“I COULD GIVE UP MANY THINGS BUT NOT THAT”

Teachers’ and pupils’ experiences of using the European Language Portfolio in assessment

Master’s thesis

Sonja Peltola

University of Jyväskylä
Department of Languages
English
May 2015

ABSTRACT

Tiedekunta – Faculty Humanistinen tiedekunta	Laitos – Department Kielten laitos
Tekijä – Author Sonja Katariina Peltola	

Työn nimi – Title "I would give up many things but not that" Teachers' and pupils' experiences of using the European Language Portfolio in assessment	
Oppiaine – Subject Englanti, ruotsi	Työn laji – Level Pro gradu -tutkielma
Aika – Month and year Toukokuu 2015	Sivumäärä – Number of pages 122 sivua + 2 liitettä

Tiivistelmä – Abstract

Arviointi on olennainen osa opetusta ja oppimista kouluissa. Kielitaidon arviointiin vaikuttavat käsitykset kielitaidosta ja aikojen saatossa kielitaidon arviointi on muuttunut lähes mekaanisesta kielitaidon osa-alueiden mittaamisesta monipuolisempaan kielenkäytön arviointiin (ks. esim. Rea-Dickins 2012: 12). Perinteisen kielitaidon testauksen juuret ovat kuitenkin yhä syvällä arviointikulttuurissamme ja vasta viime vuosikymmeninä on alettu painottaa testeille vaihtoehtoisten arviointimetodien tärkeyttä. Johtava ajatus nykypäivän kouluarvioinnissa on, että kaiken arvioinnin tulisi edistää oppimista. Lisäksi oppilaiden tulisi olla aktiivisesti osallisina arvioinnissa.

Yksi monipuolinen työkalu kielitaidon arviointiin on Eurooppalainen kielisalkku, joka koostuu oppilaiden omista tuotoksista. Salkkutyöskentelyssä arviointiin osallistuvat opettajan lisäksi myös oppilaat. Tämän tutkimuksen tavoitteena oli selvittää kielisalkkua käyttäneiden opettajien sekä oppilaiden käsityksiä kielitaidon arvioinnista, kielisalkun eduista ja haitoista sekä oppilaan roolista arvioinnissa. Tutkimuksessa haastateltiin viittä englannin- ja ruotsinopettajaa sekä kymmentä ala- ja yläkoululaista oppilasta.

Tutkimuksessa selvisi, että sekä opettajat että oppilaat kokevat kielisalkun mieluisaksi ja hyödylliseksi arvioinnissa. Oppilaat saavat näyttää osaamistaan ja salkkutyöt antavat monipuolisemman kuvan kielitaidosta. Toisaalta oppilaiden keskuudessa kokeiden ja arvosanojen merkitys korostui ja heistä moni oli sitä mieltä, että arviointi on opettajan tehtävä ja että oppilaiden itsearvioinneilla ei ole kovinkaan suurta merkitystä arviointiin. Oppilaat myös usein määrittelevät taitonsa saamiensa arvosanojen perusteella. Kielisalkun käyttöä tulisi lisätä suomalaisissa kouluissa, sillä se on selkeästi motivoiva ja monipuolinen työkalu oppimista edistävään arviointiin. Lisätutkimukselle on yhä tarvetta ja esimerkiksi koearviointiin ja kielisalkkuarviointiin tottuneiden oppilaiden näkemyksiä olisi hyvä verrata, jotta kielisalkun hyödyistä saadaan entistä vahvempaa näyttöä.

Asiasanat – Keywords Language assessment, assessment for learning, the European Language Portfolio, teacher and student perceptions
Säilytyspaikka – Depository Kielten laitos, JYX

TABLE OF CONTENTS

1 INTRODUCTION

Twenty years ago J.C. Alderson (1993: 2) noted that: "recently more applied linguists have taken an interest in the area [of language testing] and a recognition has grown that testing need not be divorced, either from teaching, or from applied linguistic theory". He, however, points out that it will take a long time until teachers, learners and academics uniformly consider language testing as something positive that enhances learning. Many people see tests as controlling and intrusive on the curriculum, as influencing learners negatively. (Alderson 1993: 2). Today, twenty years later, a glance at the current assessment literature reveals that although language testing is still often considered a field of its own the impact of assessment on learning is commonly acknowledged. Moreover, Turner (2012: 65) points out that the field of language testing and assessment is evolving and classroom assessment is gradually seen as something more than an offshoot of traditional large-scale testing. She states that the focus is now on the uniqueness of classroom learning and on the teacher's role as an assessor. The interest in teaching and learning is probably the most explicit trend in the field of assessment in the 21st century.

Thus, the concept of assessment has extended over the last couple of decades. Today it is not enough just to score learning results or give a grade at the end of a teaching period but assessment needs to be an active and continuous action in the teaching and learning process. Besides the teacher, also the learner him- or herself as well as his or her peers may be involved in the assessment process. (Keurulainen 2013: 37-38.) There is, however, still a long way to go until researchers, teachers and learners fully understand the positive effect language assessment can have on learning. More research and time is needed as old habits and teaching philosophies remain strong. Keurulainen (2013: 38), for example, argues that the change and expansion in the concept of assessment is still in progress in the everyday routines in Finnish schools. He states that in many cases the teacher is the prime assessor and the main functions of assessment are to control what the students have learned and to give grades.

In the era of assessment it is important to step back and ponder why assessment is needed, for whom, what is assessed and how. As Atjonen (2007: 6) points out, we are currently living in the age of an assessment boom where everything from teachers and schools to productivity and learning environments are being assessed. She questions whether all assessments are done solely for the purposes or learning or if the motives are more political, economic or even nominal. This type of criticism has clearly enhanced the growth in interest in the assessment for learning paradigm, which is also the leading paradigm of the present thesis. In this study assessment is reviewed in small-scale, classroom environment and the underlying aim is to raise awareness of one alternative language assessment method, the European Language Portfolio, that contributes to learning.

The European Language Portfolio (ELP) is a document in which learners can record and reflect on their language learning and intercultural experiences (Council of Europe 2011). A reasonable amount of research has been conducted about the ELP especially in the beginning of 21st century when the first ELP experiments were conducted in Finland. Most of the research consists, however, of different types of experiment reports and fewer studies have been written about the effects the ELP has on learning or on assessment. Still, for example Lammi (2002) has studied how the ELP affects learner motivation, learner autonomy and self-reflection skills. Thus, there is a clear need for research that studies not only teachers' but also learners' perceptions about the ELP in language assessment.

The aim of the present thesis is twofold. Firstly the purpose is to examine teachers and pupils views on language assessment and on the ELP as a language assessment tool. Secondly, the underlying aim of the study is to raise awareness of the ELP and the assessment for learning paradigm since also in my own experience language assessment in many school still relies heavily on traditional exams and teacher-led assessments. Assessment should be more interactive and always aim to enhance learning. The previous experiments show that the ELP is

a versatile tool for language teaching learning and assessment (see e.g. Kohonen and Pajukanta 2003).

In the present study five English and/or Swedish teachers and ten pupils were interviewed. Three pupils were from primary school and seven from lower secondary school. All the interviewees had used the ELP and they were asked about their conceptions of the use of the ELP in language assessment; what does the language portfolio assessment include, how does the ELP function as an assessment method, what more does it bring to language assessment, what is the pupils' role in assessment and whether there are any differences between using the ELP in English or Swedish language learning and assessment. The qualitative data was analysed through content analysis.

Hence, the present thesis discusses the current assessment practices and raises awareness of practices of using the ELP in classroom assessment. In the first chapter I will present some general information about language testing and assessment. The terminology is clarified after which the history of the concept of language ability is accounted for. In addition, the different uses of language tests are introduced and some considerations about assessment quality are discussed. In the second chapter I move on to the classroom context. The concepts of classroom assessment and assessment for learning are explored and the role and purpose of assessment in Finnish schools are discussed. Moreover, the European Language Portfolio is examined more profoundly. The research questions, data collection and method of analysis are clarified in chapter four and in chapter five the results of the study are analysed and discussed. Finally, the discussions are concluded in chapter seven.

2 LANGUAGE ASSESSMENT

During the last couple of decades the concept of assessment has expanded beyond traditional testing and measuring of skills (Hildén 2009: 33). The role of

assessment is shifting from testing and pure measurement towards a new assessment culture that endorses learning (Inbar-Lourie 2008: 287). The changes are not in progress only in Finland but also in other countries (Keurulainen 2013: 38). For example in the United Kingdom, the Assessment Reform Group (ARG) has over the last few decades gathered research from around the world to get insights into how assessment can truly promote learning (Gardner 2012: 1). Indeed, the concept of *assessment for learning* has become widely popular not only in language assessment but also in the field of assessment in general. The new assessment culture, like the current learning cultures, considers intelligence multi-faceted and emphasizes diverse and individual learning opportunities. (Inbar-Lourie 2008: 287).

Before looking at the current or desired state of language assessment today, it is important to view the history of language assessment and testing, and discuss the fundamental consideration of language assessment. In the following sections the field of language assessment is discussed in more detail. Firstly, the terms test, measurement, assessment and evaluation are defined in order to clarify the varying uses of assessment terminology. Secondly, the history of language assessment as well as the recent changes in the field are introduced briefly. This includes a short review of the concept of language ability and how it has been defined over the years. Thirdly, the different uses of language tests as well as the two commonly acknowledges testing approaches: norm- referenced and criterion-referenced testing are examined. Finally, the factors that contribute to assessment quality are elaborated.

2.1 Defining measurement, test, assessment and evaluation

The changes in the field of language testing and assessment have both engendered new terminology and created a need to redefine the existing terminology. According to Bachman (1991: 18, 50), the terms *measurement*, *test*, *evaluation* and *assessment* are often used interchangeably because they can involve similar activities in practice. In assessment and testing literature test, evaluation

and measurement are, nonetheless, in many cases delineated as separate terms and also Bachman (1991: 18) argues that distinguishing the three terms from each other is necessary for proper language test development and use. He defines measurement as follows: "the process of quantifying the characteristics of persons according to explicit procedures and rules". Thus, numbers are assigned to people's attributes and abilities and the observation procedures must be replicable later on (Bachman 1991: 19-20). Douglas (2010: 5), nevertheless, reminds that there can be measurement without a test. He notes that a teacher may for example give grades, and hence, order students along a scale based on several of sources of information, such as homework assignments, performance on classroom exercises and out-of-class projects, and no testing is included.

Although measurement does not necessarily involve a test, a test is one form of measurement that "quantifies characteristics of individuals according to explicit procedures" (Bachman 1991: 20). The factor that delineates a test from other measurements is that a test obtains a specific sample of an individual's language use. Inferences about certain abilities must be supported by specific samples of language use language and that is why language tests are needed. (Bachman 1991: 20-21.) Tests are often used because it is believed that they ensure fairness and enable comparisons of students against external criteria better than less standardised forms of assessment (e.g. Douglas 2010: 5-6). In addition, Douglas (2010: 9) argues that well-designed tests provide teachers "a second opinion" which confirms or sometimes also disconfirms the teachers' perceptions of their students' language performance.

The term assessment seems to lack a proper definition in the assessment literature (e.g. Bachman 1991: 50). Most of the books and articles covering language assessment do not include any definition of the term. Lynch (2001: 358), however, defines assessment as "the systematic gathering of information for the purposes of making decisions or judgements about individuals". He sees assessment as a superordinate term for a variety of methods and practices that

assist in the information gathering process. These methods include measurement and tests but also many non-quantitative procedures, such as portfolios and informal teacher observations (Council of Europe 2001: 177, Lynch 2001: 358). Thus, all tests and measurement procedures are types of assessment but, in essence, the concept of assessment involves much more than only quantitative measuring. (Council of Europe 2001: 177).

Lynch (2001: 358–359) illustrates the relationship between assessment, measurement and testing with three circles (see Figure 1) where the outer circle depicts non-measurement and non-testing forms of assessment. The figure is a rather simple representation of the complex relationships between the terms but it, however, gives a clear overview of the term hierarchy and clarifies the fact that assessment includes both qualitative and quantitative information gathering procedures. Thus, assessment is not equal to testing.

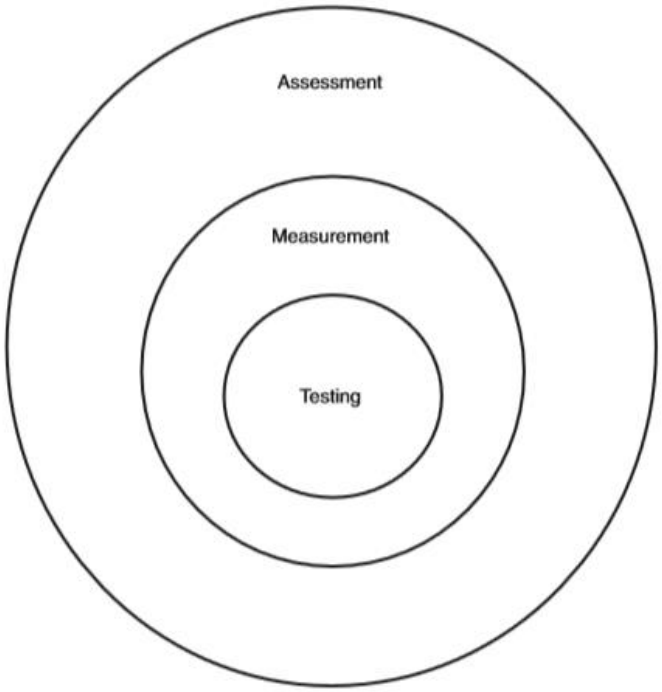

Figure 1: Assessment, measurement and testing (Lynch 2001: 359)

Whereas assessment refers to decisions about individuals, evaluation concerns also larger entities like schools and educational policies (Atjonen 2007: 20). Evaluation includes assessment but it involves also evaluation of other factors than only language proficiency. In a language programme, for example, the effectiveness of used materials and methods, the type and quality of produced discourse, learner or teacher satisfaction, and the effectiveness of teaching could be evaluated in addition to a learner's language ability. (Council of Europe 2001: 177.) Thus, in both evaluation procedures and assessment procedures information is gathered to make decisions but the range and the purpose often differ. Evaluation does not, however, necessarily involve testing, and tests, on the other hand, do not necessarily have to be evaluative (Bachman 1991: 22). According to Bachman (1991: 22-23), tests often have either a pedagogical or a merely descriptive function, which does not involve any evaluative decision making, but evaluation occurs only when test results are used for making a decision. Thus, tests serve an information-providing purpose whereas evaluation serves a decision-making purpose (Bachman 1991: 23).

In the present thesis the focus is on gathering information about individual learners, and hence, the focus is on language assessment. As mentioned above, assessment includes tests and measurement procedures but also qualitative information gathering procedures, like portfolios. The term most often used in older language assessment literature is testing but from the 1990's onwards the term assessment has become more and more common. This is not only a matter of terminology but it also reflects a cultural change in the field of language learning and assessment. Thus, the term testing occurs also in the first sections of the present thesis as the history of defining language ability is discussed. Later on, in chapter three, assessment is viewed from the aspect of language learning.

Using the term language assessment is not, nonetheless, so straightforward either. The term is broad and there are plenty of different types of language assessment. Different professionals use differing terms such as diagnostic

assessment, classroom assessment, formative assessment, dynamic assessment, alternative assessment or authentic assessment (see e.g. Turner 2013, Hildén 2009: 33, Alderson 2005, Lynch 2001) depending on their preferences. As Hildén (2009: 33) acknowledges, the range of terminology used in the field of assessment has extended during the last couple of decades as several alternative approaches to assessing language performance have been promoted. Moreover, sometimes the different terms are used interchangeably.

The quintessence of this matter here is not, however, to explicitly discuss all the different types of assessment but to understand the core idea behind them. Alternative assessment, as the different assessment methods are often referred to, were born when the interest towards finding alternatives to 'traditional' tests began to increase in the 1990´s (Douglas 2010: 73, Lynch 2001: 360). A distinction between the traditional testing culture and alternative assessment approaches is often emphasized (Fox 2008: 97, Lynch 2001: 360). Moreover, the need to find more suitable assessment methods for classroom contexts to replace the practices applied from large-scale testing (i.e. testing large numbers of learners for example in standardised international exams) have brought the attention to the purpose of assessment (Turner 2012: 65). The proponents of alternative assessment approaches promote assessments that are among other things extensions of usual classroom learning activities, related to real-life contexts, rated by human beings instead of computers and prioritising the learning process more than the product of learning. (Douglas 2010: 73.)

In the present thesis the term classroom assessment is used to describe the small-scale assessment processes which take place in classroom contexts. Of all the different alternative assessment approaches classroom assessment seems to be the most appropriate term to describe the context of the present study. Moreover, in the present study classroom assessment is used as an umbrella term for all the different alternative assessment approaches that promote learning, including both formal and informal assessment procedures.

2.2 Defining language ability over time

People's language skills have been assessed and tested through the ages whenever and wherever languages have been learned or the level of language ability has affected any decisions made about one's future. Before the 1960's language testing was not studied much since nobody considered it very complicated. From the beginning of the 20th century until the 1970's language testing was strongly influenced by psychology and psychometrics where qualities of different indicators, such as reliability of a test, were surveyed through statistics. (Huhta and Takala 1999: 179–180.) During the last three decades substantial progress has been made with the research and understanding of language development and language assessment. These advances have been forwarded by many distinguished language testers like Bachman, whose model of language ability has had a significant influence not only on language testing but also on second language acquisition research. (O'Sullivan 2011: 2.) Moreover, since the 1970's also other sciences, such as sociology and anthropology, have begun to affect language assessment (Huhta and Takala 1999: 180).

Although language assessment is linked to many different disciplines the relationship between language assessment and language learning is probably the strongest and most evident. During the years language testing and assessment have changed and awareness of the challenges of assessment has increased as the concept of language and language skills has changed (Huhta and Hildén 2013: 160). According to Bachman (1991: 2), the relationship between the disciplines is reciprocal; language testing contributes to and is contributed by research in language acquisition and language teaching. Language tests can, for example, provide useful information about the success of teaching and learning or about the usefulness of different language teaching methods (Bachman 1991: 2-3). On the other hand, information gained from language acquisition research and language teaching practices can be useful in test development.

In order to be able to assess language ability and interpret the results meaningfully one has to understand what language ability is (Huhta and Takala (1999: 182, Bachman 1991: 3–4). During the years many frameworks of language ability have been presented and unfortunately, there is still no one theory that could explicitly explain what language ability is and how to properly test it. It is rather a combination of different theoretical models that has influenced the current understanding of language ability and the way how language assessment has developed. (Huhta 1993 cited in Huhta and Takala 1999: 182, Huhta and Hildén 2013: 160-161). Some of the most influential models of language ability are described in the following paragraphs.

2.2.1 The traditional view of language ability

The traditional way to view language ability is to divide it into four skills: listening, reading, speaking, and writing, and moreover, into various components, such as grammar, vocabulary and pronunciation (Bachman and Palmer 1996: 75). This originates from structural linguistics and emphasizes the idea of language ability being composed of several elements (Huhta and Takala 1999: 183). Lado (1961: 25) postulated that language consists of different elements that are "integrated in the total skills of speaking, listening, reading and writing". The different elements, such as intonation, stress, morphemes, words, and arrangements of words can be tested separately but still they are always integrated in language (Lado 1961: 25). Furthermore, Lado (ibid.) pointed out that the skills do not improve evenly. One may be more advanced for example in reading than in writing, and hence, all four skills need to be tested.

The model of viewing language ability in terms of the four skills was significant in language testing during the second half of the 20th century (Bachman and Palmer 1996: 75). Bachman and Palmer (1996: 75-76), nevertheless, argued that the model was inadequate, too theoretical as it does not take account of actual language use. They suggested that rather than being part of language ability, the four skills need to be considered realisations of purposeful language use. Hence,

instead of considering for example speaking an abstract skill, it should be identified as an activity that is needed in specific language tasks and described in terms of actual language use (Bachman and Palmer 1996: 76). The influence of this traditional view of language ability is still clear today as will be discussed later in the thesis.

2.2.2 Models of communicative competence

In the 1980's the four skills model of language ability was challenged by new models of language ability, that is the models of communicative competence (Fox 2012: 2933). The social context of language use was recognised and researchers began to emphasise the dynamic interaction between the situation, the language user, and the discourse in communicative language use. In fact, authenticity became a desired quality in language testing. (Bachman 1991: 4.)

In the 1980's Canale and Swain introduced a framework for communicative competence that is one of the most well-known views of language ability in applied linguistics (Canale and Swain 1980, Huhta and Takala 1999: 184). Their model included three components: grammatical competence, sociolinguistic competence and strategic competence (1980: 28). Also, a couple years later Canale added a fourth component, namely discourse, to the framework (Fulcher and Davidson 2007: 208). Canale and Swain (1980: 29) considered their framework a model of knowledge which would be evident, by implication, in actual communicative performance. Thus, they made a clear distinction between communicative competence, which is a model of knowledge, and communicative performance, which is a realisation of the competences (Canale and Swain 1980: 6). What comes to language assessment, Canale and Swain (1980: 34) proposed that language test should not only include tasks that require knowledge about the language (i.e. competence) but also tasks where test takers need to demonstrate their understanding in actual communicative situations (i.e. performance).

Also Bachman (1991) introduced his model of communicative language ability. The model was based on the work conducted by Canale and Swain but he extended the model by adding more components and subcategories, such as pragmatic and organisational components (Fox 2012: 2933, Bachman 1991: 81, 87). Bachman (1991: 81) aimed to explain how the several components interact with each other and with the context of language use. The three main components in Bachman's model were language competence, strategic competence and psychophysiological mechanisms. Each of these competences included several subcategories and components but in essence language competence represented the knowledge of language (including organisational and pragmatic competences), strategic competence meant the capacity that connects the knowledge of language with a context and the language user's knowledge structures and psychophysiological mechanisms referred to the neurological and physiological aspects of using a language. (Bachman 1991: 84, 107-108.)

A few years later Bachman and Palmer published a refined version of Bachman's model. The revised model presented some of Bachman's ideas more precisely and focused more on teaching of language testing. (Fulcher and Davidson 2007: 45.) Some of the changes were minor but like McNamara (1996: 72, 74) points out, Bachman and Palmer added affective schemata to their model which was a significant development from Bachman's model. By including the new component Bachman and Palmer recognised the effect of emotions on individuals' language use as well as on their language test performance (Bachman and Palmer 1996: 65–66). This was the first attempt to explicitly associate language use with affective factors in second language communication (McNamara 1996: 74).

2.2.3 Other views of language ability

In addition to the models of language ability presented above also other researchers and linguists have proposed their own models. For example in 1995

Celce-Murcia, Dörnyei and Thurrell introduced a model of communicative competence specifying the content of the different competences of language ability. Moreover, one of most recent advances has been the concept of interactional competence. (Fulcher and Davidson 2007: 49.) This reflects the modern thinking that language ability is not an individual's inner quality but rather something that is built in interaction (Huhta and Hildén 2013: 160–161).

The Council of Europe has also conducted a lot of research about language ability and in 2001 they published the Common European Framework of Reference for Languages: Learning, Teaching and Assessment (CEFR, Council of Europe 2001). The function of the CEFR is well explained in the following:

> It describes in a comprehensive way what language learners have to learn to do in order to use a language for communication and what knowledge and skills they have to develop so as to be able to act effectively (Council of Europe 2001: 1).

As implied in the citation, the CEFR is based on a communicative view of language. The model of communicative competence that the CEFR embraces is based on three basic components: linguistic, sociolinguistic and pragmatic. Each of these competences further includes various skills, knowledge and know-how. Language learners are seen as social agents whose language competencies are activated when they actually use a language (Council of Europe 2001: 1, 9, 13–14). In summary, some of the fundamental ideas behind the CEFR are communicative language proficiency, learner-centredness and action-oriented approach to language learning (Council of Europe 2001: 9, Little 2009: 1–2).

Since its publication the CEFR has influenced the language teaching and language assessment all over Europe. For instance, the CEFR levels are referred to in language curricula and textbooks in many European countries (Little 2009: 2). Also in Finland the NCC is based on the CEFR (POPS 2004). Moreover, Huhta and Hildén (2013: 161) note that the scales of CEFR are so widely used that almost all major international language tests have had to balance their result in relation

to the CEFR levels, for commercial reasons at the very least. Thus, the CEFR has at least to some extent succeeded in its aim to provide a common basis for language learning, teaching and assessment (Council +of Europe 2001: 1). Nonetheless, the CEFR has been criticised for example for its lack of theoretical accuracy and explicitness and some do not consider the CEFR ideal for test development (O,Sullivan and Weir 2011: 16, 26; Alderson et al. 2004, cited in O'Sullivan and Weir 2011: 16).

According to Huhta and Takala (1999: 183), the numerous sociolinguistic views that since the 1970's were executed as communicative language teaching and assessment have also had a major influence on language assessment.

2.3 The effects of the varying views of language ability on language assessment

The concept of language ability has extended significantly during the last few decades and the focus has clearly shifted from knowledge of a language to the ability to use a language. Today the communicative view of language ability dominates the field of language learning and teaching and this shows also in assessment. (Huhta and Hildén 2013: 161.) In chapter three of the present study the effects of these changes on classroom assessment are discussed more thoroughly. Nevertheless, when the traditional view of language ability prevailed, many so-called objective tests formats, such as multiple-choice and true-false tests, were very common (Fox 2012: 2931). As Huhta and Hildén (2013: 162) point out multiple-choice questions were commonly used to test receptive language skills, namely listening and reading. After this era test methods such as cloze tests (words omitted from a text), C tests (parts of words omitted from a text) and dictation became more popular as these methods were supposed to recognise the importance of the context. Finally, when the communicative view of language ability begun to gain in popularity, more subjective testing methods such as essays and oral interviews became approved and desired methods. (Fox 2012: 2932.)

Today the range of assessment methods is wide and tasks where learners' have to produce language themselves are favoured. The aim in testing is often to simulate real-life situations and context. (Huhta and Hildén 2013: 162.) It is commonly acknowledged that, as Douglas (2010: 20) expresses it, "language is never used in a vacuum". People do not just merely speak, write, read or listen but they use languages for different purposes, in different contexts, with different people, and the way how languages are used varies in these contexts (Douglas 2010: 20). This makes language assessment challenging and creates one of the fundamental dilemmas of language testing:

> the tools we use to observe language ability are themselves manifestations of language ability. (…) Thus, one of the most important and persistent problems in language testing is that of defining language ability in such a way that we can be sure that the test methods we use will elicit language test performance that is characteristic of language performance in non-test situations. (Bachman 1991: 9)

Hence, language assessment is challenging and there is no one testing method that would give the most reliable and thorough description of someone's language ability (Huhta and Hildén 2013: 162, 167). Bachman (1991: 8) points out that it is nearly impossible to identify all the skills and other factors that influence language performance in a testing situation. Huhta and Hildén (2013: 162) exemplify this problematic situation by listing factors which affect the assessment of speaking and writing: the given assignment, the assessment scale and the definition of language ability which the scale is based on, as well as the assessor's experience, strictness, interpretation of the assessment criteria and understanding of language ability. Furthermore, in speaking tests the possible interlocutor can affect the test taker's performance (McNamara 1996: 86).

Indeed, one perpetual problem in language assessment is that it is nearly impossible to separate language ability from other abilities (Huhta and Hildén 2013: 178). Even personality can affect a test performance as an extrovert personality, for example, is undoubtedly useful in oral communication tasks (Huhta and Hildén 2013: 178). It is easy to understand the desire of language

testers and assessors to test only language ability, especially in large-scale assessment, in order to achieve standardisation and conformity but in the light of what is currently known about language ability the desire seems rather unreasonable (Paran 2010: 3). Paran (2010: 3, 5) argues that the standardisation of tests means also narrowing of vision and reminds that language teaching involves much more than learning only language. Thus, testing only language is not always even desirable.

The concept of communicative language learning is constantly evolving and expanding. One change happening at the moment is to view language learning as language education. In language education concepts such as learner autonomy, learner commitment, learner responsibility, self-assessment and student-centred learning are highly valued. (Kohonen 2005b: 26, 28.) These concepts are present in current classroom assessment literature, which will be discussed later, but in my view these principles are not yet widely applied in practice. Nevertheless, language learning and teaching is about much more than only language.

In conclusion, language ability is not just a simple quality to be measured but a multi-faceted concept that involves also other than purely linguistic factors. As Fulcher and Davidson (2007: 50) point out, test scores can provide only limited information, under set circumstances and for a specific purpose. Also, as discussed earlier, the current understanding of language learning suggests that it may not be even relevant to try to assess only language ability. The main thing is, however, that one has a clear understanding of language ability and the affective factors as well as of the scoring system (Bachman 1991: 8). Moreover, the purpose and the audience of a test influence for example the selection of the test method the scoring procedure (Fulcher and Davidson 2007: 50). Next, two approaches to interpreting test results and the different uses of language tests are introduced.

2.4 Norm-referenced and criterion-referenced assessment

The results of a test can be interpreted by using either of the two referencing approaches: norm-referencing or criterion-referencing. When an individual's test performance is compared with the performance of the other test takers, the approach applied is norm-referenced. Thus, the test results are interpreted in relation to a norm formed by a group of other test takers. Alternatively, when an individual's test performance is interpreted with regard to a certain level or given criteria, the interpretation is criterion-referenced. (Bachman 1991: 72.)

Norm-referencing is often used when the purpose of testing or assessment is to compare candidates, that is, in selection situations and competitions. For example, in entrance examinations the selection of students is often based on comparison between the candidates' test results and the students with the highest scores are accepted in. (Keurulainen 2013: 41, 44.) Hence, as Hughes (2003: 20) points out, norm-referenced tests do not provide information about a test taker's language ability but rather about how skilful he or she is compared to others. Norm-referenced tests are most appropriate when the number of participants is large (Huhta and Takala 1999: 219).

Criterion-referenced assessment is most appropriate when assessing learning results and language ability. Criterion-referenced tests provide information about what a test taker can do in the tested language and so when test takers' performances are assessed against a criterion level of ability they are not ranked from the best to the worst (Bachman 1991: 74–75, Hughes 2003: 20–21). Hughes (2003: 21) summarises the commonly acknowledged virtues of criterion-referenced assessments as follows: "they set meaningful standards in terms of what people can do… and they motivate students to attain those standards".

According to Huhta and Hildén (2013: 163), the use of criterion-referenced assessment has increased as communicative assessment, and furthermore,

testing of language production skills, have become more common. Moreover, the Common European Framework of Reference (CEFR), which combines the communicative language view and criterion based assessment, has increased the popularity of criterion-referenced assessment (Huhta and Hildén 2013: 161, 178). Indeed, in many countries the CEFR proficiency levels are applied to language curricula (Little 2009: 2). Using CEFR scales, or criterion-referenced measures on the whole, is however said to be challenging because the scales describe language ability in a very general level (Bachman 1991: 338, Huhta and Hildén 2013: 179). Also, criterion referenced tests have been criticised for their lack of agreed procedures and accuracy which threaten the consistency of assessment (Hughes 2003: 22). Huhta and Hildén (2013: 178–179), however, note that for example in school contexts the CEFR scales be modified for each course to meet the requirement of specificity. Consistency and other qualities of assessment are discussed late in this section.

2.5 Uses of language tests

The purpose of assessment affects the referencing approach of a test but also many other factors. Huhta and Takala (1991: 189) state that in order to make any interferences of test results an assessor needs have a clear understanding of not only the target of assessment, that is, language ability, but also of the purpose of assessment. Bachman (1991: 54), however, argues that the purpose of a test is the most significant factor in language test development and result interpretation. This is because the purpose of a test delineates the specific skills or components of language ability that are to be tested. The purpose of a test can vary from very general to very specific and the target of testing might be one or several skills and components. For example, if one was to design an admission test for entrance to a language programme, the test designer would have to define the skills needed for succeeding in the language programme. (Brown 2012: 5979.)

Language ability can be tested in various contexts and for various purposes. Furthermore, the information gained from testing can be used for making various

decisions about people and programmes. (Brown 2012: 5979.) There are several ways to categorise language tests according to their purpose but the main uses of language tests are often said to be the assessment of language learners, the evaluation of language programmes and test use in research (Bachman 1991: 54, Brown 2012: 5979). In the scope of the present thesis the focus here lies on the assessment of learners, which, according to Brown (2012: 5080), is also possibly the broadest of the three categories.

The assessment of learners can be for example proficiency, achievement, selection, entrance, readiness, placement, diagnostic, progress, attainment or aptitude testing. Many of these categories have overlapping features and often it is difficult to distinguish all there categories from each other (Bachman 1991: 70, 77; Brown 2012: 5980). Of all the different categorisations Brown (2012: 5980) introduces in my opinion a very extensive and useful framework for discussing test uses for the assessment of learners by separating four points at which assessment is used: gatekeeping assessment before receiving a work or a study place, placement and diagnostic assessment at the beginning of a period of study, progress assessment during a period of study, and achievement assessment at the end of a period or study. Although this framework offers no unambiguous viewpoint for categorising the different uses of assessment it provides a useful tool for understanding the basic ideas behind the numerous categories.

In the context of the present thesis it is most relevant to discuss progress and achievement assessment. Progress assessment is executed during a period of study and used to measure student progress. The content of the assessments is typically based on the course objectives and syllabus. (Brown 2012: 5981.) A more familiar term for progressive assessment in assessment literature is probably formative assessment, which also refers to continual, interactive assessments of student progress. The information received from assessment is used to recognise learners' needs and tailor teaching accordingly. (OECD/CERI 2005: 21). Thus,

formative assessment is not only a measure of progress but aims also at enhancing student learning (Brown 2012: 5981).

Formative assessment is often contrasted with summative assessment which measures what learners have achieved at the end of a course or a semester (OECD/CERI 2005: 21). A lot of summative assessment is achievement assessment linked to language courses when the purpose is to assess how well a student has met the learning objectives of a course (Council of Europe 2001: 186, Hughes 2003: 13). The results of summative tests can be used for example to give a grade, to decide whether a student is ready to proceed to the next course or whether a student has achieved the required level of proficiency to complete a programme (Brown 2012: 5981, Council of Europe 2001: 186). Sometimes achievement tests can also further formative assessment as they can be used to measure student progress (Hughes 2003: 14).

In conclusion, the purpose of assessment influences the content, the criteria and other essential components of assessment (Huhta and Hildén 2013: 177). Huhta and Hildén (ibid.) outline that in an ideal situation the intended purpose of a test and the actual use of the test are equivalent to each other. They claim that changing the purpose after a test has been made may be impossible and, furthermore, lead to false conclusions. Thus, it is easy to agree with Bachman (1991: 54) who states that the purpose is the most important consideration in language testing.

2.6 Quality in assessment

Many decision are made and actions taken based on the results of language assessments and tests. A high stakes test can have a notable effect on one's life but also low stakes tests and assessment have consequences. For example, a wrong placement decision may result in a student being placed in a course that is too demanding for him or her (Douglas 2010: 9). That is why it is important

that assessment is of high quality. Quality in assessment refers often to three concepts: validity, reliability, practicality (Huhta and Takala 1999: 211). There are also other important qualities for test developers to consider, for example authenticity, impact and interactiveness of a test, but the three above-mentioned factors are often recognised the most fundamental considerations in language testing and assessment (Bachman and Palmer 1996: 17, Council of Europe 2001: 177).

In assessment reliability means consistency of measurement (Bachman and Palmer 1996: 19). Hence, a reliable test yields the same or similar result if the test was to be repeated (Jones 2012: 352). Nevertheless, as noted earlier, it is not only the abilities the test designers want to measure that influence a test performance but also other factors, such as lack of motivation, unclear instruction or unfamiliar test tasks have an effect (Douglas 2010: 10, Bachman 1991: 160). Allowing for all the possible errors of measurement it is impossible for a test to be perfectly consistent (Bachman 1991: 160, Douglas 2010: 10). Still, it is important to recognise the potential errors and aim at minimizing the effects of the errors in order to maximise reliability (Bachman 1991: 60).

A degree of reliability is required for test results to be meaningful, that is valid, but high reliability alone does not always indicate high validity (Jones 2012: 352). Validity pertains to the relevance of inferences drawn from test results (Douglas 2010: 10). Hughes (2003: 26), on the other hand, states that a test is valid if it measures accurately what it was intended to measure. Hence, reliable results are a precondition for valid interpretations of test results (Bachman 1991: 289).

Jones (2012: 357) notes that in reality reliability and validity are pursued within the limits of practicality. On that account, if the implementation of a test would involve more resources than is available the test would not be practical and hence will not be implemented (Bachmand and Palmer 1996: 35–36). Huhta and Takala (1999: 215) remind that although the resources are often small in small-scale

assessments, practicality should not be valued over the reliability and validity. They state that if the results of an assessment cannot be trusted or interpreted, the time used in the assessment process is wasted. The key factor is to find a balance between all the different qualities (Jones 2012: 357).

Reliability and validity are important considerations in all testing and assessing practices but the concepts have slightly different meanings when moving from large-scale testing towards classroom assessment. As Fulcher and Davidson (2007: 33) point out, the contexts of classroom assessment and large-scale testing are very different. They (2007: 24–35) list several key aspects of assessment in the classroom showing that it is not always so straightforward to take the principles from large-scale testing and apply them directly to classroom practices. Firstly, teachers often have a broad understanding of the abilities and skills of their students as they continuously have a chance to observe them On the contrary, in standardised large-scale tests the test developers do now know the test takers. Secondly, in learning contexts students do tasks that are related to previously studied issues while in large-scale tests the tasks are designed separately to produce as comprehensive picture of the test taker as possible. Thirdly, the working methods applied in classrooms may include for example group work, and the assessor can be not only the teacher but also the learners themselves or their peers. In contrast, in standardised testing individual work is the only option and the tasks are assessed only by qualified assessors.

Thus, some adjustments have to be made when applying validity and reliability, originally defined large-scale testing, to the classroom context. Turner (2012: 68) confirms this by stating that in both contexts there is a need for information gained from assessment but the uses of the information are unalike. In classroom assessment learning is the main goal, and hence, the interpretations drawn from the results are valid only if they enhance learning (Fulcher and Davidson 2007: 35). Turner (2012: 68) concludes, that the concepts of reliability and validity need to be appropriately redefined for classroom-based assessment but in order to do

that more research about the realities of assessment practice in the classroom is needed.

The most fundamental issues in language testing, and hence the roots of language assessment, have now been introduced. The traditional language testing literature concerns mostly standardised, often large-scale, testing which shows also in this first chapter of the present thesis but the principles and theories introduced are the same either discussing standardised large-scale testing or small-scale classroom assessment. Moreover, as interest in classroom based formative assessment began to increase only one or two decades ago, many teachers have had to utilise large-scale assessment literature and apply those methods to the classroom (Fulcher and Davidson 2007: 23, Rea-Dickins 2011: 10).

Next, I move towards the context of classroom where teachers, learners as well as other people involved in teaching and education work together. In the following chapter the focus is on assessment methods that aim at promoting learning. First, I will discuss classroom assessment in general and cover the foundations of alternative assessment that embraces the ideology of assessment for learning. Secondly, the principles of assessment in the Finnish school system will be introduced. Also a couple studies about assessment practices in Finland will be examined. Thirdly, some classroom assessment practices are introduced. Fourthly, the European Language Portfolio (ELP) will be discussed and some research about the use of the ELP will be presented. Finally, the chapter is concluded with a review of the future directions of classroom assessment and the European Language Portfolio.

3 THE ROLE OF ASSESSMENT IN LANGUAGE LEARNING

The ideas of more learner- and learning-centred assessment are evident now in the 21th century. The interest towards alternative assessment methods and classroom assessment has clearly grown and research in the field of classroom

assessment has increased significantly during the past two decades (Read-Dickins 2011: 12). According to Inbar-Lourie (2008: 285, 288), this has led to the birth of a new assessment culture which highlights the connection between assessment and learning. As discussed in the previous chapter, the construct of assessment has moved beyond testing and standardised measurement by identifying the social aspect of assessment, understanding the meaning of assessment and ensuring all learners the same possibilities in assessment (Rea-Dickins 2011: 12). The new assessment culture does not exclude traditional testing cultures but broadens the concept of assessment by increasing variety in the assessment data to include several assessment tools and sources of information, also learners (Inbar-Lourie 2008: 288). Assessment should be considered a pedagogic tool which can be used to improve learning and engage students in the language learning process (Rea-Dickins 2011: 12).

Moreover, Inbar-Laurie (2008: 293–295) argues that although the new assessment culture, or at least parts of it, are slowly being endorsed as options for testing cultures, it has not been applied to practice due to issues of power and willingness. He questions the assessment authorities' motives and teachers' willingness to seriously adjust to the new assessment culture and undertake the change process from the old culture to the new one. Likewise, Rea-Dickins (2011: 12–13) criticises the authorities' obsessive needs to receive measurable learning outcomes. Also teachers are attempting to balance between the two cultures (Rea-Dickins 2011: 12). Furthermore, there is a lack of information for teachers as the principles of the new assessment culture have not been considered properly in the language teaching literature. Some texts and handbooks for teachers still apply information from large-scale testing. (Fulcher and Davidson 2007: 23.) There is still a long way to go before the new assessment culture is clearly visible in practice, at all levels (Rea-Dickins 2011: 12).

Although the implementations of the new assessment culture may not yet show in practice, many of the principles of alternative assessment have been a serious

topic of interest and discussion under the titles *alternative assessment* or *alternatives in assessment* since the 1990's. The terminology varies but the main idea is to present alternatives to traditional testing (Fox 2008: 97). Also, it is important to remember that it is not the learning activities changed into assessment methods that make a difference but rather the ideas and perspectives behind their use. Portfolios, diaries and peer assessment, which are often considered alternative assessment methods, can also be used to measure, rank and externally monitor learners. These assessment methods can be called alternative assessment methods only when they represent the notions of the new assessment culture. (Fox 2008: 99–100, Lynch and Shaw 2005: 264–265.)

3.1 Classroom assessment

As mentioned before, there are various terms to describe the alternative assessment methods embracing more context-related, classroom-embedded assessment practices that at least to some extend try to replace or expand the traditional testing culture (Davidson and Leung 2009: 395). From now on the focus in the present thesis will be on classrooms, on the assessment processes in which teachers and learners are involved. Here classroom assessment represents the same basic principles as any alternative assessment approach but the context is now classrooms.

In classroom assessment teachers and students work together to plan, collect, analyse and use the information they have gathered by using several assessment tools and methods. Assessment is embedded in the learning process which is a social event. Teachers and learners collect and share information in order to meet the needs of all learners. (Katz and Gottlieb 2012: 161). Davidson and Leung (2009: 400–401) point out that classroom assessment covers various types of assessments from formal, planned assessments, for example portfolio work, to informal, unplanned classroom observations. They further note that feedback as well as self- and peer-assessment are also important elements of all assessment in the classroom. Thus, classroom assessment is a rather complex process which

includes student-teacher interaction, self-reflection and various types of assessment procedures all aiming to improve learning. (Gardner 2012: 3).

3.1.1 Assessment for learning

One of the fundamental consideration of classroom assessment is that it should enhance learning. Although Bachman and Palmer (1996: 19) consider that tests are meant mostly for measuring, that should not be their purpose in classroom context. Measuring may be necessary for example in the cases of placement and proficiency tests, but in classrooms measuring does not benefit either teachers or students. It is useful only for bureaucrats who need statistics. Thus, I believe that all classroom activities, whether they were tests, other assessment practices or learning activities, should promote learning.

Indeed, also Black and Wiliam (2012:11) outline that "Assessment in education must, first and foremost, serve the purpose of supporting learning." Also summative assessment can be used to provide useful feedback and enhance learning (Davidson and Leung 2009: 397). In fact, teachers should use various methods to gather information about their students and also to assess that information (Katz and Gottlieb 2013: 163). Davidson and Leung (2009: 399) conclude that all assessments have to be continual and rooted into the processes of learning and teaching, not only at the end of each learning period. Again, they emphasise that it not about the methods used but the way how the methods are used. The leading idea should always be that all practices enhance learning (ibid.)

A good example of how the assessment for learning principle could be executed in practice is provided in the formative assessment development project conducted by Black and Wiliam in the turn of the new millennium. Their extensive review on previously published formative assessment research had revealed that formative assessment has a significant, positive effect on student achievement and they wanted to apply the effective practices to classrooms.

Their project involved 48 teachers of mathematics, science and English in the United Kingdom. In one part of the project the teachers were encouraged to develop ways to make formative use of summative tests. One example of this was when the students were asked to use the colours of traffic lights to mark the key topics of the upcoming test. Thus, the students had to reflect their own learning and develop learning strategies for covering the topics that they did not yet master. (Black and Wiliam 2012: 13, 15, 19.) The project indicates how assessment, also summative, can improve learning. As will be pointed out later in the present study, this type of experiments should definitely be conducted also in Finland. Summative exams should always be used formatively in schools.

3.1.2 Assessment practices in the classroom

"Teaching involves assessment" (Rea-Dickins 2004: 249). Assessment, especially formative, is present in teachers' everyday work as they gather information about their learners' progress and assess learning outcomes and performances (ibid.). All this gives teachers tools to adapt their teaching and meet the needs of individual students (OECD/CERI 2005: 21). Nevertheless, when asked about classroom assessment, teachers tend to tell about the formal assessment procedures that they use. The observation-driven approaches are often underplayed although they are clearly an essential part of everyday classroom practices. (Rea-Dickins 2004: 249.)

Many other researchers confirm Rea-Dickins' (2004) remarks. Huhta and Takala (1999: 197) claim that formative assessment that supports teaching is the most generally used form of assessment but as it is stated in the study of OECD/CERI (2005: 21), the most visible assessments in schools are summative. One reason for this is said to be the accountability requirement since schools often have to provide evidence of student achievement (OECD/CERI 2005: 21). Thus, formative and observation-driven assessments are strongly in evidence in everyday classroom practice but the formal assessment procedures are the centre of attention. I believe that another reason for this is that formal assessment

procedures appear more authoritative and explicit to people outside the school. It is hard to examine assessment when it is embedded in teaching and can happen spontaneously during classes. Parents, researchers and authorities urge to see grades that are based on standardised tests and assessments that they see as reliable and valid. In my opinion teachers' professionalism as assessors is somewhat undervalued. Be that as it may, Hill and McNamara (2011: 395), among other researchers, have acknowledged that classroom teachers are given more and more responsibility for assessment today. Next, one study concerning classroom assessment practices and the reasoning behind the practices is presented.

Cheng, Rogers and Hu (2004) conducted a comparative survey about English as a second or foreign language (ESL or EFL) teachers' assessment purposes, methods and procedures. The 267 from Canada, Hong Kong and China filled in an extensive questionnaire. Cheng, Rogers and Hu (2004: 367) found out that the teacher assess for several purposes which can be categorised into three constructs: student-centred, instruction-based and administration-based purposes. Likewise, when Cheng, Rogers and Hu analysed the practices the teachers used in assessing reading, writing and speaking/listening categorised the findings in the case of each skill to instructor-made assessment, student conducted assessment and standardised testing. Many different assessment practices were found and some practices were used more in one country than in others. For example student summaries of what is read and short answer items were the most common reading assessment strategies in Canada and Hong Kong whereas in Beijing, China, multiple-choice items and other formatted assessment methods were reported more. Student constructed assessments, such as journals, portfolios and peer assessment were most popular in Canada and least used in China. (Cheng, Rogers and Hu 2004: 370, 372, 378–379.) Cheng, Rogers, and Hu (2004: 378) explain the differences in the assessment practices with several factors including teaching experience, nature of the courses, teachers' knowledge of assessment, the influence of external testing, and the general teaching and learning environment. Thus, it is not only cultural differences but also teachers'

background and some very practical constrains, such as the number of students in a class, that can affect a teacher's uses of assessment methods.

The interest towards classroom assessment has increased during the last decades but as Rea-Dickins (2004: 251) points out, there is not much research that considers the instruction-embedded perspective of assessment. Seven years later Hill and McNamara (2011: 396) still confirm the view and state that are relatively few studies have actually researched the processes of classroom assessment. They see a need for comprehensive classroom assessment research and challenge researchers to study not only what the teachers do but also the things the teachers look for in assessment processes, what theories they base their assessments on and whether the learners share the same understandings. People should learn to see classroom assessment as a concept including all kinds of assessment from unplanned, unconscious and embedded assessment to planned, deliberate and explicit assessments. (Hill and McNamara 2011: 416–417.) The nature of classroom assessment is indeed multifaceted and complex. In order to fully implement the principles of the new assessment culture are recognise the formative assessment in classrooms more research is needed. Next, classroom assessment is viewed in the Finnish context.

3.1.3 Assessment in Finnish schools

In Finland the National Core Curriculum (NCC) sets the guidelines and principles for assessment in schools but the methods and execution of assessment in practice are decided in the curriculums of municipalities and individual schools (Luukka et al. 2008: 55). The NCC in Finland is an intricate system which foundations are deep in the history of the nation's culture, communication and exercise of power (Hildén 2011: 7). The aims of language education in Finland are related to language skills, cultural skills and learning strategies (POPS 2004: 138–142). The essential contents of language education are described in the curriculum through language use situations, focal points in grammar, cultural

skills, communication strategies and learning strategies (POPS 2004: 138–142, Hildén and Takala 2005 316).

It is stated in the current NCC for basic education, established in 2004, that the main purpose of assessment is to direct and encourage studying and describe how well a student has gained the goals set for learning and growth. Assessment should also help students to form a realistic understanding of their learning and contribute to the growth of their personalities. (POPS 2004: 262.) In addition to the cognitive and knowledge-related goals, the NCC highlights on-going feedback, versatile assessment and assessment criteria which is shared with the students and their parents (POPS 2004: 262–263, Huhta and Tarnanen 2009: 2). Furthermore, developing students' self-assessment skills is one important task of basic education. Students should be guided to assess their learning skills in order for them to see their own progress and set themselves learning goals. (POPS 2004: 264.) Hence, in Finland also students are supposed to have an active role in assessment. Altogether, based on the guidelines presented in the current NCC, it seems that assessment in Finnish schools is more or less in line with the 'assessment for learning' ideology, and moreover, with the new assessment culture – at least in theory.

The Common European Framework of Reference (CEFR) has influenced language teaching as well as language assessment all over Europe and has done so also in Finland (Huhta and Hildén 2013: 161). In the current NCC, the guidelines for language assessment are based on the CEFR (POPS 2004, Luukka et al. 2008: 56). The challenge in using CEFR scales in language assessment has been that they describe language ability on a rather general level, and hence, some claim that they are not ideal for assessment. Thus, when the CEFR was applied to the scales of the Finnish curricula, some clarifying descriptions were added, for example descriptions of learner errors and lacks in performance. (Huhta and Hildén 2013: 179.) It was, however, contradictory to the current understanding of language ability and learning to add learner errors to make the

scales more specific. The relevance of learner errors is questionable and what could have been added instead are examples of learner productions at each level of the scale. Yet, the Finnish CEFR modification is an extension of the original six-point scale as each level is divided into sublevels, for example level B1 is divided into B1.1 and B1.2. Today all language teachers in Finland are obligated to refer to the proficiency levels when grading their students. (Tarnanen and Huhta 2011: 131.)

In the Finnish comprehensive school, comprising the age group of 7–15-year-olds, students are assessed thorough the school year but according to the law each student is entitled to a school report at least once a year, that is, at the end of each school year (Opetushallitus n.d.). In reality primary school students usually receive a report twice a year but in lower secondary school, which is includes students aged between 13 and 15, the school year is divided into four to six learning periods and students receive a school report at the end of each period. The turning points are when a student finishes primary school, at the age of 12, and when a student finishes the whole comprehensive school, at the age of 15. Hence, the criteria for "good mastery" (that is grade 8 on a scale from 4 to 10) at the end of primary school and at the end lower secondary school are provided for second/foreign languages in the NCC for basic education (POPS 2004: 140, 142; Huhta and Tarnanen 2009: 2).

In the school report assessment can be verbal until the seventh form but after that numerical assessment is required (Opetushallitus n.d.). A grade describes the level of proficiency but verbal assessment can also include an account of student progress and the learning process. Assessment in the school report can thus be verbal, numerical or a combination of the two. (OPS 2004: 262.) The teacher is responsible for the assessment and gives the grades but students, peers or even parents can be involved in the assessment processes (Opetushallitus n.d.; Tarnanen, Huhta and Pohjala 2007: 381). The school reports are, however, only

one way of giving assessment feedback and assessment should be given mainly in other ways (Opetushallitus n.d.).

Although students are assessed thorough the time they spend in schools, Tarnanen, Huhta and Pohjala (2007: 382) argue that questions and problems related to language assessment are not well-known in Finland, not even amongst teachers. As Huhta and Tarnanen (2009: 3) state, "Finland is not a very testing-oriented country". Assessment has never had a perceptible role in the field of languages, or in any other fields of the Finnish society (Tarnanen, Huhta and Pohjala 2007: 382). Besides the Matriculation Examination, which students take at the end of upper secondary school, there are no nationwide high-stakes tests in Finland (Huhta and Tarnanen 2009: 3). Moreover, Tarnanen, Huhta and Pohjala (2007: 383) point out that there is very little public discussion about assessment let alone its quality. They list a couple of reasons behind the lack of discussion and state that language assessment as a discipline is rather new in Finland and there are only a few language assessment experts in the country. Also, since assessment is often included in teaching, it easy to ignore its importance. Some people believe that assessment just comes naturally to teachers, as a part of teaching. As a result, the knowledge of assessment is often very superficial and the problems of language assessment are often neglected. (ibid.)

As Tarnanen, Huhta and Pohjala (2007: 384) note, teachers in Finland have a huge responsibility for assessment. This means that all teachers should have a good knowledge of different assessment theories and practices. During the last two decades the national curriculum has changed and the emergence of criterion-referenced assessment, and the CEFR, have influenced the principles of language education and assessment in Finland significantly. Moreover, these changes have created new requirements for language teachers. (Tarnanen, Huhta and Pohjala 2007: 384–385.) Tarnanen and Huhta (2011: 131) believe that it is very demanding for many teachers to use the proficiency levels in assessment and especially in

grading since neither teachers nor students are used to assessing in terms of levels. In addition, they find that grading at the end of a learning period is often based on comparing students with each other, and hence, it can be problematic to combine this type of grading with the proficiency levels. I find that if comparison really is the basis grading in schools, the situation is rather alarming. Other students' performances should not have any effect on the assessment of individual students. The change from norm-referenced assessment to criterion-referenced assessment is huge but necessary and teachers need support and training in order to this change.

It seems that language teachers are aware of the lack in their assessment skills as surveys of teachers' needs for updating training show that language teachers wish to get more training in assessment (Tarnanen, Huhta and Pohjala 2007: 383). In a like manner, my own experience is that assessment is covered only superficially in universities' teacher training. Nevertheless, without proper training and clear nationwide assessment principles and practices one could argue that assessment in Finnish schools is quite informal and based on schools, or even teachers', individual interpretations of what counts and what is important when assessing language skills. In fact, Tarnanen, Huhta and Pohjala (2007: 385) point out that despite the national curriculum the assessment practices may vary quite a lot depending on the school and the teachers. A teacher's assessment can be based on grammar-oriented exams whereas another teacher might use self-assessment as well as various oral and written activities (ibid.). Likewise, Hildén (2009: 39–40) criticises the lack of agreed rules for weighting the different subject specific learning objectives and cross-curricular aims. Many studies also show that language assessment is not consistent throughout the country. The ways how different teachers see language ability and weight different skills in assessment vary which signals that the changes towards criterion-referenced thinking have not yet reached the practice. (Tarnanen, Huhta and Pohjala 2007: 385–386.)

Althought the current NCC encourages teachers to advance varied assessment practices (Tarnanen and Huhta 2011: 131), it is not self-evident that the variety of assessment methods is used and endorsed in language classrooms across the country. Huhta and Hildén (2013: 164) remind that the assessment of written production skills have long traditions in the Finnish school system. They argue that although the focus in language assessment has shifted from grammatical correctness towards communicative ability, the methods of assessment have not changed significantly. Indeed, Tarnanen and Huhta (2011: 130) argue that the role of grammar is still significant in language teaching and that written skills and summative exams are emphasised in assessment.

Jorma Kauppinen from the Ministry of Education and Professor Jouni Välijärvi, the head of the Finnish Institute for Educational Research, state in a news article (Vähäsarja 2014) that although the traditional exam is only one form of assessment, it has strong traditions in the Finnish schools. Kauppinen and Välijärvi would like to encourage teachers to try new assessment methods and Välijärvi even suggests that at least in some schools exams could be renounced entirely. Välijärvi and Kauppinen acknowledge that it is hard to abandon old practices and that also students can be very conservative. Kauppinen recalls cases where some upper secondary schools tried exam-free learning periods but the students got offended and begun to protest. Students do not have much experience on other assessment methods and exams are often considered an easy method for assessment. (Vähäsarja 2014.)

As mentioned, the study of language assessment at schools has been a rather neglected topic in Finland. There are not many studies that report the assessment practices or processes used in Finnish schools. Nevertheless, Luukka et al. (2008) conducted a large study which aim was to discover 9th grade students' and their language teachers' literacy practices, pedagogical practices, media use and the way how the current language teaching faces the challenges set by the modern society (Luukka et al. 2008: 15, Tarnanen and Huhta 2011: 131). Language

teachers' assessment and feedback practices were also covered in the study. Luukka et al. (2008) provide a rather comprehensive and unique review of the assessment practices in Finnish schools examining the views and opinions of both teachers and students.

The survey conducted by Luukka et al. (2008) was conducted in 2006. Altogether 1720 students, 417 Finnish teachers and 324 foreign language teachers participated in the study (Luukka et al 2008: 35–36). In the present thesis only the results regarding foreign language teachers' and students' responses to questions about assessment and feedback practices are reported and analysed. The questions included themes such as who assesses pupils' knowledge, whose assessment affects pupils' grades, what is emphasised in the assessment of language skills and how feedback is given (Tarnanen and Huhta 2011: 132, 143–146).

The study by Luukka et al. (2008: 123) shows that teachers are the main assessors in schools. A vast majority of teachers aknowledged that teachers conduct assessment more frequently than any other party (Tarnanen and Huhta 2011: 132). Self-assessment, however, appears to be rather common as well whereas peer assessment is used rarely. Furthermore, the study reveals that self-assessment and peer assessment have hardly any influence on grades. It becomes apparent in the study that teachers do not entirely trust the students' ability to assess their own skills reliably and, on the other hand, some students do not trust their own self-assessment abilities either. Only 58 percent of the foreign language teachers thought that students are able to assess their own skills realistically. The corresponding percentage for students was 69 but, then again, about 30 percent of the students felt insecure about their self-assessment skills. (Luukka et al. 2008: 124–125.) Luukka et al. (2008: 125–126) suggest that the reason for the students' insecurity might be that they do not have much experience of self-assessment or that they have noticed that their self-assessment has had no influence on their grades.

When the teachers were asked about the skills and competencies they weight in the final assessments of the 9th graders, the five most important ones were ability to communicate in everyday situations, grammar and language structures, listening comprehension, interactive speaking and reading comprehension. Moreover, when giving grades, exams and active class participation have the biggest effect. The students held the same opinion to some extent but they emphasised diligence, active participation in class and success in exams more than the teachers. The most significant differences were related to language skills and competences that teachers consider important as only a few percentages of the students believed that listening and reading comprehension are valued. (Luukka et al. 2008: 128-132.) Thus, in general the students seemed to have an idea of what their teachers assess but they did not have a clear understanding of all language related aspect that are being assessed.

Luukka's et al. (2008: 139–142) study proves the importance of giving and receiving feedback. Most of the students and teachers, who participated the study, marked that they think that feedback affects student's behaviour and learning. Nevertheless, a third of the students felt that they did not receive enough feedback. Both teachers and students mentioned teacher comments and corrections in students' written work as the most common way of giving feedback. Teachers also tend to give personal feedback during lessons and go through answers with the class while returning exams or written work back to their students. The teachers' and students' responses were rather similar except that only 46 percent of the students reported receiving personal feedback during classes although over 70 percent of the teachers reported giving personal feedback often. Luukka et al. (2008: 142), however, believe that this difference is due to the fact that although teachers give feedback to their students often, the individual students in the class receive it relatively seldom as there are so many students in the classes.

On the whole, Luukka et al. (2008: 154) came to the conclusion that assessment and feedback seem to have a significant impact on what the students feel they can. Exam grades and success in school influence the way how students assess their abilities although it seems that students' experiences of using foreign languages in everyday situations, also outside school surroundings, matter the most.

The study by Luukka et. al (2008) shows that teachers have a leading role in language assessment. It seems, however, that also students are involved in assessment in Finnish schools (Tarnanen and Huhta 2011: 140). Still, it appears that the possibilities of self-assessment are not fully exploited and self-assessment may sometimes be only superficial. Student self-assessment has no effect on grading, and hence, doing self-assessment can feel demotivating or even useless. On the other hand, teachers do not trust their students' self-assessment skills. Tarnanen and Huhta (2011: 141) conclude that self-assessment, let alone peer-assessment, not fully integrated into assessment practices in school. Moreover, they point out that many students are not entirely familiar with the assessment criteria. As pointed out by Luukka et al. (2008: 120), assessment can motivate students if it is target-oriented and the assessment criteria is shared by both students and teachers, and hence, assessment supports learning. On the other hand, if assessment is based on student control and pursuing of good grades, the motivation is only superficial.

All in all, the purpose of language assessment is to support learning, motivate learners and provide them with a realistic picture of their skills and needs of improvement (Tarnanen, Huhta and Pohjala 2007: 381). Furthermore, Tarnanen and Huhta (2011: 131) point out that if responsibility for assessment, the assessment criteria as well as learning objectives are all shared by teachers and students, the students learn more about assessment, and hence, are better able to participate in assessment that enhances learning. They believe that the Finnish school system allows the implementation of the principles of the new assessment

culture but educators need to be responsive to the opportunities. Nevertheless, based on what is discussed so far, it seems that the range of assessment methods used in classrooms is quite limited, and that the students' role is assessment is still minor (Luukka et al. 2008, Vähäsarja 2014). It appears that the aims and methods exist but teachers, students as well as other educators lack the courage or motivation to abandon the old traditions and try something completely new in language assessment. One alternative to implement versatile language assessment, student participation and lifelong learning is to use the European Language Portfolio.

3.2 The European Language Portfolio

The new language teaching and assessment culture, which also the Council of Europe promotes, crystallises in the following concepts: linguistic and cultural diversity, intercultural learning, lifelong, plurilingual learning and learner autonomy (Little 2009: 1, Council of Europe 2011). Especially learner autonomy, that is learners adopting learning to learn skills and taking responsibility of their own learning and assessment, is emphasised in today's language teaching and assessment (Luukka et al. 2008: 56, Little 2009: 1–2). Kohonen (2005a: 8) notes that also the importance intercultural language learning has changed the goal of language learning from communicative competence to intercultural communicative competence.

As noted previously, the changes is the field of language assessment have also increased the popularity of alternative assessment methods. These methods endorse the new assessment culture and as such the term *alternative assessment* is often connected to some more creative and authentic assessment methods, such as portfolios, diaries, learning projects and self-assessment. Indeed, Fox (2008: 99) remarks that for long portfolio assessment has been associated with alternative assessment especially in the second language research literature. Portfolios are said to be a means of connecting assessment and learning seamlessly to a unity that manifests the assessment for learning philosophy (Little 2009: 4). Fox (2008:

99), however, adds that it is not the form or format of portfolios that makes them alternative assessment tools but rather the way how portfolios are used.

All the principles of the new teaching and assessment culture are fostered in the European Language Portfolio (ELP), which integrates language learning and assessment. It embraces authentic assessment, supports learners' spontaneous language learning, considers assessment as a process and, most importantly, requires criterion-referenced self-assessment (Kohonen 2005a: 8, Kohonen 2005b: 28). The ELP is not, however, just any portfolio but it is a product of large research and development processes in the field of language teaching and assessment conducted by the Council of Europe (Luukka et al. 2008: 56). The Council of Europe developed the ELP to forward international mobility but for individual learners the purpose of the ELP is to support multilingualism, multiculturalism, learner autonomy and life-long learning (Kantelinen and Hildén 2012: 1).

The development of the ELP, as it is known today, began in 1991 and since then the ELP has been validated and registered in more than 30 European countries (Kohonen 2005a: 11, Kantelinen and Hildén 2012: 1). The ELP is strongly linked to the other product of the Council of Europe, the CEFR (Kohonen 2005a: 11). Indeed, in the ELP learners assess their work and skills in relation to the proficiency levels of CEFR. This enables comprehensibility and transparency of assessment as the proficiency scales are known all around Europe. Furthermore, all the models of the ELP registered in different European countries must adhere to the set principles and guidelines of the ELP (Council of Europe 2011). Although the ELPs can be diverse and personal, the jointly agreed principles and components must be present in all national and individual ELPs in order for them to be recognisable and intelligible all around Europe (Kantelinen and Hildén 2012: 2).

The ELP is a collection of a learner's work that the learner himself or herself has compiled, reflected on and assessed (Kohonen 2005b: 26). Consequently, it is the

learner's property (Council of Europe 2011). The ELP, however, always consists of three parts complementary to each other: the Language Passport, the Language Biography and the Dossier (Kohonen 2005a: 11–12). The Language Passport includes the learner's criterion-referenced self-assessments of his or her skills in different languages at different stages of the learning path. It also includes information about the learner's intercultural experiences (Kohonen 2005a: 12). The aim is that the learner updates the Language Passport regularly in order to indicate progress in learning (Council of Europe 2011). In the Language Biography the learner reflects on and assesses his or her own learning and ponders his or her language learning experiences. The actual portfolio tasks are presented in the Dossier. It consists of a learning folder, which includes all the learner's portfolio work, and of a reporting folder, where the learner reports and describes his or her language skills with the best pieces of work that s/he has chosen. (Kohonen 2000: 15–16, Kohonen 2005a: 12.) The pieces of work in the reporting folder can be for example written stories, poems, songs, plays, video clips from communication situations abroad, teachers' assessments or anything that substantiates the skills presented in the Language Passport and the Language Biography (Kantelinen and Hildén 2012: 3).

Thus, the ELP has a pedagogical and a reporting function as it on one hand develops the learner's learning to learn and self-assessment skills but on the other hand reports the learner's language skill based on the CEFR (Kohonen 2005a: 10–11). Kantelinen and Hildén (2012: 3) remark that the ELP is a tool for life-long learning and it adjusts to different types of language learning paths. As Kohonen (2005b: 26) summarises, the ELP presents the learner's language learning and learning experiences in different languages, cultural skills as well as overall progress in language learning.

3.2.1 Language Portfolio assessment

The European Language Portfolio assessment (*salkkuarviointi* in Finnish) refers to the manifold assessment conducted by the teacher and the reflective self-

assessment conducted by the learner while working on his or her portfolio (Kohonen 2000: 17). Hildén and Takala (2005: 318–319) remind, however, that the ELP combines teaching, learning and assessment in a way that it is not justified to talk about the ELP only as an assessment method. They further note that the pedagogical and reporting functions of the ELP should not be separated too much as the two functions follow each other in cycles. Moreover, in the ELP both formal and informal learning are recognised and learning in school is not separated from learning outside of school. The idea with the ELP is that all learning is important. (Kantelinen and Hildén 2012: 3.)

Since the main aims of the ELP are to enhance learner autonomy as well as interactive and communicative language learning, self-assessment and peer assessment have a significant role in the ELP work (Kantelinen and Hilden 2012: 3). Self-assessment in the ELP is executed in relation to the proficiency levels of the CEFR (Little 2009: 3). Assessment is viewed positively and learners abilities emphasised when the learners' assess themselves with *can do* statements that are provided for listening, reading, spoken interaction, spoken production and writing (Council of Europe 2001: 25–27, Eurooppalainen kielisalkku). For example, in level A2 in spoken interaction learners assess their skills though statements such as *I can apologise and reply to an apology, I can order something to drink and eat* and *I can start and end a discussion* (Eurooppalainen kielisalkku). According to Little (2009: 3), the checklists are used to identify learning objectives, observe progress in learning and assess learning outcomes, that is, formative, constant reflection of learning. Moreover, in the Language Passport learners assess and summarise their overall language proficiency regularly (Little 2009: 3). The checklists in the ELP enable even young learners to recognise progress and set new goals for learning (Hildén 2011: 14). Indeed, the self-assessment does not concern only language skills but also other skill and competences, for example cultural knowledge and learning to learn skills (Hildén and Takala 2005: 317).

Self-assessment is, however, itself a skill that has to be learned and practised. Little (2009: 3) discusses three self-assessment related concerns that have arisen during the ELP pilot projects. First, some suspect that learners do not have enough knowledge of assessment to assess themselves. This, as Little (2009: 3) comments, derives from the idea that teaching and learning are somehow separate from assessment, and hence, self-assessment would be something that learners do by themselves, separate from the learning process. Second, some educators fear that learners overestimate their skills, and third, learners can be induced to deceitfully include somebody else's work in their ELPs. The second and the third concern originate from the conception that learners cannot participate in their own assessment since in many cases formal examinations direct learners' possibilities in the future. This can, nonetheless, be avoided if the ELP and its reflective activities are given a central role in learning, and thus, learners gradually learn self-assessment skills and become proficient in assessing their performances with regards to the CEFR scale. (ibid.) Likewise, Hildén (2011: 14) notes that the students who have become familiar with the curriculum aims in their studies are able to assess their progress and skills rather unanimously with their teachers and peers. She adds that that self-direction and responsibility progress naturally when learners are guided and encouraged to take notice of their progress and keep their portfolio work for themselves.

Self-assessment and reflection skills are essential for life-long learning and it is vital that students learn to assess their performances realistically (Alanen and Kajander 2011: 68). When learners conduct self- and peer assessments they develop become aware of their progress in learning, which, furthermore, facilitates the forming of individual learning targets (Katz and Gottlieb 2012: 165). Alanen and Kajander (2011: 68–69) state that self-assessment affects also a learner's self-esteem and confidence in taking control of his or her learning process. Douglas (2010: 75), however, reminds that it is important that learners receive enough practice of self-assessment. Alanen and Kajander (2011: 69) emphasise that the key to successful self- and peer assessment is clear assessment criteria, which is taught to the learners, and continuous, guided practice. They

further note that the different kinds of self-assessment forms are only meant to guide and facilitate self-assessment in the early stages and, in time, the assessment should be more free. The same applies to reflection skills too. Alanen and Kajander (2011: 69–70) suggest that organised reflection and dialogues between teachers and students are cornerstones of autonomous self-reflection.

In conclusion, the ELP is so far the most advanced pedagogical tool that combines the common educational aims, strategic skills and cultural knowledge (Hildén 2011: 15). Indeed, Kohonen (2000: 35) remarks that the use of the ELP enables the development of the new language teaching and assessment culture. He lists eleven points which the use of the ELP endorses:

- assessment is an essential part of teaching
- learners are viewed as individuals
- learners reflect on their progress in the long term
- learners assess also each other's work together
- information for teaching and assessment practices is gathered in multiple ways
- there are several possible viewpoints instead of one "correct answer"
- information gained from assessment is used in improving learning
- individual learners can affect the content of learning and assessment
- studying supports the development of learners' reflective, self-assessment and critical thinking skills
- learners are encouraged to work together
- learners' skills are compared with the set learning goals and their previous progress, not with their peers

Thus, the ELP takes into consideration all the aims of alternative assessment in the new era.

Despite the versatility and other good qualities of portfolio assessment, the use of portfolio assessment, and alternative assessment methods in general, has been sometimes criticised. The criticism focuses most often on the reliability and validity of portfolio assessment as well as on the costs of the time-consuming assessment method (Fox 2008: 104). As was previously stated, traditional testing adheres to quality measures such as validity and reliability, that is, in short, procedures which assure that the assessment method in question assesses what it is claimed to assess and that the scoring is consistent (Lynch and Shaw 2005: 266, 268). Therefore, if alternative assessment methods are to be used as meaningful assessment methods, their validity and reliability need to be assessed too. The critics of portfolio assessment, however, think that the same criteria of reliability and validity used with traditional language tests should be applied to alternative assessment (Lynch 2001: 360, Fox: 2008 105). This is a very peculiar demand as the core ideas of alternative assessment methods and traditional testing are very unlike. Lynch (2001: 360–361) exemplifies this by stating that within the alternative assessment paradigm measurement and tests are not used for gathering information for decision makers. Furthermore, she emphasizes that since there are no separate components to be measured, one cannot apply the traditional criteria. She suggests that the validity of alternative assessment methods should be judged against criteria that are relevant to the principles of alternative assessment. Indeed, as Fox (2008: 105) notes, the distinctive and varying nature of the information gained from portfolio assessment, and other alternative assessment methods, which challenges the traditional assessment quality criteria.

What comes to reliability of the ELP, Hildén and Takala (2005: 319) argue that since the ELP is a collection of student work conducted at different times, in different situations and in different ways, incidental factors such as a learner's mood or clarity of instructions become insignificant, and hence, the ELP can be considered a reliable method. Furthermore, when assessing validity, many proponents of alternative assessment argue that the features of authenticity of, for example, portfolio assessment are adequate for proving validity (Fox 2008:

105). Lynch (2001: 364–366) suggests that the validity framework for alternative assessment should include careful reasoning of questions such as "Are the perspective of all affected participants in the portfolio assessment process being taken into account?" or "What specifically is done as a result of the assessment?". Hildén and Takala (2005: 320) find that the authenticity, flexibility and realism of the ELP work make the ELP meaningful and effective. They further note that the ELP is personal, contains examples of functional language ability and that the portfolio tasks are connected to different sorts of communication situations. This, in their opinion, means that the ELP can be considered a fair assessment method for individual learners. Nonetheless, Hildén and Takala (2005: 320) admit that the ELP cannot fulfil the requirements of commensurate effectiveness but traditional testing in steady testing conditions might be sometimes appropriate. These types of tests can also be included in the ELP (ibid.).

It must be noted here, however, that the criticism towards the quality of alternative assessment methods, particularly portfolio assessment, has been more common in high-stakes contexts than in classroom contexts (Fox 2008: 105). It is inevitable that if or when portfolio assessment is used in high-stakes contexts it sets new requirements for the assessors, let alone resources. Although portfolios often follow a set structure and guidelines, they are always individual pieces of work and cannot be rated the same way as traditional tests. In portfolio work there is more room for individual variation. There are no right or wrong answers in the same way as there might be in, for example, multiple-choice tasks. That is, however, the whole point of portfolio work and portfolio assessment. Language ability is a tool for communication and cannot be tested comprehensively with standardised written tests. In my opinion, alternative assessment methods should be used more widely but when norm-referenced traditional testing is required alternative assessment methods are not the most suitable ones. Nevertheless, accountability is rarely the purpose of assessment in classrooms.

3.2.2 The ELP in Finland

In Finland ELP research has been quite active since the 1990s and ELP experiments have been conducted around the country (Kantelinen and Hildén 2012: 1). Viljo Kohonen and Ulla Pajukanta coordinated the first ELP project in Finland in 1998–2001 (Kohonen 2005b: 44) and since then particularly Kohonen has been active in developing and promoting the ELP. In 2011 the Finnish National Board of Education ordered a model for the Finnish ELP and the work was conducted in cooperation between four Finnish universities in 2011–2012 (Salo et al. 2013: 38, Perspectives on the European Language Portfolio for the comprehensive school in Finland 2014: 1). Today the Finnish versions of the ELP (FinELPs) comprise three models aimed at different classroom levels: FinELP for grades 1–3, FinELP for grades 4–6 and FinELP for grades 7–9. These include pupils in primary and lower secondary school, aged from seven to fifteen years. The FinELPs are founded on the current NCC and, of course, follow the guidelines and principles of the ELP. The three models can be downloaded from the Finnish ELP website, and thus, they are available for all language teachers and learners around the country. (Perspectives on the European Language Portfolio for the comprehensive school in Finland 2014: 1–2.)

Although quite a lot of research about the ELP has been conducted in Finland, the research has been mostly ELP teaching experiments conducted in the beginning of the 21st century when the ELP project was launched in Finland. Salo et al. (2013: 38) note that despite the various trials and experiment conducted during the launching project, the use of the ELP is still rather limited and the ELP is not very well-known. Furthermore, they acknowledge that the ELP is used mainly by teachers who have been actively involved in the development process of the FinELP. This was also my observation when I tried to find teachers, who are using the ELP, for the present study; it was very challenging to find lower secondary school language teachers who use the ELP. Kantelinen and Hildén (2012: 1) believe that the reason behind this is that there has been no full version of the ELP available nationwide, and thus, the idea of the ELP has seemed rather

fragmented. The electronic versions of the ELP were not published until 2013. In addition, Salo et al. (2013: 38) profess that since teachers have no personal experiences of using the ELP it might feel foreign to them. Nonetheless, various ELP experiments and development processes have been well documented and there are several reports available about the use of the ELP in language classrooms at different levels.

Finland participated in a European ELP piloting project in 1998 by executing a three-year action research project (Kohonen 2005a: 26). In total, 420 students and 22 language teachers from lower secondary school, upper secondary school and vocational schools in Western Finland were involved in the project. The data of the experiment consisted of teachers' and students' written reports, classroom observations, students' ELPs, assessment seminars and questionnaires made by the Council of Europe. Thus, the data was versatile and teachers developed their skills in using the ELP thorough the project. (Kohonen 2000: 25–26, Kohonen and Pajukanta 2003: 8–10.)

In summary, the ELP received quite positive feedback and comments from both teachers and learners. Kohonen (2005a: 27) reports that the experiment showed that the ELP is a pedagogically justified tool for promoting self- and peer assessment. Indeed, increase in learner autonomy was noticed during the three-year project. The teachers considered the ELP as a new tool and a challenging opportunity for developing independent and socially responsible language learning. Teaching reflection and self-assessment skills, however, proved to be burdensome although the workload eased as students learned to help each other and took responsibility of their own learning. Moreover, some students did not like portfolio work and preferred teacher-directed learning. The perquisites for successful use of the ELP appeared to be, among other things, teachers' enthusiasm and knowledge of the principles of the ELP, consistent and persevering guiding of students to understand the assessment criteria as well as

the idea of the ELP, feedback sessions and efficient integration of the ELP into the curriculum. (Kohonen 2005: 27–30.)

Kolu and Tapaninaho (2000, 2003) participated in the above-mentioned project and reported their experiences of using the ELP in English and Swedish teaching in two lower secondary schools in Tampere, Western Finland. Before the project the teachers were slightly worried about the amount of work that the ELP would require and they were also pondering whether self-assessment would help their pupils to become independent and responsible language learners (Kolu and Tapaninaho 2000: 83). During the project Kolu and Tapaninaho (2000: 87) noticed that before pupils really become autonomous learners, a lot of guidance and practice of reflection and self-assessment is required. Moreover, they found the CEFR scale too broad and generic, and thus, added one more level, A3-level to the scale. The CEFR scales proved to be rather difficult for pupils to understand since the CEFR levels are not equivalent to the learning targets of the courses. (Kolu and Tapaninaho 2000: 89–90.) In Swedish language learning the teachers decided that pupils would not assess their portfolio work according to the CEFR scale since they were only just starting to learn the language. In Finland most pupils start learning Swedish in lower secondary school and reaching even the first level, A1, during the first year is challenging for many students. Thus, they used only verbal feedback and traditional numeric assessment. (Kolu and Tapaninaho 2000: 98.)

The three-year project resulted in many positive experiences. Kolu and Tapaninaho (2000: 98, 103) noticed for example that reading each other's work was very beneficial for the pupils. Kolu and Tapaninaho found that by observing their own work and the work of others, pupils learned to set realistic goals for themselves and reflect their own skills. After getting used to peer assessment also the pupils found peer assessment useful for the same reasons as the teachers. Moreover, the teachers observed that as peer assessment became a habit, the

social atmosphere and pupils' cooperation skills improved. (Kolu and Tapaninaho 2003: 44–45.)

During the three years pupils also learned to take responsibility of their own learning and begun to understand that language ability consists of several skills which require a lot of work and patience. (Kolu and Tapaninaho 2000: 103 and 2003: 35, 48). Furthermore, assessment became a natural part of learning and the teachers experienced it easier because it was now continuous (Kolu and Tapaninaho 2000: 103). The pupils found especially checklists useful and they gradually learned to use the lists in assessment (Kolu and Tapaninaho 2003: 51). The checklists and self-assessment faced, nevertheless, also quite a lot of criticism during the project as the pupils felt that there were too many forms and lists to fill in. Many pupils got bored with continuous self-assessment. (Kolu and Tapaninaho 2000: 99 and 2003: 53.) In addition, the CEFR scale was found too broad. In comprehensive school pupils often reach A2-level, sometimes B1-level, and hence, it is difficult for pupils to notice their progress when they stay at the same level for a year or two or even longer. This was most apparent in Swedish language learning in which only some pupils barely reach level A2 during the three years they spend in lower secondary school. (Kolu and Tapaninaho 2000: 100 and 2003: 37.) Thus, the broad scale is not very motivating for pupils but helps in setting realistic and concrete learning targets (Kolu and Tapaninaho 2000: 100–101).

From the teachers perspective the most challenging issue with the ELP was the workload and persistence that the new way of working causes. First of all, teachers need to internalise the principles of the ELP and let go of old habits. They need to study the CEFR scale and develop their teaching and assessment forward. There are no previous practices but the teachers just need to believe in themselves and apply the ELP to the classroom practices. (Kolu and Tapaninaho 2000: 89, 102.) Secondly, the teachers need to motivate their pupils to learn a new way of working and assessing. Many pupils have strong views of what learning

in school is and not all pupils are willing to change the practices especially if it requires much work. (Kolu and Tapaninaho 2000: 55–56.) Thus, the beginning is hard and a lot of work is required before the ELP work functions smoothly. The project conducted by Kolu and Tapaninaho shows that there are challenges with using the ELP, especially in the beginning, but once both teachers and pupils adapt to the new way of working, the results are rewarding. This, however, is the case with everything new, and one should not be scared of the challenges of the beginning.

Based on the good experiences gained from the three-year piloting project, Finland participated in a European ELP implementation project by organising a national ELP mentoring project in 2001. During the project, which involved teacher training departments from six Finnish universities, local three-year ELP subprojects were executed around Finland. (Kohonen 2005a: 31–32.) In her article, Hildén (2004) reports the experiences of teachers in the capital region. In the subproject three mentors guided twenty language teachers in six schools, including primary schools, lower secondary schools and upper secondary schools. The teachers' observation in this subproject were very similar to what Kolu and Tapaninaho (2000, 2003) had observed. Most pupils learned to assess themselves quite realistically and they found self and peer assessment useful. Nevertheless, as time went by, the pupils' motivation to assess themselves began to weaken and pupils began to criticise the need for continuous self-assessment. (Hildén 2004: 94–95.) All in all, the teachers thought that pupils responded to the ELP mainly positively, although some preferred more teacher-directed learning (Hildén 2004: 95). Hildén (2004: 95) points out that some pupils may be so used to the traditional setting where the teacher is responsible for all assessment and distribution of work that the pupils shun the ELP work. This is very likely since traditionally pupils have been responsible only for doing their homework and studying for exams. It is no wonder that the ELP feels foreign at the beginning.

For teachers the project appeared to be burdensome but rewarding. The teachers reported that especially the continuous assessment, correcting of pupils' portfolio work and a new way of working made the implementation of the ELP hard. On the other hand, many teachers acknowledged that using the ELP has provided them with a more practical perspective on developing their teaching and the whole school community. Co-operation with other language teachers appeared to be valuable asset in the project, and thus, the project encouraged teachers to work more together. (Hildén 2004: 97.)

The subproject reported in the article by Hildén (2004) continued still for a year, and thus, the article did not include the final results and thoughts of the teachers. Nevertheless, the main advantages and challenges of ELP appear to be the same in most ELP studies. The age group of the pupils involved in the ELP projects may also affect the results as this far the focus has been on ELP experiments and studies involving pupils under 18 years old. One could assume that learner autonomy would be easily adopted by older students who have more developed cognitive skills. Indeed, the ELP can be used and utilised in many ways and in different forms, also in adult education (Saarinen 2010: 12). European Language Portfolio experiments have been conducted not only in comprehensive school but also at universities of applied sciences, for example.

In her article Saarinen (2010: 12–18) presents ways to use the ELP in language courses at the universities of applied sciences and summarises the advantages and challenges of using the ELP in universities of applied sciences, institutions in which the aims of vocational education are integrated into language teaching. Saarinen (2010: 14–16) acknowledges that the ELP enables and contributes to authentic assessment and student autonomy, but moreover, the ELP work enhances also other skills, such as time management, problem-solving skills and thinking and reasoning skills. She has also noticed that the ELP enables teachers to develop their teaching.

Saarinen (2010: 17–18) admits that the new role of a student as an autonomous learner, who takes responsibility of his or her own learning and regularly reflects his or her own work, can be challenging for some students to adapt. She further notes that some students feel that assessment is a teacher's job and that a teacher should also be the one setting the learning objectives. Indeed, some students want ready-made answers and take a rather passive role in learning. Furthermore, Saarinen (ibid.) notes that there are students who have very poor understanding of themselves and their abilities, sometimes due to bad leaning experiences in the past. Thus, Saarinen (2010: 19) concludes that it requires time and effort to adjust students to this new way of thinking and learning. The downside is, however, that at many universities of applied sciences the amount of contact lessons varies enormously between different groups and sometimes the groups can be so big that continuous feedback and assessment becomes burdensome and the time is very limited (Saarinen 2010: 17).

It has become evident that abandoning old habits and adopting a new teaching and learning paradigm is challenging for both teachers and learners. Moreover, the studies show that no matter the age of a student, self-assessment and reflection skills need to be practised, and the process of adopting a more active role as a student requires time and effort. All studies, however, show that in the long run learner autonomy develops and learners begin to understand the extent of language ability. Assessment is easier when it is integrated into learning and teachers and students work together to create a comfortable and personal learning environment (Kolu and Tapaninaho 2000: 103). For these and many more reasons the ELP should be used in language teaching and learning. Moreover, the ELP supports the aims of the national curriculum and is an excellent tool not only for assessment but also for developing learners' learning skills and cooperation skills (Salo et al. 2013: 39). For teachers, on the other hand, starting to use the ELP can be a great opportunity to increase cooperation between the schools' language teachers.

Some challenges have come up during the two major ELP projects executed in Finland. Teachers have found the projects rather burdensome, it has been difficult to integrate the ELP into language teaching, learners have criticised the amount of self-assessment and the levels of CEFR appear to be too generic and hard for learners to understand (e.g. Kohonen and Pajukanta 2003: 24–28. Since the first ELP project, the CEFR scales have been adapted so that more sublevels have been added. I also believe that when the ELP becomes more common in Finland and teachers and learners become accustomed to the principles of the ELP, the workload will diminish and teachers will find routines that suit their learners. The amount of self-assessment and reflection is one of the practices that should be adjusted to the needs of the learners. If self-assessment becomes a forced activity, the main idea of it will be lost. All in all, it is important to remember that the ELP is not extra workload added to previous practices but the ELP is to replace old practices (see Salo et al. 2013: 39).

3.3 Future directions for language assessment

As discussed above, the concept of assessment has expanded notably over the last few decades. The main purpose of assessment is to enhance learning and learners are assessed against a set criteria, not against each other. Nowadays there are also other actors than only a teacher in the assessment process: the teacher has an important role but also the learner and his or her peers are involved in the process (Keurulainen 2013: 37). Indeed, in the era of the new assessment culture assessment is integrated into teaching and learning. Assessment is considered a pedagogical tool (Rea-Dickins 2011: 12).

In December 2014 the Finnish National Board of Education approved a new National Core Curriculum, which will be implemented in the autumn of 2016 (POPS 2014). Assessment has a more visible role in the new curriculum and it seems that finally more attention is paid to the meaning of assessment. One small but significant difference can already be seen at the title of the assessment section. The title in the current curriculum is ‘Assessment of the learner’ whereas in the

new curriculum the section is called 'Assessment of learning'. Thus, the focus will be more on the learning process in the future. The new curriculum continues to emphasize self-assessment and the use of a variety of assessment methods even more. It also includes much more detailed descriptions of, for example, the principles of assessment and the nature of assessment. Moreover, there are more precise aims for assessment at different stages of a student's learning path. Most importantly it is clearly stated in the new curriculum that the emphasis is on assessment that enhances learning and that students or their performances should not be compared with each other (POPS 2014: 48). Thus, it can be said that the new curriculum endorses the principles of the new assessment culture. It is immensely important that it is now clearly stated in the NCC that the main aim of assessment is to enhance learning. Although time, training and also some courage is required before the changes will actually materialise, I consider these changes significant.

The European Language Portfolio is one form of assessment and learning that embodies the aims of the current paradigm of language learning, teaching and assessment but also the aims of the national curriculum. The use of the ELP is now recommended in the new curriculum for the first time. The ELP is mentioned in several parts, for example in the principles of language assessment. In the assessment sections for each foreign language the ELP is used as an example method to be used in the assessment of both learning and learning skills.

The concept of language ability is still undergoing a significant change and Huhta and Hildén (2013: 181) predict that assessment methods will change in the future as well. They state that the current assessment methods cannot comprehensively and reliably assess all the aims of modern language education. International communication skills and authentic language use should be included in the assessment processes better (ibid.). Likewise, Turner (2012: 75) believes that classroom assessment will develop and stimulate more research in the future. Inbar-Lourie (2008: 296) upholds an idea that in the future external

and internal classroom assessment data are used "in a non-threatening manner" to create socially interactive learning environments which constantly reflect on their practices. Moreover, he suggests that teachers would interact and work in collaboration with external language assessment experts in order to gain mutual expertise.

Huhta and Hildén (2013: 181) as well as Inbar-Lourie (2008: 296) predict also that technology will have a more powerful role in language assessment in the future. Many international language tests can already be done online and Huhta and Hilden (2013: 181) suspect that that will be the tendency with other tests and assessments as well. Moreover, the current technology could provide the means for on-line feedback. This would not have to be restricted only to classrooms but also different international language tests could provide online feedback in the future. (Huhta and Hildén 2013: 181–182.) Time will tell what the future of assessment will be but technology will certainly affect language assessment practices and research already in the near future.

In the following chapter the present study will be introduced in more detail. The research questions are covered first and after that the data collection as well as the method of data analysis will be viewed.

4 THE PRESENT STUDY

The present study is a qualitative, phenomenographic study aiming to collect teachers' and pupils' experiences and opinions about using the European Language Portfolio in language assessment. Sumner (2006: 249) defines qualitative research as following:

> Associated with a variety of theoretical perspectives, qualitative research uses a range of methods to focus on the meanings and interpretation of social phenomena and social processes in the particular context in which they occur.

Although, the dichotomy between qualitative and quantitative is not considered as straightforward as it was a couple of decades ago, qualitative research is often related to studying subjective meanings through which people interpret the world (Sumner 2006: 249, Hirsijäri and Hurme 2008: 21). Moreover, Sumner (2006: 249) outlines that in qualitative research the researcher focuses on understanding the actor's, e.g. the interviewee's, perspective on the research topic. Indeed, in the present qualitative present study the aim is to understand language teachers' and pupils' perceptions of language assessment and of the ELP.

The present qualitative study takes a phenomenographic approach to examining the ELP as an assessment method. Limberg (2008: 612) states that "phenomenography is a research approach aimed at the study of variation of human experiences of phenomena in the world". Hence, it is about studying how different people see and experience different phenomena and processes (Limberg 2008: 612). In the present study the phenomenographic approach is present since the main idea of the study is to learn about teachers' and pupils' conceptions of the ELP and assessment. It is their experiences and conceptions that are key questions here.

The research questions are introduced next. In section 4.2 the methods of data collection are described, followed by an account of the participants of the study. Finally, the method of analysis is explained before the results discussed in section 5.

4.1 The research questions

The overall aim of the study is to review primary and lower secondary school English and Swedish language teachers' and pupils' conceptions of language assessment. Furthermore, the focus is on teachers and pupils who have used the ELP in language teaching and learning and the aim is to examine these teachers'

and pupils' experiences and opinions about the ELP in language assessment. As stated earlier, the ELP is more than an assessment method. It integrates teaching, learning and assessment, and thus, it is difficult and maybe even fallacious to talk about the ELP only as an assessment method. Thus, to be precise, the present thesis focuses on the assessment aspects of the ELP.

Previously it was reported that not only teachers but also students can be very conservative when it comes to assessment (Vähäsarja 2014). Students in Finland are used to being assessed for their performance in language tests and other ways of assessing seem strange. Moreover, the study conducted by Luukka et al. (2008) showed that teachers and students appear to have different understandings of what teachers assess, and that students would like to have more feedback from their teachers. The same study examined also self-assessment practices which seemed to be still rather superficial (Tarnanen and Huhta 2011: 141). On the other hand, the previous ELP studies suggest that self-assessment has been conducted to the extent that many pupils have begun to protest against it (see e.g. Kolu and Tapaninaho 2000: 99 and 2003: 53).

All these assumptions and results from previous studies are reflected in the present study. Assessment is reviewed from the teacher and the student perspective and the study examines different aspects of assessment practices: assessment methods used, the nature of assessment, the pupils' role in language assessment, and the advantages and disadvantages of using the ELP. Teachers of English and Swedish were chosen for the interviews which made it possible to examine the applicability of the ELP to different languages. In short, the present study seeks to answer the following questions:

- How do English and Swedish language teachers and pupils find the ELP as an assessment method?
    - what is ELP assessment in their opinion?
    - how are pupils' language skills assessed?

- what is the pupils' role in language assessment?
- are there any differences between using the ELP in the assessment of English and Swedish language learning?
- why to use the ELP?

4.2 Data collection

Like in most phenomenographic studies, the data for the present study was collected in interviews (Limberg 2008: 612). According to Hirsijärvi and Hurme (2008: 35), in interviews the research subject, that is, the interviewee is considered an active party who can create meanings and express his or her thoughts freely. Tuomi and Sarajärvi (2009: 73) find flexibility the greatest advantage of interviews. They state that the interviewer can among other things ask for clarification, repeat questions and reformulate phrasing or the order of the questions at any time of the interview. This is not possible in many other data collection methods, for example in questionnaires. Interviews enable more in-depth discussion about the research topic, and hence, are a better method for studying teachers' and pupils' experiences and thoughts of assessment. (Hirsijärvi and Hurme 2008: 35.) The need for an interactive data collection method also excluded other methods, such as observations and self-reports.

Interviews are also a much used method which many people are familiar with. Hirsijärvi and Hurme (2008: 11) point out that interviews are often considered pleasant by both parties. They further note that interviews are close to everyday practices, and hence, people often know what to expect when they receive a request for an interview. Yet, another reason for choosing interviews was that I had previous experience of using this data collection method. As Hirsijärvi and Hurme (2008: 35) state, a lack of skills and experience can be problematic when conducting interviews.

The interviews were semi-structured interviews based on chosen themes. Like usually in semi-structures interviews also in the present study the themes and the questions were planned beforehand but the sequence and form of the questions varied from one interview to another (Hirsijärvi and Hurme 2008: 47). Some questions were also added during the interviews if the discussion led to a new and interesting point. Thus, in semi-structured interviews the interviewees' views and opinions are not restricted by the interviewer or the structure of the interview but the interviewees have a chance to express and clarify themselves more freely (Hirsijärvi and Hurme 2008: 48).

The outlines of the interviews for teachers and pupils were tested in two pilot studies before the actual interviews. Eskola and Suoranta (2008: 88) highly recommend pilot studies also because it is important to test the interview practices and equipment before the actual interviews. The first pilot interview was done with a teacher who previously taught in an upper secondary school and used the ELP but who currently has a different type of teaching job, still involving the ELP. The other pilot interview was done with a lower-secondary school pupil. Originally the pilot interview was supposed to be a pair interview but unfortunately the other interviewee got ill and was not able to participate on the prearranged day. The outlines for both interviews included the same themes but the questions are somewhat different. The questions were formulated in consideration of the respondents' expertise and point of view. Based on the pilot interviews some minor modifications were made to the wording of questions before the actual data collection begun. The outlines of the teacher and pupil interviews can be found in Appendix 1 and 2.

4.2.1 The data collection process

The data collection process began in January 2014 when several teachers were contacted by email and the pilot interviews were conducted. Finding English and Swedish language teachers who are using the ELP was challenging and some of the teachers were very busy but finally five primary and lower secondary school

teachers of English and/or Swedish agreed to an interview. Some of the teachers had also pupils who were willing to participate in the study, and thus, for practical reasons the teachers recruited altogether ten the pupils for the interviews.

The interviews were held during February and March 2014. Before the interviews the principals of the schools were contacted and asked for a permission to conduct a study at their school. Also the parents of all the interviewed children filled in a permission form confirming that their child can participate in a recorded interview. At the end of each interview the teachers too signed a consent form giving a permission to use the collected data in the present study. One of the teachers was interviewed at her home, according to her own wish, but all the other teachers and pupils were interviewed at their schools.

The teachers were interviewed individually whereas pupils were interviewed in small groups of two to three pupils. Hirsijärvi and Hurme (2008: 63) acknowledge the advantage of group interviews when research participants are young children. They state that children are often shy, and hence, group interviews often result in more confidence to talk. Moreover, Eskola and Suoranta (2008: 94) note that in group interviews the participants can support each other, recall things that happened in the past and evoke memories together. Although the pupils in the present study were not young children but rather teenagers aged between 11 and 14, it was to be expected that they had not been interviewed before and that they might feel nervous talking to a stranger. Moreover, as it was discovered during the pupil pilot study, the topic of the interviews was somewhat challenging and the support of a peer was considered valuable. Thus, group interviews were chosen to make the young interviewees feel more comfortable and also to rouse more discussion during the interview. On the other hand, the teachers in the study were interviewed individually since they were expected to have more confidence in talking about language

assessment as it is a part of their everyday work. Teachers also often have previous experience of interviews.

All interviews were recorded and the group interviews were also videotaped. Videotaping is often recommended in group interviews because it helps later in transcribing and analysing process when the researcher needs to know who is talking at which point (Hirsijärvi and Hurme 2008: 63). The pupils were interviewed in pairs or in groups of three but one pupil was interviewed alone since the other pupil, who was also supposed to participate, did not show up for an unknown reason.

4.2.2 Participants

All teacher participants teach English currently or have taught it before and four of the teachers have taught or still teach also Swedish. Only one of the teachers teaches English and German and has never taught Swedish. All teachers have wide knowledge of the ELP and they have used it for many years. Nonetheless the teachers' current practices in using the ELP varied. Some of the teachers had integrated the ELP in their everyday teaching whereas others were not presently using the ELP as much as they had in the past or as they were planning to in the future. For example a change of workplace had affected some teachers' (3/5) use of the ELP, at least temporarily.

When reporting the results of the interviews in the following sections the teachers are referred to by pseudonyms. The teachers behind these pseudonyms are introduced in the following Table 1. All information in parenthesis refers to the teachers' previous experience or current experience which only has a minor role in the teachers' current work.

Table 1: The interviewed teachers

Pseudonym	Gender	Teaching experience	Languages taught	Type of school	Location of the school
Tarja	female	20 years	English, Swedish	Lower secondary school, (primary school)	Western Finland
Eila	female	24 years	English, (Swedish)	Primary school	Western Finland
Ulla	female	38 years	English, German	Primary school	Southern Finland
Osmo	male	about 20 years	Swedish, (English)	Primary school, lower secondary school, upper secondary school	Central Finland
Anna-Maija	female	30 years	English, Swedish	Lower secondary school, (primary school)	Western Finland

In addition to the teachers, ten pupils participated in the present study. The pupils were interviewed in five groups, although one group finally formed of only one pupil. Like the teachers, also the pupils will be referred to by pseudonyms in the following sections. The pupils of the five groups are introduced in Table 2. The reported amount of ELP experience is based on the pupils' own responses and conceptions

Table 2: The interviewed pupils

Group	Pseudonym	Gender	Age	ELP experience	School
1	Kaarlo Conny Matias	male male male	14 14 14	some experience, especially of self-assessment	lower secondary school, Western Finland
2	Henri Stella	male female	12 11	3 years, very familiar with the ELP	primary school, Southern Finland
3	Sofia	female	12	4 years, very familiar with the ELP	primary school, Southern Finland
4	Maiju Oona	female female	13 13	some experience, have used parts of the ELP occasionally over the past 3–4 years	lower secondary school, Central Finland
5	Elsa Hilla	female female	14 13	some experience, have used parts of the ELP occasionally over the past 3–4 years	lower secondary school, Central Finland

4.3 Method of analysis

The main aim of qualitative data analysis is to create clarity into the data and to reduce the data in a way that the fragmented pieces of information become meaningful (Eskola and Suoranta 2008: 137). The most common way to start analysing interview data is to transcribe it, and that was the first phase of the analysis in the present study too (Hirsijärvi and Hurme 2008: 138). After the interviews were transcribed, the data was read and reread for several times in order to get to know to data and see whether some new themes would arise.

The framework used in the data analysis was content analysis. According to Tuomi and Sarajärvi (2009: 14–15), content analysis is an appropriate method for analysing data for example in phenomenographic studies. Content analysis can be considered a single method but also a theoretical framework for qualitative, or even quantitative, analysis. When talking about content analysis as a qualitative method of analysis its purpose is to organize data and create a clear, concise description of the studied phenomenon. (Tuomi and Sarajärvi 2009: 91, 103, 106.)

In the present study the content analysis was guided by the theoretical framework of the study. In theory guided content analysis some main themes or concepts are taken from the theoretical framework to further analyse the ideas that arise from the data. (Tuomi and Sarajärvi 2009: 117). As the themes of the interviews were derived from the theoretical framework of the present study, it was obvious that the framework would also affect the analysis. Moreover, it felt natural to start to analyse the data according to the pre-set themes and see if any other relevant themes arouse from the data.

Thus, the data was analysed according to the themes of the interviews since after carefully reviewing of the data, the original themes seemed still valid. There were many interesting topics covered in the data but in order to find a clear focus the pre-set themes were the ones that gave answers to the research questions and were covered comprehensively in the data. Based on the data, some modifications were made to the original themes as some subthemes were made main themes and vice versa. Thus, the themes that clearly arouse from the data were given the most emphasis in the categorisation of the data. When the themes were set, the aim was to find connections between the interviewees' responses but also to examine the responses of the interviewees as a whole. This was done in order to understand the wider context and to ensure that nothing relevant to the themes was omitted.

The analysis of the results is covered in the following section. In that section the findings of the present study are reported, analysed and discussed. First, the themes are covered from the teachers' perspective. The main themes are the ELP assessment and its added value, assessment procedures, the pupils' role in assessment, the use of the ELP in English versus Swedish language learning and assessment and the advantages and disadvantages of using the ELP. Second, the pupils' responses are reported and analysed according to similar themes. Finally, the teachers' and the pupils' responses are compared with each other.

5 THE ELP IN CLASSROOM ASSESSMENT

The aim of the study was to examine teachers' and pupils' conceptions of language assessment and particularly the use of the ELP in assessment. The interview data is reported and discussed first from the teacher perspective, then from the pupil perspective and finally the two perspectives are contrasted.

Thorough the following sections extracts from the interviews are used to support the analysis. The extracts are numbered and separated from the surrounding text with spacing. The interviews were conducted in Finnish and the extracts in italics are the exact words of the interviewees. After each extract the same text is translated into English. Sometimes the Finnish extracts, however, involve expressions or sayings that are peculiar to the Finnish language, and thus, the translations may not be exactly identical on the word level but the core idea is always the same.

5.1 Teacher perspective

In this section the teacher interviews are analysed. The teachers' responses are categorised under five main themes that cover the teachers' perspective of Language Portfolio assessment. These themes are the ELP and its added value, assessment procedures, the pupils' role in assessment, the use of the ELP in

English versus Swedish language learning and assessment and the advantages and disadvantages of the ELP.

5.1.1 The ELP assessment and its added value

In the interviews all teachers were asked questions about the concept of the European Language Portfolio assessment as well as about the value the ELP adds to assessment, that is, why to use the ELP in assessment. When defining the ELP assessment, all five teachers rather unanimously emphasised learner-centredness. The teachers reported that the ELP assessment is based on pupils' own work and that it includes self-assessment and reflection. This was effectively summarised by Anna-Maija in her comment:

(1) *Eli se on oppilaasta lähtöisin, oppilaan omia tuotoksia.*
So it is from the pupil, his/her own work. (Anna-Maija)

Osmo stressed learner-centredness also through the choices the pupils make when processing their portfolio work. According to him, the ELP assessment involves assessing the processes the pupils conduct. During the processes pupils have to make decisions on how to proceed, when to ask for help, from whom to ask for help, how to keep to the set schedule and so forth. Moreover, pupils can choose how they want to show their skills within the framework the teacher has set. The chances for choices also mean that the portfolio work is authentic and meaningful for the pupils.

Similar topics arose also during the discussions about the value the ELP adds to language assessment. The teachers were again quite unanimous about the benefits the ELP brings. All teachers mentioned that the ELP is a great tool to **notice progress** in learning. The learning objectives become more challenging each year and sometimes it is very hard for the pupils to notice that their skills are developing if they, for example, received the same grade every year. Osmo exemplified this by saying that the advances that pupils make have to be

verbalised so that a pupil realises what s/he has learned. Also, as Ulla pointed out, pupils can see their learning curve by viewing at their portfolio. On the other hand, the ELPs provide also teachers opportunities to see what their pupils' have learned.

Most teachers (3/5) mentioned that the ELP work and assessment **extends pupils' understanding of language ability** and learning. Eila explained that when pupils are conducting portfolio work and assessing their own and their peers' work they learn to understand that learning languages means learning many different skills. Pupils begin to see language learning as a much wider concept, something more than the mean of their exam results. Anna-Maija shared this view by noting that sometimes pupils think that it is enough if they only memorise some grammar rules and vocabulary. She sees that the ELP is a good tool to change that kind of thinking.

All teachers agreed that one of the greatest benefits of the ELP is that it **embraces versatility** and provides pupils opportunities to show their strengths and skills that would not perhaps otherwise surface during lessons. As Tarja noted, the portfolio work brings forward different, and possible even new, sides of pupils. This is how Eila described the value of the ELP work:

(2) *… et niille [oppilaat] annetaan suurempi mahdollisuus tavallaan niinku pärjätä ku siinä vaan siinä suorittamisputkessa, siinä numeroputkessa…*

… so they [pupils] are given a greater chance, like, to do well than when just taking exams and getting grades… (Eila)

Moreover, the ELP is a great tool for **differentiating** teaching and assessment, but also for **motivating and activating pupils** who are usually bored or passive during normal classes. Tarja had observed how portfolio work had inspired some otherwise passive pupils to really work and learn. Eila thought that the ELP offers possibilities also for skilled pupils to whom the tasks in the textbooks and workbooks are too easy and demotivating. Ulla emphasised the usefulness of the

ELP when teaching and assessing pupils who start learning English as their third language at the age of 11. She told that the groups are very heterogeneous and some pupils know English already quite well when they begin the learning process. Thus, the ELP work can be designed adequately challenging for all learners. Osmo added that when pupils can show their strengths it also boosts their self-esteem, which is one of the main aims of basic education.

The teachers felt that the ELP supports their assessment work by promoting language learning as a process that involves more than only language ability. Anna-Maija, for example, said that the ELP is **an important asset for all assessment**. Likewise, Osmo found that portfolio assessment is a more comprehensive method than traditional summative tests as it shows pupils' progress in many ways. Tarja clarified that when using the ELP, it is easy to include skills such as learning skills, cultural skills and cooperation skills, that is, skills listed in the NCC, in teaching and assessment. These other skills need to be built into the learning methods in order to be able to teach and assess them. Likewise, Eila believed that the ELP improves the quality of assessment and makes it more justified. She believes that there are different types of talent, and that pupils may surprise their teachers with a piece of portfolio work which really shows their talent. Thus, when viewing pupils' portfolio work teachers have much versatile data on which to base their assessment. Teachers learn a lot about their pupils when reading or listening their pupils' work.

These ideas are very much in line with what for example Kolu and Tapaninaho (2000: 103) noticed during their ELP experiment. They stated that the ELP work provided them with more information about their pupils. They learned to know their pupils strengths and weaknesses better and assessment became easier because it was integrated in learning. Personally I find this to be one of the greatest benefits of the ELP. When pupils can affect the content, mode, format and/or the style of their portfolio work, they can use their strengths and show theirs skills in a unique way. Furthermore, the work is more personal.

All teachers seemed to be very convinced about many benefits of using the ELP in assessment. Nevertheless, it must be noted here that pupils, as well as teachers, are all different and the ELP work may not motivate everyone or expand the knowledge of every pupil. Osmo, for example, mentioned that not all pupils are pleased or ready to take responsibility of their own learning and assessment. This is how he commented on the topic:

(3) *… joittenkin [oppilaiden] on kauheen vaikee irrottautua siitä, että, että opettaja ei sanokaan, että mitä pitää tehdä, ku sit taas osa lähtee heti innoissaan…*

… it is very hard for some [pupils] to let go of the idea that, that the teacher tells them what to do, whereas some start the work very eagerly… (Osmo)

The same thing has been brought forward also in the earlier ELP trials. Kolu and Tapaninaho (2000: 36) as well as Saarinen (2010: 17–18) pointed out that adopting a new way of working is not easy for everybody. Likewise, Hilden (2004: 95) noted that some pupils' clearly prefer more teacher-directed learning. The ELP is, however, a multifaceted tool that can be used in many ways and forms to meet the requirements of individual learners, and the fact that pupils, some more that others, are so attached to the traditional teacher-directed learning methods should not be an impediment to the use of the ELP. On the contrary, it should challenge teachers and learners to change the way of learning.

5.1.2 Assessment procedures

All five teachers shared the idea of dividing language ability into writing, speaking, listening and reading skills in assessment. Two of the teachers reported that they had made efforts to emphasize all skills equally in assessment and for example in exams they test all four skills. The skills were not, however, equally present in pupils' portfolio work as the teachers had different practices. Ulla and Tarja told that their pupils' ELPs consist mostly of written work but the other skills are then assessed in other ways. Osmo, on the other hand said that if there are for example three pieces of work in a pupil's ELP, at least one of them has to be an oral production. Eila and Tarja emphasised that the ready-made exams

provided by the different publishers of teaching and learning materials do not cover all skills equally but teachers need to pay attention to versatile assessment. Furthermore, Anna-Maija pointed out that different pupils have strengths in different skills. Thus, as Eila's comment shows, assessment has to include all aspect of language learning:

(4) *… pitäis olla kaikista näistä [puhumisen, kirjoittamisen, lukemisen ja kuuntelun taidot] opettajalla sit se mistä arvioidaan ja sit vielä ne opiskelutaidot ja sitte vielä se kulttuuri…*

…a teacher should have a record of all of these [speaking, writing, listening and reading skills] and then also of learning skills and then also culture… (Eila)

In addition to the ELP, the teachers use other assessment methods. They all (5/5) reported that they use summative tests, which supports the claim made earlier in the present study about the dominating role of exams in Finnish schools (see e.g. Vähäsarja 2014). Nevertheless, most teachers (3/5) said that they had either reduced the amount of exams per year or replaced some of the written exams by for example oral exams. Indeed, oral exams or observations of oral tasks were used by all five teachers as well. Only Osmo reported organising actual oral exams where pupils answer the teacher's questions, interview each other and have guided dialogues in pairs, whereas the rest of the teachers told that they assess oral skills only by observing and listening to their pupils during lessons. Tarja, for example, mentioned that sometimes she tells her pupils that she is now going to walk around in the classroom and observe the pupils' oral skills. According to her, this tends to motivate pupils to do the given tasks properly since often pupils have a quite relaxed attitude towards speaking activities. They tend to consider speaking activities as 'free time'.

The teachers' practices of assessing oral skills actually embrace the new assessment culture as the assessments are embedded in the instructions and in the classroom activities. The teachers observe their pupils while they are engaged in meaningful conversations.

Thus, none of the teachers used only portfolio work in assessment, and speaking skills were often excluded from the ELP work. Each teacher used the ELP a bit differently. Ulla told that her pupils' portfolios include both written work and exams, and thus, the pupils have all their assessment material, except for oral tasks, in their ELPs. The other teachers' seemed to conduct ELP work besides exams and other assessment methods, at least at the moment of the interview. Their pupils did not all have physical portfolios but they executed different assignments according to the principles of the ELP. Indeed, assessment methods and tasks can be used in various ways. Tarja illustrates this idea well in her comment:

(5) *…ei sen välttämättä tarvi olla kielisalkkua jos sä teetät kirjotelman, mutta parhaimmillaan siihen liittyy ne, niinku ne kielisalkun vaiheet, että on sitä valinnanvapautta ja seurataan sitä prosessia ja sit arvioidaan porukalla…*

… a piece of writing does not necessarily have to be a part of the ELP but in an ideal situation it involves the, like, the phases of the ELP, so that there is freedom of choice and the process is monitored and assessed together… (Tarja)

Therefore, no task or method is a part of the ELP automatically, but again, it is the implementation and use of the task which makes it either 'normal' assessment or ELP assessment. The teachers clearly aim to assess their pupils diversely and assess also other skills than only language ability. When assessing their pupils' ELP work, many of the teachers (3/5) reported that they view the whole process and assess not only the pupils' language skills but also the way how the whole process was conducted. The other two teachers mentioned also that the pupils' study and learning skills affect the assessment.

The teachers were also asked about the nature of their assessment processes, that is, what their assessment is like. One clear pattern arose from the teachers' answers. The teachers assess and give written feedback periodically, for example when a course finishes or when pupils have returned a written assignment, whereas all teachers assess their pupils' performances orally and continuously during lessons. The teachers have discussions, give feedback and guide their

pupils continuously. Eila particularly emphasised that she aims to give encouraging feedback to her pupils. Thus, it appears that the teachers' feedback practices aim at enhancing their pupils' learning. The reason for giving written feedback only periodically is certainly practical as it takes more time than oral feedback which can be for example an encouraging comment or a piece of advice given in class.

5.1.3 The pupils' role in assessment

The pupils' role in assessment is emphasised in the current NCC as well as in the ELP. One of the aims of the present thesis was to study the ways how pupils are involved in their assessment. All five teachers found the pupils' role important and they emphasised the importance of interaction and discussion. By discussing assessment, language learning and the learning objectives the teachers engage their pupils in the thinking and assessment processes. This is how Anna-Maija described the pupils' role:

(6) *Must on hyvin tärkeetä, et oppilas on koko aika tietonen siitä, et kuinka häntä arvioidaan, miksi näin, ja mitä enemmän voidaan yhdessä keskustella sen parempi*.

I think it is very important that a pupil is all the time aware of how s/he is being assessed, why so, and the more we can discuss it together the better. (Anna-Maija)

The teachers were unanimous about the importance of pupil-teacher interaction, both whole class and individual discussions. Two of the teachers reported that they also have personal conversations with their pupils at the end of each semester. The aim of all the discussions is not only to involve pupils in the assessment processes but also to help them to see language learning as a broad concept. When communicating the assessment criteria and learning outcomes, the pupils also receive a more realistic idea of their skills and knowledge. For teachers the advantage of open interaction is that they learn more about their pupils and about how the pupils see themselves as language learners. This information, as Anna-Maija pointed out, affects teaching and everything else that

is done during classes. She further noted that it would be great if pupils participated more in assessment, but then the pupils' assessment skills would have to be more advanced.

All teachers agreed, again, that self-assessment is an important skill that needs to be taught. Self-assessment does not require the use of an ELP but the teachers considered the ELP to be a good tool for learning self-assessment. Like Osmo and Tarja noted, when pupils are guided to assess their own work, they learn to pay attention to factors that can affect the quality of their work. In primary school the focus is more on letting the pupils to get used to assessing their own learning and the self-assessment is more about their studying skills than language skills. With older pupils, who are a bit more mature and have more developed cognitive skills, the self-reflection can be more analytical and regard also the pupils' language skills. Osmo, however, reminded that if the pupils are not taught self-assessment skills or if they are not used to doing self-assessment, it is hard also for older students. This is what has been noted in earlier studies as well (see e.g. Douglas 2010: 75, Alanen and Kajander 2011: 69). In addition, Osmo emphasised that the assessment needs to be well-planned and specific rather than general, because otherwise pupils will easily get frustrated and will not take the assessment seriously. Nevertheless, when the pupils' self-assessment skills improve, the pupils are able to analyse their learning rather easily. Anna-Maija exemplified this by stating that about 90 percent of her pupils are able to assess their own skills rather accurately.

Ulla and Eila, who both teach in primary school, had very established assessment practices that involved also pupils. In addition to all discussions, they ask their pupils to assess their language and study skills regularly. After the pupils have done the assessment, Ulla and Eila add their own comments and then the assessment forms are also shown to the pupils' parents. Furthermore, Ulla and Eila reported using also peer assessment. Their pupils assess each other's work and learn from that. Eila emphasised the positive effect peer assessment has on

the atmosphere in the class. It is encouraging for the pupils to hear positive feedback from their peers. Also Tarja had experience of using peer assessment and she had noticed that it is more exciting for the pupils when also their peers assess their work. These observations are in line with the earlier reported experiences of Kolu and Tapaninaho (2003: 44–45). They also noticed how peer assessment affected the social atmosphere in the class.

In conclusion, the teachers were very unanimous about the importance of teacher-pupil discussions and self- assessment. They all had noticed that often pupils are surprisingly good at assessing themselves if they just are given the appropriate guidance and the opportunity. The teachers' observations are very similar to the ones made in earlier ELP experiments and studies. For example Hildén (2004: 94–95) noted that most pupils learned to assess themselves quite realistically during the ELP trials. Thus, it seems that if teachers only let pupils to get involved in the assessment processes, the pupils really have the ability to become good at assessing themselves. Moreover, when the pupils are involved in the processes, assessment is made more visible.

5.1.4 Using the ELP in English and Swedish language assessment

None of the teachers thought that the ELP would suit better to English or Swedish language learning and assessment. They found no significant differences between the use of the ELP in either of the languages. Anna-Maija said that she uses the ELP less in Swedish classes simply because there are less Swedish lessons than English lessons in the timetable, and hence, there is less time to devote to the portfolio work. Here is Anna-Maija's comment on the differences:

(7) *... salkku antaa lisäarvoo noin yleensä, et on se sit mikä kieli tahansa, ni kyllä se antaa sitä, jos sä jaksat sen työn ja vaivan nähdä, ni kyllä sä saat, se tukee sitä arviointii niin hienolla tavalla…*

…the ELP adds value in general, no matter what the language is, it will add value if you just take the trouble to use it, you will get, it supports the assessment greatly… (Anna-Maija)

Similarly, Osmo and Eila pointed out that the use of the ELP is not dependent on any specific language since the ELP is the way how the languages are taught, learned and assessed. Osmo added that it is more dependent on individual pupils and their ways of learning. For some pupils the ELP suits better than others.

Some of the teachers (3/5), however, admitted that the level of language ability may have an effect on the use of the ELP. Osmo and Tarja mentioned that since pupils are often more advanced in English than they are in Swedish, it can be more effortless to start processing the ELP work in English. Many pupils are able to express themselves fluently in English but in Swedish it is more limited. Ulla also added that she finds the ELP especially useful with pupils who start learning English as a third language at the age of eleven. The learning starts from the basics but some of the pupils can already speak English and know quite much, so the ELP supports the teaching, learning and assessment of all the pupils in these heterogeneous groups.

Furthermore, the ELP can affect the way how the pupils see themselves as English or Swedish language learners. Osmo mentioned that often pupils compare their skills in English with their skills in Swedish and especially at the end of comprehensive school, when most pupils have studied Swedish for three years, they feel that they cannot speak Swedish and that their language skills are poor. Then when they look at the checklists and learning objectives together with the teacher, the pupils begin to see all the things they actually can do with the language. On the other hand, Anna-Maija noted that the structure of the Swedish language is simpler than the structure of the English language which requires another type of approach to teaching, learning and assessment. She has observed that pupils are often very confident about their skills in English but in lower secondary school when the learning material becomes more and more challenging some pupils face problems. Therefore, it might be reasonable to say

that the use of the ELP can help to form a realistic view of one's skills, and especially in Swedish language learning that can have a positive effect even on pupils' self-images.

Eila and Osmo reported also that as the ELP encourages individual pupils to show their skills and cover topics related to their interests, the use of the ELP increases variety in English and Swedish language learning. It enables a wider recognition of specific linguistic areas or specific linguistic traits as pupils can, for example, search for information that they are interested in and seek an aspect that they find interesting. Osmo admitted that this happens usually only in English because many pupils use English in their spare time and they might have more real-life needs for the language. Generally pupils do not have this kind of contact with the Swedish language outside of school. Nevertheless, Osmo noted that in English pupils may bring more current vocabulary and topics forward, and thus, teach the teacher something new.

In summary, the ELP works well no matter the language. The ELP is more of an approach to language teaching, learning and assessment than any language specific teaching material. The use of the ELP enables allowing for the different language specific characteristics of Swedish and English or any other language. The value that the ELP may add to Swedish language learning is that it can encourage pupils and increase their interest in the language. Often the problem is that pupils do not have any contact with the language outside of school and that they do not find the language useful or interesting. The ELP might be one solution to this problem. On the other hand, pupils are often interested in learning English and one can see, read and listen to English almost everywhere, at any time. The English speaking world is wide and diverse and the ELP can be a great tool for focusing on the topics and issues that the pupils find useful and interesting. Therefore, the difference may lie in the benefits that the ELP can bring to either language.

Thus, it appears that the ELP can be used in the teaching, learning and assessment of any language and that, of course, justifies the promotions of the ELP as an approach rather than as a single, language specific method. The use of the ELP has been experimented and studied in the teaching and learning of different languages but there are no studies where the benefits of the use of the ELP in different languages would have been compared with each other. Kolu and Tapaninaho (2000, 2003), for example, used the ELP in English and Swedish language learning but they did not actually compare the use between the two languages. Nevertheless, the fact that the pupils only began learning Swedish during the trial period meant that the pupils did not use the checklists of the common reference levels in their self-assessment until during the third year of their Swedish studies (Kolu and Tapaninaho 2003: 36). Kolu and Tapaninaho (2000: 98 and 2003: 36) state that since it often takes long before the pupils reach the first level, A1, they did not consider the reference levels useful in the beginning. Hence, the level of language ability affected the way how the ELP was applied.

An interesting observation in Kolu's and Tapaninaho's (2003: 36–37) experiment was that when the pupils began to use the checklists and reference levels, most pupils overestimated their skills. In the present study Osmo's observations were quite the opposite as he stated that often pupils feel that they have very poor skills in Swedish. Still, on the other hand, as Anna-Maija pointed out, pupils tend to overestimate their skills in English. No generalisations can be made based on these individual observations but it seems that there really is a need for teaching and practising self-assessment skills in order for the pupils to form a realistic view of their skills in different languages.

5.1.5 Thoughts of the advantages and disadvantages of the ELP in assessment

During the interviews the teachers were asked to describe their feelings and experiences of using the ELP. None of the teachers used any negative adjectives to describe their experiences but they used adjectives such as interesting,

wonderful, natural and eye-opening. The teachers had clearly found a way of teaching that suits their teaching philosophies and motivates both them and their pupils. Here is how Eila and Tarja expressed their thoughts:

(8) *…et monesta muusta asiasta luopus, siitä [Kielisalkku] en.*

… so I could give up many things but not that [the ELP] (Eila)

(9) *… mä koen niinkun tän omaks tavakseni opettaa ja huomaan oppilaissaki sen kuinka ne hirveesti virkistyy kun ne saa tehdä jonkun työn jossa saa edes johonkin asiaan itse vaikuttaa…*

…I feel that this is my way of teaching and I have also noticed that my pupils cheer up when they get to do a piece of work where they have an opportunity to affect at least something… (Tarja)

It is obvious that the ELP represents values and principles that are important to the teachers and that is why they find it inspiring and useful. Ulla said that the ELP includes issues that she would also otherwise want to do, but now she has an officially accepted tool for it. Likewise, Osmo outlined that he thinks that it is important that pupils have opportunities to consider what interests them, and he wants to offer his pupils insights into the many purposes and uses of languages.

Most teachers (3/5) mentioned that reading pupils' portfolio work is interesting and pleasant. For example, Ulla said that it is very interesting to read pupils' own productions because they are so different. Moreover, Eila stated that it can be eye-opening if one only has time to pay attention to individual pupils. She believes that a teacher learns a lot about his or her pupils when reading their productions. In a similar manner, Tarja stated that the ELP assists in getting closer to one's pupils, and on the other hand, lets pupils to bring their personalities forward.

The teachers did not find any significant disadvantages in using the ELP but a few problems were brought forward during the interviews. Two of the teachers felt that sometimes the workload is burdensome. It requires time and effort to

read and assess pupils' ELP work and Ulla also pointed out that sometimes she worries that since some of her pupils are conducting the ELP in two languages, it might be too burdensome for the pupils as well. Anna-Maija, on the other hand, mentioned that she has had to study a lot in order to internalise the principles of the ELP and the CEFR.

The same challenges have been brought up also earlier by for example Kolu and Tapaninaho (2000, 2003), Hildén (2004: 97) and Saarinen (2010: 17). Kolu and Tapaninaho (2000: 89, 102) reported that a lot of work was required to first study the CEFR scale and then learn the principles of the ELP. They found that adopting a new way of thinking demanded quite a lot not only from the pupils, but also from the teachers. Moreover, Hildén (2004: 97) and Saarinen (2010: 17) mentioned that the continuous assessment and correcting of portfolio work are often considered burdensome. It must be noted here, however, that most of these observations come from ELP starting projects or experiments when all the practices have still been very new to the teachers. I would assume that if a teachers gradually adopts the principles of the ELP work and plans the work carefully, it is possible to decrease the amount of work. This is also what Osmo pointed out during his interview. He emphasised the importance of setting clear deadlines and spreading the workload over a longer period of time.

Another challenge that was brought forward in the present study was, rather surprisingly, the expectations and pressure set by the pupils' parents. Two of the teachers found that some parents are very 'traditional' in the sense that they want their children to have summative exams and word quizzes at school. Sometimes the parents only want the best possible grades for their children and the idea of the ELP seems foreign to them. The two teachers reported that they have to have test results of their pupils' skills in order to show the pupils' parents how the children have been assessed. Furthermore, the prejudices against the ELP do not come only from the parents' side. Eila mentioned that also some language teachers underestimate the idea of the ELP. Many teachers consider the ELP to

be something fun an extra for which they do not have time. Thus, Eila noted that the disadvantage of using the ELP can be that in some schools one is alone with the ELP work.

In conclusion, the teachers found the use of the ELP mostly beneficial, and although the ELP work might feel burdensome sometimes, they still feel that it is worth the work. It was a very interesting but unfortunate observation that some parents and teachers are so strongly attached to the traditional ways of teaching and assessing that they do not see the value of the ELP. This indicates that a lot of work still needs to be done before the ELP is recognised as a valid approach by all teachers, pupils and parents in the country. Nevertheless, the teachers in the present study reported having found an approach, a concrete tool that contributes to assessment and helps pupils to become masters of their own learning. They advise other teachers to gradually integrate the ELP in the teaching and assessment processes. As Eila's comment shows, the benefits are worth the work:

(10) *Ihan vaan rohkeesti, pieni pala ja antaa mennä ja kokeilla ja sit kun huomaa, että joku ilahduttaa ja hämmästyttää, ni sitten tietää, että on perillä.*

Just bravely, a small piece at a time, and let go and try, and when you notice that something delights and amazes, then you know you are there. (Eila)

5.2 Pupil perspective

What follows are the pupils' thoughts and views of language assessment and the use of the ELP. The pupils' answers are categorised into four themes which cover the topics discussed earlier from the teacher perspective. The pupils were not asked about the differences between using the ELP in English and Swedish language assessment since none of them had experience in using the ELP in both languages. Otherwise the themes are the same as with the teachers.

5.2.1 ELP assessment

When the pupils were asked to explain in their own words what they think ELP assessment is, most pupils answered that it means assessment of pupils' own portfolio work. The pupils were slightly hesitant with their answers except for one, Elsa, who defined ELP assessment as a compilation of a pupil's own work that is assessed on the basis of the language and the process, for example whether the pupil did the work independently or received help. One of the pupils did not know what the ELP is but after her friend gave her some examples of the ELP work that they had done, she remembered and understood the topic of the interview.

To receive some clearer ideas about the ELP assessment from the pupil perspective they pupils were also asked to approach the topic by comparing the ELP and traditional exams with each other. This evidently clarified the idea and the pupils had very insightful thoughts about the differences. The pupils found that there is variation in the situation and in the content. In groups 2, 3, 4 and 5 the pupils thought that exam situations are more stressful as one has to know certain topics and the time is limited. In groups 4 and 5 the girls noted that when conducting portfolio work one can always ask for help or consult a dictionary. Sofia, on the other hand, mentioned that they often have the possibility to write portfolio tasks alone at home and then the work might be only presented and assessed with others in class. She noted that exams are always done in class when the whole class is present.

The pupils' perceptions of the differences connect language assessment to the situational factors quite strongly; exams are always done in class, portfolio work can be done at home. Moreover, the assessment conditions seem to be relevant as the pupils noted that exams stressful. One has to know certain things and be able to do all the tasks within the given time. It is, however, most likely that pupils are given deadlines also when conducting ELP work and they are responsible for producing a piece of work that shows a certain area or areas of

their language and study skills. Although there might be more freedom of choice when considering for example vocabulary, the pupils still need to be able to use to words correctly and create a purposeful piece of work. Thus, one could argue that the ELP work requires often more from the pupils than traditional exams. Still, the pupils did not find the ELP stressful. It could be that what the pupils actually find stressful is the uncertainty of whether they have studied the "right" things or not. The pupils feel that they need to know the topics, words and phrases that their teacher has chosen and found important. On the other hand, it may be that the pupils do not consider the ELP to be a "real" assessment method, and hence, they regarded the ELP work more relaxed.

In groups 1, 2 and 3 the pupils paid attention to the content and nature of the assessment tools. In group 1 Conny and Kaarlo explained that when assessing ELP work one assesses language ability in general whereas in exams the assessment is more limited to a certain area or skill. Here is how Conny described the differences:

(11) *Kokeissahan kysytään tiettyjä asioita siltä tietyltä alueelta. Sit kielisalkussa taas, siinä vähän laajemmin niitä kaikkia.*

In exams certain things are asked from that certain area. Then in the ELP things are assessed more extensively. (Conny)

Likewise, Sofia (group 3), pointed out that in the ELP work one can use any words but in exams one has to know the exact words that are required. Sofia as well as Henri and Stella (group 3) and Kaarlo, Matias and Conny (group 1) emphasised that the ELP work is more free in the sense that one can implement the work the way s/he wants and show his or her strengths and personality. In exams the tasks are set, and moreover, they consist of several tasks, perhaps of different types.

The pupils were also asked whether they think the ELP work provides a realistic appraisal of their language ability and skills. This divided the pupils into two

groups representing opposite point of views. On one hand, the pupils in groups 1, 2 and 3 said that the ELP is a better assessment method than exams because of the freedom of choice and the possibility to show one's strengths. The boys in group 1 were somewhat hesitant but then decided that if one is just honest when doing self-assessment, the ELP should work fine.

On the other hand, the pupils in groups 4 and 5 found that the ELP assessment does not provide a realistic appraisal of a pupil's language ability because the pupil can ask for help when doing the work. Hence, they believed that the ELP is not as good an assessment method as exams, which show pupils' skills more realistically. As Hilla's comment reveals, the pupils believed that if one can receive help, his or her actual ability to use the language cannot be assessed truthfully:

(12) *…melkein sama ku lunttais jossain kokeessa jos niinku tekee jotain ja sitte saa kattoo jostain apuu eikä oikeesti osaa, ni sitte se saa siitä, ni ei se oo sillee…*

… it is almost the same as if you cheated in an exam, so if you like do something and then look for help and you don't actually know it, and then you'll get, it is not like… (Hilla)

Furthermore, Maiju and Oona thought that the ELP might work in English but in Swedish they would not be able to produce more than some pictures with a few words, and that would not be enough for assessment. Despite this, the pupils in groups 4 and 5 agreed that the ELP work could provide an approximate of a pupil's language ability and could be a good addition to exams. As Elsa concluded, it would be good to assess how well pupils do in situations where there is no time pressure.

The pupils' comments on how exams provide a realistic appraisal of a learner's language ability raise questions about the pupils' view of language learning, language use and language ability in general. It seems that the pupils do not have a full understanding of what language ability is. I base this claim on two things here. Firstly, the pupils mentioned that the use of a dictionary can falsify the

assessment. This indicates that the pupils believe that knowing the translation of a word in a foreign language is the same as knowing the word and being able to use it. Nevertheless, looking a word up in a dictionary is not enough to produce a fluent piece of writing or speech, and it is very likely that a teacher can notice when a pupil uses a word or phrase that does not yet belong to his or her active vocabulary. Moreover, it should only be a positive thing if a pupil is very eager to use a dictionary and learn new words.

Secondly, the pupils' comments imply that the language learning in school is not well connected to real-life language use. The pupils found that if one does not have a time limit or if one can ask for help, the assessment does not provide a realistic appraisal of one's language ability. Of course, it is a good aim for all language learning to be able to produce both oral and written text without constantly referring to a dictionary or consulting another person, but in real-life communication situations one can often ask for help and overcome communication problems with the help of body language, a friend or a dictionary, for example. Thus, the pupils seem to be very strict about the norms of what is counted as good language ability. They believe that one has to survive on his or her own if s/he is to say that s/he can speak a language.

Furthermore, all this raises more questions about why the pupils' have this kind of conceptions about language ability and language learning. Again, one could argue that the long tradition of using mainly exams in language assessment has affected the pupils' views. At this point one can of course only speculate the reasons but it could be that the main reason behind these believes is that the pupils' skills and knowledge have been tested in exams thorough the time they have spent in school. Exams, however, bear very little resemblance to real-life needs. Yet, the pupils have learned that exam tasks are the measures of language skills. For example, Maiju's and Oona's comment about how the ELP would not work in Swedish language assessment because of their limited skills shows how accustomed the pupils are to exams. If the pupils' skills are indeed limited, the

pupils cannot produce any better language in exams either. Still, they do not criticise the use of exams in the beginning of their language studies but only the use of the ELP.

The old traditions are still strongly present in people's minds today. I would argue that it is not only the school practices and teachers which can affect pupils' views but also the pupils' parents might unintentionally affect their children's conceptions when, for example, telling their children to study for exams and learn lists of words by heart. It was already noted earlier how some of the teachers in the present study had noticed how parents can have very persistent opinions about assessment. The school is not a closed environment but it affect everybody at least at some point of their lives and thus the changes have to happen in a wider scale. Although teachers might be enhancing the new assessment ideology, it takes time before people's attitudes and thoughts change.

In general, the pupils' comments were rather well in line with their experience of the ELP. Of course the amount of respondents is too small to make any statistically significant conclusions, but the pupils in groups 2 and 3 who had much more experience of the ELP than the other pupils were the ones who found the ELP useful and thought that ELP work represents one's skills better than exams. Still, although the pupils in group 1 had rather limited experiences of the ELP they also thought that the ELP could bring up the strengths of individual learners'. On the contrary, the idea held by the pupils in groups 4 and 5 is very much alike what Välijärvi and Kauppinen stated about the conservative attitudes and the strong tradition of exams in Finland (Vähäsarja 2014). Exams are easily seen as the only 'right' assessment methods because it has been used for so long and these pupils did not have much experience of the ELP. All the pupils in the present study found, however, some kind of need for an assessment method like the ELP.

5.2.2 Assessment procedures

The pupils understood the concept of language ability slightly differently but some of them had rather advanced ideas of what language ability is or what it includes. Most pupils (7/10) mentioned vocabulary and grammar or language structure but all pupils emphasised the importance of being able to communicate. Kaarlo and Matias (group 1) believed that language ability involves skills and knowledge that are needed to have a conversation and understand native or near-native speakers. Similarly, the pupils in groups 4 and 5 stated that it is important to be understood by other people but also manage in everyday situations, for example abroad, or later in the future, at work. Kaarlo also remarked that it is useful to be able to search for information in English. It was interesting that also the pupils who stated that an exam can provide a realistic view of one's language ability emphasised the importance of communication skills. It could be that these pupils think that knowing grammar and vocabulary, which are often tested in exams, are enough to be able to communicate in the language. Alternatively, they might be just familiar with the fact that knowing a language means more than knowing only grammar and vocabulary but they have not really understood how it shows in language learning and assessment.

The pupils were also asked to mention factors that affect their assessment and grades. It was not surprising that all pupils listed active participation as one of the factors that affects their assessment. This is consistent with the study conducted by Lukka et al (2008: 132) where the pupils were reported to believe that their teachers emphasise participation more than the teachers actually did in their answers. Here the same pattern repeated. Here is, for example, what Sofia answered when she was asked about the factors that affect one's grade.

(13) *Aktiivisempi aina tunneilla ja sitte totanoin harjottelee paljo enemmän juttuja, ei jätä vaan läksyjä tekemättä tai jotain.*

More active in class and then one has to practise a lot more things and not neglect homework, or anything. (Sofia)

All pupils clearly believed that their activity and participation in class affect their assessment.

The pupils seemed to be quite satisfied with the assessment methods and feedback practices that their teachers were using. Exams were used in every class but also word quizzes, small essays and listening and reading comprehension tasks were mentioned by a few pupils. Classroom observations were listed only in one group. In group 1 Conny, Kaarlo and Matias outlined that they would like to have their oral skills tested more often, in actual oral exams. They found that their assessment is currently often based on their written work.

In general the pupils reported that they receive feedback quite rarely. Nobody mentioned receiving feedback continuously or too much. In group five, Elsa and Hilla did not comment on the feedback practices but Maiju and Oona from group 4 said that their teacher always gives them encouraging feedback when they feel that they cannot do something or that they have poor skills in Swedish. Their teacher sometimes also writes comments on the margins of their exam papers. Also the pupils in groups 1 and 3 seemed to be satisfied with the current situation. The pupils in group 1 reported receiving oral feedback from time to time, but other types of feedback rarely. Matias, however, remembered that sometimes they have had assessment conversations with their teacher. Sofia mentioned receiving some written feedback from her teacher occasionally. On the other hand, Henri and Stella (group 2) stated that they most often receive feedback in the form of a grade. From their exam grades or final grades they can then conclude what their teacher thinks about their abilities. Nevertheless, of all the pupils, only Henri said that he would like to receive more feedback, at least sometimes.

In the study conducted by Luukka et al (2008: 139–140) it was discovered that marking corrections or writing comments to pupils' written work and reviewing the answers of an assignment or an exam together with the whole class were the

most common ways of receiving feedback from teachers. For some reason only three pupils in the present study mentioned receiving written feedback or comments from their teachers and none of the pupils referred to any whole class feedback discussions. Generally speaking, the pupils listed only a few feedback practices. I would argue that the pupils actually receive feedback more than they reported because it is very hard to believe that there are language classes where teachers never go through exam tasks with the whole class or never write any comments on their pupils' written work. Moreover, only one pupil would have actually liked to get more feedback, the others were satisfied. Thus, it is reasonable to assume that maybe the pupils just did not remember how they have received feedback. It could also be that they do not consider brief comments or whole class discussions feedback but normal communication.

To conclude, the pupils had received written and/or oral feedback from their teachers and they reported also that several assessment methods are in use. The pupils did not have very strong opinions about either the used methods or the feedback practices. For example, although the pupils in group 1 told that they receive feedback rather rarely, they did not urge to change anything. They were satisfied with the current situation. This in general might suggest that the pupils are used to having a rather passive role in the decision-making processes in school. The pupils' role will be analysed next.

5.2.3 The pupils' role in assessment

In most groups (4/5) the pupils stated that they can affect assessment by working hard and being active in class. Only Elsa (group 5) observed that pupils can influence language assessment by taking responsibility of their own learning and asking the teacher if something is unclear. She, however, noted that pupils may not be ready to take full responsibility, but maybe some. Kaarlo (group 1) had a similar idea but he emphasised the pupils' role in planning the content of learning. He outlined that pupils can affect the topics to be studied and a bit also

the way the topics are covered. Kaarlo evidently had the most advanced idea of pupil involvement.

On the other hand, Maiju's (group 4) comment on the pupils' role reveals the other point of view that was evident in all groups:

(14) *No jos se tekee töitä hyvin, saa hyvät kokeest, numerot, se on vähän niinku, ei se oikeen muuta voi tehä.*

Well, if s/he works hard, gets good grades, it's like, there is not much else s/he can do. (Maiju)

Likewise, Henri and Stella (group 2) pointed out that one has to work hard and also show interest in learning. It is not enough if one simply does well in exams but s/he has to try his or her best and try to improve all the time. Thus, all pupils clearly signalled that good grades demand active participation and hard work.

The pupils' were also asked whether they feel that they are allowed to participate in or affect assessment satisfactorily. In groups 1, 2 and 5 the pupils agreed that it depends on the pupil. If the pupil works hard then s/he can affect the assessment. Kaarlo (group 1) noted, for example, that a pupil can influence the assessment until the grade is given.

(15) *No kyl se aikalailla sillee niinku, et opettaja yleensä sanoo sen jonku niinku arvosanan ja sit sä et paljoo voi vaikuttaa siihen enää ku se arvosanan on annettu, mut ennen sitä sä voit tehdä sen mitä sä haluat.*

Well, it is pretty much like that the teacher usually tells the grade and then there is not much you can do about it after the grade has been given, but before that you can do as much as you want. (Kaarlo)

Hilla (group 5) admitted that she is not active in class and does not show her skills so her teacher cannot know how good she actually is in Swedish. Thus, she believed that class participation affects the way how her teacher sees her skills. Sofia (group 3) was satisfied with the current situation as she felt that there are

already so many things that she as a pupil can affect. She did not, however, give any examples.

In group 4, however, Maiju and Oona were slightly confused about the question. First they answered that yes, pupils can adequately affect the assessment and their opinions are considered, but then they began to doubt their answer. Maiju and Oona questioned the value of pupils' opinions. As Maiju explained, it would not be fair if pupils' own assessment would affect their grades because then some pupils would deliberately overestimate their skills in order to receive better grades. It would not be right if pupils' opinions influenced their grades.

Although Elsa and Kaarlo had good examples of how a pupil can affect assessment, it seems that the pupils in general do not know how they could be involved in the assessment processes. They all emphasised the pupil's role as a hardworking learner who does his or her homework and is active during lessons. It seems that the pupils do not see any other form of participation even possible. This probably means that none of them has been involved, for example, in deciding assessment methods, planning exam tasks or discussing assessment related questions, such as their own language skills. It is the teacher who decides the practices and pupils can either do their part and work hard or do only the things that are required to get a satisfactory grade.

All pupils had done self-assessment at least a few times. In groups 2 and 3 the pupils felt that self-assessment is quite challenging. They all said that they often have to think very carefully about the self-assessment questions and that they are not always sure what to write. Conny, Kaarlo and Matias (group 1) had used the ELP checklists in self-assessment and they found the lists useful, although Kaarlo added that some topics or skills are easier to assess than others. Elsa (group 5) commented that it is both easy and difficult as it is easy to overestimate or underestimate oneself. In comparison, Maiju and Oona said that it is easy to assess their own skills since they have poor skills in everything. I repeated the

question asking whether they actually felt that way and Maiju answered that she really thinks that she cannot produce anything in Swedish. She succeeds well in exams but after the exams she always forgets everything. Oona agreed with Maiju and stated that it is easy to find weaknesses.

It can be only a coincidence in a small data like this, but the pupils who were experienced in assessing their skills (the pupils is groups 2 and 3) found self-assessment challenging and emphasised the thinking process the assessment requires. The other pupils, who had less experience in doing self-assessment, did not have any clear opinions about the easiness or difficulty of self-assessment. Moreover, two pupils, Maiju and Oona, had not perhaps learned to assess their skills properly. Hence, it seems that the more experienced pupils had maybe more advanced self-assessment skills than the others. At least they described self-assessment more insightfully than the other pupils. This would be in line with what has been discussed earlier; self-assessment skills are like any other skills and need to be practiced (see, e.g. Douglas 2010, 75). In addition, self-assessment skills affect a learner's self-esteem and confidence positively (Alanen and Kajander 2011: 69) which sides with the claim that maybe Maiju and Oona did not yet have much experience of self-assessment.

Maiju and Oona also thought that self-assessment is not useful. They believe that it has no effect on anything but on the teacher's conception of their pupils. A pupil might seem conceited or alternatively a teacher might feel pity for a pupil depending on what the pupil has written on the self-assessment form. Maiju and Oona both admitted that self-assessment would be good if they did not have to show their assessments to anyone. They said that if the self-assessment form needs to be given to somebody else, nobody can write about the things s/he is good at. Thus, they are afraid that their teacher or their peers might consider them conceited if they emphasised their skills. Nevertheless, they acknowledged that self-assessment could be a good practise. Oona also added that self-

assessment could be useful in situations where many pupils point out the same problem or difficulty as a teacher can then revise the problematic topic.

All other pupils considered self-assessment useful. In groups 1, 2 and 5 the pupils thought that self-assessment provides important information for teachers as the teachers can then modify their teaching based on the information they have received. This is how Elsa explained the advantages of self-assessment:

(16) *Että siitä ni saa [opettaja] myös käsityksen mitä mieltä oppilas on omista kielitaidoistaan, että onks sillä jotain ongelmii, jotka ei välttämättä näkyis päällepäin.*

So from that s/he [teacher] can also get more information about what the pupil thinks about his/her language skills, so like if s/he has some problems that do not show in class. (Elsa)

Moreover, the Elsa and Hilla (group 5) as well as Henri, Stella (group 2) and Sofia (group 3) believed that it is useful to reflect one's knowledge and skills as it is important to recognise strengths and weaknesses, and thus, receive a more realistic conception of oneself. Henri, however, pointed out that sometimes self-assessment can have a negative effect if one writes a poor assessment of oneself and the teacher then gives a lower grade because s/he sees that the pupils does not expect a better grade.

All in all, some pupils had rather mature ideas about why self-assessment is useful. They found that it is important to know one's own strength and weaknesses in order to focus on one's own learning but also to provide useful information for teachers. Although two pupils felt that self-assessment is not useful because it does not affect anything, also they admitted that it could be a good thing if they did not have to show their self-assessments to anyone. Thus, the comment maybe signalled a lack of strong self-esteem rather than a feeling of self-assessment being useless. They were not used to assessing their own skills and as teenagers they probably were a bit insecure.

Nevertheless, the pupils' responses implicate also that the pupils underestimate their teachers' skills and professionalism in a sense that they probably do not understand how well their teachers actually know them. Many of the pupils' comments showed that the pupils were slightly worried about how a single assignment or feedback would affect their teachers' conceptions about them. In my view, however, teachers observe their pupils constantly and they might even teach the same pupils for years, so it most likely that the teachers have a quite good understanding of their pupils' skills and personalities. A teacher will notice when a pupil has received some help when writing an essay, for example, and a teacher will think no less of a pupil who overestimates or underestimates his or her skills while practising self-assessment. It is interesting that the pupils are so worried about this. The teachers' opinion clearly matters.

On the other hand, the pupils did not seem to trust themselves either. Some of the pupils were afraid of overestimating or underestimating their language skills when doing self-assessment and some said that it is easy to find only weaknesses. Moreover, some of the pupils were not convinced that everybody would assess themselves truthfully. They suspected that some pupils would write better self-assessments to receive a better grade and some would not have the courage to speak well of themselves. These factors show that the pupils are still young, maybe a bit immature and lacking self-confidence. It appears that although the pupils doubt that their teachers could be fooled by writing falsified self-assessment, they still trust their teachers more than themselves. The pupils clearly think that if their self-assessment is not equivalent to their teacher's opinion, the teacher is right and they have either overestimated or underestimated their skills.

Peer assessment was brought forward in most of the interviews when the pupils' roles were examined. An interesting observation was that Maiju and Oona as well as Conny, Kaarlo and Matias, who did not have much experience of peer assessment, found the idea interesting but said that one could not tell a classmate

anything negative. Maiju pointed out that nobody would have the courage to say what they really think, and hence, anonymous feedback would be good. Moreover, Elsa pointed out that peer assessment could lead to a feeling of being bullied, especially if one did not succeed well in the assessed task. It could be argued that all these comments and feelings illustrate the insecurity that teenagers often experience. On the other hand, the comments also show that the pupils are not experienced in assessing their peers. They would probably feel differently if they had started peer assessment in primary school and continued it for several years. For example Sofia, who had about three years of experience of peer assessment, found it very useful and pleasant.

To conclude, it seems that the teachers have the authority to decide the assessment methods and give grades. The pupils' role, on the other hand, is to work hard in order to earn good grades. Thus, the pupils are more a part of the different assessment processes than active participants in them. Nonetheless, all pupils noted at the end that self-assessment is useful. It is only the insecurity of being able to assess oneself truthfully that bothers some of the pupils. In my view the pupils' responses show that many of the pupils do not have very advanced self-assessment skills yet, and that is why they have doubts about their role in the assessment processes.

5.2.4 The advantages and disadvantages of the ELP in assessment

The pupils in groups 2 and 3 had more experience of the ELP than the other pupils and one could argue that it showed in the pupils' answers when they had to describe the advantages and disadvantages of using the ELP. Henri and Stella (group 2) considered the ELP work more free and fun because there is no similar kind of pressure as there is in exam situation. They can use their imagination and show their skills. Here is how Stella described her experiences:

(17) *...se tuntuu silt et mä haluun tehä sitä ja mä haluun niinku näyttää opettajalle et mä, niinku mitä mä osaan ja se tuntuu hyvältä…*

> … it feels like I want to do it and I want to, like, show my teacher that I can, like, what I can do, and it feels good… (Stella)

Also, Sofia (group 3) told that she likes doing ELP work and that it is nice to look at the ELP and see what she has done in previous years and how her skills have improved. She also found it nice to ponder on what to still practise and what she already can do. Neither Sofia nor Stella were able to say anything negative about the ELP but Henri noted that his grades are getting lower and it is harder for him to get the same grades as he used to. This probably cannot be counted as a disadvantage of the ELP but Henri's comment could indicate that the ELP has not helped him to notice his progress. His self-analysis is based on his grades.

Elsa (group 5) and Conny, Kaarlo and Matias (group 1) all thought that the ELP could be good for language assessment. Elsa thought that these kinds of alternative assessment methods should be more visible in schools since they would provide better opportunities for example for pupils who are shy or who have learning disabilities. Conny, Kaarlo and Matias emphasised the possibility to show their strengths when doing portfolio work. They could not find anything negative either, except that Elsa and Hilla thought that the ELP would not function as the only assessment method because it does not provide a realistic appraisal of a pupil's language ability.

Maiju and Oona (group 4) stated that ELP work is fun because it is not "real teaching". They thought that if assessment would be based only on the ELP work they would have fun in class as they could work and talk with their friends and they would not have to stress about exams. Oona also theorised that she could actually learn more if she had to for example look words up in a dictionary herself. She could remember the words better. Maiju, nonetheless, added that she would probably not learn better and she was worried that she would not learn any grammar if they were just doing portfolio work and no-one was teaching. Thus, the girls did not have a clear idea of how the ELP might actually work in

language learning and assessment. They understood that the ELP work would be somehow less serious and burdensome but more fun.

To summarize, the pupils had very positive experiences of the ELP and the pupils who did not have much experience of ELP work considered the ELP a positive boost to school assessment. Not all pupils were willing to accept the ELP as an only assessment approach but they thought that exams are needed. All pupils, however, noted that the ELP would bring something new to school assessment. Something new that would acknowledge some previously neglected skills, or alternatively, pupils who do not do well in exams.

Yet, one issue that was apparent in every interview, at least at some point, was the importance of grades. When discussing the pupils' role in assessment, grades were often mentioned, but also here, for example, Henri brought up the meaning of grades. The pupils seem to define their knowledge and skills on the basis of their exam grades and the grades in their school reports. Moreover, when reviewing the pupils' part of the research data, it becomes evident that exams are an important measures of language assessment for the pupils. Exam results are an indicator of a pupil's language skills. Similar ideas were present also in the study conducted by Luukka et al. (2008: 154). They found that assessment and feedback have a clear impact on what the pupils feel they can. Especially exam grades had an effect on how the pupils assessed their abilities.

It appears that exams and grades go hand in hand and both have traditionally had great influence on pupils' self-esteem and possibilities for further studies. Exam grades have, at least traditionally, weighted a lot when deciding the final grades. Moreover, good grades have been, and often still are, considered to indicate success in life. Parents tell their children to study hard so that they can get good grades and get in to a good upper-secondary school or the kind of school that can provide them a good education for the desired occupation. Thus, in the Finnish society grades matter and one's skills are assess based on his or her school

grades. Pupils are socialised to the system where good grades provide more opportunities in the future.

Of course the matter can be considered from a more practical perspective as well. Although a grade does no tell much about ones skills, it is an easy way to assess and compare pupils with each other. Written final assessments would certainly require more time and work than giving a grade. Moreover, it can be that pupils appreciate grades because they consider grades simple and clear signs of their language skills. Also, some pupils might enjoy comparing their grades with their friends. Nonetheless, it can be argued that giving grades derives from the traditional normative assessment practices of comparing pupils with each other. Now, when the pupils' skills are assessed against criteria levels, it would be appropriate to abandon grades and start giving pupils more descriptive assessments by utilising the CEFR scales more.

5.3 Comparing the teacher and pupil perspectives

The teachers clearly had a broader idea of the ELP assessment and its benefits than the pupils, but that was to be expected. All the teachers had wide knowledge of the ELP and they had been conducting the ELP work for several years. Some of the teachers were not using the ELP at the moment of the interview as much as they had in previous years or as much as they were planning to in the future. Thus, only the pupils in groups 2 and 3 had conducted ELP work for about three years. The other pupils had done separate pieces of portfolio work and some self-assessment but they did not have much experience of the ELP.

Both the pupils and the teachers defined the ELP assessment as assessment of pupils' own work. Moreover, some of the pupils mentioned that in ELP assessment pupils can show their strengths and personalities, which also the teachers mentioned. Otherwise the pupils had a very practise-oriented view of the ELP assessment when comparing it with exam. The pupils listed the lack of

time limit and stress and the possibility receive help. These aspects were not mentioned by the teachers which might indicate that the teachers' perspective was different from the way how pupils viewed the ELP. On the other hand, the difference may have been due to the way how the issue was approached. Teachers were asked to describe the ELP assessment whereas the pupils were asked to approach the topic also by describing the differences between exams and the ELP.

What was interesting was that four of the pupils stated that the ELP work does not provide a realistic appraisal of a pupil's skills and abilities. They thought that only exams can do that. This was in contradiction with the opinions of the rest of the pupils and all the teachers. The pupils who had more experience of the ELP believed that since one can use his or her skills more freely in the ELP work, it represents their language skills realistically. The teachers also emphasised the versatility of the ELP work and variation in assessment in general. In my view the pupils who believed that exams provide realistic results did not just have enough experience of the ELP or other types of assessment methods, or they simply were so used to taking exams that they saw the exams as the only right method.

The pupils had rather varying views about what language ability is but some of the pupils had very advanced ideas. Although the pupils were not asked about whether they see language ability consisting of the four skills they stated that in addition to knowing vocabulary and language structures it is important to be able understand or to be understood by others. Some pupils also pointed out that it is important to be able to have a conversation or use the foreign language later in life, for example in working life. The differences between the teachers' and pupils' answers became clearer when the used assessment procedures and feedback practices were discussed. All teachers and pupils listed exams, and then word quizzes, essays and listening and reading comprehensions tasks were mentioned by some individual pupils and teachers. Nevertheless, all teachers

reported assessing their pupils' oral skills, often by observing their pupils, but only two pupils in group 4 mentioned classroom observations. None of the other pupils said anything about the assessment of oral skills. Actually the boys in group 1 stated that it would be good if they also had oral exams. The reason for this disparity in answers could be the assessment method. The teachers reported assessing oral skills mostly by observing their pupils and it might be that the pupils did not consider observations assessment. As it has been pointed out, the pupils seemed to have a quite traditional views of assessment and since classroom observations are no formal testing situations, it can be that the pupils did not recognise observations as proper assessment.

The same kind of disparity concerned the feedback practices. All teachers reported giving oral feedback to their pupils continuously during classes. Written feedback was given more periodically. The pupils' responses varied a bit but none of the pupils thought that they received feedback often. Indeed, the pupils in group 1, for example, stated that they receive feedback rarely but sometimes they have had assessment discussions with their teacher. Furthermore, the pupils in group 2 found that they receive their feedback in the form of a grade and one of the pupils hoped that he would get more encouraging oral feedback from his teacher. This disparity is interesting but could be explained again with the fact that the teachers give mostly oral feedback and that feedback may be in the form of guidance or a quick comment, and hence, the pupils do not consider those to be feedback. It could be that they regard it natural and they see that feedback practices are some more obvious and concrete acts, like separate feedback conversations. It may be that if the pupils had been given examples of possible feedback methods, they would have mentioned also other practices.

The study conducted by Luukka et al. (2008: 139–140) showed similar results. In their study a third of the pupils found that they did not receive enough feedback. Moreover, only half of the pupils reported receiving personal feedback from their

teacher, whereas over 70 percent of the teachers reported giving personal feedback (Luukka et al 2008: 142). In the present study the pupils were satisfied with the current situation with the feedback practices but they reported receiving feedback rather rarely although the teachers reported giving ongoing oral feedback. As Luukka et al. (ibid.) analysed, a possible explanation for the differences between the teacher and the pupil perspective could be that although the teachers felt that they give oral feedback continuously, an individual pupils may receive it more rarely. The pupils share the teachers' attention.

When discussing the pupils' role in assessment the teachers emphasised the importance of interaction. They reported that they discuss assessment criteria, learning goals and the concept of language ability with their pupils and try to give them tools for assessing their language skills. The pupils, on the other hand, identified that their role is to be hard-working and active pupils. All the pupils mentioned active participation and hard work when they were asked about the ways how they could affect their assessment. It seems that the pupils did not see themselves having an active role in the assessment processes in the sense that they could in some way impact on, for example, the assessment methods, but they rather thought that they can affect their grades by working hard. Two pupils were even slightly amazed by the question as they were wondering whether pupils should even be able affect their assessment. Still, there were pupils, for example in groups 2 and 5 who thought that pupils could take more responsibility for their own learning and aim at improving their language skills. Also in group 1 Kaarlo mentioned that pupils can sometimes influence the topics that will be studied. Thus again, the teachers and the pupils had somewhat different point of views and it seems that the teachers and the pupils have different aims. The teachers want their pupils to reflect their own learning and take responsibility for learning whereas pupils aim to receive good grades.

Most pupils (8/10) considered self-assessment useful. The pupils thought that self-assessment can provide important information for teachers and that it is

good for pupils to know their strengths and weaknesses. The two pupils who did not find self-assessment so useful stated that it has no effect on anything, and moreover, it feels disconcerting to assess one's own strengths. Similar thoughts were presented when the possibility of peer assessment was discussed in groups 1, 4 and 5. The pupils believed that it would feel awkward and difficult to assess their peers work. Elsa also added that it could increase bullying and that it should be a teacher's job to assess pupils.

The teachers, on the other hand, all agreed that self-assessment is important and that it needs to be taught to the pupils. It is a skills that develops. None of the teachers said that the pupils' self-assessment would explicitly affect their grading but they reported that it is important to discuss it with their pupils and make sure that they understand what their assessment consists of and why they are given a certain grade. Furthermore, the teachers noted that the way how pupils see themselves affects the teachers' work. Nevertheless, for example, Anna-Maija reported that about 90 per cent of the pupils are able to assess their own skills rather accurately. This is slightly contradicting with the study conducted by Luukka et al. (2008: 125) where it was found that the teachers did not trust their pupils' self-assessment skills. In the present study the teachers did not state that the they would not trust their pupils' skills in self-assessment but some teachers even pointed out that many pupils are able to assess their skills well. Then, on the other hand, the pupils' self-assessment did not seem to have any direct effect on the grades in the present study either. It might be that the teachers think that their pupils are not enough mature to analyse their skills comprehensively or it can also be that the teachers merely trust their own professionalism more.

Finally the teachers and the pupils described their feelings and the advantages and disadvantages of using the ELP. Some of the pupils did not have much experience of the ELP but then they focused more on the advantages and disadvantages than their feelings. The teachers used the adjectives natural, nice, inspiring, interesting and eye-opening. The pupils described the ELP work as

fun, nice, stress-free and emphasised the fact that in the ELP work pupils can show their strengths, and the work in general is more free. The pupils did not find any significant disadvantages except for that one pupil was worried that she would not learn grammar rules by conducting ELP work and a couple other pupils believed that the ELP would not provide a realistic image of one's skills. The teachers reported different types of disadvantages. They viewed the ELP from their own perspective and mentioned the workload and other peoples' attitudes towards the ELP. Thus, the two parties' different perspectives clearly affected their answers.

It can be concluded that both teachers and pupils enjoy the ELP work and consider it to be a good assessment method. Some pupils favoured traditional exams but sill admitted that the ELP is would be useful. In general, the teachers' views and answers did not drastically contradict with the pupils' answers. In my view the most significant differences were caused by the 'invisible' work that many teacher do inside and outside the classroom. The teachers modify their teaching, observe their pupils, communicate with their colleagues and the pupils parents and do many small things that the pupils see as natural elements of classroom activities. The pupils might not notice all the work the teachers actually do for them. In addition, it seems that the pupils are not fully aware of how well their teachers actually know them and they are worried that teachers do not notice if a pupils underestimates his or her skills, for example. The pupils are not all so sure about their self-assessment skills but they value their teachers' opinions.

Moreover, the pupils, as well as the teachers to some extent, consider exams important. The results of the present study also suggest that pupils see only formal assessment methods as proper assessment. They do not consider for example classroom observations to be a real method for assessing their oral communication skills. That is also a likely reason for why the pupils' thought that their writing skills are assed more than their oral skills. A pupil does not get a

grade from classroom observations but from exams s/he receives a numeric result that indicates his or her level of language ability. Grades from formal assessment methods are more concrete evidence.

Altogether, two notable difference can be found between the teachers' and the pupils' answers. Firstly, as mentioned, it seems that the teachers' aim in involving pupils in the assessment processes is to let the pupils take responsibility of their own learning whereas the pupils' aim to get good grades. Many times the pupils discussed the importance of exams and grades. Secondly, the teachers emphasised reciprocity in assessment and in teaching in general whereas the pupils often consider the teacher as the judge of their language ability. This was evident for example in Kaarlo's commet (example 15) when he said that when the teacher has given the grade there is nothing one can do. In addition, the pupils in group 5 though that their self-assessment, or pupils' opinions in general, should not affect the assessment of their skills. Although most of the pupils found self-assessment useful, the idea that the teacher has the responsibility for the assessment is sometimes noticeable in the present study. Like in the study conducted by Luukka et al (2008), also here the teachers' role in assessment is significant.

In conclusion, when comparing the teachers' and the pupils' answers it is important to remember that these two groups have very different backgrounds and perspectives. It is natural that the teachers, who have a university degree and more importantly many years of teaching experience, see language assessment differently than the pupils. In comparison, the pupils' cognitive skills are still developing and in my opinion children often know more than they realise. Still, it is the teachers' job to teach the pupils self-assessment skills and involve them in the learning and assessment processes by discussing for example the learning material, learning strategies, the concept of language ability and language assessment. Only by involving pupils in the processes they can actually take responsibility for their own learning and be motivated to develop their skills. It

is not an easy task and it will take time but those are the type of factors that enhance learning and promote the ultimate goal of lifelong language learning.

6 CONCLUSION

The aim of the present study was to examine pupils' and English and Swedish teachers' perceptions of language assessment, the use of the European Language Portfolio (ELP) in language assessment and the pupils' role in assessment. In general, both pupils and teachers thought that the ELP is a very useful and versatile method for language assessment. Especially the teachers found that the ELP is an important asset for all assessment as it shows pupils' progress, extends pupils' understanding of language assessment and embraces versatility.

There was more variety in the pupils' responses but also the pupils considered the ELP beneficial as the ELP work is more free than exams. In exams one has to know certain vocabulary and topics whereas in the ELP one can express him- or herself more freely and use his or her strengths. Thus, the ELP provides better opportunities for success. Most of the pupils (6/10) thought also that the ELP gives them a chance to show their skills but, on the other hand, some pupils believed that the ELP does not provide a realistic appraisal of a learner's skills. When conducting portfolio work pupils can look words up in a dictionary or ask for help, and hence, they believed that the ELP assessment is not as truthful as exams. Nevertheless, the pupils admitted that the ELP could be a good addition to exams. This implicates that the pupils' conceptions of language ability are somewhat limited. It can be that the extensive use of exams as assessment methods in Finnish schools has affected the pupils' conceptions.

Altogether, also the differences in the amount of ELP knowledge and experience have most likely influenced the teachers' and pupils' answers. All the teachers in the present study had wide experience and knowledge of the ELP whereas most of the pupils had very little experience of the ELP. Only three of the ten pupils

had done ELP work more or less continuously for about three to four years. Thus, most of the pupils were more used to taking exams which are in a way clearer and more explicit assessment methods than the ELP. From exams pupils receive grades which are more familiar to them than the criteria levels of the CEFR.

In the present study it became evident that the ELP is more of an approach to language teaching, learning and assessment than any language specific teaching tool. The ELP can be used equally well in English and Swedish language assessment, and at any level. The level of the ELP work just needs to be adjusted to the skills and the age of the pupils, but that is the same with any type of assessment method. The ELP can, however, bring different types of benefits for different languages. It can, for example enable the recognition of language specific traits or individual pupils' interests and language needs.

The results of the present study indicate that teachers still have a leading role in assessment. Pupils are involved in the assessment processes by conducting self-assessment, for example, and all the teachers in the study considered pupil involvement important. Nevertheless, some of the pupils seemed to value teacher-led assessment. The pupils' thought that they can affect their own assessment by working hard and being active in class. It seemed like the pupils could not even think of any other way of being involved in the assessment processes. This indicates that the pupils did not have any experience of being able to influence, for example, the used assessment methods or the content of exams.

It could be that the teacher's role is so significant because pupils are socialised to consider teacher as an authority who decides the content, materials and methods of learning. Moreover, the long tradition of using summative exams in language assessment is undoubtedly one factor supporting the popularity of teacher-led assessment. As Välijärvi stated in the news article (Vähäsarja 2014), pupils do not have much experience of other forms of assessment than exams. One could even claim that exams guide language learning in schools; first one has to learn new

vocabulary and structures and then after the learning period one's skills and knowledge are tested in an exam. That pattern is rooted in the minds of most teachers and pupils, but also the pupils' parents. It is likely that this is the main reason for the popularity of exams today, and furthermore, for the significant role of the teachers in assessment. Pupils are the test takers who have no power on the assessment process.

Thus, change is needed. Traditional exams do not fit the modern assessment culture. The aim of school assessment should be to enhance learning via multiple assessment methods. Summative exams are not an appropriate method for assessing learners' communicative language abilities authentically or purposefully. A test situation, with set tasks and a time limit, cannot be ideal for assessing one's language ability. Provided that exams are the main form of assessment in Finland, are pupils' then learning the skills they are assessed for, or vice versa, are all the learnt skills being assessed? As pointed out in the present study, the views of language ability always affect the ways of assessment. Thus, one may ask what the current assessment methods used in Finnish schools tell about teachers' as well as pupils' views of language ability.

Of course, exams can be of many types and, exams can and should be used also formatively. Pupils should be involved in the exam planning and grading processes so that they learn to see what is important, why are they being assessed and how. Thus, they would gain a greater understanding of language assessment and learn to see versatile assessment processes as tools that enhance their learning.

There is a clear need for further assessment studies that would provide more information about different language assessment methods that could be used in schools. More research data is needed especially about exams since they are commonly used. The strengths and weaknesses of exams need to be studied properly but also formative uses of exams should be promoted and examined.

Research always raises awareness and it is vital that all educators, pupils and their parents understand the meaning of versatile assessment that includes both formal and more informal assessment practices.

The main aim of the present study was to examine the use of the ELP in language assessment. In the study six teachers and ten pupils were interviewed, and thus the participants had an opportunity to discuss and clarify their perceptions in their own words. The study, however, has some limitations. Firstly, the amount of participants is not sufficient to make any generalisations about the use of the ELP. Secondly, most of the pupils did not have much experience of the ELP, and thus, their answers to some of the questions were based on their assumption of what it would be like if the ELP was used more in their assessment. Thirdly, as the pupils were interviewed in groups or pairs, one must considered the possible effect of group pressure. Teenagers especially are often concerned about what their peers think about them and that is why one can only assume that all the pupils answered the questions truthfully and left nothing unsaid because of a fear of embarrassment, for example. Finally, all the teachers had used the ELP for many years and even if they were not using it at the moment, they were planning to add the elements of the ELP in their teaching in the future. Hence, it was very likely that all the teachers would find the ELP very useful and pleasant. In order to get a wider understanding of the advantages and disadvantages of using the ELP, one could interview also teachers who have used the ELP but do not use it anymore for some reason.

Thus, the use of the ELP should be studied more from the teachers, but also from the pupils' point of view. It would be very interesting to examine pupils' perceptions of language assessment by studying pupils who have never used the ELP and pupils who use the ELP. Nevertheless, it seems that not only research but also promotion of the ELP would be needed since the ELP is still unknown for many teachers, pupils and parents (see, e.g. Salo et al. 2013: 38). The ELP is a foreign tool for many teachers and it is obvious that adapting a new way of

teaching, or even thinking, feels burdensome and difficult. In my view, many teachers simply do not have a clear idea of what the ELP actually is, and hence, they cannot see the benefits of adopting it. Also, in the present study two teachers mentioned that the role of the ELP as a valid tool is sometimes questioned by other teachers and pupils' parents. The ELP is considered to be something extra and fun, and it is not taken seriously. Nonetheless, as the present study shows, many teachers and pupils think that the ELP work is fun but also useful. The ELP enables pupil involvement and differentiation. Teachers clearly need ELP training and pupils and parents need more information.

In conclusion, language assessment is an intricate issue that should be discussed more in Finland. Exams alone are not enough to assess learners' language skills diversely. There is no single assessment method that would assess a learner's language ability comprehensively and be valid in the sense that standardised tests are expected to be but the ELP is a great approach to language assessment embracing versatile and purposeful assessment.

The use of the ELP will be encouraged in the new NCC (see POPS 2014) and the ELP will surely become more popular in Finland in the future. In addition, to the forthcoming official recommendation, there are also several other factors that, in my opinion, encourage the use of the ELP in the future. Firstly, the assessment culture is gradually changing, and as it reaches the practise, there is a clear need for assessment methods such as the ELP. Secondly, the Finnish versions of the ELP are finally available online, and hence, the examples, self-assessment grids and other templates are available to everyone. Thirdly, as the name suggests the ELP is a product developed by the Council of Europe and adheres to internationally agreed principles. The ELP is related to the CEFR and increases transparency in language learning and assessment in Europe. Finally, the ELP is well studied in Finland and the development process has continued for more than a decade which makes the ELP more than a transient trend.

Thus, the perquisites for the ELP to become more common in Finland exist. More promotion, teacher training and time is, however, needed. Old traditions and habits are hard to break. The ELP is one remarkable approach that embodies the new assessment culture by including versatile proof of learners' abilities, involving learners and their peers in the learning and assessment processes, assessing the work and progress of individual learners and providing tools for authentic, life-long learning. All in all, the European Language Portfolio conjoins language learning and assessment.

7 BIBLIOGRAPHY

Alanen, R. and Kajander, K. (2011). Reflektio ja itsearviointi. Opettajan mielestelyä vai kriittistä oman toiminnan arviointia? In R. Hildén and O-P. Salo (eds.), *Kielikasvatus tänään ja huomenna. Opetussuunnitelmat, opettajankoulutus ja kielenopettajan arki.* Helsinki: WSOYpro Oy, 65–82.

Alderson, J. C. (1993). The state of language testing in the 1990s. In A. Huhta, K. Sajavaara and S. Takala (eds.), *Language testing: New openings*. Jyväskylä: Institute for Educational Research, 1–19.

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

Atjonen, P. (2007). *Hyvä, paha arviointi*. Helsinki: Tammi.

Bachman, L. F. (1991). *Fundamental considerations in language testing* (2nd impression). Oxford: Oxford University Press.

Bachman, L. F. and Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Black, P. and Wiliam, D. (2012). Assessment for learning in the classroom. In J. Garder (ed.), *Assessment and learning*. London: SAGE Publications Ltd, 11–32. doi: http://dx.doi.org.ezproxy.jyu.fi/10.4135/9781446250808.

Brown, A. (2012). Uses of language assessment. In *The Encyclopedia of Applied linguistics*, 5979–5985. doi: 10.1002/9781405198431.wbeal1237.

Canale, M. and Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* [online] 1, 1–47. http://ibatefl.com/wp-content/uploads/2012/08/CLT-Canale-Swain.pdf.

Chen, L., Rogers, T. and Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: purposes, methods and procedures. *Language Testing* 21 (3), 360–389. doi: 10.1191/0265532204lt288oa.

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Council of Europe (2011). European Language Portfolio. Education and Languages. http://www.coe.int/t/dg4/education/elp/ELP-REG/Default_EN.asp. (26 September, 2014)

Davidson, C. and Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly* 43 (3), 393–415. doi: 10.1002/j.1545-7249.2009.tb00242.x.

Douglas, D. (2010). *Understanding language testing*. London: Hodder Education.

Eskola, J. and Suoranta J. (2008). *Johdatus laadulliseen tutkimukseen*. Jyväskylä: Gummerus Kirjapaino Oy.

Eurooppalainen kielisalkku. http://kielisalkku.edu.fi/fi. (1 October, 2014).

Fox, J. (2012). Language assessment methods. In C. A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*. Oxford: Blackwell Publishing Ltd, 2931–2936. doi: 10.1002/9781405198431.wbeal0606.

Fox, J. (2008). Alternative assessment. In E. Shohamy and N. H. Hornberger (eds.), *Encyclopedia of Language and Education*, Vol. 7. *Language Testing and Assessment* (2nd edition). New York: Springer Science+Business Media LLC, 97–110.

Fulcher, G. and Davidson, F. (2007). *Language testing and assessment: an advanced resource book.* London: Routledge.

Gardner, J. (2012). Assessment and learning: Introduction. In J. Gardner (ed.), *Assessment and learning* (2nd edition). London: SAGE Publications Ltd, 1–9. doi: http://dx.doi.org.ezproxy.jyu.fi/10.4135/9781446250808.

Hildén, R. (2004). Opettajien näkemyksiä eurooppalaisen kielisalkun kokeilusta Helsingin alueella. In J. Loima (ed.) *Theoria et* praxis. Viikin normaalikoulun julkaisuja / Viikki TTS Publications. Helsinki: Helsingin yliopiston Viikin normaalikoulu, [online] 76–100. http://www.vink.helsinki.fi/files/Theoria_et_praxis_edited_04.pdf. (11 August, 2014).

Hildén, R. (2009). In pursuit of validity: an empirical pilot on validating the Finnish version of the European Language Portfolio developed for upper

grades of compulsory basic education. *Melbourne papers in language testing* [online] 14 (2), 32-71. http://wayback.archive-it.org/1148/20130404065629/http://ltrc.unimelb.edu.au/mplt/papers/14_2_2_Hilden.pdf. (11 August, 2014).

Hildén, R. (2011). Muutosten tarpeet ja mahdollisuudet kielten opetuksessa. In R. Hildén and O-P. Salo (eds.), *Kielikasvatus tänään ja huomenna. Opetussuunnitelmat, opettajankoulutus ja kielenopettajan arki.* Helsinki: WSOYpro Oy, 6–18.

Hildén, R. and Takala, S. (2005). Kielisalkulla kohti selkeää ja monipuolista arviointia. In V. Kohonen (ed.), *Eurooppalainen kielisalkku Suomessa*. Helsinki: WSOY, 315–326.

Hill, K. and McNamara, T. (2011). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing* 29 (3), 395–420. doi: 10.1177/0265532211428317.

Hirsijärvi, S. and Hurme, H. (2008). *Tutkimushaastattelu. Teemahaastattelun teoria ja käytäntö*. Helsinki: Gaudeamus Helsinki University Press.

Hughes, A. (2003). *Testing for language teachers* (2nd edition). Cambridge: Cambridge University Press.

Huhta, A. and Hildén R. (2013). Kielitaidon arvioinnin metodologisia vaihtoehtoja. In A. Räsänen (ed.), *Oppimisen arvioinnin kontekstit ja käytännöt*. Raportit ja selvitykset 2013:3. Opetushallitus, Helsinki: Juvenes Print, 159–186.

Huhta, A. and Takala, S. (1999). Kielitaidon arviointi. In K. Sajavaara and A. Piirainen-Marsh (eds.), *Kielenoppimisen kysymyksiä*. Jyväskylä: Soveltavan kielentutkimuksen keskus, Jyväskylän yliopisto (Jyväskylän yliopistopaino), 179–228.

Huhta, A. and Tarnanen, M. (2009). Assessment practices in the Finnish comprehensive school – what is the students' role in them? *LED2007: Refereed conference proceedings of the 2nd International Conference on Language, Education and Diversity*. Hamilton, New Zealand: Wilf Malcolm Institute of Educational Research (WMIER), University of Waikato.

Inbar-Lourie, O. (2008). Language assessment culture. In E. Shohamy and N. H. Hornberger (eds.), *Encyclopedia of Language and Education*, Vol. 7. *Language Testing and Assessment* (2nd edition). New York: Springer Science+Business Media LLC, 285–299. doi: 10.1007/978-0-387-30424-3_182.

Jones, N. (2012). Reliability and dependability. In G. Fulcher and F. Davidson (eds.), *The Routledge Handbook of Language Testing*. London, New York: Routledge, 350–362.

Kantelinen, R. and Hildén R. (2012). Eurooppalainen kielisalkku vihdoin suomalaiseen perusopetukseen! *Kieli, koulutus ja yhteiskunta*. [online] Lokakuu, 1–5. https://jyx.jyu.fi/dspace/bitstream/handle/123456789/40208/eurooppalainen-kielisalkku-vihdoin-suomalaiseen-perusopetukseen.pdf?sequence=1. (15 September, 2014).

Katz, A. and Gottlieb, M. (2012). Assessment in the classroom. In C. A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*. Oxford: Blackwell Publishing Ltd, 161–169. doi: 10.1002/9781405198431.wbeal0042.

Keurulainen, H. (2013). Pelisääntöjä arviointipäätösten tekemistä varten. In A. Räsänen (ed.), *Oppimisen arvioinnin kontekstit ja käytännöt*. Raportit ja selvitykset 2013:3. Opetushallitus, Helsinki: Juvenes Print, 37–60.

Kohonen, V. (2000). Eurooppalainen kielisalkku kielenopetuksen kehittämistyössä: suuntaviittoja ja mahdollisuuksia. In Kohonen, V. and Pajukanta, U. (eds.) *Eurooppalainen kielisalkku – kokemuksia EKS-projektsin alkutaipaleelta.* Tampere: Juvenes-Print – Tampereen yliopistopaino Oy, 9–44.

Kohonen, V. (2005a). Eurooppalaisen kielisalkun kehittämistyö ja tavoitteet: miten kielisalkku voi edistää kielikasvatusta? In Kohonen, V (ed.) *Eurooppalainen kielisalkku Suomessa*. Helsinki: WSOY, 7–44.

Kohonen, V. (2005b). Arviointi autenttisemmaksi: miten Eurooppalainen kielisalkku voi edistää kielikasvatusta? In Jaatinen, R. and Lehtovaara, J. (eds.) *Arviointi autenttisemmaksi. Kieltenopettajien kertomuksia opetuksen ja arvioinnin kehittämisestä*. Pirkanmaan ammattikorkeakoulu julkaisusarja A. Tutkimukset ja selvitykset. Nro 8. Tampere: Pirkanmaan ammattikorkeakoulu, 25–61.

Kohonen, V. and Pajukanta, U. (2003). Eurooppalainen kielisalkku -kokeiluprojektin tulosten koontaa. In Kohonen, V. and Pajukanta, U. (eds.) *Eurooppalainen kielisalkku 2 – EKS-projektin päätösvaiheen tuloksia.* Tampere: Juvenes-Print – Tampereen yliopistopaino Oy, 7–31.

Kolu, U. and Tapaninaho, H. (2000). Kokemuksia salkkutyöskentelystä yläasteella. In Kohonen, V. and Pajukanta, U. (eds.) *Eurooppalainen kielisalkku – kokemuksia EKS-projektin alkutaipaleelta*. Tampere: Juvenes-Print Oy –Tampereen yliopistopaino, 83–114.

Kolu, U. and Tapaninaho, H. (2003). Eurooppalaisen kielisalkun kokeiluja ja tulosten loppuarviointeja yläasteella. In Kohonen, V. and Pajukanta, U. (eds.) *Eurooppalainen kielisalkku 2 – EKS-projektin päätösvaiheen tuloksia.* Tampere: Juvenes-Print – Tampereen yliopistopaino Oy, 33–66.

Lado, R. (1961). *Language testing: The construction and use of foreign language tests: a teacher's book.* London: Longmans.

Lammi, K. (2002). *Kielisalkku lukion ruotsin kielen opiskeluun motivoinnissa: opiskelijoiden ja opettajan kokemuksia.* Jyväskylä: Soveltavan kielentutkimuksen keskus.

Limberg, L. B. (2008). Phenomenography. In *The SAGE Encyclopedia of Qualitative Research Methods*, 612–615.
doi: http://dx.doi.org/10.4135/9781412963909.n316.

Little, D. (2009). *The European Language Portfolio: where pedagogy and assessment meet*. Strasbourg: Language Policy Division DGIV/EDU/LANG 19. [online]. http://www.coe.int/t/dg4/education/elp/elp-reg/Source/Publications/ELP_pedagogy_assessment_Little_EN.pdf.
(6 March, 2014).

Luukka, M-R., Pöyhönen, S., Huhta, A., Taalas, P., Tarnanen, M. and Keränen, A. (2008). *Maailma muuttuu - mitä tekee koulu? Äidinkielen ja vieraiden kielten tekstikäytänteet koulussa ja vapaa-ajalla*. Jyväskylä: Jyväskylän yliopisto, Soveltavan kielentutkimuksen keskus.

Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing,* 18 (4), 351-372. doi: 10.1177/026553220101800403.

Lynch, B. and Shaw, P. (2005). Portfolios, power and ethics. *TESOL Quarterly* 39 (2), 263–297. doi: 10.2307/3588311.

McNamara, T. (1996). *Measuring second language performance*. [online]. http://languages-linguistics.unimelb.edu.au/academic-staff/tim-mcnamara. (12 January, 2014).

Opetushallitus (n.d.). Todistuksenanto-oikeus ja yleistietoa todistuksista. Säädökset ja ohjeet. http://www.oph.fi/saadokset_ja_ohjeet/ohjeita_koulutuksen_jarjestamiseen/perusopetuksen_jarjestaminen/perusopetuksen_oppilaan_arviointi/todistuksenanto_oikeus. (23 September, 2014).

Paran, A. (2010). More than language: the additional faces of testing and assessment in language learning and teaching. In A. Paran and L. Sercu (eds.), *New perspectives on language and education: Testing the untestable in language education*. Clevedon, GBR: Multilingual Matters, 1–13.

Perspectives on the European Language Portfolio for the comprehensive school in Finland (2014). [online] http://kielisalkku.edu.fi/wp-content/uploads/2014/06/Perspectives_on_-ELP_15.3.2014.pdf, 1–6. (8 October, 2014).

POPS/Perusopetuksen opetussuunnitelman perusteet (2014) Finnish National Board of Education [online]. http://www.oph.fi/download/163777_perusopetuksen_opetussuunnitelman_perusteet_2014.pdf (14 May, 2015).

POPS/Perusopetuksen opetussuunnitelman perusteet (2004). Finnish National Board of Education [online].

http://www.oph.fi/download/139848_pops_web.pdf
(27 August, 2014).

OECD/CERI (2005). *Formative assessment: improving learning in secondary classrooms.* Paris: OECD Publishing.

O'Sullivan, B. (2011). Introduction. In B. O'Sullivan (eds.), *Language testing: theories and practices.* Basingstoke; New York: Palgrave Macmillan, 1-12.

O'Sullivan, B. and Weir, C. J. (2011). Test development and validation. In B. O'Sullivan (eds.), *Language testing: theories and practices.* Basingstoke; New York: Palgrave Macmillan, 13-32.

Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language testing* 21 (3), 249–258. doi:10.1191/0265532204lt283ed.

Rea-Dickins, P. (2008). Classroom-based language assessment. In E. Shohamy and N. H. Hornberger (eds.), *Encyclopedia of Language and Education*, Vol. 7. *Language Testing and Assessment* (2nd edition). New York: Springer Science+Business Media LLC, 257–271.

Rea-Dickins, P. (2011). Foreword. Formative assessment: Scoping the horizons. In D. Tsagari and I. Csépes (eds.), *Classroom-Based Language Assessment.* Language Testing and Evaluation Series, vol. 25. R. Grotjahn and G. Sigott (general eds.). Frankfurt am Main: Peter Lang, 9–14.

Saarinen, K. (2010). Kielisalkkutyöskentelyn mahdolisuuksista amattikorkeakoulussa. In K. Saarinen (ed.), *Kielisalkkutyöskentelystä on moneksi*. Tampere: Tampereen ammattikorkeakoulu, 8–21.

Salo, O-P., Kalaja, M., Kara, H. and Kähkönen, K. (2013). Kielisalkku kielikasvatuksen työvälineenä – Jyväskylän normaalikoulun kielisalkkuhankkeen taustoja ja tavoitteita. In van den Berg, M., Mäkelä, R., Ruuska, H., Stenberg, K., Loukomies, A. and Palmqvist, R. (eds.) *Tutki, kokeile ja kehitä.* Suomen harjoittelukoulujen julkaisu 2012. Helsinki: Yliopistopaino, 35–56.

Sumner, R. (2006). Qualitative research. In *The SAGE Dictionary of Social Research Methods*, 249–251. doi: http://dx.doi.org/10.4135/9780857020116.

Tarnanen, M. and Huhta, A. (2011). Foreign language assessment and feedback practices in Finland. In D. Tsagari and I. Csépes (eds.), *Classroom-Based Language Assessment.* Language Testing and Evaluation Series, vol. 25. R. Grotjahn and G. Sigott (general eds.). Frankfurt am Main: Peter Lang, 129–146.

Tarnanen, M., Huhta, A. and Pohjala, K. (2007). Mitä on osaaminen? Kielitaidon arviointi vastaajana. In S. Pöyhönen and M-R. Luukka (eds.), *Kohti tulevaisuuden kielikoulutusta*. Kielikoulutuspoliittisen projektin loppuraportti. Jyväskylä: Jyväskylän yliopisto, Soveltavan kieltentutkimuksen keskus, 381–412.

Tuomi, J. and Sarajärvi, A. (2009). *Laadullinen tutkimus ja sisällönanalyysi*. Helsinki: Kustannusosakeyhtiö Tammi.

Turner, C. A. (2012). Classroom assessment. In G. Fulcher and F. Davidson (eds.), *The Routledge Handbook of Language Testing.* London, New York: Routledge, 35–78.

Vähäsarja, S. (2014). *Tulevaisuuden koulun kehittäjät: Perinteinen koe on turhan hallitseva – kokeettomuuttakin voisi kokeilla rohkeasti*. Yle uutiset [online]. http://yle.fi/uutiset/tulevaisuuden_koulun_kehittajat_perinteinen_koe_on_turhan_hallitseva__kokeettomuuttakin_voisi_kokeilla_rohkeasti/7426939. (2 September, 2014).

8 APPENDICES

Appendix 1: The interview questions for teachers

Opettajat

- Taustatietoa
  - mitä salkkuarviointi sisältää?
  - kuinka kauan olet toteuttanut salkkuarviointia
    - onko jatkuvaa
    - miten salkkuarviointisi on muuttunut vuosien aikana?
    - käytätkö eri kielissä
      - onko eroja, millaisia?
    - millaisia eroja luokka-asteiden välillä

- Salkkuarviointi
  - miten salkkutöitä arvioidaan
    - mitä arvioit? (mihin huomio kohdistuu? oikeakielisyys, kehitys…?)
    - millaista arviointi on? (suullista, kirjallista, jaksoittaista, jatkuvaa?)
    - jaotteletko kielitaitoa osa-alueisiin arvioinnissasi?
    - painottuvatko kaikki kielitaidon osa-alueet tasapuolisesti?
  - miten toteutuvat reliaabelius ja validius?
  - kuinka paljon oppilailla on valinnanvapautta salkkutöissä
    - miten se vaikuttaa arviointiin?
  - miten taitotasoasteikkoja käytetään
    - onko helppoa sijoittaa oppilas asteikolle salkkuarvioinnin perusteella (kriteeriviitteinen arviointi vs. numeroarviointi todistukseen)?
  - käytätkö lisäksi muita arviointikeinoja?
    - miksi?

- Oppilaan rooli arvioinnissa
  - miten oppilas osallistuu arviointiin?
  - itsearvioinnin toimivuus
  - auttaako salkkutyöskentely oppilaita ymmärtämään ja pohtimaan omaa oppimistaan/kielitaitoaan?
  - miten hyödynnät kielitaidon tarkistuslistoja?

- Omat kokemukset ja tuntemukset
  - miltä salkkuarviointi tuntuu?
  - mikä on tärkeää salkkuarvioinnissa (käytännön kannalta)? mitä opettajan on tehtävä, että salkkuarviointi onnistuu?

- miten salkkutyöt auttavat arvioinnissa?

- mitä lisäarvoa salkkuarviointi tuo englannin /ruotsin kielen arviointiin?
- salkkuarvioinnin hyvät ja huonot puolet

- Haastateltavan taustatiedot
  - ikä, kokemus opettajana, opetettavat aineet, koulun koko

Appendix 2: The interview questions for pupils

Oppilaat

- Taustatiedot
  - ikä, missä kielissä salkutellut, kuinka paljon (kenties jo alakoulusta asti?)
  - millaisia kielisalkkuja olet tehnyt
  - millaisia eri arviointikeinoja opettajasi ovat käyttäneet?
  - mistä asioista kielitaito koostuu? (mitä pitää osata, että osaa hyvin jotakin kieltä?)

- Salkkuarviointi (kokemukset ja käsitykset)
  - miten ymmärrät salkkuarvioinnin
  - minkä verran sinulla on kokemusta salkkuarvioinnista
  - mitä mieltä olet salkkuarvioinnista; onko se hyvä kielitaidon arvioinnissa?
  - miten salkkuarviointi eroaa muista arviointimenetelmistä
  - koetko, että salkkuarviointi antaa realistisen kuvan taidoistasi

- Oppilaan rooli
  - miten voit itse vaikuttaa arviointiin?
    - saatko vaikuttaa tarpeeksi?
    - millaista kehitystä olet huomannut omassa kielitaidossasi?
    - millaista palautetta olet saanut opettajalta kielitaidostasi? missä muodossa?
  - ovatko taitotasoasteikot tuttuja?
    - oletko käyttänyt taitotasoasteikkoja?
    - miltä tuntuu arvioida omaa kielitaitoa asteikon avulla?
      - onko oman kielitaidon arvioiminen helppoa? (helpottavatko kuvaukset omien vahvuuksien ja heikkouksien löytämisessä)
      - oletko täyttänyt itsearvioinnin tarkistuslistoihin, missä merkitään omaa osaamista? miten tarkistuslistat ovat auttaneet näkemään omaa oppimistasi?
  - miltä tuntuu arvioida luokkakavereidesi töitä?

- mitkä ovat salkkuarvioinnin hyvät ja huonot puolet