

THE 1999 NATIONAL EVALUATION OF THE ENGLISH LANGUAGE SKILLS OF FINNISH 9TH FORM PUPILS

Eeva TUOKKO*

National Evaluations in Finland

According to new school legislation, adopted in Finland in 1999, schools are subject to external evaluation organised by the National Board of Education [4, 21§]. Evaluations are either sample-based (at least 5 % of the age group) or include all pupils studying the subject. Mathematics and Mother Tongue (Finnish and Swedish, the second national language in Finland) are evaluated alternately every second year while other subjects are evaluated less often according to a plan designed by the Ministry of Education [1, p. 69].

The principal objective of the national evaluations is to examine the extent to which the aims set in the 1994 Framework Curriculum have been achieved. After the decentralisation of administration in the 1990s the power and the responsibility for designing curricula was transferred to local authorities and to schools themselves. Although the national core curriculum sets a common framework and establishes certain minimum requirements, local needs and preferences can also be taken into account, which may lead to a large variety of curricula with a certain degree of difference in terms of content and interpretation. Therefore, it is of great importance for decision-makers to be informed whether the spirit of the law about educational equality in different parts of the country, between the genders, and among Finnish and Swedish speaking pupils is fulfilled. On the other hand, the information obtained is also made available to developers of curricula to be considered in forthcoming revision. Information about the results is also conveyed to the local authorities, which makes it possible for them to see the relative position of their school(s) in the country.

* M.A., *Special Planner, Finnish National Board of Education*

The writer now works on the national evaluation of the Swedish language (the second national language) at the Finnish National Board of Education. Two years ago she was in charge of a similar project on the English language, the procedure and results of which are reported here. She is also preparing a doctoral dissertation on the subject.

Objectives of comprehensive school English studies

The general objectives of foreign language studies are as follows:

“The student

- gets along in the language he is learning in everyday language communication;
- knows ways to communicate that are characteristic of the target language and its culture;
- receives information about the countries, people, and cultures of the language area and has an open mind towards different cultures and its representatives;
- develops his study skills alone and in groups;
- develops his ability to evaluate himself, and learns to be responsible for his studies;
- experiences the teaching and study as meaningful, emotional, and challenging; and
- becomes interested in foreign languages and cultures.”

[2, p. 74]

The general objectives are defined as skills (gets along in the language), knowledge (knows ways to communicate; receives information...), and attitudes (has an open mind...; becomes interested in foreign languages and cultures; experiences the teaching...). In addition, learning-to-learn skills are emphasised (develops his study skills; ... ability to evaluate himself).

The contents are presented separately for forms (2)3 – 6 of primary school and 7-9 of lower secondary level. The aim is that at the end of basic education (= comprehensive school 9th form) the skills of students having chosen the foreign language as the A2 language would be equivalent to those obtained in the A1 language. At the end of comprehensive school it is expected that

“... the student

- understands speaking that concerns everyday things and that is delivered at a normal tempo;
- can participate in a conversation on ordinary things by applying natural and fluent pronunciation, accent, rhythm, and intonation;
- can understand fairly easy written language that discusses general things;
- has assimilated vocabulary central to the language, main phrases, and basic structures;
- can produce short, narrative or descriptive written texts, for example, with help, if necessary;
- knows ways to communicate that are peculiar to the target language and culture;
- has assimilated knowledge about the countries, peoples, and cultures of the language areas.”

[2, p. 75]

The content descriptions emphasise everyday communicative situations. The expectations are expressed in “can do”-form, which contains the idea of functional and communicative competence, i.e. being able to act properly both as the transmitter and the receiver of the message.

The Framework Curriculum does not define topics in detail. Consequently, they have to be determined both for local curricula and national assessments by making inferences of what the concepts of “everyday communicative situations” are as well as what has been said in the general part of the Framework Curriculum about integrated subject matter to be dealt with during the lessons. Efficient language learning is expected to be supported by the student learning to define his/her own learning objectives [2, p. 77].

The Framework Curriculum does not pay attention to the special characteristics of different languages. In the English language, for example, being able to use idioms and appropriate vocabulary in certain (everyday) communicative situations may be more important than grammatical correctness. No criteria for defining the level of language proficiency have been given. Thus, they have to be decided upon by the teachers in schools, and included in the local curricula, which may lead to very different definitions. Since autumn 1999 schools have had access to a manual published by the National Board of Education where the level of mark 8 (showing average skills on the scale of 10 – 4, the latter one indicating failure) has been defined for the final stage of basic education. These criteria were not available, however, for the teaching of the students taking part in the spring 1999 national evaluation in English.

Preparation of the assessment of the English language

The preparations for the assessment to be conducted in spring 1999 were started a year before by appointing an expert group, consisting of representatives from the National Board of Education, the University of Jyväskylä, the Trade Union of Education in Finland OAJ, the Federation of Foreign Language Teachers in Finland SUKOL, and the Association of Teachers of English in Finland. It was decided that the group would supervise the work of a group of test makers to be appointed later, and that a special planner be employed to take care of the editing and administrative work as well as reporting. The traditional four skills were to be tested: listening and reading comprehension, oral and written communication. In addition, a grammar sub-test was to be included and, depending on the time available, also a separate or an integrated vocabulary test. Background information about the schools, teaching, and pupils and their attitudes were to be gathered using questionnaires directed to the heads of schools, teachers and pupils.

Designing the tests

The group of test makers, consisting of active comprehensive school teachers and teacher trainers started work at the beginning of the autumn term 1998. The basic target of their work was to operationalise the objectives of the Framework Curriculum, i.e. to interpret the meaning of the curriculum statements about language skills into concrete functions, in other words, what the pupil was expected to be able to do with the language. In the first meeting the division of work was defined, and the timetable fixed. It was decided that to cover as wide an area of learning as possible, three versions of the written test would be needed with one so-called anchor text in each sub-test of different skills. This number of versions was regarded as sufficient in order to obtain comprehensive information on various skills. The estimate was that material for about five versions would be needed for pre-testing in case some of the texts or items proved unsuitable for one reason or another. Six versions of the oral test were pre-tested, and the same number was used, with minor changes and definitions, in the final test. In addition to the manuscripts, test specifications, i.e. rubrics, topics, intended functions or text type, correct answers or suggestions for scoring answers of different levels of correctness, test type, the language of the tasks, and the estimated level of difficulty of

each item, were to be supplied by the test makers. Listening comprehension sub-tests were to be compiled first due to the time needed for recording. The results of the pre-tests were expected to be available by the end of the year, to allow enough time for selecting the most appropriate items, compiling the tests, having them printed and recorded, planning the answer sheets and instructions for schools and teachers, and packing and mailing the parcels.

Sampling

In order to ensure a representative sample, a decision was made to have a 10 % sample of the age group. For selecting the participating schools, three different sampling strata were used. The first stratum was based on the number of schools in each province to guarantee a representative sample on a national level. The second stratum consisted of sampling a representative number of schools in areas with different economic backgrounds using the European Community Structural Fund objective regions as the criterion. The third stratum consisted of the type of municipality, i.e. urban, densely populated, or rural.

The sampling resulted in 124 Finnish speaking and 19 Swedish speaking schools attending the assessment. The number of pupils taking part in the test was, however, dependent on the size of the school. The heads were asked to draw up an alphabetical list of all the 9th form pupils in their schools. The sampling for the written test was done as follows:

Number of 9 th form pupils	Method of sampling
> 31	All pupils participated
31 – 66	Every third pupil was left out (two thirds participated)
67 – 100	Every second pupil participated
< 100	Every third pupil participated

The final sample consisted of 5027 Finnish-speaking (8 % of the population) and 614 Swedish speaking (17 %) pupils. It was not possible to have all of them take the oral test. Therefore, a suggestion was made that in every school a quarter of the sample pupils would participate. The pupils were to work in pairs, and it was up to the teachers to see that pupils of roughly equal skills would work together. Furthermore, the schools were offered the opportunity to order the tests for free for the students of the whole school if only some of them were included in the sample. However, only the results of the sample pupils were evaluated by the National Board of Education.

Procedure

The media were informed about the assessment and its timetable in September 1998. The heads of the sample schools learned about their participation at the beginning of December. Enclosed with the information was a questionnaire enquiring about the size of the age group and the groups the pupils were taught English in, the way the groups were formed, funding, opportunities to give remedial teaching, and the number and type of optional English courses.

The oral test was conducted as pair work. Each pair was allowed 15 minutes to prepare for the test, and another 15 minutes was reserved for the test itself. Six different versions of the oral test were available. The order of the versions sent to schools was drawn by lot. If, for example, the first school on the list received version number 4, and three pairs were supposed to take part, the school was sent versions 4, 5 and 6. The next school with e.g. two pairs would get versions 1 and 2, and the next school versions 3, 4, etc.

The written tests were taken on the next day. During the first 45 minutes – the normal length of a lesson in Finland – the reading comprehension and grammar sub-tests were taken. During the 15-minute break the pupils completed a questionnaire, and the second lesson was used for taking the listening comprehension and writing (composition) sub-tests. Out of the three versions of the written tests not more than two types of versions were sent to one single school.

Structure of the tests

The pupils taking part in the assessment had been studying English either for almost seven years (A1 language) or for almost five years (A2 language) depending on what kind of language programme they had chosen in primary school. Practically all Swedish-speaking pupils study English as their A2 language due to the relatively close resemblance of their mother tongue and English (as opposed to Finnish, usually the A1 language). The programmes to choose from are shown in Figure 1. Since the main objective of this assessment was to find out to what extent the pupils had achieved the aims mentioned in the Framework Curriculum, it was important that the coverage of the test be as wide as possible. This can be achieved in two ways. Either tests are administered so frequently that in the course of time most of what is supposed to be learnt will be evaluated, or a host of skills and knowledge should be tested for the evaluation to yield enough information. Since there is no decision by the Ministry of Education about when English will be

evaluated again, the latter procedure was chosen. That is why the size of the sample was increased, and three versions of written tests were prepared. It was hoped

that 5-6 versions could have been produced, but this time there were not enough resources for it.

PATHS THAT DESCRIBE THE OPPORTUNITIES FOR THE STUDY OF LANGUAGES

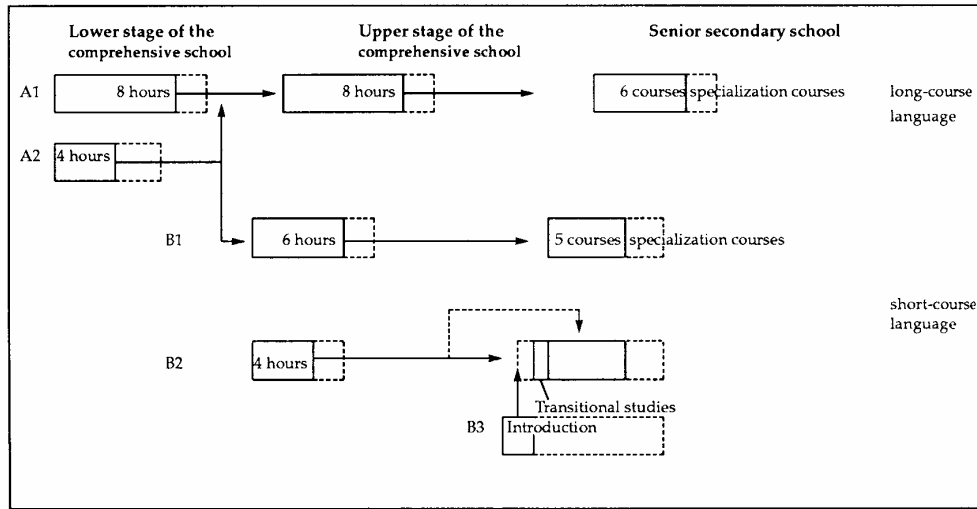


Figure 1. [2, p. 21]

The test was compiled in Finnish and then translated into Swedish. Since the sample also included a group with Same as their mother tongue, a Same translation was also produced.

Oral test

Each version of the oral test contained three types of tasks: structured dialogues (A), reading passages (B), and a discussion/debate (C). Pupils were given instructions in their mother tongue about what to deal with and/or what message(s) to convey.

Productive oral skills were assessed by the teacher using a 6-grade scale with 0 standing for more or less complete silence and 5 for excellence. Figure 2 shows the outline of the test, where it can be seen, for example, that Dialogue A1 was an anchor task common to all participants and that Dialogue A3 appears in versions 1, 2 and 4. Reading passages B1 + B2 appear in versions 1, 3 and 5, and the Discussion topic C1 in versions 1 and 5.

Task	Version					
	1	2	3	4	5	6
1	A 1					
2	A 2	A 3	A 2	A 3	A 2	A 2
3	A 3	A 4	A 5	A 4	A 3	A 5
4	A 6	A 7	A 8	A 9	A 10	A 8
5	B 1 + B 2	B 3 + B 4	B 1 + B 2	B 5 + B 6	B 1 + B 2	B 3 + B 4
6	C 1	C 2	C 3	C 4	C 1	C 2

Figure 2 . Outline of the oral test

Reading comprehension

The reading comprehension tasks differed from each other regarding both the text type and the task type. Common to each of the versions was Task 1 with five multiple choice statements with three options given in Finnish. The rest were unique items, i.e. each of them only appeared in one version. In addition to multiple choice with options either in Finnish or in English, short-answer questions, true/false statements, and translation of words and word groups from English into Finnish were used.

Listening comprehension

The anchor task consisted of nine short dialogues tested using multiple choice questions/statements with four options in English. The unique tasks were also tested using true/false tasks, short-answer questions in Finnish, and dictation.

Grammatical structures

When testing grammatical structures, the idea of rotation was applied to some extent. The first 20 multiple choice items were common to all versions. Versions 1 and 2 had ten more multiple choice items (21-30) in common whereas only six of them, items 25-30, were included in Version 3. Since the sample was big, the rotation of items yielded more information of the pupils' knowledge of grammatical structures, and will provide the test makers with pre-tested items to be used in future assessments. All the gap-filling items differed from each other in all versions, as did

the sentences to be translated into English.

Writing

In each of the three versions there were three rubrics for the pupils to choose from. One of the rubrics was common to all versions. Two out of three titles were guided. Pupils were expected to observe the guidance, and consequently either produce a descriptive, narrative (or a combination of both), persuasive, or argumentative piece of writing.

The assessment instructions the teachers were provided with were rather open. Three features were supposed to be focussed on: contents (comprehensibility / fluency, versatility of ideas, demands of the task type), grammatical structures (accuracy / correctness), and vocabulary (versatility). Each feature was to be assessed using a 6-grade scale. The suggested length of the piece of writing was about 100 words, and 1-2 (out of 30) points were to be penalised due to too short (less than 25 words) or too long (more than 125 words) a production. If the production clearly dealt with something other than what the title implied, 2-5 points were penalised.

Validity and reliability

As mentioned before, the Framework Curriculum does not define very accurately what to teach and what level to reach. While the test was being made, a manuscript of the manual for average proficiency in English was available. However, since the test had already been

prepared by the end of 1998, the criteria could not be taken fully into consideration. Therefore the test must be regarded as an interpretation of the Framework Curriculum made by the test makers, experienced comprehensive school teachers, and the expert group.

To determine the concurrent validity of assessment procedures, the relationship between the test results and an external criterion, the latest school marks in English, was computed. This yielded a correlation of 85.

To assess the internal consistency of the test the Cronbach α coefficient was computed on the sub-tests and the test as a whole. Table 1 displays the results of the α coefficient.

Table 1. Internal consistency (Cronbach α of the test)

Test	Internal consistency (Cronbach α coefficient)
Reading comprehension	.78
Listening comprehension	.75
Grammatical structures	.92
Total test	.94

Overall results of the test

Figure 3 shows the distribution of scores, given as percentages of maximum scores, in the whole test. The overall mean percentage score was 64.

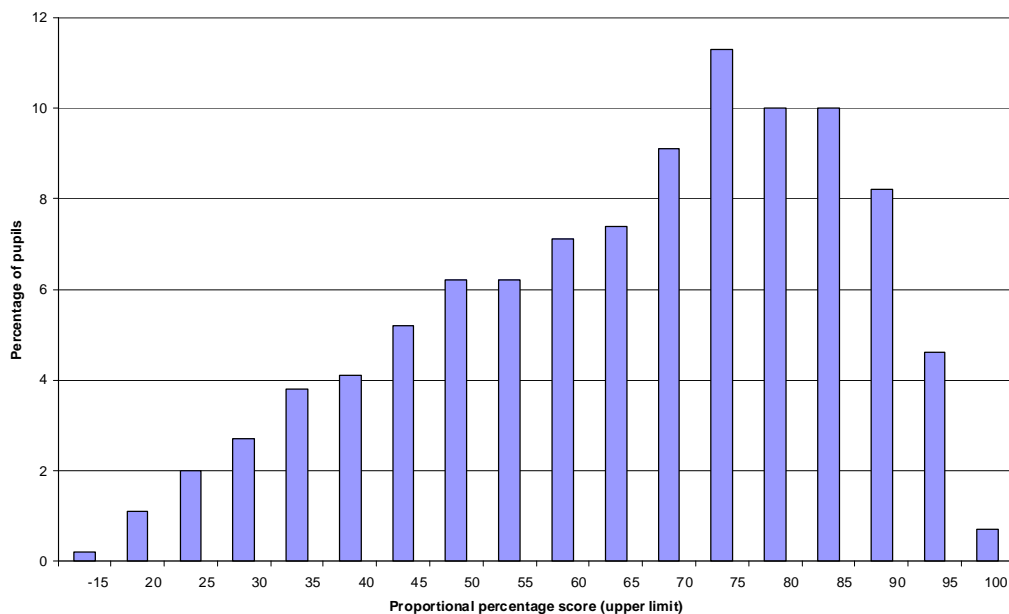


Figure 3. Distribution of the overall percentage scores

The mean percentage score corresponds roughly to what was expected. No one pupil scored full points, but a very good result (percentage score over 85) was achieved by 14% of the sample pupils. On the other hand, the level of proficiency at 45% or lower was considered to fall short of a reasonable expected level of attainment, and about 19% of the pupils were accordingly regarded as under-achievers.

Table 2 shows the overall mean percentage scores and standard deviations in the sub-tests.

Table 2. Overall mean percentage scores and standard deviations in the sub-tests

Sub-test	Mean percentage	St. Dev.
Reading comprehension	69.7	21.25
Listening comprehension	58.3	21.07
Grammatical structures	64.6	21.88
Writing	61.3	24.55
Speaking	65.1	22.65

Receptive skills

Reading comprehension sub-test yielded the highest mean percentage scores and listening comprehension the lowest. The difference between the mean percentage scores of the two sub-tests was 11.5, which means a difference of about one mark in the Finnish marking scale of 4 – 10.

Table 3 shows the percentage of pupils who scored less than 45%, between 50% -80%, and more than 85% of the maximum points.

Table 3. Percentage of pupils grouped according to their achievement in receptive skills

Skill	Percentage < 45%	50% - 80%	> 85%	of the subject matter
Reading comprehension	15	42	29	
Listening comprehension	29	52	11	

The analyses of item results have not been made yet. That is why only general assumptions can be made on what the cause of difficulty in the listening comprehension test was. Some teachers commented on the recording having been too fast. One source of difficulty may have been the dictation test with perhaps too many gaps to fill in.

Grammatical structures

The grammar test mean percentage score 64.6 is closest to that of the whole test. More than 85% of the subject matter was covered by 21% of the pupils while 22% remained under the percentage of 45. Somewhat

fewer than 50 % of the pupils covered 50% - 80% of the subject matter.

Productive skills

Table 4 shows the percentage of pupils whose scores were < 45%, between 50 – 80% and > 85% of the maximum.

Table 4. Percentage of pupils grouped according to their achievement in productive skills

Skill	Percentage of pupils covering	< 45%	50% - 80%	> 85%	of the subject matter
Writing		22	45	14	
Oral		25	48	16	

A sample was taken on both the written and oral products of the pupils to rate them to see if there was a difference between the rating of teachers and external raters. About 700 compositions were re-read, and the oral products of about 160 pairs, i.e. 320 pupils were listened to. There were three raters altogether, and each product was rated by two of them. In addition, about ¼ of the products were rated by a native speaker.

Criterion-related assessment was not introduced in the 1994 Framework Curriculum. Although writing and speaking are official objectives of teaching and practising, no guidance is given about levels of proficiency to be aimed at. It has been up to the teachers to decide about the level since a coherent and well-established tradition of evaluating those skills has been missing. Due to lack of time in preparing the tests, proper criteria for assessing productive skills could not be produced. Thus, the instructions submitted to the teachers were rather vague.

Before the re-rating, however, an attempt was made to produce more precise criteria for the raters.

The work was done both for assessing written and oral production by Sauli Takala, Research Professor of the University of Jyväskylä and a member of the expert group of this evaluation project. Writing was rated for contents, structure of the text (organisation and clarity), linguistic form and orthography according to a 6-grade scale (Appendix 1). When the ratings of written production made by the teachers and raters were compared, it was noticed that the teachers' ratings were approximately ½ point (max. 15) higher than those of the external raters.

Oral skills were re-rated for fluency, grammar/structures, vocabulary, pronunciation and interaction (Appendix 2). Moreover, instead of assessing the whole test as one entity as the teachers had been advised to do, each of the tasks was assessed separately, and the pupils' final result was the average

of the points gained in each task.

Overall results of the test in different skills by gender

The test results show that girls scored on average 67% of the maximum while boys scored 61%. In a previous evaluation, based on the 1985 Framework Curriculum, the corresponding figures were 66% and 59% [3, p. 359]. Both the level in general and that of boys show a slight improvement.

Table 5 displays the percentage scores in different skills by gender. The biggest differences are found in writing (10 percentage units) and grammatical structures (almost 8 percentage units) in favour of girls. The boys' result in writing is about 85% of that of the girls', and the result is roughly the same, 88%, when considering speaking skills. Boys seem to have scored better in receptive skills than in productive ones.

Table 5. Percentage scores in language skills by gender

Skill	Mean percentage	
	Boys	Girls
Reading Comprehension	66.8	72.5
Listening Comprehension	56.2	60.3
Grammatical structures	60.6	68.4
Writing	55.9	66.4
The whole test	60.6	67.1
Speaking*	60.6	68.9

* Is not included in the results of the whole test

Figure 4 shows that when the percentage scores are below 45, the number of boys is over-represented. This group is represented by 25% of the boys, but only 13% of the girls. An excellent result (> 90) was reached by 7% of the girls and 4% of the boys.

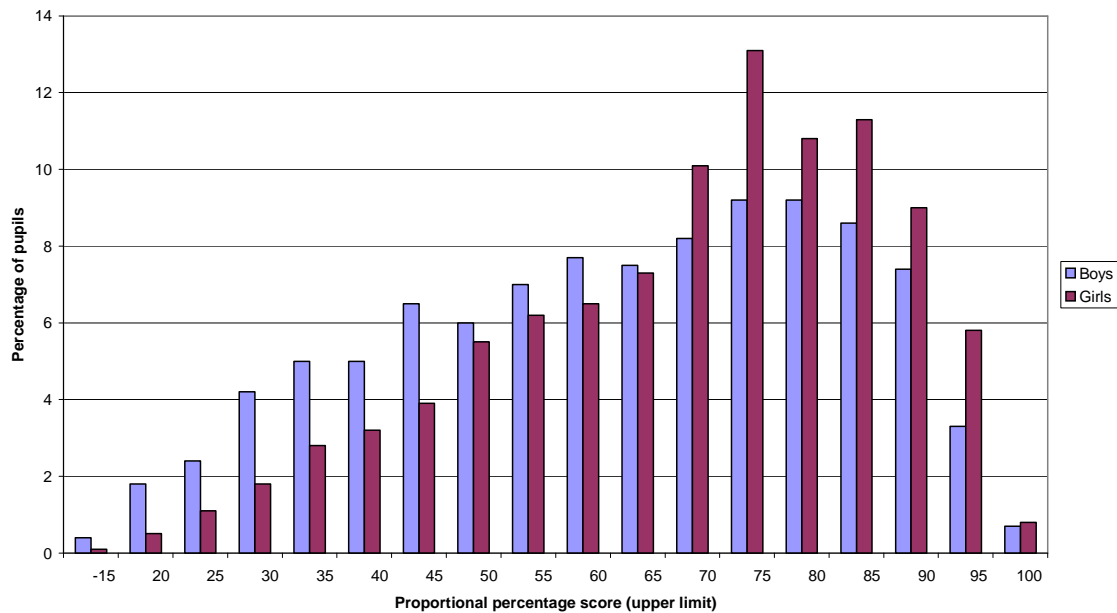


Figure 4. Distribution of the overall percentage scores by gender

Overall results of the test by mother tongue

Swedish belongs to the Indo-European, Germanic languages as does English whereas Finnish belongs to the Fenno-Ugric language group. There is quite a lot of

resemblance both as for structure and vocabulary between Swedish and English, and therefore it is understandable that Swedish-speaking pupils score better in tests of English than Finnish-speaking pupils. This test proved to be no exception.

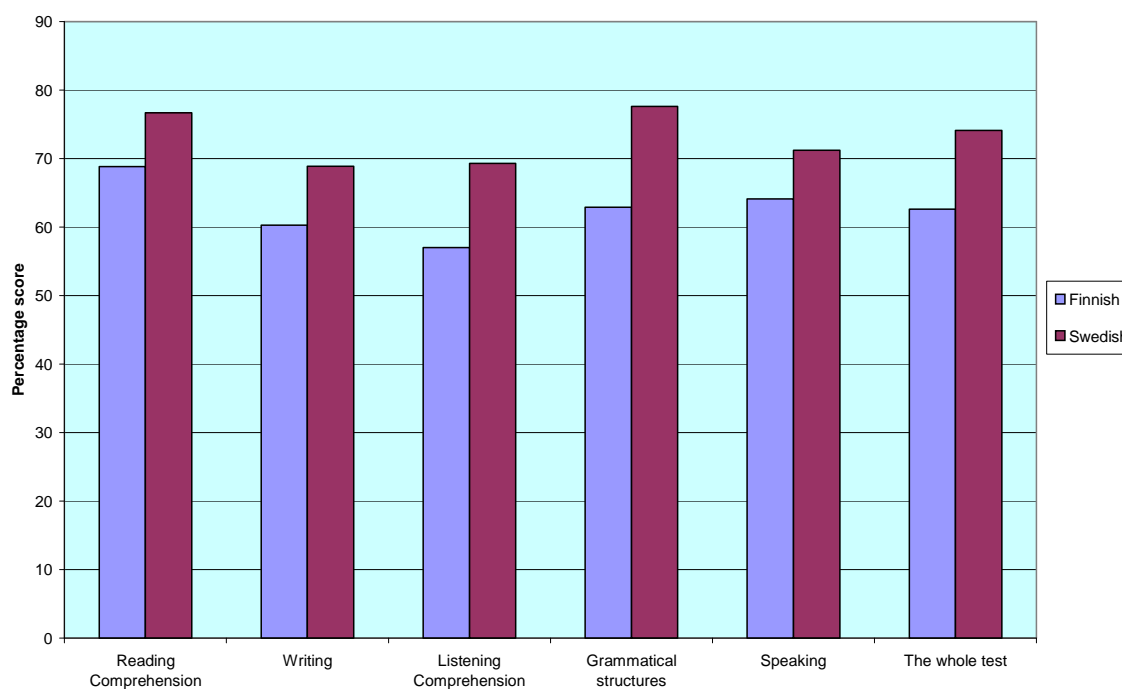


Figure 5. Percentage scores in different language skills by mother tongue

Figure 5 shows that there was a 12 percent unit difference in favour of Swedish-speaking pupils in the overall results. The biggest difference occurred in knowledge of grammatical structures. The differences in productive skills seemed to be smaller than in receptive skills. Whether there is a difference in the marking scales of the teachers is not known because the writing and oral tests have not yet been rated from that point of view. However, teachers in Swedish-speaking schools seem to be stricter in giving marks in certificates [5, p. 94].

Overall regional results

One of the objectives of the assessment was to study if pupils, regardless of their place of residence and the economic status of the region, have equal access to education. That is why the test results were grouped by provinces and the European Community Structural Fund objective regions.

There are six provinces in Finland, one of them, the

Aland Islands, being practically Swedish-speaking. The tendency was for the test scores to get lower towards the eastern and northern provinces of Finland, and this trend has been found in previous studies too. In these areas the unemployment rate is high, which tends to lead to economic difficulties and decreased budgeting by municipalities on education.

Resources were allocated by the EU to certain disadvantageous regions according to the European Community Structural Fund's objectives 2 (social and economic degeneration of urban areas in crisis), 5b (promoting rural development), and 6 (development of regions with extremely low population density). In addition, there were regions with no support (Appendix 3). When the test results are grouped according to EC regional objectives, it can be seen that the Objective 6 regions cover most of the area of the underachieving provinces of Lapland, Oulu, and East Finland. On the other hand, the region in need of no support covers a big part of the province of South Finland with the metropolitan area included.

Table 6. Percentage scores by provinces and EC objective regions in different language skills

Skill	Provinces							Regions by objectives			
	SF	WF	EF	Oulu	Lapland	Aland	All	No support	Objective 2	Objective 5b	Objective 6
RC	72.9	68.0	66.5	66.1	65.5	77.6	69.7	73.8	67.6	67.3	65.2
LC	61.5	57.3	54.6	54.2	53.1	65.6	58.3	62.3	57.0	56.7	52.7
Struct	67.9	62.9	62.0	60.7	58.6	75.7	64.6	69.3	62.2	62.3	58.7
Written	62.8	59.8	61.9	61.1	56.4	72.6	61.3	64.6	57.3	59.9	58.9
Whole	66.8	62.3	61.7	61.0	59.4	73.5	63.9	68.1	61.1	62.1	59.3
Oral x)	66.3	64.2	64.8	64.2	60.6	77.6	65.3	68.6	58.6	66.0	62.0

x) is not included in the concept of the whole test

Tasks measuring operations / functions in different language skills

When compiling the test, the idea was to make it as authentic as possible by incorporating different language functions into the tasks, i.e. having the pupils read and listen for different purposes. However, assessing knowledge of basic grammatical constructions was also considered essential.

In general, girls scored better than boys in all functional tasks, as did Swedish-speaking pupils compared with Finnish speaking ones. An exception

was understanding concepts such as cause, effect, etc., which presupposes the skill of drawing conclusions. There were seven items of this type in the test, and especially in one of them, where interpreting a bus timetable was required, the boys' result was significantly higher. When comparing the results of speakers of different native languages (either Finnish or Swedish), it was seen that it had been remarkably easier for the Swedish-speaking pupils to understand connections between different parts of the text. The smallest difference found referred to understanding details.

Table 7. Percentage scores of operations in reading comprehension

Type of reading comprehension	Number of items	All	Girls	Boys	Finnish	Swedish
Reading for main ideas	10	75.6	76.5	70.3	72.5	81.1
Skimming: scanning to locate specific information	7	70.3	75.7	64.5	69.3	77.4
Reading for important detail	12	56.8	64.8	54.7	56.2	61.7
Understanding concepts such as cause, effect, etc.	7	65.9	58.5	67.6	65.7	77.6

[According to 6, p.73]

As for listening comprehension skills, finding the main message seemed to be the easiest operation for all participants. Understanding details was most difficult

for speakers of Finnish; for speakers of Swedish it also proved to be difficult.

Table 8. Percentage scores of operations in listening comprehension

Type of listening comprehension	Number of items	All	Girls	Boys	Finnish	Swedish
Listening for main ideas	3	67.3	69	65.6	66.5	73.4
Determining speaker's attitude / intentions	7	54.8	57.7	51.6	53.3	65.3
Listening for specifics	18	51.5	51.2	48.7	47.9	65.4
Making inferences and deductions	3	61.2	66.3	55.8	60.1	69.4

[According to 5, p. 98]

To obtain detailed information about the pupils' knowledge of grammatical structures, multiple choice items were used quite a lot for test economy. Translation was included to find out about how accurately basic grammatical constructions were used

in everyday sentences. However, it has not been possible to analyse the translation products yet. Conjunctions seemed to be the easiest items, and translation as a whole the most difficult. However, the number of conjunction items was rather low, so only

very basic ones were tested. Boys in general scored worse than girls, and so did Finnish speakers compared with Swedish ones. The big difference in recognising

and producing grammatical structures is explained, as mentioned before, by the resemblance of Swedish and English.

Table 9. Percentage scores of grammatical items

Structures	Number	All	Girls	Boys	Finnish	Swedish
Adjectives and adverbs	4	66.1	70.7	63.2	65.8	76.2
Use of articles	10	63.8	67.3	60.3	62.0	78.4
Translation	18	53.2	57.5	48.7	51.3	66.8
Conjunctions	5	73.6	78.1	68.8	71.8	86.3
Pronouns	13	61.2	65.0	57.4	59.2	76.2
Verbs	15	61.4	64.8	57.8	59.0	78.2

The following text types were represented among the composition topics: persuasive, persuasive/descriptive, descriptive/narrative, narrative/descriptive, and argumentative. The first topic yielded the lowest results (8.3 points / 15 p.) whereas the last one yielded the highest (10.5 / 15 p.).

The results of different rubrics indicate that 63% of the pupils choosing the persuasive topic belonged to the half of the pupils gaining lower points. The narrative/descriptive topic was the favourite of 58% of the pupils in the two upper quartiles, while about 15% of the pupils in the best quartile had chosen the argumentative topic.

If, again, attention is paid to the connection between the overall test results and to the selection of the rubric, it appears that the persuasive topic was chosen by more

than 50% of the pupils in the lowest quartile. The argumentative topic was chosen by 11% of the pupils in the highest quartile but, on the other hand, its representation in the lowest quartile was only 8%.

The relationship between the teachers' marks in English and the test results

Although the national test is not used for the same purposes as a school test, it is interesting to know how closely teacher assessments agree with the national test results. The correlation between the marks and the overall scores (0.80), and the marks and the scores of different sub-tests proved to be very strong. The correlations between different variables also proved to be high (range 0.63 – 0.80).

Table 10. Correlation of the girls' and boys' marks to different skills and the whole test

	Whole test	Reading comprehension	Listening comprehension	Structures	Writing	Speaking
Girls	0.79	0.64	0.60	0.78	0.72	0.71
Boys	0.79	0.67	0.66	0.76	0.73	0.72
All	0.80	0.66	0.63	0.78	0.74	0.73

In general, the national test seemed to assess the pupils' English skills quite well if the marks are regarded as criteria. The school mark explains 64 % of the variation of success in this test.

The highest correlation was observed between the marks and the results in the grammar and writing sub-tests. One explanation may be that in schools more emphasis is placed on teaching those skills, and they

are perhaps even assessed using tasks of the same type as the ones in the national test. As for speaking, arranging specific oral tests is probably rather rare in schools but judging by its fairly high correlation with the marks, a conclusion could be drawn that it is practised in school and the result is included in the mark under the criterion "work / performance during the lessons".

REFERENCES

1. * * * - National Board of Education *Framework for Evaluating Educational Outcomes in Finland*. Evaluation 8/1999. Helsinki: Yliopistopaino, 1999
2. * * * - National Board of Education *Framework Curriculum for the Comprehensive School 1994*. Helsinki: Painatuskeskus, 1994
3. Pasanen, U, - M. *Englannin ja ruotsin kielen taidoista peruskoulun päättövaiheessa*. (About the skills in English and Swedish at the final stage of comprehensive school) In: Jakku-Sihvonen, R., Lindström, A., Lipsanen, S. (Eds.) *Toteuttaako peruskoulu tasa-arvoa?* (Does comprehensive education lead to equality?) Opetushallitus, Arviointiraportteja 1/96, 352-362, 1996.
4. Perusopetus L. - *Perusopetuslaki 21.8.1998*. (Law about basic education)
5. Tuokko, E.. - *Peruskoulun 9. vuosiluokan englannin (A1-kieli) oppimistulosten kansallinen arviointi 1999*. Oppimistulosten arviointi 3/2000. Helsinki: Opetushallitus, 2000.
6. Weir, C. - *Understanding & Developing Language Tests*. Trowbridge: Redwood Books, 1993

Appendix 1**Criteria for Assessing Compositions****Contents**

- 5** – the contents cover the title very well and the subject matter is dealt with in a varied way
- 4** – the contents cover the title well and the subject matter is dealt with in a fairly varied way
- 3** – the contents cover the title well enough and are not disturbingly one-sided
- 2** – the contents cover the title only in part and only concentrate on a few facts
- 1** – the contents hardly or only partly cover the title
- 0** – no answer

Structure of the Text (organisation and clarity)

- 5** – the text is well organised (a clear introduction, good treatment of the topic, good ending and a clear division into paragraphs) and easy to read
- 4** – the text is well organised and easy to read
- 3** – slight imperfections in the organisation (less clear introduction or ending, need for improvement concerning division of paragraphs) which somewhat disturb readability
- 2** – obvious imperfections, which may cause problems in understanding (parts of) the text
- 1** – very poor organisation, the text may remain badly unfinished
- 0** – no answer

Linguistic Form

- 5** – varied use of grammar and vocabulary; very few mistakes

4 – fairly varied use of grammar and vocabulary; not too many mistakes

3 – partly one-sided use of grammar and vocabulary; a few mistakes which may cause problems in understanding the text

2 – restricted use of grammar and vocabulary; lots of mistakes causing problems in understanding the text

1 – very one-sided and faulty use of grammar and vocabulary causing serious problems in understanding the text

0 – no answer

Orthography

5 – spelling and punctuation nearly faultless

4 – minor spelling and punctuation mistakes

3 – uncertain spelling and punctuation, which may to some extent cause problems in understanding the text

2 – serious spelling and punctuation mistakes causing considerable problems in understanding the text

1 – extremely poor spelling and punctuation; hard to understand

0 – no answer

Appendix 2**Criteria for Assessing Speech****Fluency**

In this context "fluency" means primarily overall impression: how well the pupil succeeds in carrying out the communication task and how comprehensive the message is, and how naturally and fluently the pupil communicates. Note: rating is naturally adapted according to the task characteristics and to what is a reasonable expected level for 9th graders.

5 – The speech is natural and fluent, self-assured, possibly even vivid and expressive. Easy and pleasant

to listen to. The communicative task is carried out fully.

4 – Speech fluency and naturalness are close to normal, even if a few unnatural hesitations and pauses occur. The speech is easy to follow. The communicative task fulfilled almost fully.

3 – The speech is occasionally rather slow and hesitant but in other places fairly fluent and natural. The speech is relatively easy to follow. The core of the communicative task is fulfilled.

2 – The speech is generally rather slow and uncertain. The speaker must occasionally search even for common words. The listener is obliged to make quite an effort but even then parts of the message are quite hard to understand. The communicative task is fulfilled only partially.

1 – The speech is disjointed, uncertain and halting. Words may come one by one. The listener must take a lot of effort and even then comprehension is difficult. Only a small part of the communicative task is fulfilled.

0 – No answer

Grammar / Structures

5 – Grammatically accurate. Only occasional mistakes, mostly slips of tongue. Uses a broad range of taught grammatical structures, and may know more structures than actually taught.

4 – Grammatical accuracy quite good, but occasional obvious errors occur, including an occasional obvious mother tongue influence. A fairly good range of grammatical structures (within what has been taught).

3 – Grammatical accuracy fairly good, but some errors are still obvious. Occasional mother tongue influence may still be quite clear. Some variety in grammatical structures.

2 – Grammatical accuracy weak and the use of structures is quite inconsistent. Strong mother tongue influence evident. Almost no variety in grammatical structures.

1 – Practically no systematic sense of the grammatical structure of English. Uses only simple, unvaried sentence structure. Errors even in simplest sentences. Inconsistent use of structures. Mother tongue influence very strong.

0 – No answer.

Vocabulary

5 – Almost fully correct and idiomatic use of vocabulary. In relation to the taught syllabus, a broad range of vocabulary, which may go beyond what has been taught.

4 – Quite correct and fairly idiomatic use of vocabulary. In relation to the taught syllabus, a fairly broad range of vocabulary.

3 – Vocabulary is adequate to cope fairly well with the tasks. Mistakes occur but they do not basically distort understanding. In relation to the taught syllabus, no particular merit in terms of the range or idiomacy of vocabulary usage.

2 – Very limited vocabulary. Inaccuracies and inconsistency in vocabulary usage. Mother tongue influence is strong.

1 – Severely limited vocabulary. Frequent inaccuracies in vocabulary even in the simplest sentences. Inconsistent use of vocabulary. Mother tongue influence very strong.

0 – No answer.

Pronunciation

5 – Almost all individual words are correctly pronounced. Sounds are unambiguous and sufficiently well articulated for easy understanding. Appropriate word stress, stress-timing, and rhythm. Foreign accent, though still evident, does not impair understanding.

4 – Individual words are only occasionally mispronounced. Most sounds are close to those of a native speaker and sufficiently well articulated for utterances to be easily understood. Foreign accent in prosodic features is fairly noticeable.

3 – Some individual words are mispronounced. Most sounds are relatively close to those of the native speaker and articulated with passable accuracy for fairly effortless comprehension. Word-stress, stress-timing and rhythm still display some obvious problems. Foreign accent in prosodic features is quite noticeable.

2 – Many individual words are mispronounced and some individual sounds poorly articulated. Mother tongue influence of prosodic features is very noticeable. Phonetic inaccuracy occasionally impairs understanding, and listening demands quite a lot of extra effort.

1 – Wrong pronunciation of words is common. Individual sounds are often poorly articulated. Mother tongue influence on prosodic features (stress, intonation, rhythm, pauses etc.) is quite disturbing. Requires considerable effort from the listener.

0 – No answer.

Interaction

5 – Handles all interaction aspects quite skilfully. Responds and interacts appropriately. Makes the best of even problematic interaction episodes. Takes

frequent initiatives. Adjusts skilfully to the interlocutor's reactions. Uses appropriate forms of address, polite social conventions appropriate to the context; behaviour, language, idiom and register are modified to communicate in ways appropriate to the context, the emotive content of the occasion and the relationship of the speakers.

4 – On the whole interacts quite effectively. Is in touch with the discussion and is able to take part at all times. While still not fully effective in interaction, is still able to communicate with reasonable impact. Takes some initiatives. Adjusts quite readily to the interlocutor's reactions.

3 – On the whole interacts fairly effectively. Is in touch with most of the discussion and is able to take part most of the time. While not very effective in

interaction, is still able to communicate with some impact. Takes minimal initiatives. Adjusts fairly readily to the interlocutor's reactions.

2 – Level of interaction low but sufficient to make some communicative interaction possible. Responds in most cases but requires some tolerance and/or assistance from the interlocutor. Does not develop points of interaction in any or almost any way. Adjusts minimally to the interlocutor's reactions.

1 – Level of interaction just about as low as possible to constitute communication. Responds in most cases but requires a lot of tolerance and/or assistance from the interlocutor. Does not develop points of interaction in any way. Does not adjust practically at all to the interlocutor's reactions.

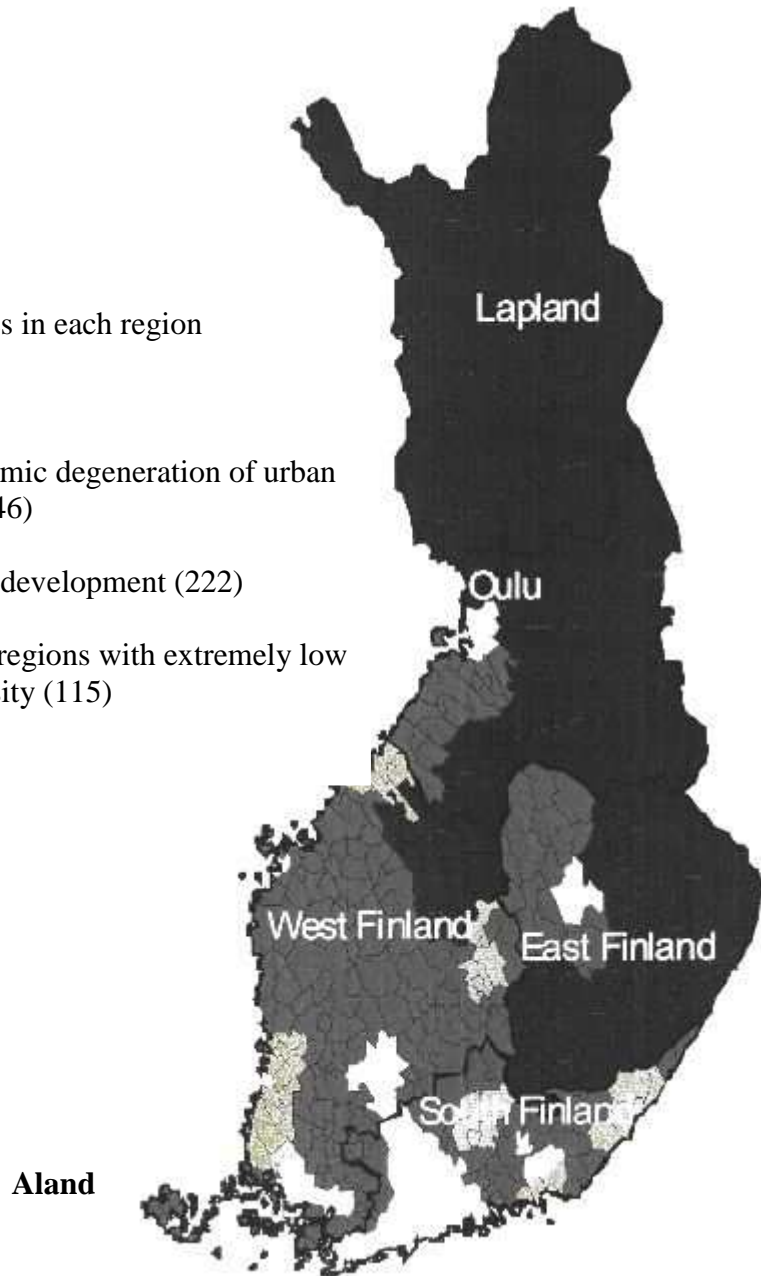
0 – No answer.

Appendix 3.

FINLAND BY EC OBJECTIVE REGIONS AND PROVINCES

Number of municipalities in each region

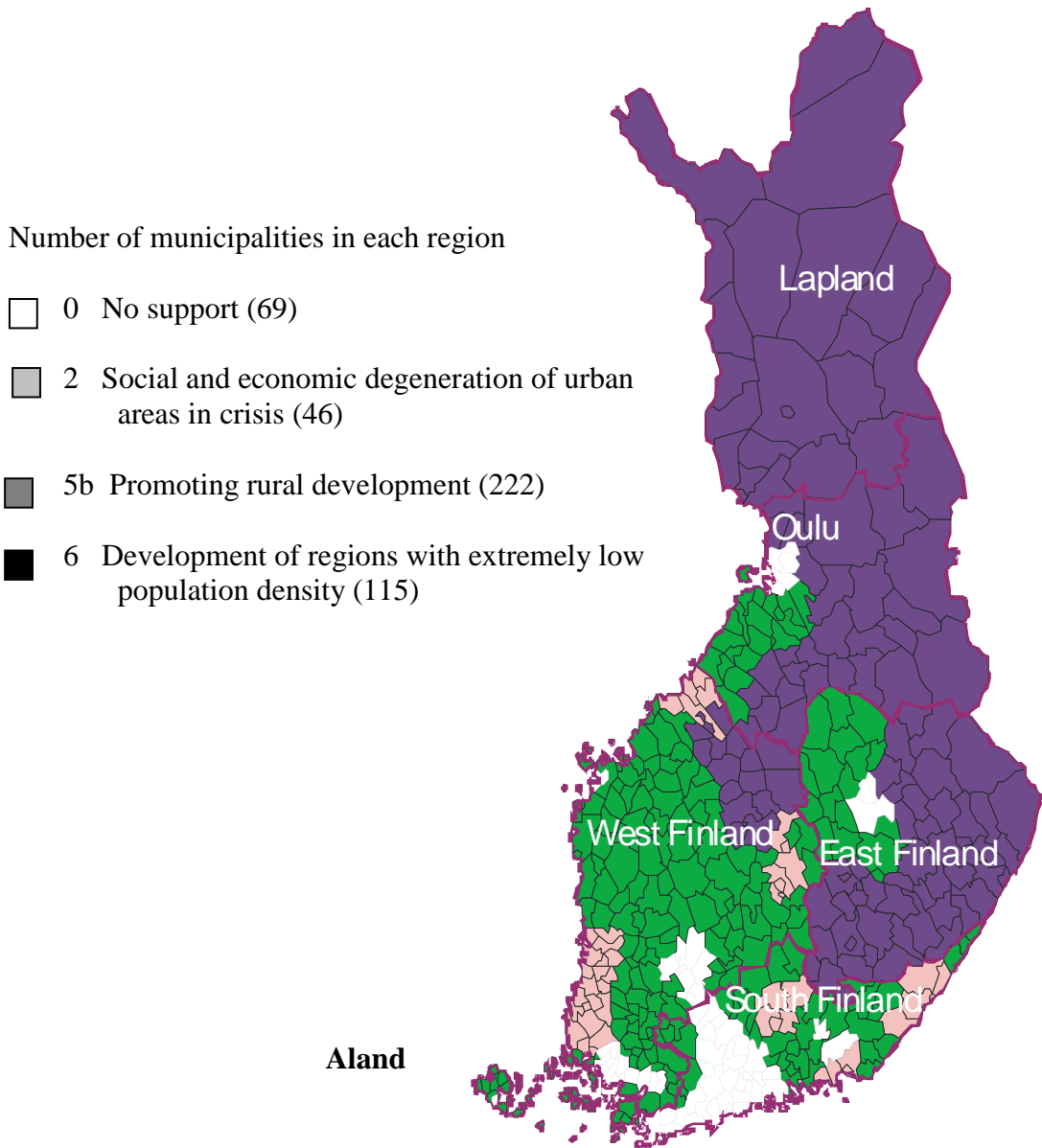
- 0 No support (69)
- 2 Social and economic degeneration of urban areas in crisis (46)
- 5b Promoting rural development (222)
- 6 Development of regions with extremely low population density (115)



National Board of Education

Appendix 3.

FINLANDA BY ECOBJECTIVE REGIONS AND PROVINCES



National Board of Education