

---

**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**

**Raili Hildén  
University of Helsinki**

**Abstract**

The purpose of this study is to explore the validation of data derived from the European Language Portfolio (ELP) in the study of English. This will be done by examining a) the relationships between different internal components of the ELP (the amount and the external appearance of the ELP, the amount and level of student comments and cultural experience displayed by the ELP) , and b) the relationships between these internal components and the school grade of English, and a number of factors included in a standardized instrument with several scales addressing the construct of learning-to-learn.

The data was derived from the implementation phase of the pilot of the Finnish version of the ELP in 2001–2004. ELP data of 33 15-year-old students were investigated. Ratings made by an external assessor on the quality dimensions of the ELPs displayed the highest correlations for most of the comparisons, while students' self-assessments did not show many significant correlations with the other variables tested.

The overall image that emerges from the Finnish ELP pilot is that individually oriented in-school-work and written language are prominent. Therefore, a broader variation with regard to both practical implementation as well as to the research methodology would be warranted.

**Key words:** language education, language teaching, language curriculum, authentic assessment, European Language Portfolio, validation

## 1 Introduction

### 1.1 Alternate forms of assessment in language education

In recent decades the concept of assessment has broadened beyond the predominant psychometric approaches to testing and measurement, and multiple alternative approaches to judging performance have been promoted. These are labeled with differing terminology depending on the focus and priorities of the proponents with various professional backgrounds including diagnostic testing (Alderson, 2005), alternative assessment (Lynch, 2003, pp. 5 - 6), authentic assessment (O'Malley & Valdez Pierce, 1996), classroom assessment (Rea-Dickins, 2001; Anderson, 2003; Airasian, 2005), educational assessment (Nitko, 1995), formative assessment (Torrance & Pryor, 2002), cognitive diagnostic assessment (Leighton & Gierl, 2007), informal assessment (Rea-Dickins, 2001), teacher assessment (Edelenbos & Kubanek-German, 2004) and, dynamic assessment (Lantolf & Thorne, 2006). Moreover, there has been a considerable growth in interest in teacher assessment and related classroom practices (for an extended account see e.g. Leung, 2005).

Despite the differences in terminology and scope, most alternative conceptions of assessment share a set of leading principles:

- a tendency of viewing language ability as a local, context-bound construct, created in socially determined activity (rather than a cognitive property of an individual in isolation) with an emphasis on an ability - of language users - in context (Chalhoub-Deville, 2003)
- alignment with regular in-class work of the teacher and the students in terms of time and effort (Huerta-Marcías, 1995 as cited in Leung, 2005; Brown 2004)
- establishing the interface between teaching and assessment with regard to instructional remediation (Bachman 1990, p.60; Alderson 2005, pp. 10 - 12)
- closeness to real-life tasks and a possibility to include them into assessment (García & Pearson, 1994, pp. 357-358 as cited in Leung, 2005)

- 
- personal contact and discussion between teacher and student, and more balanced decision making based on multiple sources across the full range of learning activities ( Rea-Dickins, 2001, p. 434)
  - capacity of displaying multiple modes of student performance including affective domain and learning experiences (Broadfoot, 2005, pp. 138 - 139)
  - promotion of students' self-evaluation (Oscarson, 1997; Dragemark-Oscarson, 2009)
  - highlighting learning process as an equal counterpart to product of learning
  - theoretical assumptions are based on constructivist and socio-cultural views of learning emphasizing the role of the student and implications of the context in shaping performance (Shepard, 2000)
  - potential for empowering students through making them aware of their learning potential (Little, 2004,) and promoting their agency in socio-culturally mediated activities (Lantolf & Thorne, 2006)
  - qualitative data and interpretative methodology of inquiry are given equal status with, or even prominence over, numerical scores and quantitative methodology (Lynch 2003, pp. 134 - 147; Davies 2003, pp. 361 - 362).

## **1.2 Validity considerations of portfolios**

A widely adopted format of alternative assessment is a portfolio, a collection of representative work samples, derived from performances in regular classroom work, projects and real-life tasks, complemented with a reflective account for the process itself, and multiple evaluative procedures prioritizing collaborative approaches and self-evaluation (Lynch 2003, pp. 119 - 122). The idea and use of portfolios spread from artistic and technical domains to education in late 1980's, first in America. In the European context, the portfolio has been strongly promoted by the Council of Europe since mid 1990's in the form of European Language Portfolio (henceforth the ELP) (Principles and guidelines, 2000). The ELP was developed during the entire 1990s, in parallel with the Common European Framework of Reference for Languages (CEFR, 2001).

Portfolios may be seen to address somewhat differently the various validity requirements. The very nature of portfolio samples cover multiple perspectives of language use situations and contexts, its overall construct representation is usually seen to be a strength. However, such mixed samples may easily run a risk of introducing sources of construct irrelevant variance (Messick, 1994) when inferences about language proficiency build on portfolio data. In view of traditional assessment theory, portfolios readily meet some aspects of ecological and face validities, because they tend to show what people really do with language. Portfolio assessments might also be strong in terms of relevance and utility dimensions (Messick, 1994) in displaying the dimensions of use of the construct. On the other hand, the current interpretation of construct validity as an overarching concept is no longer easily tackled with portfolio. The same vagueness shadows consequential validity implications of portfolio assessments (Messick, 1989). Value based conclusions and social consequences are highly context dependent, and therefore issues of uncertainty prevail, whenever portfolio samples are proposed as evidence crossing diverse educational settings.

The traditional validity facets were further rearranged and adapted for language testing by Bachman (1990) and Bachman & Palmer (1996) with slight terminological and conceptual modifications. Bachman and Palmer introduce the concept of test usefulness that comprises construct validity, reliability, interactiveness, authenticity, impact and practicality. Portfolios have promise in view of authenticity and interactiveness, provided there is a well-informed compilation of meaningful data. On the other hand, portfolios give rise to some doubts concerning practicality, especially as a part of regular school instruction, due to time constraints during the process and afterwards at the assessment phase of the entries (Brown, 2004). The impact of portfolios and the degree of context-dependence also requires verification.

There are some interesting new conceptualizations of validity from a unitary model towards a more relative view. Weir (2005) prefers to speak about validation instead of validity, consonant with his view to establish validity by empirical conduct of validation to ensure fairness. Efforts

made a priori to compiling a portfolio should ideally cover areas of content and theory-based dimensions of validity. Context-bound validity considerations include defining demands in terms of linguistic and non-linguistic task demands, while theory-based validity factors refer to internal processes of task execution. In educational context a priori validity definitions are often dictated by the curriculum framework within which the portfolio is practised.

A posteriori forms of validation include exploring scoring validity, criterion related validity and consequential validity (Weir, 2005). Scoring validity roughly corresponds to the concept of reliability (Weir, 2005, p. 43), traditionally regarded as a necessary, but not sufficient condition of validity. This is the domain where portfolio is at its most vulnerable with regard to the validity of inferences. Certain problems of reliability seem to pertain to alternative forms of assessment, and portfolio is therefore often rejected as a primary source of evidence in large-scale settings despite its indisputable benefits in other domains (Gredler, 1995; Shapley & Bush, 1999; Gomez, 1999; Klein, 1995 among others). Low reliability makes comparisons difficult (Gomez, 1999).

A number of studies also report drawbacks of portfolio assessment related to their practicality and feasibility in terms of cost (Gomez, 1999), time resources for the assessment process proper (Klein, 1995), as well as in-service training needed to guarantee a defensible level of uniformity of interpretation (Janisch, Liu & Akrofi, 2007). Nevertheless, there is evidence to support teachers' ability to assess their own students' work reliably in local settings (Supovitz, MacGowan & Slattery, 1997), and there is evidence for increased levels of reliability across large numbers of students in statewide settings, especially with efficient scoring systems and well trained scorers (Koretz, 1992).

In the last decade the notion of validity, and the subsequent role of reliability as its subcomponent, has been placed in the framework of the larger context of test and testing practices underlying its fairness and overall ethicality of their consequences (McNamara 2000, pp. 72 - 77). The focus on the social character of language assessment has its roots in consequential validity (Messick, 1989), later on elaborated by proponents

of socio-cultural approaches to educational research (Kunnan, 2005; Moss, Girard & Haniford, 2006, pp. 137 - 144), and critical language testing (Shohamy, 2001). We have not seen much work of the study of validity in this line yet, but portfolio unquestionably has great potential to respond to a range of challenges of social and cultural diversity in assessment.

Recently, proposals have also been made as to compromised reliability and accustomed interpretations of generalizability claims in favor of contextualized manifestations of assessment practices initiated by Moss (1994). The present author feels, however, that we should not confuse the fundamental quality requirements of assessment as defined in testing and measurement theory. It goes without saying that for instance formative and summative assessments are of different rigor in terms of replicability, accountability and potential of generalizing across situations and individuals. Especially a tool like portfolio with a dual function of showing the process and reporting its outcomes deserves equally careful consideration of both sides of the coin: the quality of the instrument and the consequences of its implementation. Even in the postmodern paradigm, a certain level of demonstrated reliability for drawing conclusions and/or making decisions for high-stake purposes remains an inescapable condition of fairness.

Traditional individual-based validity models have justifiably been criticized by socio-culturally oriented research. However, as long as schools assign grades to individuals and as long as tertiary institutions and work places cannot accept every applicant, we are faced with the selective function of assessment. If the portfolios are not an integral part of assessment evidence, their role in teaching and learning will tend to be marginalized reflecting the logic of washback effect (e.g. Alderson & Wall, 1993). Thus, if we wish to truly promote portfolios, it is advisable to take them into account when grading students. This is, in fact, an inherent part of the reporting function of the portfolio (Principles and guidelines, 2000; Kohonen, 2008). Grades, again, are used for the more or less high-stake evaluative purposes of selection and placement, and are consequently subject to the traditional, more rigorous demands of reliability, generalizability and accountability than is the case with formative types of classroom assessment. Therefore, quality standards pertaining to any

numeric assessment artifacts, must be applied to portfolio likewise. This line of reasoning from evidence (Mislevy, Steinberg & Almond, 2003) does not remove any of the obvious merits of portfolio assessment, including the ELP, but it is a step further on the way towards higher levels of awareness, more consistent interpretation and understanding, and fairness, at the level of individuals and at the level of society.

## 2 Context of the study and research questions

### **2.1 The role of assessment of language education in the context of compulsory education in Finland**

In the Finnish educational context, language teaching and learning are regulated by national core curricula, influenced by a number of, social, political, theoretical and practical factors. Both general basic (and secondary education) are free to all citizens, and school grades are mandated by the society at large as indicators of socially important attainments.

The general aims set for school education in Finland address a range of components: language skills, cultural skills and learning strategies in addition to the additional overall, cross-curricular goals of school education. The levels targeted in language skills at the end of each language syllabus are indicated in terms of a language proficiency scale based on the Common European Framework (Hildén & Takala, 2005, 2007; NCC = National core curriculum for basic education, 2004 ). The goal descriptions for cultural and strategic skills are far more loosely formulated. The cross-curricular goals state explicitly that basic education should

- provide an opportunity for diversified growth and learning
- enhance the development of a healthy sense of self-esteem and capability of further study for becoming involved citizens of a democratic society
- support each pupil's linguistic and cultural identity and the development of his or her mother tongue.

- 
- awaken a desire for lifelong learning
  - transfer cultural tradition from one generation to the next
  - assume the mission to create new culture
  - develop the pupil's ability of critical evaluation.
- (NCC, 2004, p. 8).

The learning-to-learn related goals have been addressed by Hautamäki et al. (2002) dividing them into categories, the following of which will be addressed by this study: means-ends-beliefs, agency beliefs, learning strategies, self-concept and academic self-esteem, academic self-concept, learning environment, significant others' perceived attitudes to school, use of information sources, group-work behaviour, future orientation and sociomoral self-concept as a student.

The subject specific goals of the study of English deal with language proficiency, strategic skills and cultural skills. Language proficiency is defined in terms of the CEFR scales. Strategic skills imply communication strategies and learning strategies defined as skills to take responsibility for various phases of one's work and capacity of reflective action including self-assessment. The cultural goals imply willingness and capacity to transcend the borders of one's own culture (Kaikkonen, 2005). The strategic skills fall on the border line between overall and subject specific goals. Certain qualities also emerging in the ELPs mirror the overall learning-to-learn goals (persistence in compiling the samples of work and finalizing them, as well as in giving comments and reasoned self-assessments).

The goal components are aggregated into a single school grade awarded twice a year at the upper level of compulsory basic education (for students aged 13 – 15 years). The grade should reflect and conform to the definitions of average "Good mastery" intended to guide teachers' grading practices and to make the grading more transparent and consistent across the country. There are, however, no agreed principles of how to weight the distinct components of subject specific and cross-curricular goals. Given the practical circumstances of a subject-driven curriculum, the grade assigned by a subject teacher is mainly based on



knowledge and effort displayed during the lessons of the particular subject, and to a certain extent on attitude and motivation.

The ELP implementation project was carried out in 2001 - 2004 as a recursive cycle of collaborative research and development work by a number of university representatives and practising teachers. The Principles and guidelines (2000) introduced by the Council of Europe were applied in this action research project, and they were promoted by intensive mentoring and in-service training. The form of the ELP as a document consisting of three parts was considered as the final target outcome, but there were no attempts to standardize the procedures of use or assessment during the process apart from informal exchange of ideas and practices introduced at project meetings.

The outcomes and scientific impact of ELP trials have been reported as academic papers (Kohonen & Pajukanta, 2000; Hildén, 2002, 2004) and as articles in professional magazines. The writer was in charge of mentoring the Helsinki metropolitan area, where we, in line with the overall agenda for implementing the Finnish ELP, sought to combine the pedagogical innovation with teachers' professional development showing how the innovation was linked to their regular work, current attitudes and beliefs. We also sought to empower the teachers to commit themselves to acting as key agents in their own everyday working contexts (as recommended by e.g. Karavas-Doukas, 1998, pp. 31- 39). By the same token, from the student point of view, the core of portfolio work is to produce a selection of language samples and to provide comments on how they came into being. The portfolio process should also foster increased student awareness and the capability of realistic self-assessment as indications of growth toward autonomous agency in learning. This aim corresponds to the calls for more constructive approaches to bridge learning and assessment (see for instance Broadfoot, 2005).

In accordance with its international counterparts, the Finnish ELP serves the twin goal of language curricula: *the reporting function* is concerned with the product aspect, while *the pedagogic process function* deals with helping students to identify their own goals, to modify their action plans and to design their work and assess its outcomes (Kohonen & Korhonen,

2007, p. 239). In this study, the individual variation of portfolio data is seen as an asset illustrating a student's personal strengths and performance in greater detail and depth. At the end of the three-year educational cycle, however, there should be sound principles for informed consensus to guide how to integrate the true potential of portfolio work into the school grade in a reliable manner which would warrant and promote its large-scale use (Takala 2002, pp. 123-127).

## **2.2 Research questions**

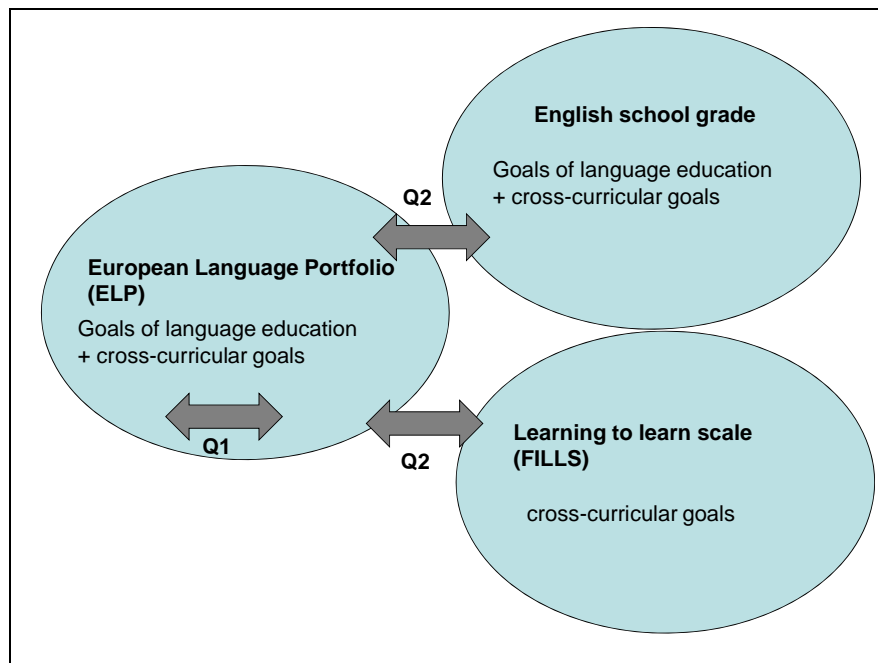
The purpose of this study is to address both the reliability and validity of inferences drawn on portfolio data, although these two are not strictly separable. Reliability is primarily addressed by investigating the agreement between ratings given by various actors involved (the student him/herself, the teacher, the external assessor), while broader conceptions of validity address the consistency of the data and the emerging correlations between a set of relevant variables, including newly constructed ones. This procedure provides insight into the content and construct facets of validity (Davies, Elder, Hill, Lumley and McNamara, 1999, pp. 31, 34). It also corresponds to the generalizations step in Kane's chain of interpretive argumentation related to score validity (Kane, Crooks & Cohen, 1999). Secondly, this study seeks to clarify issues of the concurrent validity (Davies et al., 1999, p. 30) of portfolios by taking a well established learning to learn scale FILLS (Hautamäki et al., 2002) as a criterion. The second set of criterion measures is attained from an external assessor who looks at certain qualities displayed in the portfolios. Finally, all the dimensions probed are correlated with the final school grade in English. Such information on concurrent validity of the ELP-based inferences will inform us about the construct of language education in action.

The main research questions were:

1. What are the internal relationships among the goal components of school curricula (linguistic and some cross-curricular goals) in the ELP?

- 1.1 What relationships are found across the goal components of the language education national core curriculum as displayed by ELPs (self-assessments of the four linguistic skills and the linguistic level of the portfolio)?
  - 1.2 What relationships are found across the cross-curricular goals displayed by the ELP (the amount and the external appearance of the ELP, the amount and level of student comments and cultural experience displayed by the ELP)?
  - 1.3 What relationships are found between the linguistic and cross-curricular goal components included in the ELP?
2. What relationships are found between components of language education and cross-curricular goals documented in the ELP and a range of external variables mirroring the same dimensions?
    - 2.1 What relationships are found between the variables displayed by the ELP data and the final school grade assigned at the end of compulsory basic education?
      - 2.1 What relationships are found between the variables displayed by the ELP data and the components of learning-to-learn ability measured by a criterion scale /FILLS)?

**Figure 1. Overview of the research questions addressing the internal and external validity of the ELP**



In traditional validity terminology, the first question primarily addresses internal consistency, while the second deals with the external type of validity.

Due to the limited number of students and their portfolios, no statistical hypotheses are presented. It is, however, reasonable to anticipate that the portfolio-based data will be positively related to the criterion measures referred to above as they all are supposed to reflect the same curricular framework.

### 3 Method

This study builds on “From the Piloting to Implementation” stage of the Finnish ELP in 2001–2004. The data is derived from the final measurements in autumn 2003 and spring 2004. The students came from the upper grades of two comprehensive schools in the Helsinki metropolitan area. Their school-related beliefs were measured with the Finnish Learning-to-Learn Scales (FILLS) instrument developed and validated at the University of Helsinki (Hautamäki et al., 2002).

The first measurement of the pupils (n=63) was carried out at the end of 2003, towards the end of their compulsory school education with a control group from the same schools. The results revealed a tendency for greater variation in beliefs between schools than between classes at the same school, irrespective of participation in ELP work (Vainikainen & Hautamäki, 2003). The study at hand focuses on the experimental group of 33 ELP students from the two schools, consisting of 13 boys and 20 girls. Due to the small size of the sample, the results only provide some preliminary suggestions to point the way toward more systematic approaches in ELP research.

**Table 1. Number and gender of ELP students**

Participants	School A	School B	Total
ELP students total	20	13	33
ELP boys	5	8	13
ELP girls	15	5	20

At the end of the experimentation period, in the spring of 2004, the students were asked to fill in the Finnish ELP version for comprehensive and upper secondary schools. The three sections proposed by the Council of Europe were included in the Finnish ELP version:

1. A Language Passport including the student’s self-assessment of his/her language proficiency using the ELP self-assessment grid by either ticking off the alternatives or by ticking off the

appropriate options (I can/ I can do with assistance/ I cannot yet).

2. A Language Biography where students were invited to tell about and reflect upon their language-related experiences, views on other cultures and language study.

3. A Dossier, a selection of written language samples across the three-year trial period along with comments.

Table 2 indicates how the target components of language education relate to respective theoretical concepts and the sources of evidence adopted for this study.

**Table 2. Concepts to be measured, tested variables and sources of evidence**

Concept to be measured	Variable Value range (Source of evidence)	n
1. Goal components of language education		
1.1 Language proficiency		
Proficiency in the English language	School grade in English (reported by the teacher)	33
Language proficiency assessed by external rater	Level of written English proficiency in the ELP samples 1-3 (Dossier)	33
1.2 Strategic skills		
Self-assessment skills	Self-assessed listening comprehension skill 1-6 (Language Passport)	20
Self-assessment skills	Self-assessed oral interaction skill 1-6 (Language Passport)	20
Self-assessment skills	Self-assessed reading comprehension skill 1-6 (Language Passport)	20
Self-assessment skills	Self-assessed oral production skill 1-6 (Language Passport)	20

Self-assessment skills	Self-assessed written production skill 1-6 (Language Passport)	20
ELP variables		
2. Cross-curricular goals		
2.1 Cultural skills	Cultural experiences: Range of cultural experience 1 - 3 assessed by external rater (Language Biography)	20
2.2 ELP related qualities		
Persistence	ELP external appearance 1-3 assessed by external rater (the entire ELP document)	33
Purposiveness, carefulness	Amount of ELP materials 1-3 assessed by external rater (the entire ELP document)	33
Amount of self-reflection	Amount of student comments 1-3 assessed by external rater (the entire ELP document)	20
Level of self-reflection	Level of student comments 1-3 assessed by external rater (the entire ELP document)	20

For the learning-to-learn variables see appendix.

The grade for English proficiency was given by the teacher following the regular practice of setting school grades in Finland on a 4-10 scale. This grade encompasses both performance and effort in and out of the classroom. The teacher did not give a grade on the ELP work itself and the weight of the ELP as a part of the grade was not explicitly defined, although it was an established part of the language study. Self-assessed level ratings were made according to the six-step Common European Framework scale (CEFR, 2001). A three-point scale (1-3) was used indicate external ratings for language proficiency. This scale links to the target level definition of good mastery in the national core language curricula approximately as follows: 1= CEFR levels A1-A2.1; 2 = CEFR levels A2.2-B1.1; 3 = CEFR levels B1.2-B2. Aggregated level indicators were used to capture the variety displayed by language samples across

three years of study. In problematic cases, however, samples deriving from the last year of compulsory education were given certain emphasis as the most updated evidence of attained proficiency.

The amount of ELP materials was determined on the basis of the number and range of language samples and the comments attached to them: a portfolio with 1–5 samples was judged as level 1; one with 5–15 samples was judged as level 2; one with more than 15 as level 3.

To assess the external appearance of the ELPs, another three-point scale was applied. At the lowest level (graded 1), the document gave an impression of unfinished and careless work (e.g. the text was hard to read, careless handwriting was used, loose sheets were dropping out of the folder, earlier versions of written samples were marked by the teacher with an unfulfilled request to correct the mistakes, some assignments were missing). An average ELP (graded as 2) reflects some effort, but among the required pieces of work were some obviously unfinished items. Portfolios rated 3 were neatly written, often by text software, and supplied with artistic illustrations or pictures showing a sense of aesthetics, which was not, of course, presumed as a prior condition.

Study-related comments were also classified on a 1–3 scale. The lowest level was assigned when the student had only ticked off a set of given options on a sheet without any comments or had written very few statements, single words or very short phrases. At level 2, the comments were usually complete sentences for all requested items. Level 3 entries included long sentences of several lines for all points required.

The conceptual level of reflection revealed by the students' comments was assessed using the SOLO taxonomy (Biggs & Collins, 1982) modified by Huttunen (1996) to capture the progression from the lowest mechanical level through the pragmatic stage to an emancipatory stage of the reflective ability to regulate one's own actions. Mechanical reflections manifest themselves in lists of practical details, such as task titles and technical details, as well as reporting sequences of work in a then-and-next mode, without stating anything about the meaning of these activities to the student's own learning. Level 2 reflections give some personal



statement of the learning experience (“This task was nice”.) but with no indication ability to draw conclusions or modify future learning plans on the basis of a particular experience. Such indicators are inherent characteristics of a level 3 reflection.

An additional three-point variable was tried out to describe the manifestations of *cultural experiences* in the ELP data, covering the duration, nature and amount of cultural encounters reported by the student in the Language Biography section of the ELP. Level 3 was assigned to descriptions such as the following:

*“I’ve been exchanging mails with foreigners through my junior high school years, and I have traveled a lot in many countries since I was very young. I participated in a school trip to Catalonia and used a lot of English there.”*

A level 2 description of cultural encounters is attained by the following:

*“I have watched a lot of movies in English, played computer games in English, and I write and read internet discussion fora in English on a regular basis. On our excursion to Catalonia we stayed with a local family for a week, got to know the family, local attractions and Spanish culture.”*

In Level 1 records no face-to-face encounters with foreigners were reported, or the encounters were of a very sporadic nature (e.g. a tourist asking the way). Culturally oriented tasks set by the teacher were carried out (e.g. seeing a film for the next lesson), but no private initiatives were accounted for.

The variables and the categorizations are obviously tentative and need a considerable effort to become useful and well-functioning tools for ELP research, as this is the first attempt in this strand in the ELP research carried out in Finland. So far, ELPs have been approached by researchers with an almost exclusively qualitative orientation. If the ELP is to be extensively disseminated nationwide and continent-wide, we need to bear in mind the same issues of validity, accountability and fairness that constrain any other form of assessment.

#### 4 Results and discussion

This section deals with the two major research questions that address (1) the relationships among the various dimensions of language education and cross-curricular goals as displayed by portfolio data, and (2) the relationships between these goals and certain learning to learn-related beliefs (as measured with FILLS), and the final school grade. Tables 3 and 4 list all the piloted ELP variables, but from learning-to-learn variables only those are included that showed at least two significant correlations with ELP variables. By the same token, only significant correlations are discussed in the text.

**Table 3: Spearman correlations between ELP variables and student beliefs (n=17) mirroring learning-to-learn capacity (Kendall's tau-b  
Approx. Sig. < .055 if the value is not given)**

	ELP amount	ELP External appearance	Amount of student comments	Cognitive level of student comments	Cultural experience	Self-assessed listening comprehension	self-assessed reading comprehension	Self-assessed oral interaction	Self-assessed oral production	Self-assessed writing	ELP linguistic level
Goal orientation (GO): Learning								-.549*			
GO: Achievem.											
GO: Avoidance					-.769**						
Means: Ability				-.473*		.516*					
Means: Chance	-.521**			-.526*							-.390*
Agency beliefs: Control motivat.								-.494*			
Academic self-concept (ASC) as a reader	.433*		.504*								.380* (tau-b sig..065)
ASC as speaker	.423*	.422*					.527*		.520*		
ASC as a writer	.443*		.571*								.380*
Parents' perceived attitude								-.568*			
Teachers' perceived attitudes								-.643**			.380*
Peers' perceived attitudes								-.653**			
Self-concept								-.597*			

and academic self-esteem											
Use of information sources (UIS): computer	.448*	.453*						-.690** (tau-b sig..102)			.426*
Group-work behaviour (GWB): Domination	.379*										
GWB: Non-participation				-.567*							
GWB: Task orientation	-.431*	-.401*		-.517*							
Future orient: Disdain towards schooling											.503**
Sociomoral self-concept (SMC) as a good student					-.517*						
SMC as indiff. student	.449*		.510*					-.599*			
SMC as a self-assured student					-.568*						

---

Research question 1 addressed the inner consistency on the ELP data mapping the internal relationships among multiple goal components. Relationships across the goal components of language education (targeted by sub-question 1.1) as stated in the national core curricula and displayed by ELPs (self-assessments of the linguistic skills and the linguistic level of the portfolio) were examined by using Spearman's rho correlation (see Table 3) and Kendall's tau-b. Among variables reflecting linguistic level of the ELP (assessed by an external rater) and student self-assessments on their listening, speaking, reading and writing skills the only significant relationship was between self-assessment of listening comprehension and oral production (Kendall's tau-b Approx. Sig. .042). Furthermore, a nearly significant correlation was found between the linguistic level of the ELP and self-assessed oral interaction by Kendall's tau-b (.055). These two findings were not, however, confirmed by the Spearman's rho correlation.

The second sub-question (1.2) dealt with the cross-curricular goals displayed by the ELP: the amount and the external appearance of the ELP, the amount and level of student comments and cultural experience displayed by the ELP. The amount of ELP data displayed a range of significant correlations with several variables related to cross-curricular goals. It was related to the external appearance of the portfolio (Kendall's tau-b Approx. Sig. .029, not confirmed as significant by Spearman's rho), the amount of student comments (Kendall's tau-b Approx. Sig. .001) and their cognitive level (Kendall's tau-b Approx. Sig. .021). Furthermore, the amount of student comments was associated with their cognitive level (Kendall's tau-b Approx. Sig. .001) and with the external appearance of the ELP (Kendall's tau-b Approx. Sig. .047, not confirmed as significant by Spearman rho). Cultural experiences did not display any significant relationships with any other variable included in the ELPs.

The third subquestion addressed the relationships between the two sets of variables, linguistic and cross-curricular goal indicators, included into the ELP. The self-ratings given by the students did not display any significant relationships with indicators of cross-curricular goals. However, the linguistic level of the ELP assessed by an external rater was related to four variables indexing cross-curricular goals: the amount of the ELP data (Kendall's tau-b Approx. Sig. .000), the amount (Kendall's tau-b Approx.

---

Sig. .000) and cognitive level (Kendall's tau-b Approx. Sig. .001) of student comments, and external appearance of the ELP (Kendall's tau-b Approx. Sig. .027). Spearman's rho correlation coefficient confirmed all other relationships as significant or highly significant, except for the external appearance. The finding suggests a more general factor encompassing the joint presence of persistence of effort and cognitive-linguistic abilities in everyday school work.

The connection between the linguistic level of the ELP and its scope and cognitive level also suggests that highly reflective students seem to have attained a higher level of written production in English than those whose reflections are meager and more superficial. This finding also suggests that the higher level of written proficiency was probably attained by diligent production and delivery of the ELP according to teacher instructions. In substance, the requirements of ELP work seem to narrow down the activity to a conscious effort of rather traditional language study at school, especially with an emphasis on the written mode.

Research question 2 addressed the relations emerging between components of language education and cross-curricular goals documented in the ELP and a range of external variables mirroring the same dimensions. The first sub-question (2.1) was about the variables displayed by the ELP data and the final school grade assigned at the end of compulsory basic education. The grade of English given by teachers showed the highest number of significant relationships with variables included in the ELP. The Spearman's rho coefficients were the following:

- the amount of the ELP .585\*\*
- the linguistic level of the ELP .637\*\*
- external appearance of the ELP .405\*
- amount of student comments .730\*\*
- cognitive level of student comments .774\*\*

These were confirmed by Kendall's tau-b, which also revealed one more almost significant relationship between the grade and student self-assessment of oral production (Kendall's tau-b Approx. Sig. .047).

---

The findings thus point towards relationships between the grade and the linguistic level of the ELP given by an external rater. The grade turned out to have significant relationships with the amount of the ELP (Kendall's tau-b Approx. Sig. .000) and its external appearance of (Kendall's tau-b Approx. Sig. .011) as well with the number of student comments (Kendall's tau-b Approx. Sig. .000) and their cognitive level (Kendall's tau-b Approx. Sig. .000). All these imply agreement between qualities of the ELP and the content of the English school grade, even if the role of the ELP was not determined numerically.

The reflective level of the comments especially tends to go together with the grade. This indicates that the students possessing higher school grades in English provided more sophisticated reflections on their learning and work samples.

Examples of responses in the Language Biography sections of the ELP to the item:

"I as a partner in pair work and group work"

Student 121:

*"I wouldn't like to correct others' mistakes, if they couldn't care less themselves." "Working with a good group, I am an active contributor. In pair work I always do my share."*

(girl, English grade 10/10, overall level of reflections 3/3 = emancipatory)

Student 122:

*"I'm lazy."*

(boy, English grade 8/10, overall level of reflections 1/3 = technical)

Research question 2.2 was focused on the relationships between the variables displayed by the ELP data and the components of learning-to-learn ability measured by a criterion scale for learning to learn abilities (FILLS). The dimensions of this scale covered by this study were goal orientations, means to achieve goals, agency beliefs, self-concepts, attitudes of significant others, certain dimensions of social and individual self-concept, group behavior and use of sources of information. Table 3 shows the significant correlations found in this study. Altogether 33 out of

---

220 possible correlations (15 percent) were significant. These are discussed below.

Among linguistic variables included in the ELP, a range of significant correlations were detected. The self-assessed listening comprehension skill was significantly related to the tendency to count on ability as a means of success. The self-assessed reading comprehension showed a significant positive relationship with academic self-concept as a speaker. The highest number of correlations among linguistic ELP-based variables were attached to self-assessed oral interaction, in other words interactional speaking. Most of the connections were negative and give a rather unexpected overall impression. This variable was negatively related to achievement orientation towards school work, as well as to control motivation. There was also highly significant negative correlations with the attitudes of significant others (parents, teachers and peers), self-concept and academic self-esteem as well as with use of computers. These findings imply that the higher the student rates his/her oral interaction skill the more achievement oriented he/she is at the cost of real learning, the less he/she wishes to be controlled at school work, the less support he/she gets from his/her parents, teachers and peers, the lower his/her self-concept as a student and the less he/she uses computers. There was also a negative correlation indicating that the higher the self-assessment on oral interaction, the less the student regards him/herself as indifferent towards school and studies. These results indicate that the goals of the ELP were not fully met despite the fact that the students had been familiarized with several procedures of self-assessment during the three-year trial. They had frequently been requested to provide free on-task comments or to fill in guided forms and tick off lists of CEFR-based can-do statements.

Self-assessed oral production had one significant positive correlation with academic self-concept as a speaker, logically enough. There were no other significant relationships between self-assessed skills and FILLS dimensions. The linguistic level of the ELP rated by an external assessor was related to a variety of learning to learn features and these findings portray a more coherent model than the self-assessments. The linguistic level of the ELP was negatively related to the student's tendency to rely



on chance in pursuit of success, which makes a lot of sense since language learning for most people is a matter of hard work rather than lucky lottery. This academic essence of formal learning may also explain the positive correlations between the level of writing emerging in the ELP and the academic self-concept of a student as a reader and writer. The positive relationship with teacher attitudes perceived by students belong to the same realm of qualities producing linguistic mastery. The ELP linguistic level also correlated with use of computers, which appears natural because most of the investigated ELPs had been produced with computers. The positive correlation with a future orientation characterized by indifference towards school work was the only relationship that did not conform to the goals of the ELP and Finnish school education, both of which set as goals to enhance will and capacity for lifelong learning. The lack of interest in future study and school work may be temporal and due to the timing of the survey at the end of a nine-year compulsory education.

The second part of RQ2 addressed cross-curricular goals emerging in the portfolio data on one hand, and through the FILLS learning to learn inventory on the other. The amount of ELP data showed the highest negative correlation with counting on chance as a means of success rather than hard work, which agrees well with the character of the ELP work as an effort over a longer period of time. The correlations with the measured aspects of academic self-esteem (as a reader, speaker and writer) all display positive connections with the amount of samples filed in the student portfolio. The positive correlation with the use of computers is obvious as indicated by the physical appearance of the files. The correlations detected between group-work behavior and socio-moral concept were more challenging to interpret and seem to some extent to run counter to some of the stated goals of the ELP: the relationship with task oriented group-work behavior was negative, whereas the correlations with dominating group-work behavior and socio-moral self concept as an indifferent student were positive. These findings may well be due to the small sample size, but they may also indicate that at the point of time of the study the ELP was still viewed as an individual endeavor and graded as such (given the highly significant correlation of the amount of ELP data with the school grade). School grades are

awarded to individuals and the performance evidence they are based on is still largely given individually. Even if cooperativeness and social responsibility are strongly emphasized among cross-curricular goals, attaining these goals is secondary as long as valid and useful instruments to assess them are largely missing.

The external appearance of the ELP was positively related to the student's academic self-concept as a speaker and, again, with his/her use of computers. The correlation with task-oriented group-behavior was negative, confirming the individually oriented approach to producing the portfolio samples. In regard to the amount of student comments, positive correlations were found with academic self-concept as a reader and writer, which conform to the mode of production of the ELP. The relationship with socio-moral self concept as an indifferent student might point at situational factors referred to earlier (the fatigue of the final year of the compulsory education) or, probably, to a sample specific factor. In the latter case the students do not attach particular meaning to school as a whole even if they have accomplished the assignments for the ELP work required from them.

The cognitive level of student comments correlated negatively both with ability and chance as means for success at school. This finding implies that the less the student relies on these effortless sources of success, the higher the cognitive level of his/her comments, which is basically as expected. The negative relationships of the cognitive level of student comments concern non-participating group-work behavior and task-orientation in group-work. The conclusion from this finding could be that reflective students are also willing to participate in group-work and they also put more emphasis on human relations than only on conducting the task. Participation and social orientation in group-work might also promote reflective abilities.

Whatever the conclusions, the findings dealing with the qualities of the ELP (amount of data and student comments and their level) are based on a very small sample and brief scales, and can therefore only be indicative at best.

The variable mirroring the nature and amount of cultural encounters reported by the students in the Language Biography section of the ELP related significantly to avoidance orientation and socio-moral self-concepts as a good and a self-assured student. In terms of cultural encounters, the sample group was fairly homogenous, since all of the students had at least a level 2 experience of cultural encounters, such as trips abroad together with their schoolmates, and in addition, most students had traveled on holidays or stayed a longer time outside of Finland (level 3). All this has probably to do with the advantageous social profile of the school district and also with the high level of professional interest and activity of the English teacher. The reverse relationship with avoidance orientations indicates, logically enough, that more frequent experiences of cultural encounters may encourage school study, language learning included.

Table 3 further illustrates that there is certain congruence between the goals set in the language core curricula in terms of co-operative orientation and student ownership of learning, gained through planning and managing one's own study efforts. Contrary to the goals of successful self-assessment skills is the fact that student self-assessments of their language proficiency had the only positive correlation with the dimension of ignorance (Viewing self as an ignorant student). In this case, ignorant students seem at least to have ignored the self-assessment training.

Student comments related positively to a number of learning-to-learn variables, among others to perceptions of the significance of school, viewing oneself as a good student and self-concept as a writer. ELP work was an integral part of regular language study in the school context, and its relevance was emphasized from the very beginning of the trial period. The performance-oriented approach to ELP work favored students inclined to put a persistent effort into school work and to take responsibility for it, and who along with this inner orientation were willing to expose their reflections in writing. The diligence of the reflective-oriented students was confirmed by the negative relationship between student comments and their tendency to rely upon sheer ability to account for success at school. Those who recorded extensive and in-depth reflections relied less on sheer ability than others. Submitting

comments also required a certain familiarity with written production (in the mother tongue), which is probably the most natural for students with a strong self-conception as writers. Unfortunately only one of the schools (A) was able to provide comments with ELPs that were filled in appropriately. Thus the results rest on this highly limited evidence.

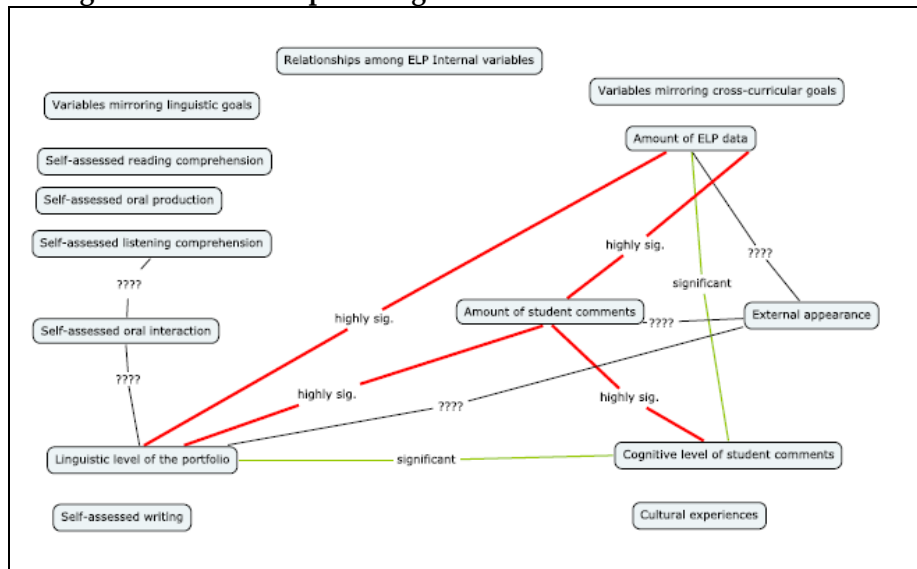
The connections discussed above were congruent both with the spirit of the ELP trial and the goals of school education. Responsible and rationally minded students may be more inclined to keep up with diligent portfolio work and to polish the external appearance of their ELP's. Since the ELP focused on written performance, it is natural that literacy-oriented students enjoy compiling this kind of sample more than less literacy-oriented classmates. Persistent practice also promotes language proficiency, and writing an extensive portfolio logically has the same effect. This might explain the positive connection between high-quality ELPs and a high level of written English. The abundance issue combines two vital, mutually related aspects of portfolio work: the contributions of a student and his teacher. If the teacher has provided rich task instructions and delivered a lot of material to support student reflection and self-assessment (checklists, forms, tables), there are more documents for the student to file and build his own records on. This probably explains the fact that the overall ELP records from school A were considerably more voluminous than the data from school B. The students with a high self-concept as readers were rated highest in terms of their written English proficiency. Cultural experiences, on the other hand, did not turn out to display any clear-cut pattern of relationships with the other variables.

In sum, language study and the content of the English grade among the ELP groups investigated appear to rest on persistent, goal-oriented work with a major orientation to written language. The process is affected by the group context and the attitudes of significant others. The focus on written language does not respond very well to the goal of curriculum, as all the linguistic skills are expected to be taken into account on an equal basis. Moreover, the level descriptions required for a good grade in English, the European levels A2-B1, emphasizes everyday communication frequently implying oral interaction and exchanges. The lack of spoken

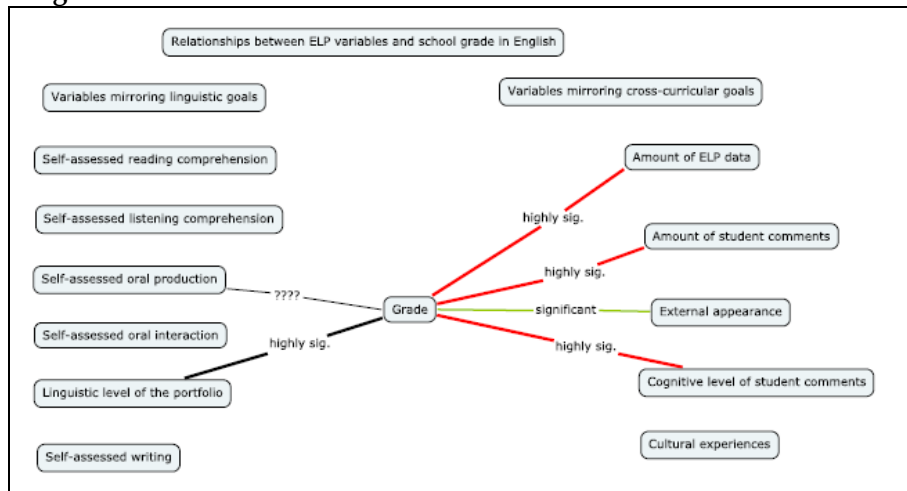
samples in the ELPs is probably due to a range of practical problems as well as lack of technical equipment or unfamiliarity in using them.

The results are illustrated and summarized in Figures 2-3 indicating the significant and nearly significant (labelled with question marks) correlations detected in this study.

**Figure 2: Relationships among ELP internal variables**



**Figure 3: Relationships between ELP variables and school grade in English**



## 5 Conclusions

The pilot study at hand was an initial attempt to broaden the scope of ELP research by exploring a set of pedagogical goals in relation to empirical ELP data, and thereby to explore concurrent validity issues of a portfolio implementation. The results reported here indicate that the various components of validity referred to in the introduction, regarding consistency across the curricular target components were confirmed by a number of expected relations among the variables. These are:

- Significant internal relationships (Figure 2) were detected in the Finnish ELP between subject-specific linguistic goals, on one hand, and a variety of cross-curricular goals of compulsory school education (diversified growth and learning), on the other.
- Among all the variables explored, external ratings on ELP qualities (amount of data, external appearance, amount and level of student comments) appeared to be the most useful indicator of inner consistency of the data (Figure 2). This outcome corroborates the internal consistency of the ELP.

- Strong correlations between with the overall school grade (Figure 3)add to the relevance of the ELP as a tool for attaining the curricular goals.
- Expected correlations between ELP variables and a range of variables mirroring cross-curricular goals of school education (learning-to-learn ability and desire for lifelong learning) contribute to the external validity of the ELP.

On the other hand, certain findings pose challenges in terms of stability and consistency (reliability) of the reporting function of the ELP:

- Contrary to expectations, student self-assessments of their language proficiency displayed low connections to other variables. Although the absence of significant relationships may be mainly due to the small sample size, it is advisable in the future to focus on the systematic training of students in self-regulation and the realistic monitoring of the development of their language skills.
- There is also an apparent need for further research to check for some puzzling outcomes of study. Are students inclined to overestimate their oral interaction skills in English? How widely accepted and implemented is the social orientation of portfolio work or other forms of language education? Is it possible to produce high-quality evidence for one's own performance with virtually no group orientation or even interest in school-work in general?
- In its current shape the variable suggested for describing cultural experience did not relate to other variables. Since these are increasingly voiced as part of the major goals of language curricula, there is an urgent need in the field of cultural skills to define the construct for school purposes in a consistent manner to enhance pedagogical adaptation and above all, fair forms of assessment. Regarding the methodology of the study at hand and future studies of the same strand, qualitative student comments on cultural encounters should be appropriately included in the analysis to complement numerical data.

As depicted by Figures 2-3, multiple ratings and replication studies are necessary to ensure reliability of the ELP, and thus enhance the validity of its use. Some of the scales constructed for this study also need to be refined and more accurately attuned with regard to ELP data. This is true e.g. for the scale and the categories piloted for the classification of student comments. As far as language proficiency is concerned, consistent use of the CEFR scales should be prioritized at each operational stage so as to guarantee national and international comparability.

As a whole, the study at hand pictures the construct of language education in the use of the Finnish ELP as being implemented in quite a traditional manner with an emphasis on written language and classroom work. The scope of ELP work could easily be broadened inasmuch as technically more advanced templates of ELP are introduced to record, for instance, spoken language and out-of-school activities in a more nuanced way, as was envisaged by the Council of Europe Principles and Guidelines (2000). The ELP seems to meet a variety of goals of language education, but it is highly desirable to assign it a more rigorously defined role as a part of the language learning process and in the school grade to be awarded.

In regard to research, further inquiry, including comparisons of the ELP across multiple educational settings, is clearly necessary for the ELP to reach the validity prerequisite for nationwide use. Even so, in approaching a holistic artifact such as the ELP by any research methodology, the results are highly susceptible to contamination from general school-bound variables unrelated to language teaching and learning as such. Despite the challenges recognized, the ELP remains a promising tool for developing autonomous language study in formal school settings, and it also lends itself to multiple assessment purposes, which is a definite advantage. Given these merits, the portfolio deserves to be promoted and refined in collaboration with multiple stakeholders to improve its overall validity and fairness.



---

References

- Airasian, P. W. (2005). *Classroom Assessment: Concepts and applications*. New York, NY: Mc Graw Hill.
- Alderson, J.C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.
- Alderson, J.C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14 (2), 115-129.
- Anderson, L. W. (2003). *Classroom assessment: Enhancing the quality of teacher decision making*. NJ: Lawrence Erlbaum Associates.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Biggs, J. and Collins, K. F. (1982). Evaluating the quality of learning: The SOLO taxonomy, structure of the observed learning outcome. London: Academic Press.
- Broadfoot, P. M. (2005). Dark alleys and blind bends: testing the language of learning. *Language Testing*, 22(2), 123 - 141.
- Brown, D. B. (2004). *Language assessment: Principles and classroom practices*. NY: Pearson Education.
- CEFR. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Council of Europe: Cambridge University Press.
- Chalhoub-Deville, M. (2003). L2 interaction: current perspectives and future trends. *Language Testing*, 20(4), 369 - 383.
- Davies, A. (2003). Three heresies of language testing research. *Language Testing* 20 (4), 355 - 368.
- Davies, A., Elder, E., Hill, K., Lumley, T., & McNamara, T. F. (1999). *Dictionary of language testing: Studies in language testing*, 7. Cambridge, UK: Cambridge University Press/ University of Cambridge Local Examinations Syndicate.
- Dragemark-Oscarson, A. (2009). *Self-assessment of writing in learning English as a foreign language*. Göteborgs universitet: Göteborg Studies in Educational Sciences 277.

- 
- Edelenbos, P., & Kubanek-German, A. (2004). Teacher assessment: the concept of 'diagnostic competence'. *Language Testing*, 21 (3), 259 - 283.
- García, E. G., & Pearson, P. D. (1994) Assessment and Diversity. *Review of Research in Education*, 20, 337-391.
- Gredler, M. E. (1995). Implications of Portfolio Assessment for Program Evaluation. *Studies in Educational Evaluation*, 21(4), 431-37.
- Gomez, E.L. (1999). [Creating Large-Scale Assessment Portfolios That Include English Language Learners: Perspectives on Policy and Practice](#). (ED446129) Retrieved 15.12.2008 from <http://www.eric.ed.gov>
- Hautamäki, J., Arinen, P., Eronen, S., Hautamäki, A., Kupiainen, S., Lindblom, B., Niemivirta, M., Pakaslahti, L., Rantanen, P. and Scheinin, P. (2002) *Assessing learning-to-learn: A framework*. Helsinki: Centre for Educational Assessment, Helsinki University /National Board of Education.
- Hildén, R. (2004). Opettajien näkemyksiä eurooppalaisen kielisalkun kokeilusta Helsingin alueella. [Teacher perceptions of European language portfolio in the Helsinki metropolitan area.] In J. Loima (Ed.), *Theoria et praxis*. Viikki Teacher Training School Publications, 1. Helsinki. Retrieved 16.11.2009 from [http://www.helsinki.fi/kasv/nokol/uusi2000/ajankohtaista/Theoria\\_et\\_praxis\\_edited\\_04.htm#\\_Toc63682351](http://www.helsinki.fi/kasv/nokol/uusi2000/ajankohtaista/Theoria_et_praxis_edited_04.htm#_Toc63682351)
- Hildén, R. (2002). ELP - Integrating theory and practices for socially responsible autonomy. In Kaikkonen, P. & Kohonen, V. (Eds.) *Quo vadis foreign language education? Reports from the Department of Teacher Education in Tampere University A27/2002* (pp. 107-121). Tampere: University of Tampere.
- Hildén, R. & Takala, S. (2005). Kielisalkulla kohti selkeää ja monipuolista arviointia. [ELP as a tool for transparent and multifaceted assessment.] In V. Kohonen (Ed.), *Eurooppalainen kielisalkku Suomessa. Tutkimus- ja kehittämistyön taustaa ja tuloksia* [The European Language Portfolio in Finland: Background and results of the research and development work] (pp. 7-44). Helsinki: WSOY.
- Hildén, R. & Takala, S. (2007). Relating Descriptors of the Finnish School Scale to the CEF Overall Scales for Communicative Activities. In

- 
- Koskensalo, A., Smeds, J., Kaikkonen, P. & Kohonen, V. (Eds.), *Foreign languages and multicultural perspectives in the European context; Fremdsprachen und multikulturelle Perspektiven im europäischen Kontext* (pp. 291-300). DICHTUNG, WAHRHEIT UND SPRACHE. Berlin: LIT-Verlag.
- Huerta-Macías, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal*, 5, 8-11.
- Huttunen, I. (1996). Metacognition in the process of development towards learner autonomy. In U. Thornberg (Ed.), *Focus on the language learner* (pp. 77-97). Uppsala: Uppsala Universitet. Centrum för didaktik.
- Janisch, C., Liu, X. & Akrofi, A. (2007). Implementing Alternative Assessment: Opportunities and Obstacles. *The Educational Forum*, 7, 221 - 230.
- Kane, M., Crooks, T. & Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice* 18 (2), 5 - 17.
- Karavas - Doukas, K. (1998). Evaluating the implementation of educational innovations: Lessons from the past. In Rea-Dickins, P., & K. P. Germaine (Eds.), *Managing evaluation and innovation in language teaching: Building bridges*. (pp. 25-50). London: Longman.
- Kaikkonen, P. (2005). Kielikasvatus koulun vieraan kielen opetuksen kehyksenä. [Language education as a framework of foreign language teaching.] In V. Kohonen (Ed.), *Eurooppalainen kielisalkku Suomessa. Tutkimus- ja kehittämistyön taustaa ja tuloksia* [The European Language Portfolio in Finland: Background and results of the research and development work] (pp. 45-58). Helsinki: WSOY.
- Klein, S. P. (1995). [The Reliability of Mathematics Portfolio Scores: Lessons from the Vermont Experience](#). (EJ511023) *Applied Measurement in Education*, 8(3), 243 - 260.
- Kohonen, V. (2008) On the significance of the CEFR and ELP in FL education: Findings from recent research. In S. Tella, (Ed.). *From brawn to brain: Strong signals in foreign language education*. Proceedings of the ViKiPeda-2007 conference in Helsinki, May 21-22, 2007 (pp. 99 - 127). University of Helsinki

- Department of Applied Sciences of Education.  
Research Report.
- Kohonen, V. & Korhonen, T. (2007). Student perspectives to the ELP: Voices from the classrooms. In Koskensalo, A., J. Smeds, P. Kaikkonen & V. Kohonen (Eds.), *Foreign Languages and Multicultural Perspectives in the European Context. Fremdsprachen und multikulturelle Perspektiven in europäischen Kontext. DICHTUNG, WAHRHEIT UND SPRACHE* (pp. 239 – 267). Berlin. LIT Verlag.
- Kohonen, V., & Pajukanta, U. (2003). Eurooppalainen kielisalkku: Kokeiluprojektin tulosten koontaa. [The European Language Portfolio: summarising the findings of the teaching experiment]. In V. Kohonen, & U. Pajukanta (Eds.), *Eurooppalainen kielisalkku 2: EKS-projektin päätösvaiheen tuloksia* [The European Language Portfolio2: Results at the end of the ELP project] (pp. 7 – 31). Tampere: Reports from the Department of Teacher Education in Tampere University.
- Korezt, D. (1992). [The Reliability of Vermont Portfolio Scores in the 1992-93 School Year. Interim Report. RAND Reprints Series.](#) (ED370983) Retrieved 15.12.2008 from <http://www.eric.ed.gov>
- Kunnan, A. J. (2005). Language assesement from a wider context. In E. Hinkel (Ed.), *Handbook of Research in Second Language Learning and Teaching* (pp. 779-794). Mahwah, NJ: Lawrence Erlbaum.
- Lantolf, J. P. & Thorne, S. L. (2006). *Sociocultural Theory and the Genesis of Second Language Development*. Oxford: Oxford University Press.
- Leighton, J. P. & Gierl, M. J. (2007). (Eds.) *Cognitive Diagnostic Assessment for Education. Theory and Applications*. Cambridge: Cambridge University Press.
- Leung, C. (2005). Classroom teacher assessment of second language development: Construct as practice. In E. Hinkel (Ed.), *Handbook of Research in Second Language Learning and Teaching* (pp. 869 – 888). Mahwah, NJ: Lawrence Erlbaum.

- 
- Little, D. (2004). Constructing a theory of learner autonomy: some steps along the way. In Mäkinen, K., P. Kaikkonen, & V. Kohonen (Eds.), *Future Perspectives in Foreign Language Education* (pp. 10-17). Research papers from Faculty of Education at the University of Oulu 101/2004. Oulu: University of Oulu.
- Lynch, B.K. (2003). *Language Assessment and Program Evaluation*. Edinburgh Textbooks in Applied Linguistics. Edinburgh: Edinburgh University Press.
- McNamara, T. (2000). *Language Testing*. Oxford: Oxford University Press.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (pp. 13 - 103). New York, NY: McMillan.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher* 23(2), 13-23.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3-62.
- Moss, P. (1994) Can there be validity without reliability? *Educational Researcher* 23(2), 5 - 12.
- Moss, P. A., Girard, B. J. & Haniford L. C. (2006). Validity in educational assessment. *Review of research in education* 30(1), 109 - 162.
- NCC (2004). *National core curriculum for basic education*. Helsinki: Finnish National Board of Education.
- Nitko, A. J. (1995). *Curriculum-based Continuous Assessment: A Framework for Concepts, Procedures and Policy*. Assessment in Education: Principles, Policy & Practice. New Jersey: Prentice Hall.
- O'Malley, J.M. & Valdez Pierce, L. (1996). *Authentic assessment for English language learners: practical approaches for teachers*. USA: Addison-Wesley.
- Oscarson, M. (1997). Self-assessment of foreign and second language proficiency. In C. Clapham & D. Corson (Eds.), *Encyclopedia of Language and Education. Volume 7: Language Testing and Assessment* (pp. 175 - 187). Netherlands: Kluwert Academic Publishers.
- Principles and guidelines (2000). *European Language portfolio (ELP). Principles and guidelines*. Council for Cultural Cooperation.

- 
- Educational Committee. Council of Europe. DGIV/EDU/LANG 33.
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: identifying processes of classroom assessment. *Language Testing*, 18(4), 429 – 462.
- Shapley, K. S., & Bush, M. J. (1999) [Developing a valid and reliable portfolio assessment in the primary grades: Building on practical experience](#). (EJ584927) Retrieved 15.12.2008 from <http://www.eric.ed.gov>
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher* 29(7), 4-14.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Harlow: Longman.
- Supovitz, J., MacGowan, A., & Slattery, J. (1997). [Assessing agreement: An examination of the interrater reliability of portfolio assessment in Rochester, New York](#). (EJ559007) Retrieved 15.12.2008 from <http://www.eric.ed.gov>
- Takala, S. (2002). Interpreting and assessing the portfolios – How to promote informed consensus? In P. Kaikkonen, & V. Kohonen (Eds.), *Quo vadis foreign language education?* Reports from the Department of Teacher Education in Tampere University A27, 123-127.
- Torrance, H. & Pryor, J. (2002). *Investigating formative assessment. Teaching, learning and assessment in the classroom*. Buckingham/Philadelphia: Open University Press.
- Vainikainen, M-P. & Hautamäki, J. (2003). *Oppimaan oppimiseen liittyvät uskomukset Malmilla ja Mankkaalla. Peruskoulun yhdeksäsluokkalaisten uskomukset syksyllä 2003*. [Learning- to-learn related beliefs among 9th graders in Malmi and Mankkaa, autumn 2003.] Koulutuksen arviointikeskuksen kuntaraportti 12/2003. Helsinki: Centre for Educational Assessment.
- Weir, C. (2005). *Language testing and validation: An evidence-based approach*. New York: Palgrave Macmillan.

## Appendix

<b>2.3 Learning to Learn</b>	<b>FILLS (Finnish Learning to Learn Scales)</b>	33
Learning motivation and learning strategies		
Goal orientation	Learning orientation Achievement orientation Performance-approach orientation Performance-avoidance orientation Avoidance orientation	
Means-ends-beliefs	Means: effort Means: ability Means: chance	
Agency beliefs	Control expectancy	
	Control motivation	
	Academic withdrawal	
	Fear of failure	
Learning strategies	Deep processing	
Self-concept and academic self-esteem		
Academic self-concept	Self-concept as a thinker Self-concept in mathematics Self-concept as a reader Self-concept as a speaker Self-concept as a writer	
	Self-esteem	
Learning environment	Classroom School Teachers	
Significant others' perceived attitudes to	Parents Peers	

---

school		
Use of information sources	Use of computer Use of books	
Group-work behaviour	Domination Non-participation Task orientation Co-operation	
Future orientation	Willingness for further schooling and improvement Disdain towards schooling	
Sociomoral self-concept as a student	Good student Indifferent student Self-assured student Rational student	