



This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Huhta, Ari

Title: Understanding self-assessment - what factors might underlie learners' views of their foreign language skills?

Year: 2019

Version: Published version

Copyright: © The Author, University of Jyväskylä/Centre for Applied Language Studies and EA

Rights: In Copyright

Rights url: <http://rightsstatements.org/page/InC/1.0/?language=en>

Please cite the original version:

Huhta, A. (2019). Understanding self-assessment - what factors might underlie learners' views of their foreign language skills?. In A. Huhta, G. Erickson, & N. Figueras (Eds.), *Developments in Language Education : A Memorial Volume in Honour of Sauli Takala* (pp. 131-146). European Association for Language Teaching and Assessment (EALTA); University of Jyväskylä.
<http://urn.fi/URN:ISBN:978-951-39-7748-1>

Understanding self-assessment

– what factors might underlie learners’ views of their foreign language skills?

Ari Huhta

University of Jyväskylä

1. Introduction

Sauli Takala had a profound influence on my career in language assessment. His course on assessing writing in the Department of Applied Linguistics at the University of Jyväskylä in 1986 was the first course dedicated to assessment that I ever attended and also the first time I met Sauli. It was a unique course in many ways. First, it was based on cutting-edge knowledge about writing assessments as Sauli had been involved in coordinating the IEA Written Composition Study (see, e.g., Gorman, Purves & Degenhart, 1988) and had just returned to Finland from the USA. Secondly, he introduced us students to something completely new at the time: word processing by using a programme called WordStar. We needed to learn word processing because the outcome of the course was a series of chapters on learning, teaching and assessing writing, to be published in a publication series targeting language teachers in Finland. The course resulted in my first real scientific publication (Huhta, 1987).

In the following years, Sauli Takala was instrumental in planning and leading several projects that entailed developing new language assessment systems, conducting research on those systems, and organising training on assessment for teachers. I was involved in many of those projects and benefited greatly from Sauli's insights and expertise. One of the most significant projects was the creation of the Finnish National Certificates in the early 1990s as a joint operation between the University of Jyväskylä and the Finnish National Agency for Education (see [https://www.oph.fi/ english/services/yki](https://www.oph.fi/english/services/yki)). The National Certificates is a language proficiency examination system intended for adults who want to have their language proficiency certified for work or study purposes. Inaugurated in 1994, the examination now has nine languages and over 10,000 test takers each year, most of whom take Finnish as a second language examination.

Although Sauli was heavily involved in creating new high-stakes language examinations such as the National Certificates and in improving existing examinations like the Finnish Matriculation Examination (see Juurakko-Paavola's account in this volume), he was also very active in promoting assessment for, rather than of, learning and teaching. He was an advocate of portfolio assessment since the early 1990s, when

the first seeds for the European Language Portfolio were sown (Little, Goullier & Hughes, 2011) and discussed the benefits of the portfolio in several publications intended for researchers, teacher trainers and teachers (e.g., Linnakylä, Pollari and Takala 1994; Takala 1992, 1995). In his view, the portfolio integrates learning, teaching and assessment in ideal ways and, thus, improves learners' agency and understanding of the entire learning process. An important benefit of the portfolio, he argued, is that it improves learners' awareness of what they can do since self-assessment is a key aspect of the portfolio (see, e.g., the European Language Portfolio; Little, 2005; Little & Erickson, 2015).

Besides his work on promoting portfolio assessment, Sauli Takala was also involved in other approaches to assessment that were designed specifically to support language learning. The most notable of these was the DIALANG project (see Huhta, Luoma, Oscarson, Sajavaara, Takala & Teasdale 2002; Alderson 2005). A key component of that diagnostic on-line assessment and feedback system is self-assessment: DIALANG does not only contain language tests but also calibrated self-assessment instruments that the users of the system can take and receive feedback that relates to self-assessment.

Self-assessment as part of more extensive approaches to diagnostic or formative assessment was, thus, one of the many areas of applied linguistics that were close to Sauli's heart. Since I myself have been very interested in self-assessment, particularly as a consequence of my involvement in the DIALANG project, it is befitting to focus on this form of assessment in my contribution to the memorial publication. The roots of the study I will be reporting here go back to DIALANG but the empirical data for it were collected several years later in a very different type of study.

2. Self-assessment in DIALANG

DIALANG is an on-line diagnostic assessment system that provides its users with feedback about the strengths and weaknesses in their language proficiency in 14 different languages and five skill areas (reading, listening, writing, vocabulary, and structures). Besides language tests, the system includes self-assessment instruments for reading, listening and writing; replying to the self-assessment task is optional but recommended. Self-assessment in DIALANG is always related to the skill to be tested: thus, if the learner wants to take a test of reading – in any of the available test languages – they are given the opportunity to self-assess their reading in that language.

The self-assessment instrument comprises 18 statements that describe specific activities related to the skill in question, for example, reading based on the CEFR (for the design and validation of the self-assessment instrument, see Alderson, 2005). Users can choose the language in which to read the self-assessment statements from a list of 18 different languages since all self-assessment instruments have been translated from the original English into the 17 other languages. This enables as many users as possible

to read the statements in a language that they know well enough to conduct meaningful self-assessment. For example, a French-speaking learner of Spanish who is not very fluent in Spanish can study the statements in French before taking a test of Spanish. The users are invited to state whether they can do (or not) the activities described in each statement. The following examples illustrate some of the English-language versions of the self-assessment statements for reading:

- (1) *I can follow short, simple written instructions, especially if they contain pictures.*
- (2) *I can read correspondence relating to my fields of interest and easily understand the essential meaning.*
- (3) *I can understand a wide range of long and complex texts, understanding fully subtleties of style and meaning which is directly stated or implied.*

Each statement has been linked to a specific CEFR (Common European Framework of Reference; Council of Europe, 2001) level. The above samples 1, 2 and 3 illustrate level A1, B1 and C2, respectively. The system treats the users' responses as if they had taken an 18-item language test covering CEFR levels A1 – C2 and calculates an estimate of their (self-assessed) CEFR level from the response data. After taking a test in the same skill, the learner is provided with feedback that reports which CEFR level most closely corresponds their self-assessment and which CEFR level they achieved in the language test. Thus, the users can see whether their self-assessed proficiency level matches the level they are assigned based on their test score.

The main function of self-assessment in DIALANG is to provide language learners with an opportunity to practice self-evaluation and improve their metacognitive skills and awareness of themselves as language learners through interacting with the self-assessment task and the related feedback. A secondary function of self-assessment is to direct the users to the most appropriate test version in terms of difficulty: based on learners' self-assessment and a vocabulary size test, the system administers them a basic, intermediate or advanced version of the test in the skill and language they chose (e.g., intermediate test of reading in Spanish).

DIALANG includes a complementary aspect of self-assessment feedback and it is this type of feedback that inspired the study reported here. Besides reporting on the match, or lack of it, between self-assessment and test result, as described above, DIALANG presents its users with an opportunity to read about potential reasons for a mismatch between self-assessments and tests. This part of feedback is titled *About self-assessment*, and its main screen is shown in Figure 1. By clicking on the links shown on the screen, learners can access more detailed information about the potential causes for the misalignment of the two.

This part of DIALANG feedback was somewhat difficult to design as we could not draw on any existing models for such feedback and relatively little research existed that could inform its content. Therefore, this feedback remained rather speculative. However, we felt that it would be important to give the users of the system

an opportunity to engage somewhat more deeply in thinking about their language skills, how they have acquired them, and how they use them. Therefore, this feedback focuses on increasing learners' awareness of their language skills, language learning, and metacognitive skills. Indeed, all feedback and information related to self-assessment in DIALANG illustrates what Hattie and Timperley (2007) refer to as self-regulation feedback that specifically focuses on improving learners' metacognitive skills. In addition to increasing learners' self-reflection with self-assessment, we also wanted to promote the value of self-assessment as an important and relevant approach to assessment in its own right and to counter the probably quite common assumption that if a self-assessment and a test do not match, the test always provides the more correct information.

I was quite closely involved in designing and drafting this feedback, which partly explains my interest in revisiting it in some way even after the DIALANG project ended in 2004, should an opportunity arise. That opportunity materialised in 2006 in the form of a research project called ToLP.

The screenshot shows a web-based feedback interface with a yellow header bar containing navigation icons (back, forward, search) and the DIALANG logo. The main title is 'About self-assessment'. Below it, a section titled 'Why self-assessment and test results may not match' contains two columns of text. The left column lists reasons for mismatch: 'For some suggestions click on the links below:', followed by a list of nine items including 'How often you use the language', 'How you use the language', 'Situations differ', etc. The right column provides explanations for each reason. At the bottom, a note states that both self-assessment and DIALANG's assessment may be accurate but reflect different aspects of language knowledge.

Reason	Explanation
For some suggestions click on the links below:	There are many possible different reasons for the difference between the level estimated by DIALANG and your own estimation.
How often you use the language	Amongst the factors listed here you may identify those that caused that difference.
How you use the language	These descriptions can help you to evaluate language tests and test scores critically. Tests give you information about your language proficiency, but you can decide how you want to use that information, in the light of the possible reasons for the difference between your estimate and that of DIALANG.
Situations differ	
Other learners and you	
Other tests and DIALANG	
You and your targets	
Tests and real life	
Other reasons	

It is worth noting that both your own and DIALANG's assessment may be accurate - they may just reflect different aspects of your language knowledge and use.

Figure 1. DIALANG feedback explaining possible reasons for a mismatch between self-assessment and test result.

3. The ToLP study

ToLP is an acronym for Towards Future Literacy Pedagogies and refers to an Academy of Finland funded research project at the Centre for Applied Language Studies at the University of Jyväskylä in 2006–2009 (see <https://www.jyu.fi/hyt/fi/laitokset/solki/en/research/projects/tolp>). The project explored mother tongue and foreign language

literacy practices among Finnish 9th grade students and their teachers in both school and out-of-school contexts. The main part of the study was a large-scale questionnaire-based survey of such practices administered to a statistically representative sample of 9th graders. A similar questionnaire was mailed to their teachers. Here, the focus is on the student survey. The main results of the project are presented in Finnish in Luukka, Pöyhönen, Huhta, Taalas, Tarmannen & Keränen (2008).

3.1 Methods

Participants: The participants were 15-year-old students in grade 9 of the Finnish comprehensive school. A sample of about 2,000 students was selected from the national population of about 55,000 students in grade 9 by using two-staged cluster sampling in which a random sample of schools was first drawn to cover all the regions of the country and different sizes of schools, followed by selecting one intact class from each sampled school. A total of 1,720 students from 101 schools responded to the survey. The response rate was 86 %; the missing 14% consisted of students who were absent from the school or engaged in other activities elsewhere in the school during data collection or who were simply unwilling to fill out the questionnaire.

Questionnaire: The questionnaire covered a wide range of questions about students' reading and writing practices in the school and in their free time, as well as questions about pedagogical practices in the school and students' attitudes towards those practices. The questionnaire was administered in the classroom during one 45 minute lesson; also the following 15 minute break could be used for the purpose if needed (in Finland, lessons last either 45 or 90 minutes and are separated by 15 minute breaks). Data collection was supervised by one of the students' teachers and/or a research assistant working for the project. For more details, see Luukka et al. (2008).

Specific questions about the basis of one's self-assessment: Our previous work on designing DIALANG feedback on self-assessment served as a starting point for designing a sub-set of questions concerning the reasons for the students' view of their proficiency in a foreign language. However, we could not systematically cover all the different factors discussed in *About self-assessment* in DIALANG because the focus of the ToLP project was on somewhat different matters. On the other hand, since the project targeted a specific group of language learners in a particular context, we could include factors that would not have been appropriate on a more general platform such as DIALANG (i.e., questions specific to the school context).

Table 1 reproduces a translated version of the questions of interest for this article. The questionnaire was administered in Finnish, the language of the school and the first language of most of the students.

Table 1. Questions about the basis of students' self-assessment.

How do you know that you have good or weak skills in a foreign language? What affects your view of your language proficiency? Circle the most suitable alternative in each line.

	This has affected my view ...				
	a lot	to some extent	only a little	not at all	I do not know
1. How easily I learn the language at school.	1	2	3	4	dnk
2. How I manage to use the language (e.g. abroad, on the Internet, when reading magazines).	1	2	3	4	dnk
3. What my teachers have said about my language skills.	1	2	3	4	dnk
4. What my friends and family have said about my skills.	1	2	3	4	dnk
5. What foreigners I have met have said about my language skills.	1	2	3	4	dnk
6. How well I succeed in exams at school.	1	2	3	4	dnk
7. How well I can use the language compared to my classmates.	1	2	3	4	dnk
8. Something else, what?	1	2	3	4	dnk

3.2. Research Questions

The research question of the study reported here were the following:

- 1) Which factors do the Finnish 9th grade students perceive to affect their view of their foreign language proficiency the most?
- 2) Are these perceptions related to students' proficiency in their first foreign language?

Students' perceptions were analysed by investigating the distributions of their responses to the questions presented in Table 1 and by conducting multivariate analyses of variance in IBM SPSS (version 24). The measure of the students' foreign language proficiency was their self-reported mark for their first foreign language in their most recent school report on the 7-point scale used in the Finnish comprehensive school. In the scale, the lowest grade 4 indicates a failure to achieve the learning goals for the term and the highest grade 10 denotes excellent achievement (Finnish National Agency for Education, 2004). Because only three students reported the lowest grade (4) as their most recent mark, they were merged with the second lowest grade (5) in the analyses reported below. Since well over 90% of the students had English as their first foreign language, their language proficiency marks refer mostly to that language.

4. Results

The results pertaining to the first research question on the factors that the teenaged learners perceived to have influenced their view of their foreign language skills are presented first followed by the findings that concern the relationship between students' perceptions and their foreign language proficiency. However, before addressing the two research questions, the findings regarding the structure of this battery of questions are reported.

First, correlations between the different questions were computed. It transpired that almost all questions correlated with each other significantly, almost always at $p < .001$ level. The only non-significant correlation occurred between success on the (language) tests or exams in the school and feedback from foreigners. The significant (Spearman rank order) correlations were rather low and ranged between .057 and .429. The fact that even correlations below .1 turned out significant was undoubtedly due to the very large sample size. The strongest correlations were the following:

feedback from friends and family & feedback from foreigners	.429
feedback from friends and family & feedback from teachers	.422
feedback from teachers & examination results-	.419

feedback from foreigners & using language in free time	.395
feedback from friends and family & comparison with classmates	.317

Overall, the correlational pattern indicates that the questions were mostly tapping rather different aspects of experience. In order to obtain a better picture of the structure of this set of questions, an exploratory factor analysis (Principal Axis Factoring with Promax rotation) was conducted. The analysis suggested that two factors underlie the students' responses, accounting for 54% of variance in their answers. The first factor related to the school and consisted of the questions on teacher feedback, examination results, ease of learning the language in the school, and comparisons with the classmates. The second factor had more to do with language use in free time and it comprised questions about feedback from foreigners, feedback from friends and family, and managing to use the language in free time. The second factor was less clear than the first one: for example, feedback from friends and family loaded on this factor only somewhat more strongly than on the first factor, possibly because many of the students' friends were also their classmates.

To answer the first research question, the distributions, means and standard deviations of the students' responses were investigated. Table 2 describes the distribution of the students' responses to each question. The last two columns in the table display the means and standard deviations (for calculating these, the responses were coded as 1 = *not at all* ... 4 = *a lot*; in addition, the *I do not know* responses were excluded because they were considered to be outside the response scale and, therefore, it was not possible to give them any meaningful numerical value). The table shows that the three elements the students thought had most strongly affected their view of their proficiency were how they had managed to use the language in various activities outside the classroom (average 3.48 on the 4-point scale), how they did on the examinations at school (3.27), and how easily they felt they had learned the language at school (3.25).

The factors that the students reported having influenced their views the least included comparison with the other classmates (average 2.63), feedback from foreigners they had met (2.73) and feedback from family and friends (2.74). However, given that 18% of the students had replied "I do not know" and that 14% had chosen "not at all" for the question about feedback from foreigners, it appears that, overall, this type of feedback was the least influential in shaping the students' views of their foreign language proficiency. The high proportion of such answers is likely due to not everybody having had a chance to meet with foreigners in the first place from whom to receive feedback.

Although the distribution of students' answers displayed in Table 2 already gives a fairly clear overall answer to the first research question, the students' mean evaluations were also compared statistically. Given the large sample size ($N = 1,343 - 1,621$, depending on the variable), it is not surprising that almost all pairwise comparisons of the mean responses to the questions listed in Table 2 turned out to be significant at $p < .001$ level. In fact, the only non-significant pairwise comparisons were

for *ease of learning the language at school vs success in the (language) exams at school* and for *feedback from family and friends vs feedback from foreigners*.

Table 2. Students' answers to the question asking them to evaluate the degree to which different factors had affected their view of their foreign language proficiency.

	This has affected my view ...						
	a lot (4)	to some extent (3)	only a little (2)	not at all (1)	I do not know	Mean (1-4 scale)	St. dev.
1. How easily I learn the language at school.	35%	51%	9%	1%	4%	3.25	.669
2. How I manage to use the language (e.g. abroad, on the Internet, when reading magazines).	57%	32%	8%	1%	2%	3.48	.686
3. What my teachers have said about my language skills.	28%	47%	19%	3%	3%	3.03	.788
4. What my friends and family have said about my skills.	16%	45%	29%	6%	4%	2.74	.813
5. What foreigners I have met have said about my language skills.	23%	28%	18%	14%	18%	2.73	1.05
6. How well I succeed in exams at school.	43%	41%	12%	3%	2%	3.27	.767
7. How well I can use the language compared to my classmates.	18%	36%	29%	11%	6%	2.63	.922

To answer the second research question concerning the relationship between the students' perceptions and their language proficiency, we examined how students with different degrees of proficiency in their first foreign language (as indicated by their most recent mark for that language) differed in their responses to the questions. Table 3 displays the means and standard deviations of the students' replies in each category of proficiency from the lowest (5) to the highest (10).

Table 3 shows a linear increase in the magnitude of values in students' responses for every question: the more proficient students considered these factors to have affected their view of their language proficiency more than the less proficient

students. The last two columns in Table 3 report the correlation (Spearman rank order correlation) between the students' foreign language grades and their responses as well as the difference, for each question, between the least able (mark 5 in the latest school report) and the most able (mark 10) students.

All the correlations in Table 3 were statistically significant at the $p < .000$ level but relatively modest in size. The strongest correlation (.278) between the school mark and the questions was found for the question on how the students had managed to use foreign languages in their free time. This was also the question in which the difference between the responses of the least and most proficient students was considerable, almost one point on the 4-point scale. The question that had the weakest relation to proficiency was the one concerning feedback from friends and family, with the correlation of .174 and a difference of slightly over half a point (0.58 to be precise) on the 4-point scale. All other correlations were in the .19 - .20 region. The question about success in language tests / examinations yielded the largest difference between the most and least proficient learners (0.92 on the 4-point scale) but the correlation between the school mark and responses to this question was only average (.197) compared to the other questions.

Although the students' perceptions of the effects of various factors on their self-assessment increased as their proficiency improved, many of these differences appear rather small and, thus, analyses of their statistical significance are in order. A multivariate analysis of variance (Manova) was conducted to establish whether the differences, overall, across the entire set of questions, were significant. As the analysis indicated that language proficiency (mark in the school report) was significant (Wilks' Lambda = 7.195, $p < .001$, effect size =.038), univariate analyses were carried out to find out to locate the differences.

Table 3. Means and standard deviations of the students' evaluation of the degree to which different factors had affected their view of their foreign language proficiency broken down by students' mark in their first foreign language (mostly English).

Proficiency (mark in the FL)		5	6	7	8	9	10	Difference 5 vs 10	Corr.
1. How easily I learn the language at school.	Mean	2.76	3.08	3.15	3.33	3.33	3.40	0.64	.191
	SD	.751	.669	.674	.602	.690	.613		
2. How I manage to use the language (e.g. abroad, on the Internet,...).	Mean	2.84	3.29	3.37	3.61	3.63	3.74	0.90	.278
	SD	.773	.766	.708	.580	.579	.587		
3. What my teachers have said about my language skills.	Mean	2.69	2.83	2.94	3.00	3.16	3.31	0.62	.209
	SD	.911	.885	.737	.780	.751	.696		
4. What my friends and family have said about my skills.	Mean	2.43	2.63	2.66	2.72	2.91	3.01	0.58	.174
	SD	.957	.807	.754	.825	.761	.761		
5. What foreigners I have met have said about my language skills.	Mean	2.16	2.51	2.67	2.75	3.01	3.01	0.85	.217
	SD	1.048	1.028	.956	1.056	.974	1.030		
6. How well I succeed in exams at school.	Mean	2.57	3.07	3.15	3.30	3.40	3.49	0.92	.197
	SD	.866	.877	.824	.672	.696	.646		
7. How well I can use the language compared to my classmates.	Mean	2.20	2.41	2.53	2.60	2.78	2.93	0.73	.193
	SD	.889	.885	.890	.921	.897	.927		
N		49	160	270	331	355	134		

The last two columns in Table 4 report the results of the univariate analyses (F-values and effect sizes) for each question. The effect sizes (eta squared values) indicate that while language proficiency is significantly associated with the students' perceptions, its effect was rather small. Even the biggest effect size, which was found for managing to use foreign languages in free time, was only .086, which means that only 8.6% of the variance in the students' answers can be attributed to their foreign language proficiency. The second highest effect size was .061 found for how successful the students had been in (language) examinations in the school. For the other factors covered in the questionnaire, the effect sizes were somewhat smaller, ranging from .033 to .046.

Table 4 also displays the results of the pairwise comparisons of the proficiency levels (i.e., different marks). The adjacent levels that were not statistically separable are greyed out. For example, for the first question (ease of learning the language at school), the students' responses at levels (marks) 5 and 6 could not be separated, neither could 6 and 7, which indicates, however, that 5 and 7 could be distinguished. Further, levels 8, 9 and 10 were also indistinguishable, but each of them was distinct from every level below 8. Also levels 7 and 8 could be separated. Overall, then, for this question, those with higher proficiency (marks 8 to 10) formed a statistically distinct group from the less proficient students (from 5 to 7), even if in the latter group a broad distinction could be made between the very weak (5) and the somewhat more proficient (7) learners. For some questions, such as feedback from teachers and comparison with classmates, the picture is somewhat more complex as there are several mutually indistinguishable proficiency groups that partially overlap, but the same principle applies: greyed out levels are not separable but those that lack any colour around them could be distinguished from the grey groups in the current study.

Table 4. Results of the pairwise comparisons of proficiency levels (greyed out levels indicate indistinguishable levels).

Proficiency (mark in the FL)	5	6	7	8	9	10	F	effect size
1. How easily I learn the language at school.							12.568	.046
2. How I manage to use the language (e.g. abroad, ...).							24.436	.086
3. What my teachers have said about my language skills.							11.400	.042
4. What my friends and family have said about my skills.							8.925	.033
5. What foreigners I have met have said about my language skills.							11.704	.043
6. How well I succeed in exams at school.							16.847	.061
7. How well I can use the language compared to my classmates.							8.886	.037

5. Discussion

This study was inspired by my involvement in designing the feedback system for DIALANG in the late 1990s and early 2000s. One of the more novel types of its feedback related to self-assessment of language skills. A key aim of DIALANG was, and still is, to support learner agency, autonomy and life-long learning by providing them with feedback on their language skills that also includes advice for further action. Having a chance to try out self-assessment in practice and to learn about it was seen as an important part of this support. While a fair amount of research on self-assessment had been carried out by the time of the DIALANG project (see e.g. the review by Oscarson, 1989, 1997), there were still many unexplored questions and some of the information presented in the feedback was based on expert opinion rather than empirical research. The ToLP project in 2006 – 2009 offered a chance to investigate one aspect of self-assessment in a particular context, namely what factors might underlie language learners' perceptions (i.e., self-evaluations) of themselves as foreign language users, irrespective of whether their perceptions were in accordance with others' views. Thus, a small set of questions targeting possible sources of such perceptions in a school context were designed and administered to a large number of Finnish 9th graders as part of a more comprehensive questionnaire.

The following research questions were of interest in the current study:

- 1) Which factors do the Finnish 9th grade students perceive to affect their view of their foreign language proficiency the most?
- 2) Are these perceptions related to students' proficiency in their first foreign language?

As regards the first research question, analyses suggested that the items in the questionnaire formed two broad factors, one that concerned activities that take place in the school (e.g., how well the students do in their studies and on tests, and what feedback they get from the teacher) and another that related to free time language use and feedback from persons other than the teacher. When we examine individual questions, both free time and school related activities were among those that were rated most highly: the highest average was found for managing to use foreign languages in free time followed by two school based activities (how well the FL exams and learning the language in the school had gone). Intuitively, this makes sense because the main foreign language that practically all 9th graders study in Finland is English for which there is a lot of exposure outside the classroom and, therefore, many opportunities to try out one's skills in both on-line and more traditional contexts of language, even at the time of the study in 2006.

Probably one of the most interesting findings concerns the result of the exploratory factor analysis that was performed to investigate the structure of the questionnaire: that separate school and free time related factors could be identified suggests that some students may derive their view of their foreign language proficiency,

and thus, its self-assessment, from what takes place at school whereas for other students their performance in so-called real life activities counts more. A related inference that can be drawn from the results is that the sources of learners' perceptions of their skills are probably quite varied, which is suggested by the rather low intercorrelations of the items in the questionnaire. For many, if not most learners, there may be no one major source of experience that has turned to dominate their perception of themselves as language learners but several factors may in fact play a role in this.

The findings concerning the second research question on the relationship between the students' responses and their foreign language proficiency indicated that the two were significantly correlated. The more proficient the students were, the higher they rated the effect of all the listed types of experience on how they perceived their FL skills. The relationship was not strong, however, and the effect sizes indicated that only 3.3 to 8.6 percent of the variation in the students' responses could be explained by their proficiency in their first foreign language (typically English). The patterns that can be seen in Table 4 suggest that often the most proficient students with marks 8, 9 or 10 in their school report rated the items as more important than those with lower marks. Seen in the context of the research instrument, namely the 4-point Likert scale, it appears that the average ratings by the two most extreme groups of students (mark 4 vs mark 10) could sometimes differ by almost one full scale point.

The reasons why language proficiency was related to the way the students responded to the questions remain unknown, and it is only possible to speculate about the causes. One potential reason relates to the learners' degree of awareness and ability or their willingness to reflect on their skills and learning. If success in language studies, achievement, and self-awareness / self-regulation are associated with each other, then the higher achieving students may also be more aware of themselves as language learners and better able to reflect on which types of experience have had a really significant effect on what they think about their foreign language skills. In contrast, the lower achieving students may have been less used to such metacognitive reflection and, not being quite sure what to answer when requested to do so, chose one of the options indicating a smaller degree of importance.

The limitations of the study include the fact that it was based entirely on a questionnaire survey and lacked a qualitative part which would have shed more light on the students' perceptions. Another limitation concerns the measure of foreign language proficiency used in this study: the students' mark in their first foreign language in their latest school report. First, it was self-reported by the students, which may introduce some inaccuracies. The second, and likely a bigger source of uncertainty, is the lack of standardisation of these teacher-based marks compared with, for example, a standardised language test. One type of uncertainty in this relates to variation between teachers and schools: teachers differ in their grading, as has been shown in national studies of educational achievement in Finland (e.g., Hildén & Rautopuro, 2017). Another type of issue concerns the fact that language marks are not based only on students' language skills but on their achievement in other goals of the curriculum that relate to the target language culture and learning to learn. Furthermore,

the teachers participating in the ToLP study reported clear variation in how much weight they give different factors in at least their final grading at the end of the 9th grade and, thus, presumably also prior to that stage. Some teachers reported even taking the students' diligence, participation, and motivation into account in their grading (Tarnanen & Huhta, 2011).

Further investigation of the topic addressed in this study can be divided in broadly two categories. First, the extensive student questionnaire data collected in the ToLP study could be utilised more comprehensively. For example, to broaden the basis of the evaluation of the students' foreign language proficiency their responses to a range of self-assessment questions could be used in combination of the external measure (mark given by the teacher) that was used in the current study. Another approach to utilising the existing data would be to conduct classification analyses that focus not on the variables (questions) but on the students. Student profiles could be extracted from their responses to the questions investigated in this article in order to find out the number and characteristics of such groups and whether the groups could be related to, for example, students FL proficiency or other relevant data that were collected via the questionnaire.

The second type of research that could build on the current study would be an entirely new investigation that would most likely be more qualitative in nature. Through narrative studies, interviews, learning diaries and other such approaches it could be possible to obtain insights into the kinds of experience that language learners consider important for the formation of their views of what they can do in a foreign language. Such research could also provide insights into how self-assessment works for different types of students in a more general sense, i.e., to what extent it relates to their more general confidence and self-efficacy. It could also increase our understanding of how self-assessment might affect the power relations between language learners and teachers.

To conclude, the current study contributes to the growing body of research on self-assessment of proficiency in a foreign language and hopefully sparks new investigation of the factors that underlie and interact with learners' self-evaluation. It is fair to assume that Sauli Takala would welcome such investigation since self-assessment was one of the many topics that was dear to him not only because of his involvement in the DIALANG project but also because he considered self-assessment to have an important role in language education more generally.

References

- Alderson, J.C. (2005). *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. London: Continuum.
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching and assessment*. Cambridge: Cambridge University Press.
- Gorman, T., Purves, A. & Degenhart, E. (Eds.) (1988) *The IEA Study of Written Composition: The International Writing Tasks and Scoring Scales*. Amsterdam: International Association for the Evaluation of Educational Achievement.

- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Hildén, R. & Rautopuro, J. (2017). On an equal footing? Comparing school grades and national evaluation results in Finland. In N. Pyry, L. Tainio, K. Juuti, R. Vasquez, R., & M. Paananen (Eds.). *Changing subjects, changing pedagogies: Diversities in school and education*, 242–259. Helsinki: Suomen ainedidaktinen tutkimusseura.
- Huhta, A. (1987). Kirjoitelmien arvointi. [Assessment of writing] In Takala, S. (Ed.) *Katsauksia kirjoittamiseen*. Julkaisusarja B: Teoriaa ja käytäntöä 5, 25–52. Jyväskylä: Jyväskylän yliopisto: Kasvatustieteiden tutkimuslaitos.
- Huhta, A., Luoma, S., Oscarson, M., Sajavaara, K., Takala, S. & Teasdale, A. (2002). DIALANG - A Diagnostic Language Assessment System for Learners. In Alderson, J.C. (Ed.) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Case Studies*, 130–145. Strasbourg: Council of Europe.
- Linnakylä, P., Pollari, P. & Takala, S. (1994). *Portfolio oppimisen arvioinnin tukena* [Portfolio as a way to support assessment of learning]. Jyväskylä: Jyväskylän yliopisto: Kasvatustieteiden tutkimuslaitos.
- Little, D. (2005) The Common European Framework and the European Language Portfolio: involving learners and their judgements in the assessment process. *Language Testing*, 22(3), 321–336.
- Little, D., Goullier, F. & Hughes, G. (2011). *The European Language Portfolio: The story so far (1991–2011)*. Strasbourg: Council of Europe. Available at <https://www.coe.int/en/web/portfolio/history>
- Little, D. & Erickson, G. (2015). Learner identity, learner agency, and the assessment of language proficiency: Some reflections prompted by the Common European Framework of Reference for Languages. *Annual Review of Applied Linguistics* 35, 120–139
- Luukka, M.-R., Pöyhönen, S., Huhta, A., Taalas, P., Tarnanen, M. & Keränen, A. (2008). *Maailma muuttuu - mitä tekee koulu? Äidinkielen ja vieraiden kielten tekstikäytänteet koulussa ja vapaa-ajalla* [The world changes - how does the school respond? Mother tongue and foreign language literacy practices at school and in free time]. Jyväskylä: University of Jyväskylä: Centre for Applied Language Studies.
- Finnish National Agency for Education (2004). *National Core Curriculum 2004*. Helsinki: FNAE. Available at https://www.oph.fi/english/curricula_and_qualifications/basic_education/curricula_2004
- Oscarson, M. (1989). Self-assessment of language proficiency: rationale and applications. *Language Testing*, 6(1), 1–13.
- Oscarson, M. (1997). Self-assessment of foreign and second language proficiency. In C. Clapham & P. Corson (Eds.) *The Encyclopedia of Language and Education. Volume 7. Language Testing and Assessment*, 175–187. New York: Springer.
- Takala, S. (1992). *Virikkeitä uutta kokeilevaan koulutyöhön* [Ideas for innovative trials in education]. Jyväskylä: Jyväskylän yliopisto: Kasvatustieteiden tutkimuslaitos.
- Takala, S. (1995). Arviointikulttuurin kehittäminen [Developing assessment culture]. In K. Salmio & K. Lindroos-Himberg (Eds.) *Askelia yleissivistävän koulutuksen arviointiin*, 6–17. Helsinki: Opetushallitus.
- Tarnanen, M. & Huhta, A. (2011). Foreign language assessment practices in the comprehensive school in Finland. In D. Tsagari, D. & I. Csepes. (Eds.) *Classroom-Based Language Assessment*, 129–146. Language Testing and Evaluation Series, Grotjahn, R. & G. Sigott (general eds). Frankfurt: Peter Lang.