, there are other ways of assessing language ability and factors of language proficiency. For example:

2.2.1 Self Assessment

"Self-assessment" (also known as self-evaluation) is an approach in which the students themselves evaluate their own work and progress. Rolheiser and Ross state that both their own studies and other research within this area shows that this method has positive effects on learning. Furthermore they argue that "self-evaluation" has an "impact on student performance through enhanced self-efficacy and increased intrinsic motivation". A similar form of evaluation is "peer-evaluation" in which students evaluate each others' work and efforts. It can be argued that feedback given in class to each other amongst the students is similar to peer-evaluation.

2.2.2 The Portfolio Model

The portfolio model could be described simply as a collection of a student's previously assessed work, which is considered in the final assessment process. However as Mueller points out, "A portfolio is not the pile of student work that accumulates over a semester or year. Rather, a portfolio contains a purposefully selected subset of student work."

2.2.3 Language Testing

In many cases a test is described as "an examination of somebody's knowledge or ability, consisting of questions for them to answer or activities for them to perform" (Test). There are a number of ways of approaching language testing. One way to embark upon this subject is to consider the skills of language proficiency that are most often tested. Aspects of language ability such as vocabulary, grammar and pronunciation are often integrated into the four main skills (e.g. NTE and many of the Cambridge exams). What is tested is usually related to the objectives of language learning, such as end goals of the students or the teacher, the content of a syllabus or a course plan. Another way to approach language testing is the purpose of a test. Harmer mentions four different types of tests, all of which have specific intentions. He states that **placement** tests are used to test a learner's level of knowledge in relation to either a syllabus or other types of teaching plans, while **diagnostic** tests are designed to reveal difficulties, knowledge gaps and inabilities. **Progress** tests evaluate learner's progress in knowledge compared to goals and expectations such as those found in a syllabus. He also mentions **proficiency** tests which are supposed to show the ability and knowledge of a learner. The latter is often used to measure ability and to make sure a learner has reached certain levels of knowledge. The proficiency test is most likely to be taken at the end of a course as an examination (Harmer 321-2). This essay focuses on the proficiency type, as these tests (the Cambridge ESOL examinations and the NTE) have such qualities.

2.3 Testing

The Cambridge ESOL Examinations are one of many international proficiency testing systems. The Cambridge tests are used on an international level by many different people for many different reasons. They are internationally recognised and are often used as a criteria for international studies or for employment; thus, they are often offered as part of a school programme.

In the United Kingdom students take subject tests, the A-levels, on the same day nationally. The results from these tests determine a student's grades, which means that in the United Kingdom grades are set on a national level. The A-level results are considered both when it comes to higher education and future employment, and thus play an important role in a student's future. The NTE, on the other hand, is a proficiency test used solely within the Swedish school system. It is not viewed as an examination test but rather as an aid in the marking process. In comparison to the A-levels, the NTE results are only used as support material by the teachers who set the final marks, which means that a student's grades are given on a local level. The NTE is therefore not viewed as important and influential as the A-levels are in the United Kingdom.

2.3.1 What is Tested?

Of all the different Cambridge examinations, the focus in this essay is on three of the “Cambridge For School” range because, as the name suggests, they are intended for use in schools. Apart from the most advanced level, the “First Certificate in English for Schools” (henceforth FCES), the “Cambridge for School” exams have been divided into three papers; reading and writing, listening, and speaking. The different papers test the appropriate micro skills of receptive ability, comprehension, production and interaction. The FCES test, in comparison, consists of four papers each covering the four main skills, in addition, there is a fifth paper which tests grammar and vocabulary knowledge. The latter is one of the differences between the NTE and the Cambridge for School Exams (Skolverket A, B, C; Cambridge B, C, D). In terms of what is tested, both testing

systems are fairly similar except for the specific focus on grammar and vocabulary in the FCES.

The NTE for the Swedish compulsory school is also divided into three parts containing oral interaction and production, receptive abilities of reading and listening comprehension and written production. The NTE for the upper secondary school tests the same skills and abilities but instead of three papers there are four, one for each skill. The NTE is more focused on the "receptive, productive and interactive competencies" (Demaret) rather than technical aspects such as grammar and vocabulary (although these elements are not disregarded completely). Moreover, the NTE is also designed to correspond to the curriculum, though it does not comprise all of the requirements (Demaret). The themes and subjects of the test items are consistent with the aspects of the curriculum which do not belong to skills or elements of proficiency. For example, themes of the written assignments could be related to the cultural aspects of the curriculum or different types of spoken English could be used for the listening comprehension part (Demaret). Moreover, the above-mentioned tests (both the Cambridge tests and the NTE) are all linked to the same levels of proficiency (CEFR), ranging from A1 to the more advanced B2 or even C1 (criteria for the different levels can be found in table 5, section 3.1.1) (Kommentarmaterial 7 and Cambridge B, C, D).

2.3.2 How Are Language Tests Constructed?

There are several aspects to consider when constructing a language test. Harmer introduces five steps which should be considered in the making of a test. First the conduct of the test needs to be taken into account; issues such as the time and place of the test and duration and time for marking it. Secondly, he points out, one has to decide what the test is actually going to test, what skills it is going to cover as well as what topics should be used and so on. Thirdly, he states that when using mixed item types, they need to be equally proportioned. He continues by emphasising the importance of "Weight[ing] the scores" (328), which basically means that all the items should carry the same amount of weight in relation to each other. The last point he makes is that of

piloting. He stresses the necessity of trying a test on other subjects before it is set to measure real test takers (Harmer 327-8). Although Harmer's advice on constructing a test is directed towards teachers creating smaller tests, the fundamental process is still the same for test construction on a larger scale. One aspect of test construction which Harmer has not mentioned is the necessity of making sure that the content and test items match up to the initial intention of the test (Cambridge A). Reliability, score meaning and validity are also fundamental factors which need to be borne in mind for all types of tests.

The construction of a language test, the size of either the NTE or any of the Cambridge ESOL Exams, takes a vast amount of effort and time to produce, it is also an expensive procedure. Both the NTE and the Cambridge ESOL Exams take around two years to construct (Cambridge A, Ramstedt 19). The Cambridge ESOL Examinations are part of an internationally recognised testing system supported by extensive research. This in combination with the wide acceptance and use of the exams, leads to the supposition that they are valid and reliable. By comparing and contrasting some of the previously mentioned factors in relation to the NTE and the Cambridge testing system a notion of the status of the NTE will become visible.

2.3.3 Validity

Validity and reliability are interconnected aspects of a test's design, construction and

purpose. As Sireci comments on Lissitz and Samuelsen's terminology regarding validity, "Any conceptualization of validity theory must acknowledge that what is to be validated is not a test itself but the use of the test for a particular purpose" (477). This relates to the definition of validity which Cambridge ESOL Examinations present in their *Principles of Good Practice*.

"Validity is generally defined as the extent to which an assessment can be shown to produce scores and/or outcomes which are an accurate reflection of the test taker's true level of ability."

Principles of Good Practice, 23.

Validity is a complex area which has changed over the years. Dörnyei highlights three kinds of validity mentioned by Chapelle in the 1999 *Annual Review of Applied Linguistics*. First there is "criterion validity" identified by the test's relationship to a comparable measurement tool. "Content validity" is also mentioned, which is concerned with the content of a test. Thirdly "construct validity" is pointed out and is described as relating to the relationship between test results and the theory on which the test is based (Dörnyei 51). Dörnyei continues by highlighting the revised 1999 AERA, APA and NMCE *Standards for Educational and Psychological Testing*, which brought together the three previously mentioned types of validation, including all of the previously mentioned qualities, under one common term, "construct validity" (Dörnyei 51).

Dörnyei continues to explain that this transformation was the outcome of a change in how validity was viewed. Instead of validity being considered a feature of the test, it has recently been seen as "considering . . . the truthfulness of the interpretation of the test scores" (51). As has been mentioned, construct validity comprises all three previously mentioned approaches to validity. However, according to Kane, this line of reasoning is complex with certain limitations. He suggests that an argument-based methodology to validation would address these limitations. He describes his own approach in terms of an interpretive argument followed by a validation argument. "The interpretive argument provides a framework for validation by outlining the inferences and assumptions to be evaluated" (Kane 8). The validity argument is then used to assess the credibility

and consistency of the interpretive argument (Kane 8). The Cambridge testing system mentions construct validity in its *Principles of Good Practice* (25), which is considered to be part of the validity argument. Reliability is sometimes seen as part of a test's validity (Dörnyei 52), which will be looked at next.

2.3.4 Reliability

When looking up the word reliability in a dictionary, one usually gets a definition such as this:

"[the] extent to which an experiment, test, or measuring procedure yields the same results on repeated trials"
(Reliability).

The *Principles of Good Practice* describes it like this:

"Reliability concerns the extent to which test results are stable, consistent and free from errors of measurement"

Principles of Good Practice, 27.

These descriptions are the basic meaning and purpose of test reliability. In *Research Methods in Applied Linguistics*, Dörnyei, referring to Bachman, states that "[if] difference in administrative procedures, changes in test takers over time, various forms of the test and difference in raters . . . cause inconsistencies, or measurement error, then our results are unreliable" (51). This explains the conscientious procedures in the administration of both the NTE and the Cambridge ESOL Examinations. They both have specific dates on which the tests take place, they also have strict guidelines on what aids the test takers are allowed. Furthermore, these tests take place in certain locations at set times. One aspect of the administration which differs between the NTE and the Cambridge ESOL Examinations is the correction procedures, in addition to who the raters/markers are.

2.3.4.1 Inter-rater Reliability and Marking Procedures

Central marking takes place with regards to Cambridge exams. The Cambridge testing system uses specialised and trained inspectors who mark parts of the tests with the marker only focusing on one part of the test. These markers go through the procedure not knowing anything about the test taker (Cambridge English First 19). The marker of the NTE (National Test of English) on the other hand often knows the students well. In addition, the teachers marking the test have not undergone any specialised training but rather rely on support material when marking. The marking of the NTE takes place locally and the individuals correcting and marking the test are most often the test takers' teachers, though there is "a strong recommendation, but no formal demand, that this be done in collaboration with colleagues" (Erickson 1).

The markers of the Cambridge ESOL Exams have two mark schemes to follow, the Task-specific Mark Scheme and the General Mark Scheme, in combination with the nine-band marking scale presented by the IELTS (International English Language Testing System) (Key English Test 13; Cambridge English First 19, 21-26; IELTS). The teachers marking the NTE are provided with support material stating certain factors which they should focus on. This material does not contain a specific marking scale for the different grading levels but the teachers are instead referred to the grading criteria found in the syllabus. Additionally the teachers can make use of a comment material to the knowledge criteria which the Swedish National Agency for Education has developed (NAFS A, B, C). Even though both systems have very similar criteria which the markers look for, the processing, interpretation and analysing of these are different. It can be argued that the marking procedure of the NTE is not as structured and clear as the system of the Cambridge ESOL Exams. However, a ten-band marking scale has been introduced to the NTE marking procedure, in connection to the new 2011 curricula and syllabus, and the new grading system (Skolverket I).

According to an investigation initiated by the Swedish Schools Inspectorate, the consistency of test results relating to the NTE (especially the written production part) may be questioned

because of deviational test results (Plunkett) and, thus, questions the inter-rater reliability of the NTE. Inter-rater reliability is concerned with the consistency of markers' ratings (Trochim, Hewell et. al.). Both marking procedure and who the marker is are factors that could explain the above-mentioned deviations. The type of assessment material is also a factor which could contribute to inconsistencies. However, in addition to the newly introduced ten-band marking scale, the Swedish Schools Inspectorate have presented recommendations relating to the deviations mentioned above, such as clarifying the support material as well as introducing test taker anonymity in the marking process (Rekommendationer). The table below compares and contrasts the marking procedures of both the NTE and the Cambridge Examinations.

Table 2. *Comparing and Contrasting Marking Procedures.*

Different aspects of the marking procedure	The Swedish National Test of English (NTE)	The Cambridge ESOL Examinations
Procedure	<ul style="list-style-type: none"> ✦ The test is collected and marked by the teacher. ✦ There are advice on collaborative marking but no definite regulations. 	<ul style="list-style-type: none"> ✦ Centralised marking takes place by trained markers. ✦ Each marker focus on one part of the test.
Marker	<ul style="list-style-type: none"> ✦ The students' teacher, possibly in cooperation with other teachers ✦ No official training is required but it is advised 	<ul style="list-style-type: none"> ✦ A specialised, trained and tested marker ✦ The markers undergo a test every two years to ensure their marking is up to date and consistent.
Test taker	<ul style="list-style-type: none"> ✦ The test taker is often known and identified to the marker. ✦ It has been suggested that the test taker's identity be unidentified. 	<ul style="list-style-type: none"> ✦ The test taker is unknown and unidentified to the marker.
Material (support material used for the marking of the	<ul style="list-style-type: none"> ✦ Provided with support material including content and language aspects to focus on, such as; the student's overall will and ability 	<ul style="list-style-type: none"> ✦ Use of Task-specific Mark Schemes, which focus on conditions featured in the task in question, as well as

written production part)	<p>of expression, content, accuracy, range, appropriacy of register and format, organisation and cohesion as well as adoption of text to target reader, etc.</p> <ul style="list-style-type: none"> ↘ Grades and levels of the test are connected to the criteria stated in the syllabus. ↘ Additional comments to the criteria could be used. 	<ul style="list-style-type: none"> ↘ General Mark Scheme focusing on criteria such as; content, accuracy, range, appropriacy of register and format, organisation and cohesion as well as adoption of text to target reader.. ↘ Following the IELTS nine Writing Band descriptors which have marks ranging from 1 to 9, each connected to specific descriptions.
*Subjectivity/ Objectivity	<ul style="list-style-type: none"> ↘ It can be argued that this method of marking is to some extent subjective, especially considering the relationship between test taker and marker and the possibility of bias. 	<ul style="list-style-type: none"> ↘ The centralised marking procedure in combination with anonymity in the relationship between test taker and marker suggests objectivity.
Summary	<ul style="list-style-type: none"> ↘ The procedure takes place locally ↘ The marker has most likely not undergone any training ↘ The test taker is often known to the marker ↘ Support material is provided which is analytical rather than systematic. 	<ul style="list-style-type: none"> ↘ The procedure is centralised ↘ The marker has undergone extensive training and is continuously tested on his/her ability to mark ↘ Test taker anonymity ↘ Use of a dual marking scheme in combination with a 9 level marking scale

**The marking of a written production and communication task is in itself quite subjective in the sense that there are no clear right or wrong answers.*

2.3.5 Score Meaning

Score meaning is another aspect of a test which is a vital and complex issue. The scores of a test must mean something for the test to be beneficial. If the scores have no meaning, then the test has no meaning. “We can add meaning to the scores by referencing them to norms for different groups or to performance levels, benchmark performance levels, or achievement levels (e.g. as in NAEP or CEFR)” (Kane 8). Score meaning is something that is important to test takers, teachers, employers, and many more. Kane's validity argument basically constitutes a structure that works as evidence of score validity. There are many factors which need to be considered and for further insight into the issue both Kane and Chapelle's articles in *Language Testing* are recommended.

Both the Cambridge Examinations and the NTE are benchmarked to the CEFR, and one can

argue that the scores represent a student's level of his or her abilities in English. In relation to the NTE, the scores are used as part of the assessment process and it can thus be maintained that the meaning of the scores denotes a level of proficiency in terms of marking. By looking at Chapelle's line of reasoning and with reference to Kane's argument for the scoring rule, one could further argue that the validity of the National Test of English (NTE) could be questioned with regards to deviations of test results connected to the written production part (23). Chapelle maintains that "the study of the criteria that raters actually use when they score" is "an important issue of language assessment" (23). When it comes to written or oral tasks she refers to McNamara (1996) by stating that "scores . . . are affected by human interpretation of a scoring rubric" (Chapelle 23). Personal interpretations of score guidelines could thus lead to inconsistencies in the test results which would then disrupt the reliability of the test.

2.3.6 Test Items

Test items can be divided into direct or indirect item types. According to Harmer, direct item types "ask candidates to perform the communicative skill which is being tested", they are also usually "as much like real life language as possible" (322). Indirect test items are described as measuring what lies beneath a student's receptive and productive skills; they are usually restricted items (Harmer 322). Another division Harmer points out is between discrete-point testing and integrative testing. He states that discrete-point testing "only tests one thing at a time" while integrative testing "expects students to use a variety of language at one point" (323).

Both of the testing systems mentioned in this text (NTE and Cambridge) use a mix of item types (Cambridge B, C, D; Demaret A, B, C). It is viewed by many that a mix of items give a good overall measurement of a student's abilities.

2.3.7 Purpose

The purpose of a test is decided by different aspects of what it is intended to measure. In this

section the intended purpose of a test will be linked to both the NTE as well as the previously specified Cambridge ESOL Exams. A language test could have many different purposes, as can be seen in part 2.2.3. The NTE as well as the Cambridge for School tests could all be seen as proficiency tests. They are designed to measure a student's knowledge of English in relation to certain levels of proficiency (Cambridge B, C, D; Demaret). However, one fundamental difference between the purpose of the NTE and Cambridge for School Exams is their intention.

The Cambridge for School Examinations tests are to be seen as examinations which will be the base for certificates of different levels of English proficiency (Cambridge A, B, C). The NTE, on the other hand, is not to be seen as an examination test which sets the final marks, but it should rather be seen as an aid in the grading process (Erickson 1). It is stated that its main purpose is to “support equal and fair assessment and grading” (Skolverket H). However, there are no guidelines on how much influence the test actually has on the grades (Skolverket G).

2.3.8 Fairness

A fair test should not discriminate against sub-groups of candidates or give an advantage to other groups. It should also be fair to those who rely on the results . . . by performing consistently and by accurately assessing the ability being tested.

Principles of Good Practice, 9

This account of fairness is taken from the *Principles of Good Practice*, which reflects a view held by many in the language testing world. As can be noticed, fairness is interrelated with validity, reliability and administration. Security is an important aspect in relation to both testing systems. Both have strict rules on how administration is to be executed as well as having confidentiality restrictions attached to them (Skolverket E, Principals 8). On top of this, there are certain guidelines for students with special needs, support material is also available for all of the NTEs and Cambridge tests (Skolverket D, Principals 9).

The issue of test taker diversity could not be found on either Skolverkets web page or the

NAFS's web page in connection to the NTE. However, Gudrun Erickson mentions some aspects arguably related to fairness. First, she states that in the development of a new test they take note of "opinions expressed by students and teachers in connection to piloting and pre-testing" (3). Second, she asserts that "all materials are developed in close cooperation with different categories of experts . . . [including] students . . . the real stakeholders" (3). It could be argued, that this information suggests fairness in the sense that different test takers and administrators have been included in the construction process.

2.3.9 The Washback Effect

The Encyclopedic Dictionary of Applied Linguistics defines the washback effect as "the effect (positive or negative) of testing on teaching" (Washback Effect). A little more in depth, Bailey refers to Hughs' model of the phenomenon in her approach to washback. Hughs makes a distinction between "participants, process and product"(9). Anyone whose "work may be affected by a test" is denoted as a "participant" and "any actions taken by the participants which may contribute to the process of learning" is denoted as a "process"(9). "Product" is further explained as "what is learned (facts, skills, etc.) and the quality of learning (fluency, etc.)"(9). Hughs then connects these components into a model describing his approach to washback.

"The nature of a test may first affect the perceptions and attitudes of the participants towards their teaching and learning tasks. These perceptions and attitudes in turn may affect what the participants do in carrying out their work (process), including practicing the kind of items that are to be found in the test, which will affect the learning outcomes, the product of that work." (*Bailey 10*)

In other words, washback could be explained as the effect a test has on

- teachers, students or even material developers;

- The development of new syllabi, new teaching methods or the introduction of test related content;
- the outcome of attitudes, new teaching methods or material such as new skills or abilities.

By looking at this description, both positive and negative washback can be described. A test can affect participants, for example a teacher or test takers, either positively or negatively in terms of their attitude towards learning English for example. It can also affect processes and products in both a positive or negative way. For example:

- ➔ If the results or content of a test influence teaching by the introduction of new, more successful material or even a more efficient method of teaching, the washback effect on processes can be deemed as positive. The new and more effective material and teaching method would further result in positive effects on the products, what abilities and skills are learned. These new skills and abilities of students can further effect the development of new material.
- ➔ On the other hand, a negative affect can be noticed when a test influences material and way of teaching which only produces students who can pass a test without actually acquiring any knowledge of the subject.

Whether or not washback exists as a phenomenon has been an area of discussion. Nevertheless, washback has generally been accepted, although there are many approaches to a definition of it. Many voices have declared that more empirical research on the area should be carried out. Mary Spratt discusses a number of empirical studies of washback in her article "Washback and the Classroom: the Implications for Teaching and Learning of Studies of Washback from Exams". However, even though a number of empirical studies are now available, she argues that the washback effect needs to be further researched because it is such a wide and general area. She highlights elements of the classroom such as "curriculum, materials, teaching methods, feelings and attitudes, [as well as] learning" (Spratt 8) which should be considered in potential studies.

3. The Study

The purpose of this study was to investigate the relationship between teaching, testing and

assessment of English in the Swedish school system. The main interest was the National Test of English (NTE) and its relationship to teachers who use it, how it relates to international testing systems and also how it relates to teaching and assessment in Swedish schools.

Initially, questionnaires were sent out in an attempt to gain as much information as possible from EFL teachers using the NTE. The first requests for participation in the study were sent out to two different groups of schools. The first group consisted of 15 schools with higher results than the national average for the NTE and the second group also consisted of 15 schools but with results lower than the national average. This distinction was made with the belief that by setting the two groups apart, certain patterns would emerge supporting or refuting different theories relating to the relationship between teaching and testing. However, I obtained no responses to my requests. Therefore I decided to discard the difference between the groups and widen my search for individual participants. Unfortunately, this resulted in only two affirmative subjects who were willing to participate, which was not sufficient for a successful study. I then had to rethink my approach to obtaining the primary data and decided to conduct interviews instead. This shift also led to a shift in the main focus of the study from the relationship between teaching and the NTE, to a focus on the test itself. An additional three subjects were interviewed and the interviews and questionnaires combined have provided me with sufficient information upon which a study could be built.

3.1 Structure and Conduct – Primary Data collection

The primary data was gathered using a qualitative research method, in the form of

questionnaires and interviews. These were selected because they enable individual opinions and experiences to emerge from chosen subjects. Teachers of EFL constituted the population of interest and the sample group consisted of five teachers working within the Swedish school system who all have some experience of the national test.

3.1.1 Subjects

The subject profiles which can be found below are based on short answer questions used in both the questionnaire and the interviews. (see Appendix I, section 3.1.2 and 3.1.3).

Table 3. *Subject Profiles*

Subject	School	Presentation
A	Compulsory School	Subject A has been teaching English for eight years . He/she teaches four classes with approximately twenty students per class. They have got digital classrooms which is positive as well as access to a wide variety of good material. Motivation is an important factor in learning and he/she thinks that it is reached by meaningful teaching where the students have the opportunity to influence its content. His/her answers in the questionnaire suggests that he/she enjoys his/her job and feel satisfied with his/her work situation.
B	Compulsory School and Gymnasium	Subject B has been teaching English for two years . He/she teaches both in the compulsory school as well as on upper secondary level. Class size varies from four students to seventeen at the most , with a total of six classes . There are differences between the compulsory school and the gymnasium, especially when it comes to technology , but on the whole he/she is satisfied with the material that they have. Because he/she teaches in a small society with small classes he/she finds him-/herself lucky, in the sense that he/she can give more time to each student. Subject B also seems happy in his/her job and is mostly satisfied with the material he/she has got at hand.

C	Vocational Gymnasium	Subject C has been teaching for almost five years . He/she teaches three classes with an average of twenty-four students per class. His/her work situation is a bit different from that of subject A and B in the sense that his/her students study to prepare for their chosen careers. He/she finds his/her students unmotivated when it comes to English, and their only aspiration is to pass the module enabling them to continue with their vocational subjects. Even though there is a lack of motivation he/she likes his/her work.
D	Gymnasium	Subject D has been teaching for ten years . He/she teaches three classes with twenty-eight students per class. All of his/her students attend a special program where English is regarded as an important factor of future professions. They are generally very motivated which he/she sees as an important factor in the learning process. Subject D seems pleased with both material and work situation.
E	Vocational Gymnasium	Subject E teaches approximately three classes with roughly twenty students per class. He/she seems happy with material and work situation although the motivation amongst the students is lacking . Subject E is a native speaker and only speaks English in class . His/her origin seems to motivate the students to converse in English. He/she seems to be happy in her work.

Neither the gender nor age of the subjects were taken into account when choosing the sample group, mainly because both were deemed as irrelevant in this particular study. Age and gender are factors which could be considered in future research. All of the subjects teach either students (ages between 16 and 18) taking English, level 1, at the Swedish upper secondary school or students of year nine (15 years of age) in the Swedish compulsory school. The Swedish compulsory school is mandatory and basically the same for everyone. The Swedish upper secondary school, on the other hand, is voluntary and the students have chosen the programme they want to attend. These subjects were mainly chosen because of their connection to the NTE, which is a compulsory test used by teachers of both age groups.

In deciding who to pick, I realised that the work situation between the teachers differs somewhat. The size of the schools varies, they teach different age groups and some of the students are attending vocational programmes while others are attending a specialised programme where English is seen as an important factor in their studies. These differences could be both beneficial and detrimental when it comes to comparing and contrasting their replies. If there is a common belief amongst all of the teachers it could be suggested that this cohesion is the result of an underlying factor which would then be of interest in this research. If, on the other hand, there is a trend suggesting that teachers of the upper secondary school feel differently about an issue than to those who teach in the compulsory school, it could suggest that the difference lies in the diverse work situations. Issues like these could be the base for further investigation with sample groups where the dissimilarities between the teachers and their work situations have been eliminated.

The subjects were ensured confidentiality and are, therefore, not referred to by name but rather by terms such as simply teacher, interviewee or Subject A, Subject B, and so on. For further protection, the schools which these subjects work at are either referred to as upper secondary school or compulsory school rather than by name.

3.1.2 The Questionnaire

The questionnaire was divided into three sections. Section one contained short answer items regarding general information such as how long the subject had been teaching, how many classes they taught, the size of the class, what material they use and so on. The second part focused on teaching where the subjects were asked to describe an "everyday" lesson and review five different models of teaching in comparison to their own. It also contained items regarding the different skills and elements of language learning. The third, main section was concerned with testing and assessment with more than fifty percent of the questions relating to the NTE in some way. It also comprised items with regards to the grading system and accompanying guidelines. The

questionnaire was initially constructed in Swedish and later translated (by the author) into English. The responses went through the same procedure.

The questionnaire was designed as a qualitative research instrument. It contained mainly qualitative items but there were also a few quantitative items with a mix of “factual”, “behavioural” and “attitudinal questions” (Dörnyei 102). Starting with the items of main interest, the open-end (qualitative), “attitudinal” and “behavioural questions”. The aim of these were to elicit personal opinions on and experiences of teaching, testing (in particular the NTE) and assessment procedures of EFL teachers working within the Swedish school system. The quantitative, short answer items were designed mostly to gather background information on the subjects and their work situations. These items were included to support or explain differences or similarities amongst the subjects' experiences and opinions. There were also a few "rank order items" (Dörnyei 107) which were included to determine what the teachers tend to highlight in their teaching, what elements they found more or less important and to instigate "clarification items" (Dörnyei 107). Appendix I provides a copy of the original questionnaire as well as the two that were answered.

3.1.3 The Interviews

The interviews were semi-structured (Dörnyei 136) in the sense that they subjects and issues which needed to be covered but not necessarily in a structured way, therefore content questions and probes were used. A number of the qualitative questions were asked at the start of the interview with the same purpose as in the questionnaire, but they also worked as "ice-breakers". Some of the original open-ended questionnaire items were used as instigators to urge the teachers to elaborate on the subject. Probes were used when the subjects needed confirmation to go into detail of the issue in question.

Two of the interviews were conducted via the internet through Adobe and Skype, both of which were recorded. The third interview was conducted via email. The two recorded interviews were carried out in Swedish and the most relevant parts have been translated and transcribed by the author. The interview which was done via email was conducted in English with the most significant information included in this investigation. The most relevant questions that were used in the interviews can be found in appendix II, table 7 and 8, along with a summary of the two transcriptions and the written interview.

3.1.4 Limitations and Analysis of Primary Data Collection

There were approximately forty requests for participation in this study sent out to different schools with an estimated two or more English teachers per school. The expected response to the requests were predicted at a low ten to fifteen percent (approximately ten respondents). However, the response was even lower than expected. Out of all of the requests, only two teachers were willing to fill out the questionnaire and participate in the study. This low response rate may be caused by a number of factors. For example, the requests might not have been forwarded to the teachers or the teachers might have found that they did not have time to participate. One teacher stated that he/she had looked at the questionnaire but found it too long and time consuming to answer, which could have been an issue for many of them. Teachers are generally busy people and during spring term, preparing for the NTE and final marking, mean they are probably a bit more busy than usual. Another issue with the questionnaire was the difference in the quality of the two responses. They were different in the sense that one was very informative and descriptive with each item addressed, while the other was short and to the point with several of the items left unaddressed (Appendix I, Table 6). Both have been included in this report because they illustrate how unreliable a questionnaire of this sort could be as well as demonstrating how vital it is with clear and precise instructions.

Flexibility is one of the reasons why interviews were chosen as a research method for the remaining subjects. An interview gives the researcher more control over what information is given. Probes can be used to elicit more information than what would have been possible in a questionnaire. Interviews also allowed items to be brought up in diverse orders, reworded and altered to suit the situation. The material and items of the questionnaire constitute a useful base for interviews and, in relation to this particular study and situation, interviews became a more reliable option. The subjects that were interviewed come from different backgrounds, they were contacted in person and agreed to participate. Two of the interviews were conducted orally through either Skype or Adobe, both of which were recorded on a computer. Extensive notes were taken mainly as backup but also as a means of recording the interviewee's thoughts and ideas. The third interview was conducted via email, which, in this case, was satisfactory. One issue to keep in mind when conducting an interview via written correspondence is the fact that the response might not be the subject's initial and natural reaction but rather an edited and constructed answer.

All of the data collected through the two different methods has been presented, analysed and incorporated into this research, because the information provided has been deemed useful in this context. However, in future research, more time should preferably be spent on contacting the subjects in person and on piloting the questionnaire in order to give an opportunity for alterations. Because of the time frame set for this study an appropriate pilot was never conducted, though it was viewed and commented on by teachers not participating in the study. A pilot study is strongly suggested for future research. Alternatively, this study could be seen as such a pilot study. The work situation for teachers and their lack of time should also be considered further. Finally, because there were only five subjects participating in this study, any theories or claims cannot be sufficiently supported. However, implications and findings from the questionnaires and interviews can be used as a foundation or starting point for future research.

3.2 The Results – A Summary of Main Item Responses

A summary of all the responses of the main items of interest can be seen in table 4, below.

These responses will be referred to in the following sections.

Table 4. *Summary of Main Item Responses*

Main Items	Summary of responses
1. Who decides the content of a lesson?	Four of the teachers answered that they did it themselves , they also mentioned that the students have some input in the content. I.e. the students can come with suggestions and express preferences wishes and then the teacher plan the content with that information in mind.
2. Who decides what material is used in class?	All of them said that they decided on all or some of it . Most of them actually said it was up to them and the students what material they use. However, one teacher stated that the "extra" material was decided upon by him/her, while what course books to use was decided upon by the Head teacher and Principal. Two of the teachers stated that they do not use any course books but instead they find and use their own material.
3. Describe how you test the students' knowledge of English and indicate what skills are tested.	<p>The teachers all assess their students continuously. Usually they use tasks and assignments related to the main skills of proficiency which have already been marked as a foundation for assessment. All but one use some sort of tests as part of the assessment process. One of the teachers uses the support material provided by the Swedish National Agency for Education.</p> <p>One other teacher stated that improvised self-made reading and listening comprehension tests were being used as part of the assessment. This teacher also stated that all of the oral tests that are used during the course are structured similarly to the spoken production tests applied in the NTE.</p> <p>* The difficulty of assessing the written aspect of proficiency might be a common issue amongst teachers. This assumption could be reflected in the re-correction the Swedish School Inspectorate has conducted on the reading and listening comprehension part as well as the written production part of the NTE, where the deviations were most noted in the written production part.</p>
4. What is your personal opinion of the NTE?	One teacher finds that the listening comprehension part is unnecessarily difficult at times, another finds the oral production part a bit too small, while a third states that the expectations of the written production part are too low.

	<p>However, the overall opinion of the NTE is positive. They all find that it tests the main skills and it is designed in a way that enables assessment on different levels. Its purpose as an assessment aid is fulfilled according to the interviewees. It supports their initial assessment of a student and it serves as confirmation that they have made right judgements of the students' level of knowledge.</p> <p>A common thread amongst the teachers' opinions is that something needs to be done about the correction procedures of the tests. A desire for an objective, centralised or external correction is voiced and it is also commented that such a solution would save time for the teachers.</p>
5. To what extent does the NTE influence the final marks?	<p>It is understood that the NTE does not influence the grades as such, but rather serves as confirmation of the accuracy of their initial assessment and marking. One teacher believe that a student should not be able to get higher grades than those achieved on the NTE because the test measures the levels of all the skills, their level of proficiency.</p>
6. Besides the NTE, what other factors are considered in the marking process?	<p>They all consider previously assessed tasks and assignments produced during the course as well as their own notes. One teacher considers previous work but states that it is the level at which the student is at by the end of the course that should be marked. It can thus be assumed that the NTE plays an important role in the marking process.</p>
7. Does your school have regulations concerning how much influence the NTE should have on the final marks?	<p>Neither of the participating schools have such regulations. One subject stated that the teacher is trusted to make a fair judgement.</p> <p>* The lack of guidelines of how much input the NTE should have on the final marks could be seen as an issue, because some schools and teachers might take the test results into account considerably more or less than others.</p>
8. Describe how you prepare the students for the NTE.	<p>In general they let the students do similar tasks to those used in the NTE. Old tests are also used as a diagnostic device. One teacher use similar tasks to the oral production part throughout the course. In addition, a few weeks before the test, the students are given reading and listening comprehension tasks. This teacher tells his/her students what to expect from the written production part, they are also given an old NTE at the end of the term in the eighth grade as a means of neutralising the test.</p>
9. Do you find the NTE a good or bad aid in the grading process?	<p>All of the teachers find the NTE to be a good aid in the grading process. As has been previously mentioned, it usually verifies their initial opinion of a student's proficiency level as well as covering the four main skills of English on all levels of the grading scale.</p>

<p>10. What is your personal opinion of the rules and regulations teachers must follow when grading?</p>	<p>One teacher had the opinion that there had been very little time to get familiar with the new grading system before it was implemented, while another found the new grading system better because it is easier to comprehend. Some of the subjects have found the grading criteria somewhat open to interpretation which arguably could lead to deviations of marks between teachers. The interviewees agree that all of the grading criteria need to be met for a student to achieve different levels.</p>
<p>11. Please describe what an “every-day” lesson would look like, how is it structured and what do you do?</p>	<p>The teachers all use a mix of teaching styles but when given a choice between a few diverse ones (see section 2.1.3) there were three in particular that took their fancy. The preferred styles were the Communicative, the Task-based and the Traditional EFL styles.</p> <p>There is no specific grammar teaching because it is rather seen as an integral part of the main skills. Communication is highlighted. They also seem to use diverse texts in their teaching in combination with corresponding tasks, written assignments and discussions.</p> <p>It appears that they make use of material they have found themselves, or that they have chosen together with the students. They also make use of films and other media as sources of material for listening comprehension and further discussions.</p>

3.3 Discussion

3.3.1 Research Question I - What do teachers working within the Swedish school system teach when they teach English, and how do they teach?

The questionnaires and interviews conducted with Swedish teachers of EFL have provided a more detailed account of what is actually being taught. The subjects seem to highlight the four main skills, and the micro skills of communication and interaction in their teaching. Another aspect common amongst the teachers was the fact that they do not highlight specific grammar instructions in their teaching. Instead of grammar being a separate element of language learning, it is seen as an integrated part of the four main skills. It can be argued that this is in cohesion with the syllabus where it is stated that teaching English should not contain a focus on the different components but rather on the language as a whole (Kommentarmaterial 6). However, one of the teachers maintained that vocabulary is the most important element and argues that it is the foundation upon which learning a language is built (Appendix II, Table 7: B12). In contrast, the other teachers consider

vocabulary to be incorporated in the other elements and it is not as highlighted.

Because there are no official guidelines or rules of how teaching English in Swedish schools should be conducted, the way of teaching is decided on a local level. In the investigating interviews five different teaching styles, based on different models of teaching (the styles and models are summaries from Cook's *Second Language Learning and Language Teaching*), were presented to the subjects (see table 1, section 2.1.3). They were then asked to identify the ones they found the most appealing and if any of them resemble their own way of teaching. Neither of the subjects chose only one style but rather a combination of two or more. One of the subjects asserted that he/she alternates between all of the styles. Cook states that teaching styles are somewhat linked to models and characteristic of learning a second language (235-6). He further proposes that "there are many area-specific models" with "no overall framework" (233), which could explain why neither of the subjects chose only one teaching style.

The interviews revealed that the subjects tend to more or less include their students in the planning of a lesson. Usually the students have some sort of input into the content of a lesson in the form of ideas and suggestions of what they would like to do. The teacher then takes their opinions into consideration when planning a lesson, but in the end it is the teacher who decides the content (Table 4, Item 1-2). In addition, the teachers seem to use diverse texts in their teaching in combination with corresponding tasks, written assignments and discussions. They also make use of films and other media, such as the internet and youtube, as sources of material for listening comprehension tasks and discussions. Both use of material and incorporation of students' wishes in their teaching corresponds to the syllabus (Curriculum 32).

A student's confidence in using English rather than it being one hundred percent correct can be seen as one of the main aims for some of the interviewed teachers, which also corresponds to the syllabus (Curriculum 32). Both subject B (Appendix II, Table 7: B11) and C (Appendix II, Table 7: C9) maintain that confidence in using the language is an important factor, subject B even states that

confidence is needed for the student to be able to progress. According to the subjects, motivation is a very important factor in successful learning. Some of the subjects believe that the notion of English as an aid in future work situations may act as motivation for learning English. In addition, the cultural aspect was seen to motivate the students (Appendix II, Table 7: C13), and thus both the professional and personal purposes and goals of the curriculum can be found as integral parts of teaching English amongst the interviewed teachers.

3.3.1.1 A) - How do they assess their students' proficiency?

The NTE is seen as a satisfactory assessment aid and all of the teachers take the results into consideration along with results from additional assessments when setting the final marks. Subject D (Appendix I, Table 6: D12) mentions feedback from both him/her and the students as part of the teaching. It was not clearly stated that self assessment was used but considering the statement one could argue that this method is at least part of his/her teaching strategy. Besides Subject D's comment, there was not much mentioned of self assessment. Instead, it seems as if teacher feedback is the preferred way. As previously mentioned (Table 4, Item 3), the teachers seem to evaluate the students' efforts continuously by assessing previously marked tasks and class performance. This leads to the assumption that they, in addition to the results of the NTE, make use of some sort of portfolio model in the assessment and grading process. However, due to the lack of empirical data this assumption can not be supported, though it does suggest a topic for future research.

In addition to the NTE, all but one of the subjects use different types of tests as part of the assessment process (Appendix II, Table 7: C3) but through feedback these tests are also supposed to advance the students' learning. Subject B uses the support material provided by the Swedish National Agency for Education. The material contains tests with keys and assessment scales. "It is nationally standardised, allowing results under the same conditions for everyone" (Appendix II, Table 7: B3). The teacher finds this material helpful in the assessment process because it is both

reliable and saves time. It is thus evident that testing is part of the assessment process for most of the subjects, though it is not central.

3.3.2 Research Question II – What are their personal opinions of the NTE?

The investigating interviews suggest that teachers believe the test measures what they themselves regard as important elements of language proficiency, and also corresponds well to the curriculum. One teacher mentions that the NTE is "such a realistic reflection of what the students would use English for in the future (the ability not just to pronounce and answer questions in English but to keep a conversation going; focus on communication and content rather than technical perfection; etc.)" (Appendix II, Table 8: E1). As can be seen in Table 4, item 4, the overall opinion of the NTE is positive with minor complaints.

The main exception to this positive notion concerns the marking procedure. The Swedish Schools Inspectorate has conducted an investigation into the correction of the reading and listening comprehension part as well as the written production part of the National Test of English (NTE). They have re-corrected a number of tests which have revealed deviations in the results (Plunkett et al. 11). The deviations of the NTE re-corrections is a cause of concern for all of the teachers. The part that had the largest proportion of deviations was the written composition where students had been given better marks by their teacher than those given by the test panel (Plunkett et al. 11). Both Subject B and E find the written production part (and, to some extent, the oral production part) particularly difficult to assess (Appendix II, Table 7: B3; Table 8:E1). Assessment of written and oral tasks are subjective in the sense that there is never only one correct answer, which could explain the perceived difficulty. The complexity of assessing the written aspect of proficiency might be a common issue amongst teachers which could explain some of the deviations.

It is up to the marker to analyse the task which to a degree contributes to variations in the outcome because of individual differences amongst markers. These differences can be minimised by sufficient support material such as assessment scales, having more than one scorer and by training (Harmer 329). For the NTE, teachers are given support material but it can be interpreted freely to some extent, as one of the interviewees pointed out (Appendix II, Table 8: E1). This aspect of interpretation was also mentioned by Subject B in relation to the grading requirements they follow (Appendix II, Table 7: B10). A more structured way of assessing and marking the written part of the NTE, similar to the method used by the Cambridge testing system, which applies two mark schemes in combination with the IELTS (International English Language Testing system) nine-band marking scale (See section 2.3.4.1), could be beneficial. However, the introduction of a ten-band assessment scale (Skolverket I) might be the solution to this issue. The teachers are also advised to correct the tests in cooperation with colleagues but whether or not this is done is left for the individual or school to decide (Erickson 1). Clear directions and regulations in this matter might be needed to ensure that inter-rater reliability is maintained. Training is also one aspect which has been suggested as a solution to the deviations (Skolverket I). However, the interviewees suggested that objective correction procedures would be the best option for solving the issues of bias and subjective test results.

3.3.2.1 A) - How do they prepare their students for the NTE?

Old copies of the NTE are used as preparation material, a diagnostic device and as some sort of placement test. As stated in Table 4, Item 8, above all, the interviewees tend to present similar tasks to that of the NTE to their students. Instructions on how it is conducted, what to expect as well as suggestions on how to approach it are presented and it can thus be argued that the students are all well prepared for the test. Whether or not a washback effect can be noticed can not be determined due to lack of data. However, Subject E mentions that he/she “teach[es] to the test” (Appendix II,

Table 8: E1) to some extent and continues to explain that this is due to the realistic use of English presented in the NTE. This study did not produce sufficient data on the phenomenon of washback, but it has suggested that further investigation into the matter could be undertaken in connection with Swedish schools and the NTE.

3.3.2.2 B) - Is the NTE regarded as a reliable and valid grading aid by the teachers who use it, and what status does it hold in contrast to the Cambridge testing system?

As has been mentioned previously, the teachers that were interviewed in this study all agree that the NTE is a good aid in the grading process. They also maintain that it tests all the main skills in a way that facilitates assessment on different levels. In addition, they stated that it corresponds to their initial assessment of the students. These factors suggest that the validity and reliability of the test are up to standard according to the interviewed teachers. However, the issues of the deviations in relation to the correction of it suggests that both reliability as well as validity can be questioned and it is something that the teachers find worrying. In comparison to the Cambridge testing system it has many similarities which would indicate that it could be counted as reliable as well as valid. However, there is one major difference between the two testing systems which undermines both the validity and reliability of the NTE. That difference lies in the marking procedure. It can be argued that the marking procedure for the NTE are unsatisfactory, both from the teachers' point of view as well as in comparison to the Cambridge testing system (See section 2.3.4.1, Table 2)

Furthermore, Subject C has used a year nine NTE as a placement test for his/her first year students at the gymnasium where he/she works. The results have indicated that some of the students who have been graded with a pass for the English module in the ninth grade, have performed well below the criteria for a pass on the test. He/she argues that these students should not have been given a pass on the module because the test indicates that their knowledge and ability is not

sufficient. The teacher is questioning why these, as he/she calls it, “charity grades and disagreeing results” have been given (Appendix II, Table 7: C3). This scenario is an issue which should be investigated further, because it indicates that the score meaning might not be valid or has been ignored.

3.3.3 Research Question III - What does the interrelation of teaching, testing and assessing EFL in the Swedish school system appear to be?

It is clear from the start that teaching, testing and assessment are deeply interrelated. The underlying factor between the three is the syllabus. All of the subjects in the study seem to believe in the syllabus and what it sets out to promote in EFL learning. They also find that the NTE, which corresponds to the syllabus, measures different levels of proficiency that they both teach and encourage. The NTE seems to be a solid part of their assessment process where it works as a verification of their own assessments. From the data gathered in the interviews there are tendencies suggesting that some of the subjects may teach to the test (NTE) to a certain degree (Appendix II, Table 8: E1; Appendix I, Table 6: A3, A8). Because the test is so closely related to the syllabus the notion of a washback effect cannot be deemed as negative.

One teacher mentions that he/she finds that the test is designed to measure the knowledge and abilities of English which the students will most likely need for future use (Appendix II, Table 8: E1). It can be argued that what Subject E mentions is one of the main purposes of English as a subject in the Swedish school system (Kommentarmaterial 10). The English syllabus is further related to the Common European Framework of Reference (CEFR) which is considered to be the standardised and internationally recognised framework for levels of language proficiency. This connection makes it easier to compare and contrast levels of proficiency in different contexts.

ALTE (Association of Language Testers in Europe) has contributed to the both the syllabus and the grade scale as well as the NTE with the "Can Do" (Cambridge G) project, which focus on what students can do rather than on their weaknesses and what they cannot do (Curriculum 39-42, Erickson 3). “The aim of the 'Can Do' project is to develop and validate a set of performance-related scales, describing what learners can actually do in the foreign language” (The ALTE Framework 3). These statements have been translated into twelve languages which “offer the chance to demonstrate equivalences between the examination systems of ALTE members” (The ALTE Framework 3). It is further stated that the 'Can Do' statements correspond to the levels of the CEFR (The ALTE Framework 7). The most basic level, “ALTE Breakthrough”, of 'Can Do' correspond to A1 (see table 5, below) of CEFR, continuing with “ALTE Level 1” corresponding to A2 and so on (The ALTE Framework 7). The "Can Do" project and the CEFR are both fundamental factors in the development of the Cambridge testing system.(Cambridge E, D).

The following table has been taken from the online version of *the Common European Framework of Reference for Languages* and it shows the different levels of proficiency which both the Swedish school system as well as the Cambridge testing system follow.

Table 5. *Common Reference Levels: global scale*

Proficient User	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
Independent User	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes, and ambitions and briefly give reason and explanations for opinions and plans.
Basic User	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and familiar information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

Common European Framework of Reference for Languages, 24.

Conclusion

This study set out to investigate the relationship between teaching testing and assessment in the Swedish school system. It has focused on testing, in particular the nationally standardised proficiency test of the Swedish school system, the NTE. Interviews were conducted as an attempt to acquire personal opinions and experiences of five EFL teachers working within the Swedish school system. The test was also compared and contrasted to the internationally recognised testing system of the Cambridge ESOL Examinations. Purpose, validity and reliability are factors of a test which were also highlighted. In addition, these factors were considered with regards to the examination of the NTE and the relation to both the interviewees and the Cambridge testing system.

It can be argued that the common factor between teaching testing and assessment of English in the Swedish school system is the syllabus. Both the syllabus and the National Test of English (NTE) have got strong connections to the Common European Framework of Reference (CEFR) and the Association of Language Testers in Europe (ALTE) which promote equal assessment of proficiency as well as offering attitudes towards assessment. The Cambridge testing system also has strong connections to the CEFR and ALTE. The study showed that the subjects of the investigation seem to teach and assess in accordance to the curricula and the syllabus. In addition, the NTE is seen as a valuable aid in the assessment process because it measures abilities and proficiency on different levels consistent with the teachers' own evaluations.

When assessing their students, all the subjects seem to employ a similar type of model: continuous assessment associated with the portfolio approach. Besides the NTE, other types of testing seem to be part of their assessment, with the exception of subject C who states that he/she does not use any tests except for the NTE. The NTE is used as part of the subjects' assessment of a student but it does not decide what final grades students actually gets. It serves more as confirmation that the initial evaluation was correct. The NTE seems to be a small part of their teaching throughout the module; they either use it as a placement test or preparation test or just for

practice. There is a notion of a positive washback effect from the NTE, but because of the small scale of this study, such a statement cannot be sufficiently supported. Washback is an issue which could be investigated further, bearing Spratt's article in mind. One of the teachers admitted that he/she "teach[es] to the test" but found that this was so because the test represents abilities and proficiency in the desired manner.

However, when comparing and contrasting the NTE with the Cambridge testing system, one deviation was made clear. The diverse manners in which the correction of the different tests are conducted. This is the foundation of the argument that validity and reliability of the NTE could be deemed unsatisfactory, especially with regards to inter-rater reliability. The validity of a test cannot be confirmed, instead it should be defended with evidence supporting any assumptions and statements made in relation to the test. Score meaning and validity are important factors in a validity argument and because of deviations in test results as well as subjects C's findings of inconsistencies between grades and results of the NTE, both can be questioned. The teachers' attitudes towards the deviational test results gives the impression that there is a need for change in the correction procedure. The overall positive feedback on the test itself is a confirmation of a well structured test but the doubts directed towards the correction issue is a matter that should be dealt with further. This study could be the base for future research on this topic.

Works Cited

- Bailey, Kathleen M. "Washback in Language Testing." *The TOEFL Monograph Series*. New Jersey: Educational Testing Service. June 1999. Web 4 April. 2012.
<<http://www.ets.org/Media/Research/pdf/RM-99-04.pdf>>
- Cambridge. See Appendix III
- Cambridge Key English First for School Handbook for Teachers*. University of Cambridge ESOL Examinations. Council of Europe. 2010. Web 3 June. 2012.
<<http://www.cambridgeesol.org/assets/pdf/exams/fce-schools/fce-schools-handbook.pdf>>
- Chapelle, Carol A. "Validity Argument for Language Assessment: The Framework Is Simple...." *Language Testing* 29.1 (2012): 19-27. Web. 23 March. 2012.
<<http://ltj.sagepub.com>>
- Common European Framework of References for Languages*. Education and Languages. coe.int/lang. Council of Europe. Web. 1 April 2012.
<http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf>
- Cook, Vivian. *Second Language Learning and Language Teaching*. 4th ed. London: Hodder Education, 2008. Print.
- Demaret, Marianne. "The "NAFS" Project." *University of Gothenburg National Assessment Project*. University of Gothenburg, 2011. Web. 10 Feb. 2012.
<<http://www.nafs.gu.se/english/information/>> See Appendix VI
- Dörnyei, Zoltán. *Research Methods in Applied Linguistics*. Oxford: Oxford University Press. 2007. Print.
- Erickson, Gudrun. "National Assessment of Foreign Languages in Sweden." 2009. *University of Gothenburg National Assessment Project*. University of Gothenburg, 2011. Web. 20 Feb. 2012.
<<http://www.nafs.gu.se/>>

- Harmer, Jeremy. *The Practice of English Language Teaching*. 3rd ed. Harlow: Pearson Education Limited, 2001. Print
- Howell, Jonathan, Paul Miller, Hyun Hee Park, Deborah Sattler, Todd Schack, Eric Sperry, Shelley Widhalm, and Mike Palmquist. *Reliability and Validity*. 2005. Writing@CSU. Colorado State University Department of English. Web. 23 March. 2012.
<<http://writing.colostate.edu/guides/research/relval/com2a5.cfm>>
- IELTS. International English Language Testing System. ielts.org. Web 3 June 2012.
<http://www.ielts.org/researchers/score_processing_and_reporting.aspx>
- Kane, Michael. "Validating Score Interpretations and Uses: Messick Lecture, Language Testing Research Colloquium, Cambridge, April 2010." *Language Testing* 29.1 (2012): 3-17. Web. 23 March. 2012.
<<http://ltj.sagepub.com>>
- Key English Tests for School*. University of Cambridge ESOL Examinations. Web. 3 June. 2012.
<http://www.lttc.ntu.edu.tw/cambridge/ms/Handbook/KET/ket_schools_handbook.pdf>
- Kommentarmaterial till kursplanen i Engelska*. Kommentarmaterial till kursplanerna. Skolverket.se. 2011. Web. 1 April 2012.
<<http://www.skolverket.se/forskola-och-skola/grundskoleutbildning/stodmaterial/kommentarmaterial-till-kursplanerna-1.122644>>
- Lightbown, Patsy M. and Nina Spada. *How Languages are Learned*. 3rd ed. Oxford: Oxford University Press. 2006. Print.
- Mueller, Jon. Portfolios. *Authentic Assessment Toolbox*. 2011. Web. 13 April 2012.
<<http://jfmuellet.faculty.noctrl.edu/toolbox/portfolios.htm>>
- NAFS. See Appendix III
- Plunkett, Arletta, Hampus Strömberg, Linn Brandelius, Sara Merenius, Gabriella Bremberg. "Lika eller olika? Omrättning av nationella prov i grundskolan och gymnasieskolan."

Skolinspektionen. Skolinspektionen.se. Web. 22 Feb. 2012.

<<http://www.skolinspektionen.se/Documents/omrattning/omratt2011-slutrapport.pdf>>

“Rekommendationer”. *Skolinspektionen*. Skolinspektionen.se. Web. 3 June. 2012.

<<http://www.skolinspektionen.se/sv/Tillsyn--granskning/Central-rattning-av-nationella-prov/Rekommendationer/>>

Ramstedt, Kristian, et.al. (in co-operation with his staff) Transl. Anders Auer "National

Assessment and Grading in the Swedish School System." 2005. *Skolverket*. skolverket.se.

Web. 5 April. 2012.

<<http://www.skolverket.se/2.3894/publicerat/publikationer>>

Rolheiser, Carol and John A. Ross. "Student Self-Evaluation: What Research Says and What

Practice Shows." *Center for Development and Learning*. cdl.org. Web. 29 April 2012.

<http://www.cdl.org/resource-library/articles/self_eval.php>

Sireci, Stephen G. "Comments on Lissitz and Samuelsen On Validity Theory and Test Validation."

Educational Researcher 36.8 (2007) 477–81. Web. 5 April. 2012.

<<http://www.ethiopia-ed.net/images/273057532.pdf>>

Skolverket. Skolverket.se. Web. 10 Feb. 2012.

<<http://www.skolverket.se/>>

Spratt, Mary. "Washback and the Classroom: the Implications for Teaching and Learning of Studies

of Washback from Exams." *Language Teaching Research* 9.1 (2005): 5–29. Web. 13

March. 2012.

<<http://ltr.sagepub.com>>

“Test.” *Oxford Advanced Learner's Dictionary*. 2011. Oxford University Press. Web. 19 July. 2012.

<<http://oald8.oxfordlearnersdictionaries.com/dictionary/test>>

Trochim, William M.K. "Types of reliability." 2006. Measurement. Research Methods Knowledge

Base. Web. 23 March. 2012.

<<http://www.socialresearchmethods.net/kb/reotypes.php>>

"Washback Effect." *Encyclopedic Dictionary of Applied Linguistics*. Ed. Keith Johnson and Helen Johnson. 1999. Blackwell Reference Online. Web. 15 April. 2012.

<http://www.blackwellreference.com/public/tocnode?id=g9780631214823_chunk_g978063121482327_ss1-1#citation>

Appendices

I. Questionnaire

PART 1

Name and type of school:

1. How long have you been teaching English?
2. How many classes do you have?
3. How many students per class?
4. How many of your students speak Swedish as their mother tongue?
5. How do you handle a difference of knowledge and proficiency between the students?
6. Who decides the content of the lessons?
7. Who decides what material is used?
8. Are you satisfied with equipment, material and classroom available to you? (*Explain briefly what makes it satisfactory or what might be improved.*)

PART 2

1. What do you think motivates students to learn English?
2. What role does motivation have in connection to successful learning?
3. What **element/s of the learning process** do you emphasize? Indicate in % how much time you think should be spent on each element.

Grammar _____ **Vocabulary** _____ **Reading** _____ **Writing** ____
Listening _____ **Communication** _____ **Other** (*specify what*) _____

4. Following is a brief summary of five different teaching styles and models in connection to second language acquisition.

A) The Academic Style - Grammar and translation. (Focus on the written language)

B) The Audio-lingual Style - 3 Steps;

- 1) *Dialogue*: Listen, repeat, act it out;
- 2) *Structure drill*: Repeatedly practice grammatical structures;
- 3) Create new dialogues by using the information acquired in step 1 and 2.

C) The Communicative Style - Communication is both the goal and means of teaching via role play and different tasks.

D) The Task-based Learning Style - Meaning-based tasks with distinct results. The teacher decides on task, the students complete them, report back and discuss the results.

E) The Traditional EFL Style (English as a Foreign Language) - 3 Steps; 1)

The teacher introduces new vocabulary and grammatical structures;

- 2) *Dialogue*: Listen, repeat and ask questions;
- 3) Controlled role play with a template.

Please specify if any of these styles appeal to you more and explain why.

5. Do any of the styles resemble the one you use in your teaching?
6. Describe what an everyday lesson would look like for you, how is it structured?

PART 3

1. Describe how you test students' knowledge in English and indicate what elements are tested.
2. What is your opinion on the English national test?
3. How much does the national test influence the final grade?
4. Also indicate what other factors are considered when you decide on grades.
5. Does your school have regulations concerning how much influence the national test can have on the final grades? If so, what do they read?
6. Briefly describe how you prepare the students for the test.
7. Do you find the national test a good or bad aid in the grading process? Please explain.
8. What is your personal opinion concerning the rules and regulations teachers have to follow when grading?

9. Are these regulations open for personal interpretation?
 10. What is your personal opinion of the new grading system?

Table 6, below, displays chosen questionnaire responses made by subject A and D.

Table 6. *Questionnaire Responses*

Questions	Subject A		Subject D	
1. Who decides the content of the lessons?	A1	I do, partially in consultation with the students.	D1	Mostly me, but the students are welcome to come with ideas or suggestions. I also decide what material we use.
2. Who decides what material is used?	A2	Same as the previous answer. I do, partially in consultation with the students.	D2	See D1
3. Describe how you test the students' knowledge of English and indicate what skills are tested.	A3	We often read whole class books of a typical Penguin character. During the reading we work with vocabulary, discuss the content and have short writing tasks related to the content. When we are finished reading we watch the movie version with English subtitles. Afterwards the students either write a reflective paper in English or alternatively they answer questions on the content. Sometimes we have a group discussion as the final task. In between I have unprepared reading and listening comprehension tests which I have constructed myself. I also assess the students through out the process.	D3	The students are continuously being tested on oral and communicative capability, also reading and listening comprehension.
4. What is your personal opinion of the NTE?	A4	I find the writing tasks and the reading comprehension, as well as the oral tasks to be very well constructed while the listening comprehension some times has parts which are unnecessarily hard. Sometimes even I have problems to keep up.	D4	It is a good aid in the assessment procedure, but I would like to have objective corrections of them.

5. To what extent does the NTE influence the final marks?	A5	After eight years I would say that the national tests often give the results the students are already on, but that doesn't mean that the test is the most important. I consider it more of a receipt which confirms that what I do and test over the years is in good agreement.	D5	I don't really know, it is just a part of everything else the student have produced.
6. Besides the NTE, what other factors are considered in the marking process?	A6	The combined efforts.	D6	Everything else that they produce.
7. Does your school have regulations concerning how much influence the NTE should have on the final marks?	A7	No.	D7	No.
8. Describe how you prepare the students for the NTE.	A8	At the end of the spring term of the eighth grade, the students get to do an old NTE, reading and listening comprehension, to neutralise and defuse it as well as using it for diagnostic purposes. A few weeks before the test we will do a few listen and reading comprehension tasks. I tell them what the writing task usually looks like. All of my oral tests, which I have used over the years, are designed the same way as those on the NTE.	D8	We go over areas of grammar where the students are insecure and we go through conjunctions. I introduce different types of texts to them and they get to practice listening and reading comprehension. I also go over learning strategies.
9. Do you find the NTE a good or bad aid in the grading process?	A9	It is good, it usually confirm what I already know about the student's knowledge.	D9	See D4

10. What is your personal opinion of the rules and regulations teachers must follow when marking?	A10	They are rules and regulations, which we simply just have to comply with.	D10	A very short time to get to know the new grading system.
11. Are these regulations open to personal interpretation ?	A11	To some extent I find that they are.	D11	? Author's note: The subject answered with a question mark which suggests that this item is not clear enough.
12. Please describe what an "everyday" lesson would look like, how is it structured and what do you do?	A12	It is very varied. I work in blocks based on movies and whole-class-books. If we are at the beginning of a block we might have a few "four-corner tasks", go over vocabulary that we encounter in the movie or the book, or we look at pictures of characters that are in them. If we are in the middle of a block we are watching movies with English subtitles. I pause in between to discuss the content or just to check that the students are following it. If we are reading we alternate between different recitation variants such as chain reading, couple reading, role play and so on. If we have finished with the movie or book we might be in a writing frame of mind which creates a creative chaos. I bounce between the students discussing phrasing and word choice while they work. If they have finished a larger paper I have collected some writing rules and linguistic problems which we need to discuss and we also practise these. Sometimes we go on youtube and listen to interviews with interesting people.	D12	I don't know if we have an everyday lesson. We discuss a lot and write a lot. Usually a short briefing followed by questions and discussions. We do the tasks followed by feedback both from me and the students.

<p>13. Which of the mentioned teaching styles appeal to you more and does any of them resemble the way you teach?</p>	<p>A13</p>	<p>I have never been a fan of ready-made solutions but rather prefer to put together a package depending on the group I am working with. If I have to choose one it is the communicative model that appeals to me the most, which would also be the model resembling my personal style the best.</p>	<p>D13</p>	<p>I use all of the models but some more and some less.</p>
<p>14. What skills or elements of the learning process do you highlight the most? Indicate in percentage how much time you think should be spent on each element, try to make it 100 percent in total.</p>	<p>A14</p>	<p>Twenty-five percent on each of the main skills; reading and listening comprehension as well as oral communication and written production. I find that vocabulary and grammar are included in the other elements, they come with the package.</p>	<p>D14</p>	<p>It is hard to say because it depends on the needs of the student.</p>

II. Interviews

This section contain questions and responses from interviews with subject B, C and E which can be found in tables 7 and 8.

Table 7. *Interview Responses - Oral*

Questions	Subject B		Subject C	
1. Who decides the content of a lesson?	B1	The Students can make suggestions and then I plan the lessons considering those.	C1	I do. We do not use textbooks instead we use my own material that I produce or find elsewhere.
2. Who decides what material should be used?	B2	I decide all the extra material that we use but it is the principal or the school board that decides what course books we use.	C2	See C1
3. Describe how you test the students' knowledge of English and indicate what skills are tested.	B3	The upper level classes have got a teacher's guide with related tests and keys. These tests are standardised nationally, allowing results under the same conditions for everyone. Because it is already prepared with grading scales and keys, I save a lot of time which is good. Though I find it hard to assess written tasks, what focus should lie on and all that. I usually set myself three specific criteria to focus on, but it is hard to assess.	C3	I don't use tests at all, except for the ninth grade NTE at the start of a module. I use it as a way to evaluate what level the students are at. Though the results that I have seen on these tests, are shocking! Out of sixty students there were sixteen that did not even come close to passing it. These students had all passed their English course in the ninth grade which is both confusing and worrying. The assessment of the previous teachers can be seriously questioned. I think that the NTE portrays the level of knowledge the students are at and the results these sixteen students got were so low that it's not possible they passed it in the ninth grade. In reality, these students have not reached the level of proficiency that they should have. I'm not criticizing the test but rather the people who correct it and those who set the final marks. We need to ask what lies behind these charity grades and disagreeing results. Maybe the teacher is pressured by the school's annual reports on grades, or maybe they have been

3. CONTINUED			C3 C O N T I N U E D	told to "do well" in connection with salary negotiations. They could feel that these results reflect if they are good or bad teachers. What ever it is, it needs to be looked into. A centralised and objective correction could be a possible solution.
4. What is your personal opinion of the NTE?	B4	It is generally good. All the grading criteria are baked in to the test and it makes it a good overall test. I just find the oral sub-test to be very small, so therefore I focus more on that assessment in class.	C4	It is really good! It tests all the base elements and it is constructed in a way that enables assessment on different knowledge levels. It strengthens the initial assessment of the students.
5. To what extent does the NTE influence the final marks?	B5	Between twenty and twenty-five percent. It works as a confirmation really. The students usually get the results and grades that they were already at. The test only confirms that my assessment was right in the first place. If someone has had bad results we usually sit down and talk to them to figure out what went wrong, and they might possibly get a second chance in proving their knowledge. I find it is easier to raise a student's grade from the NTE than it is to lower it.	C5	The fact that teachers do not consider the test results when they are deciding the grades is very bothersome. Students should not be able to get higher grades than what they got on the NTE. Because it tests all the basic requirements with a level of difficulty enabling a fair grade for all levels. Teachers are the problem both when it comes to final grading as well as when it comes to correction of the test. I just wonder what the underlying factor is that is contributing to the false grades. I call these "charity-grades". All the criteria need to be met for a grade to be set, which is something some teachers seem to ignore or miss.
6. Besides the NTE, what other factors are considered in the marking process?	B6	Things that have been documented throughout the school year like notes and assignments and that.	C6 C6	The NT is a good base, but I tend to assess my students as the course is going, through tasks and assignments they have handed in. I give them feedback and let them know areas they need to practice on and improve. But at the end of the day, it is the level of knowledge that the student is at when they finish their course that should be

6. Continued			C	graded. I don't set their final marks because it is done by the teacher they have for the last bit of the module. You see, in our school the module that usually runs over one year, runs over two years. I assist the other teacher if there are any uncertainties, if necessary we usually go through everything the student have done.
7. Does your school have regulations concerning how much influence the NTE should have on the final marks?	B7	No, there are no conditions concerning the NTE. The teacher is trusted to make a fair judgement.	C7	- Subject C was not asked this question -
8. Describe how you prepare the students for the NTE.	B8	They get to do similar tasks but the oral part is a bit tricky.	C8	I don't do that because it is the teacher they have in the second year that does the NTE and not me. I like to think that my teaching will provide them with the tools they need in the future anyway.
9. Do you find the NTE a good or bad aid in the grading process?	B9	It is a good aid! The deviations in the correction of tests are a bit worrying. One has to trust that all the teachers assess fairly. I think that external or centralised correction would be helpful and it would also give us teachers more time, which is something we do not have a lot of.	C9	Like I have said before, it is a good base, and it works as a confirmation that the initial assessment was right.
10. What is your personal opinion of the rules and regulations teachers must follow when marking?	B10	The old grading system was unclear, I found it hard to see what requirements for each criterion were required. The new system is better, it is clearer, but sometimes it is also hard to interpret the conditions. For example what do they mean with "good knowledge" (Sw: "goda kunskaper")?	C10	There are specific criteria that need to be met for different grades, and all of these criteria must be fulfilled! A student cannot get a pass if not all of the criteria are met and they can't raise their grades unless all of the goals are fulfilled. That's just the way it is.

<p>11. Which of the mentioned teaching types appeal to you more and does any of them resemble the way you teach?</p>	<p>B11</p>	<p>Type C and D** - I think communication and oral production are important factors, it helps the student to get over their fear of making mistakes which in turn enables them to achieve more in other areas as well. The books that we are using follow the task-based model which is good because it covers all the elements of language learning, including grammar. I don't focus on grammar but rather see it as an aid to the other elements.</p>	<p>C11</p>	<p>Note by author: Subject C was not asked this specific question, but by the way he/she describes his/her teaching it sounds like the Task-based style or the Traditional EFL style (See section 2.1.3 for a brief summary of the diverse teaching styles).</p>
<p>12. What skills or elements of the learning process do you highlight the most?</p>	<p>B12</p>	<p>*** Vocabulary is the most important element – without the words you cannot move on, vocabulary is the base for all the other elements. It is hard to teach them grammar which has got just as many exceptions as there are rules, students are just not interested! As I mentioned before, communication and oral production is also something I highlight. So maybe twenty each on vocabulary and communication, and fifteen each on reading, listening and writing so grammar gets maybe five - ten percent.</p>	<p>C12</p>	<p>I am a dedicated opponent to grammatical tasks. Grammar is baked into the other elements and I do not put explicit time and effort into it. Grammar drilling and tasks, such as filling in gaps and translation, does not work. This is my opinion anyway.</p>
<p>13. Please describe what an "everyday" lesson would look like, how is it structured and what do you do?</p>	<p>B13</p>	<p>Well like I have said already, we work with the book or discuss or maybe watch a movie. Sometimes they make presentations and so on. We tend to do a lot of oral tasks or just have discussions. With year eight I have basically only had the oral and communicative elements. And I only teach in English with them. In the beginning it was quite quiet but they eventually overcame the embarrassment of making a mistake, and now all of them talk. I have noticed a great evolution amongst these</p>	<p>C13</p>	<p>I usually present a text, which they either listen to or read or both, then there are questions to the text which they either answer individually or in a discussion. They all get the time they need to finish the text . Then I usually have them write an analysis, summary or maybe an ending of their own, which gives them the opportunity to place themselves at their own level. I give feedback on the grammar they need to work on for each student individually. I also have them write papers and create PowerPoint presentations on</p>

13. Continued	B13 C O N T I N U E D	C13 C O N T I N U E D
---------------	--	--

* Following is an approximate summary of the probe used in the interview with subject B (based on the questionnaire). "Indicate in percentage how much weight you put on the test when deciding the final marks".

** C – Communicative Style and D – Task-based Style (see Section 2.1.3 for a brief summary of the styles)

*** Following is an approximate summary of the probe used in the interview with subject B (based on the questionnaire). "Indicate in percentage how much time you think should be spent on each element, try to make it 100 percent in total".

Table 8. Interview Responses - Written

Question	Subject E	
1. What is your personal opinion of the National test?	E1	<p>I think the test is a very good measure of the students' fluency and abilities in English. In America there's always a lot of talk and criticism of "teaching to the test" -- that so much pressure has been placed on teachers to get their students to perform well on standardized tests that lessons center on nothing but training for said test. I admit without shame that my lesson plans are often centered around training for the national test, but that's because I think the test is such a realistic reflection of what the students would use English for in the future (the ability not just to pronounce and answer questions in English but to keep a conversation going; focus on communication and content rather than technical perfection; etc.).</p> <p>My criticisms of the test are two-fold. One is the grading instructions (bedömningsanvisnarna) are not as helpful as they could be for the written and spoken portions of the test. To judge a person's ability to speak or write in a second language is so subjective, and I guess I would like a more methodical way of breaking the essay/conversation down. In the Skolverket's <i>bedömningsanvisningar</i> (The assessment supportmaterial provided by the Swedish National Agency for Education - enterpretation by author) they use vague ideas such as "avancerat ordförråd" ("advanced vocabulary" - enterpretation by author) (how advanced is advanced?) and judge each text/conversation in several paragraphs of flowing text where you can't be sure which individual sentences are, say, at the VG-</p>

1. CONTINUED	E1 C O N T I N U E D	<p>level that matches the students grade, and which are perhaps better or worse than that. I've read through my colleague's <i>bedömningsanvisningar</i> (See above) for the test in "svenska som andra språk 1" (Swedish as a second language 1) and think they have actually given me much more insight into how to judge a text's quality. But this is perhaps more of a criticism of my own training in judging the quality of "subjective" assignments than of the test itself!</p> <p>My other criticism of the test is the very low standard expected of the student's written English. Again, this might be a criticism of the course requirements rather than of the test, but when I see what is acceptable at the G-level, I wonder what we actually think we are training the students for. To write a job application, a letter asking to rent an apartment, etc. with a quality that poor seems to me to be a bit pointless. I would argue with myself and say that very few of my students will need to write in English in the future, but one wonders why writing should be included in the course at all in that case.</p>
2. How do you prepare your students for the test?	E2	<p>We spend about 6 weeks (which for us is just 6 lessons) before the test doing exercises that are either taken from old NP or that I've designed to simulate NP. The reading- and listening comprehension tests are, of course, very straight forward, and my goal is therefore for them to see what the tests look like and get an idea of their level. I also give them tips about how to do their best, avoid common mistakes, and avoid nervousness (leave nothing blank! Spelling and grammar don't matter on the reading and listening parts! etc.). This is even more important for the speaking and writing parts, as I've found students' poor performance on these tests is often due more to misunderstanding what is expecting of them (for instance, that they are meant to have an interactive conversation -- not just take turns talking -- on the speaking test) than on lack of ability. In this case I actually do a mock-test with the students in pairs and then we listen to examples of students doing the test.</p> <p>For the writing portion I take my computer with a projector and we write a text "together" as a class, everyone giving tips as to how we can make the text better. During these "training sessions" we focus on content, organisation and variation (making "un-boring" sentences) and not on spelling and grammar -- I figure it's "too late" to work on that at this point in the year and I also want to make sure the students know that content and communication is the focus.</p>

<p>3. When deciding the final grades, how much do you rely on the National test and what other factors do you consider?</p>	<p>E3</p>	<p>I usually have an idea of what grade I want to give each student even before the national test, and luckily the test normally matches my own judgement! There are criteria in the course plan that aren't tested in the test (culture and literature; ability to "bearbeta en text" – (work on a text – Interpretation by author)), so assignments I've given in those areas are also factors when I set grades. It's also, of course, possible for the student to get a passing grade on the test as a whole but not fulfill all of the criteria for a passing grade. Every year I have a couple of students who get IG in the course despite having G on the test -- that usually results in some aggressive phone calls, but if they pass the test because they're gotten good grades on 3 out of 4 parts but are so afraid of speaking that they have never said a word of English, well!</p>
<p>4. Besides the National test, how do you assess your students' proficiency, and do you use tests (if so what sort of tests) or other methods?</p>	<p>E4</p>	<p>Sometimes we take the discussion as a whole or half class, but some students are so reluctant to speak in that environment that I get nothing out of them until we do the pair exercises. Normally when we do a writing exercise, they will have two chances to work on it -- a first draft, and then after getting comments and tips from me, a second draft. I use very few "tests" in that sense, but rather combine both training and testing in the same exercises. I do give them one written test at the beginning of the year, when we do a unit on English specific to their study program. I test them on vocabulary from their own study program (for example, plumbing words for the plumbers, from a list that we've compiled together) as well as their ability to answer basic questions about texts I've taken from professional plumbing/electrician/painter etc. magazines, and to write a text describing a typical day at their apprenticeships.</p>
<p>5. What would an "every-day" lesson look like for you?</p>	<p>E5</p>	<p>I do not use a textbook nor follow a typical textbook-type lesson plan. We work in themes (Like "Program English," where they work with English related to their study program, or "Health and Food," where each text or film or discussion has something to do with health and food!) and the assignments are quite a bit "bigger" than you would find in a normal English textbook. That is, instead of doing several different things in one lesson, we might spend 3-4 lessons working with a film -- watching it for 2 lessons, discussing it in pairs or as a class during the third, writing a summary or reaction to the film in a fourth. So most lessons are aimed at one specific language skill. I NEVER speak Swedish when I'm teaching; only English. Sometimes the lesson will go off track because the students start asking me about something else, especially about America; I let it go that way as long as they speak English while we're talking about it.</p>

6. Do you find the National test a good or bad aid in the marking process? Please explain what it is that makes it either good or bad.	E6	See E1
7. Who decides what material is used?	E7	See E5
8. Who decides the content of the lessons?	E8	Author's note: By looking at E5 , it can be argued that the content of the lessons are mainly decided by the teacher, but there is also, to some extent, student input.

III. Web – Cambridge, NAFS and Skolverket

- Cambridge ESOL Examinations

A) <http://www.cambridgeesol.org/about/standards/exam-production.html> (Accessed 23 March 2012)

B) <http://www.cambridgeesol.org/exams/ket-schools/index.html> (Accessed 22 March)

C) <http://www.cambridgeesol.org/exams/pet-schools/index.html> (Accessed 22 March)

D) <http://www.cambridgeesol.org/exams/fce-schools/index.html> (Accessed 22 March)

E) <http://www.cambridgeesol.org/about/standards/cefr.html> (Accessed 27 March)

F) <http://www.cambridgeesol.org/about/standards/alte.html> (Accessed 27 March)

G) <http://www.cambridgeesol.org/about/standards/can-do.html> (Accessed 27 March)

- Demaret: The National Assessment Project (NAFS)

A) http://www.nafs.gu.se/prov_engelska/exempel_provuppgifter/engelska_ak5_exempeluppg/
(Accessed 22 Feb. 2012)

B) http://www.nafs.gu.se/prov_engelska/exempel_provuppgifter/engelska_ak9_exempeluppg/
(Accessed 22 Feb. 2012)

C) http://www.nafs.gu.se/prov_engelska/exempel_provuppgifter/engelska_a_exempeluppg/
(Accessed 22 Feb. 2012)

- Skolverket - The Swedish National Agency for Education

A) <http://www.skolverket.se/prov-och-bedomning/nationella-prov/ak6/engelska-1.125269>
(Accessed 24 Feb. 2012)

B) <http://www.skolverket.se/prov-och-bedomning/nationella-prov/ak9/arskurs-9-engelska-1.105692>
(Accessed 24 Feb. 2012)

C) <http://www.skolverket.se/prov-och-bedomning/nationella-prov/2.1185/Engelska> (Accessed 24

Feb. 2012)

D) <http://www.skolverket.se/prov-och-bedomning/2.5255> (Accessed 25 Feb. 2012)

E) <http://www.skolverket.se/prov-och-bedomning/nationella-prov/2.1101/2.1107> (Accessed 25 Feb. 2012)

F) <http://www3.skolverket.se/ki03/front.aspx?sprak=SV&ar=1112&infotyp=8&skolform=21&id=EN&extraId=> (Accessed 30. March 2012)

G) <http://www.skolverket.se/prov-och-bedomning/2.5601/bedomning-av-prov-1.144390#listAnchor1.64797> (Accessed 24 Feb. 2012)

H) <http://www.skolverket.se/prov-och-bedomning/nationella-prov> (Accessed 22 Feb. 2012)

I) <http://www.skolverket.se/prov-och-bedomning/2.5601/bedomning-av-prov-1.144390#listAnchor1.174207> (Accessed 3 June. 2012)