TESTING COMMUNICATION SKILLS IN MIXED-ABILITY GROUPS

Viljo Kohonen (University of Tampere)

Abstract

This paper will first discuss some aspects of education, evaluation and communication with regard to the instructional goals of the Finnish comprehensive school. Problems involved in designing summative tests for mixed-ability groups will be discussed next, and an example will be given of a test of English administered to 7th-graders (N=104). It is suggested that communication is a "graded" property that is possible at various levels of accuracy. The essential criterion in scoring should be comprehension. If the development of communicative skills can be seen as a continuum ranging from "zero communication" to virtually bilingual competence, learners can and should be encouraged to work towards the quality of skills that is within their reach.

1. Some views of education and evaluation

1.1. Evaluation to improve learning

It is occasionally salutary to ask oneself the creative question, why test at all? Thought-provoking ideas have been offered by Bloom (1981), who makes a distinction between two major functions of education: selection and developing learner talent. Bloom points out that education has traditionally had a selective function, whereby the basic task of education was the "identification of the few who were to be permitted to enter and complete the secondary school academic program and then be admitted to higher education" (p.2). In this view, a central task of education is to classify learners for various "streams" of learning, under the assumption that learners are different and the school must select those that are capable of higher education. This guiding function is, of course, still one task of the school. But Bloom argues that it has become less important in developed societies in which a major proportion of the age group completes secondary education and there is an increasing demand for education for all citizens in a complex society. In this situation, education should primarily serve the development of the individual. Education should develop those characteristics in all students which will enable them to live effectively in a changing society. Thus the school should devote major resources to increasing the effectiveness of individuals rather than predicting and selecting talent:

> Education must be increasingly concerned about the fullest development of all children and youth, and it will be the responsibility of the schools to seek learning conditions which will enable each individual to reach the highest level of learning possible for him or her. (p.3)

Bloom further points out that the rapid change in modern society requires education to continue throughout life. Education must thus be seen as a lifelong process that never fully ceases (cf. Faure et al., 1972). Now,

these views of education are closely connected with three philosophies in evaluation: the selective, criterion-referenced and "learner-supportive" evaluation (cf. Warries 1982). The central question is what the learner's score is compared to: whether the scores of his peers, a pre-defined mastery level, or his previous learning.

The selective philosophy is, of course, well-known to all teachers and is deeply rooted in differential psychology and the statistical trend for the "normal" distribution of human abilities in large groups. It is connected with the selective function of the school. What is considered excellent, moderate or poor performance is determined by the empirical distribution of the scores. By definition, an "excellent" performance by some learners implies less excellent performance by some others. One can ask with good reason what are the affective consequences of the comparisons for the less able learners (cf. Bloom 1971a), and how necessary such comparisons really are in education. Information about rank order is obviously needed for admission to vocational institutions, educational programs or universities, i.e., in instances of restricted intake to further education. But such information is not needed within the institution itself.

In the criterion-referenced philosophy, the learner's performance is compared with a required level of mastery, based on the objectives of instruction. Instead of competing with his peers, the learner competes with the objectives of the learning task (cf. Takala 1985; Carroll 1971; Bloom 1971b, 1976; van der Linden 1982). This philosophy represents a vast improvement in educational thinking. Learning is seen as an individual effort of the learner towards the objectives. There are, however, some shortcomings: the definition of the behavioral domain is problematic , as is the definition of the required level of mastery (cf. Takala 1985). Further, in school conditions with fixed numbers of teaching periods available, it seems unavoidable that a number of learners will not reach the mastery level, while others will go beyond the common objectives. A problem for fast learners is, however, whether they have enough incentive to work for the "extended" objectives, after reaching the common objectives. Is there a danger perhaps that they may not work as hard as they could?

This is where the third, learner-supportive philosophy of evaluation would seem to offer some promising possibilities. In this approach, the learner "competes" with himself, to reach his personal objectives and augment his learning in comparison with his previous level of learning. This is essentially what Bloom seems to suggest in his second view of education: enabling each individual to reach the highest level of learning possible for him. This thinking leads to individualized learning whereby learners take on "learning contracts" that are reasonable within their total learning situation. In a sense, any learner that has efficiently utilized his potential is a "good" learner. This suggests bigger contracts for fast learners and smaller ones for slow learners. Such contracts should not, however, be forced on the learners, but should be based on their own responsibility and thus gradually lead to an attitudinal growth towards their being autonomous learners.

It is perhaps impossible, and even undesirable, to define exactly

what the "efficient" use of the potential means. Rather than viewing it as a "closed" phenomenon, it should be seen as an unfolding potential which depends on genetic, environmental and individual factors. It refers to the learner's total learning situation and includes ability factors, learning opportunity, home and school support, and the individual's needs and motivation (cf. Strevens 1980). We should beware of labelling our learners too hastily, as teacher expectations may involve well-known "self-fulfilling prophesies" (cf. Rosenthal and Jacobson 1968; Elashof and Snow 1971; Rosenthal and Rubin 1978; Babad et al. 1982). We should rather aim towards an open-ended, basically supportive view of our learners and encourage them to go as far as they wish and are able to go in their learning, within the possibilities allowed by their total learning situations.

1.2. On goals of education and foreign language instruction in Finland

In the new Comprehensive School Law (1983), the educational goals are defined in terms of a number of desirable learner characteristics to be attained by the end of the obligatory school. Specifically, the school should aim to educate pupils so that they are

- well-balanced
- responsible
- autonomous
- creative
- cooperative
- peace-loving

and physically healthy individuals and members of society. In addition, the school must provide education in morals and good manners and impart knowledge and skills that are necessary in their lives (§ 2). Language teaching aims are specified in the recent syllabuses (1982), whereby the general aims of foreign language teaching are defined as the "useful basic skills needed in everyday situations", a "positive attitude to language study" and a "continuous maintenance and development of the skills after school". Language teaching should also impart knowledge about the cultural background, increase international understanding, and further the pupil's ethic, aesthetic and socio-emotional growth. Language teaching is thus seen as language education, aimed at increasing the learner's skills of self-expression. Every language teacher is also a language educator. The syllabus contents include common objectives for all, and extra materials are offered for fast learners.

1.3. Learner communication as a "graded" property

Even though all learners work on the same common objectives, it is obvious that not all of them will reach the same level of communicative skills within the same amount of instructional time - there will always be learners for whom languages are easy and those for whom they are difficult. Thus they will end up with various levels of proficiency at the end of school. It is particularly the abstract rule system that poses difficulties for slow learners, while fast learners are still able to acquire both accurate and fluent language skills. Due to the historical tradition, we have been too much concerned with accuracy as the language teaching objective; we all know, of course, that communication is possible even with erroneous code. Errors will variously hamper communication and may

even cause it to break down in some more serious cases, but the essential criterion for communicative success must be **comprehension** and not grammatical accuracy.

Communication is a "graded" property which is possible at various levels of grammatical and lexical accuracy, ranging from mere non-verbal communication to fluent verbal and non-verbal communication. In fact, it is deficient lexis that hampers comprehension more than inaccuracies in pronunciation or grammar (cf. Littlewood 1984; Dulay et al. 1982). Accuracy should obviously be the goal to work towards, but it is not an end in itself. Equally important is that learners should have something to say and the courage to use the language, even at the risk of making mistakes. Accuracy is nothing more than a tool for efficient communication.

If accuracy may not be reached by all learners, everybody should be able to communicate something in the foreign language nonetheless. Thus some acceptable level of communication should be within everybody's reach. Now, communicative thinking would seem to offer a possibility for reconciling variation in learner ability and instructional objectives. We might consider how far it is advisable to try and push the abstract system with slow learners, knowing that communication is not a simple matter of "success or failure" in interactive face-to-face situations. As Henry Widdowson (1979) has repeatedly pointed out, it is a matter of negotiation about meanings. Thus, if I do not understand what is said to me I can always ask for clarification, just as my interlocutor can do when I cannot make my meaning clear. Successful communication requires courage, perseverance and communicative self-confidence. And such attitudinal objectives should also be included in our teaching and testing syllabuses.

Foreign language **teaching** should naturally provide normative feedback to learners, since they need to know how they are progressing in their learning tasks. But language **testing** need not be geared strictly to requirements of accuracy for **all** learners. Thus slow learners can - and should - be given some credit for any comprehensible language and effort. But by the same token fast learners should be required to attain high standards of accuracy, and they should be encouraged to work to this end. This means, then, accepting different learning outcomes for different learners, depending on their ability (and motivation) to learn languages.

2. Testing language skills in mixed-ability groups

2.1. Some guidelines for test design and scoring[1]

With the advent of the new Finnish school laws (from August 1985 onwards), mixed-ability teaching will be a reality in foreign languages and mathematics in the upper comprehensive school (grades 7-9) as well. This means that all learners will be offered the same common objectives within the same amount of time, and their skills will be tested using the same tests for all. The change has been preceded by a number of teaching experiments in various schools, in order to develop pedagogical solutions, testing techniques and learning materials. In our teaching experiments in Tampere, we have found that the following principles can serve as useful guidelines for designing and scoring tests for mixed-ability groups.

(1) What to measure? This is, of course, the basic question of the validity of the test. The central question is, what we mean by "knowing a language" and how we operationalize our views in concrete tests. Summative tests will inevitably indicate what we regard in our syllabuses as "worth learning". In communicative thinking this must be essentially the ability and willingness to put into a communicatively meaningful use whatever amount of language the learner has acquired. Tests must therefore be designed so that they give the learner opportunities to show what he can do with his language. This brings in the requirement of authenticity or, in any case, an attempt to simulate real-life tasks of language use, which is the goal for which we must prepare our learners. Tests must also be fair. That is, they should not contain any unnecessary gimmicks or tricks aimed at "catching" weaknesses in the learner's skills. While testing is always artificial to some extent, we should attempt to reduce artificiality as far as it is possible within the practical constraints of the classroom situation and the available resources.

(2) What kinds of tests to use? This question is a corollary to the above question of validity. Communicative skills can be measured by various kinds of tests, calling into play different aspects of communicative ability. We should therefore aim at using various types of tests, in order to get a many-sided coverage of the skills involved. Any single test type (or a limited number of types) will inevitably give a narrow measure of the skills. We should not lose sight of the communicative relevance of the tests. This is also important because of the well-known backwash effect of tests on teaching: our testing procedures should encourage communicatively oriented classroom work. Analytic tests could be used as formative tests of the learners' knowledge of the discrete points of grammatical structures. Integrative tests, on the other hand, would be better suited as summative tests, enabling learners to show what they can do with their language skills. This would provide a communicatively balanced basis for giving grades in the school reports.

(3) At what level of mastery? Following the familiar Bloomian taxonomy (cf. also Takala 1985), language processing can be thought of in terms of three levels:

- recognition of language items, understanding of meaning units
- mechanical skills, limited recall, guided production
- creative, autonomous, personal use of language in communication

For slow learners, simple recognition is already quite demanding, and they seem to need cues to guide their productive use of the language. Recognition of discrete structural forms and meanings by answering multiple-choice items is, of course, a relevant part of language competence, but this is rather a limited view of the variety of the language skills needed in real-life contexts. Such contexts will typically require open-ended, independent interpretations of meanings. Besides, if we think of language processing as a creative reconstruction of meanings, we also have to provide open items to measure such skills. Perhaps comprehension can be measured more directly by open tasks, since a learner who is able to answer open questions will usually pass multiple-choice items as well, while the converse is not necessarily true (for the well-known distinction between recognition and recall, cf. Anderson 1980). In this sense open questions are more economical, as they measure language processing at a deeper and more demanding level.

(4) How to score learner productions? Scoring problems are bound to arise in mixed-ability groups where learner responses will scatter over a wide range in terms of both quality and quantity of the language. To cope with the variation, some scale is necessary in scoring; mere 0/1 will not do justice to all. Besides, communicative thinking also suggests that language use is not a simple "wrong/right" matter; the same contents can usually be expressed in a number of possible ways, and what is involved is rather a scale of acceptability and communicative efficiency. The basic criterion for the application of the scale must be comprehensibility: has the learner understood the gist of the text? or would a speaker of the target language understand the intended meaning? If the answer is positive - despite incorrect spelling and bad grammatical mistakes - the learner has managed to interpret or process the message successfully and deserves some credit for it. Obviously, a zero will be given for a blank answer as well as for an answer that is clearly not interpretable. Thus, on a scale 0 - 3, for example, 3 points will be given for a "perfect" answer, 1 point for a comprehensible attempt, and 2 points for a variety of combinations of partly incorrect language and deficient message contents. There is no reason why slow learners could not answer in their mother tongue in listening and reading comprehension tasks if they have difficulties in writing. What is measured is their ability to understand target texts, and this can equally well be controlled by answers in the mother tongue. There is no reason to be dogmatic about the use of the mother tongue. More important is the communicative quality and meaningfulness of the tasks, and the learner's motivation for and involvement in completing them.

(5) How to transform the scores into grades? Communicative thinking leads to a criterion-referenced orientation, with communicative efficiency and the degree of attainment of the objectives as the criteria. The specification of the mastery level is a difficult task, and the application of the criteria to the evaluation of the learner language is always a matter of subjective interpretation. This entails problems of both intra-rater and inter-rater reliability. But such problems should not be exaggerated. It is a well-known problem that validity and reliability are difficult to maximize in the same test. While multiple-choice tests eliminate problems of inter-rater reliability, their validity is limited as a measure of the communicative use of the foreign language. Of the two important statistical properties, validity must be given the first priority in communicative testing. On the other hand, any communicative test must also meet sufficient requirements of reliability. Reliability can be increased by specifying the criteria and ensuring their consistent application by different raters.

In school contexts, evaluation usually involves a conversion of the scores into some scale of grades in the school reports. As noted above (Section 1.1.), relative grading should not be considered so "necessary" while the students are in school - society and higher institutions will need such information only in school-leaving reports. More important than learner comparisons is the tenet that evaluation should be geared to improving student learning. It seems very difficult to convert the learner-supportive evaluation philosophy into any commensurable scale of numerical grades that could be applied consistently to all in mixed-ability groups, but this might function as a framework for self-assessment by the learner himself. One possibility for converting the scores into school grades within a somewhat loose criterion-referenced orientation would be

to describe the pass grades (in the Finnish system, grades 5 to 10) at three broad levels of performance:

(1) level 1: "pass level" (grades 5 and 6)
(2) level 2: "common objectives" (grades 7 and 8)
(3) level 3: "extended objectives" (grades 9 and 10)

For the lower grades, accuracy standards are less important as long as communication is comprehensible, while accuracy is an important criterion for the two highest grades (9 and 10). Fast learners will thereby be encouraged to work both for accuracy and fluency. They can take on more demanding learning contracts by aiming towards qualitatively higher standards of language use. In addition to the written tests, the grades are also affected by other factors, such as oral skills (tested mainly through continuous assessment), contribution to classroom work, effort to learn, and diligence in doing homework. Language testing is thus seen as part of the school's wider educational task and the grades are also based to some extent on the educational goals discussed above (Section 1.2.).

2.2. Some empirical findings of a written English test (7th grade)

One aim of the Tampere experiments has been to develop possible testing techniques for mixed-ability groups. An attempt towards this aim is the test in English (given in Appendix 1), administered in Autumn 1983 in grade 7 (i.e., after some 300 lessons' instruction in English, learner age 14 years, N=104). The test consisted of the following parts:

(1) contextualized vocabulary test (20 items)
(2) contextualized past tense recognition and production (16 items)
(3) cued interview - ask five questions (5 items)
(4) cued dialogue completion (5 moves to complete)
(5) telling about past events (6 scorable points)
(6) reading comprehension test (5 open questions in Finnish)

The total number of items was 57, with the maximum score of 90 points. The statistical results of the test are given in Appendix 2.

In the vocabulary test, the first eight items were aided rocognition in context: the learners were given the Finnish word in brackets, with the English counterparts in the box on the exam sheet. They had to be able to identify the right word out of a large selection. In items 9-15, the Finnish words were still given, but the English counterparts had to be recalled. Items 16-20 were a cloze test, with acceptable-word scoring (0/1). The statistical analysis showed that the test functioned in a satisfactory way: the items were progressively more and more difficult (solution percentages from some 80 to some 30 per cent), and the reliability index was high (.94, internal consistency, Cronbach's alpha coefficient).

Test 2 (past tense formation) was easy (average solution percentage 76.4%) but functioned quite well, becoming more difficult towards the end (reliability index .90). Test 3 (cued interview) was scored on a 0-2 scale, to give some credit for partially comprehensible answers. It proved similarly quite easy (71.1%) and functioned well (reliability index .88). Tests 4 (dialogue completion) and 5 (telling about past events) were scored partly on a scale ranging from 0 to 3. They were more difficult (66.6% and 55.8%), but functioned well, particularly test 5, where the reliability index (.93) was good considering that the test consisted only of

6 items. It seems that open questions are an efficient measure of language skills. Even though they are more laborious to score, the amount of work is compensated to some extent by the fact that the test need not be long in order to be reliable. The reading comprehension test had two items (4 and 2) which were too easy and actually measured pragmatic knowledge rather than comprehension of the text; its reliability remained poor (.65). The reliability index of the total test battery was very high (.98), indicating consistent discrimination. Thus it can be said that the test as a whole functioned well statistically.

The frequency distribution of the scores shows that the test was relatively easy. The shape of the diagram is, in fact, a typical mastery learning curve which is positively skewed. Thus a majority of the pupils obtained quite high scores: the average score was 60.5 (with s= 23.8), which is 67.3 per cent of the maximum total score (90 points). The conversion of the scores into the grades was done so that the cut-off points for the grades were certain proportions of the learner score out of the maximum total score. The cut-off points for the four broad levels of the grades were as follows:

score:		level	grade
below 24	points	Fail	(4)
24 - 56	"	Pass	(5 - 6)
57 - 78	"	Good	(7 - 8)
79 - 90	"	Very good	(9 - 10)

It will be seen, then, that the scale was more lenient at the lower end of the score distribution, while very good grades required accurate language use as well. The cut-off points were arbitrary and based on traditionally used proportions of the learner scores out of the total score. They were thus not based on mastery descriptions, while attemps were made to ensure that the common objectives were reached by those learners who got a "good" grade (above 7), i.e. beyond 57/90 points (thereby getting a minimum of 63% out of the total score).

These results indicate, then, that the present test is one possible way of measuring written language skills in mixed-ability groups. But it is just one possibility that needs to be developed further. In retrospect, it is still too much oriented towards testing grammatical structures (in this case, the past tense). It would seem more advisable to test the mastery of analytic grammatical structures in formative tests and thereby shift the emphasis in summative tests towards evaluating the use of global skills. As noted above, this would give a communicatively balanced basis for the grades in the reports. To make the tests communicatively more meaningful for the learners, more opportunities should also be provided for a personal and creative use of the language.

Another shortcoming is the lack of a systematic evaluation of the spoken skills. Any written test is bound to give only indirect evidence of the learner's abilities for face-to-face interaction. As this would involve arranging interview tests, which are quite laborious in large learner groups, it has not been possible to do it so far. Thus the testing of spoken skills has remained more informal, on the basis of teacher observation during the lessons. Perhaps it will be possible to make such observations more systematic and thereby more reliable by developing criteria for continuous assessment.

3. Discussion

When a foreign language is taught to the whole of the age group learner performance will be graded in terms of both the quality and quantity of the language produced, as it is in the mother tongue communicative skills as well. On the other hand, communication is also a "graded" property in the sense that it is possible at a variety of accuracy levels. Errors will variously hamper communication, but language is so redundant that it "tolerates" a fair amount of mutilation before communication really breaks down. Besides, in normal interactive situations, it is usually possible for the interlocutor to check his comprehension by asking. Language use is thus a matter of negotiation about meanings. While language teaching should naturally be normative and thereby have accuracy as an aim, accuracy should not be seen as an end in itself, but rather as a tool for efficient communication.

These ideas about learner variation and communication could lead to a somewhat individualized approach whereby different learners take on, as it were, different learning contracts in terms of the quality of language skills aimed at. Thus fast learners should be encouraged to work for both an accurate and fluent command of the language, thereby utilizing effectively their potential for learning an abstract rule system. For slow learners, on the other hand, it would seem equally advisable to attempt to put into communicative use that amount of language which is within their reach, thereby giving them positive experiences of being able to use the language in a comprehensible way. Communication is thus a relative concept, and everyone should be encouraged to proceed as far as possible within his total learning situation. How far such an individualized approach can be realized in practice depends on the constraints imposed by classroom realities. But it would seem to be a possibility that is worth exploring.

A second important point is the role of affective factors in language use. Language use always involves an element of ambiguity tolerance and risk-taking. Comprehending unheard (and unseen) messages and producing one's own messages requires negotiation skills, perseverance and willingness to attempt communication. Using an acronym, communication could be described as a "SWAP" phenomenon, where S stands for "having something to say", W for "willingness to attempt communication", A for "ability to do so", and P for "personal payoff from so doing". If such attitudinal aspects are to be learned in school, we should be aware of their existence and think of pedagogical ways of developing them in classroom work. This will involve conscious attention to the quality of the learning atmosphere, and it will lead to improvised, personal language use and an extensive treatment of texts, both aural and written. Communicative competence includes an element of communicative self-confidence. This point is also made by Dulay et al., who point out that "All things being equal, the self-confident, secure person is a more successful language learner" (1982:75). A problem with testing is that such attitudinal attributes do not easily lend themselves to objective and reliable measurement. But the problems of quantifiability should not dictate our testing procedures, if we are to evaluate our learners' total communicative skills. As noted above, validity should be given the first priority in testing, while ensuring sufficient reliability at the same time. We need to be critical about the kinds of skills measured by our tests.

An further important point about communicative testing is that it should not undermine the learner's self-confidence and discourage his attempts to use the language. We should therefore try to give him positive feedback about the development of his skills, rather than pinpointing errors in his performance. Concentration on errors will put into focus what the learner does not know, thus leaving aside the important aspects of what he already knows and how much he can communicate with the foreign language. Communicative testing should allow room for both "hard" and "soft" aspects of evaluation. We must bear in mind that summative tests will inevitably tell learners what we regard as "worth learning" in communicative skills, and they will exercise a powerful backwash effect on teaching. Our tests should encourage communicatively oriented classroom work. In practice we have to make compromises due to existing resources. But such compromises should be made with an awareness of what could be done, and how much of it can be done with available resources.

In mixed-ability groups, tests should be "tailor-made" for the different levels of learner performance. In our approach, the notion of "graded tests" refers to a deliberate and careful design of the items at the hierarchical levels of language processing, ranging from simple recognition to free production. Emphasis has been shifted onto integrative skills, thus enabling learners to show what they can do with their language. A second important point is that learners should be allowed and encouraged to proceed as far as they can. Thus they can attempt to solve any items as far as they wish. This is in accordance with our conviction that we should avoid imposing limits on learner performance. We prefer grading to be tactful, something that takes place automatically in the learner's mind when he sees that he cannot solve the items any further.

Communication is not just a matter of knowing the language system. It is equally importantly a matter of having something to say and the courage to attempt communication, a willingness to engage in discourse. There are numerous possible and acceptable ways of sharing in cross-language communication. It is our task as teachers to help our learners to find their ways and get positive learning experiences from their discovery of the foreign language.

---

1 I am grateful to the experimenting teachers at Kaukajärvi and Pikkola Comprehensive Schools (Arja Haavisto, Meeri Halttunen, Ulla Hartiala, Irmeli Kalli, Arja Kauppila, Kaisu Kuronen, Ulla Liljeström, Pentti Niemistö, Ulla Nordberg, Ulla Tammi, Riitta Tuukkanen) for lively discussions at several joint seminars.The paper reflects, indeed, the shared insights arrived at in these discussions. I am also grateful to Mr Ian Gurney for comments on this paper, and to Mrs Jaana Rantanen for help with the computer processing of the empirical data.

Bibliography

Altman, H. and V. James (eds). 1980. Foreign language teaching: meeting individual needs. New York: Pergamon Press.

Anderson, J.R. 1980. Cognitive psychology and its implications. San Francisco: W.H. Freeman and Company.

Babad, E.Y. et al. 1982. "Pygmalion, Galatea and the Golem: investigations of biased and unbiased teachers". Journal of Educational Psychology 74: 459-74.

Block, J. (ed.) 1971. Mastery Learning. New York Holt, Rinehart & Winston.

Bloom, B. 1971a. "Affective consequences of school achievement". In Block, J. (ed.) 1971. 13-28.

Bloom, B. 1971b. "Mastery learning". In Block, J. (ed). 1971. 47-63.

Bloom, B. 1976. Human characteristics and school learning. New York: McGraw-Hill.

Bloom, B. et al. 1981. Evaluation to improve learning. New York: McGraw-Hill.

Carroll, J.B. 1971. "Problems of measurement related to the concept of learning for mastery". In Block, J. (ed). 1971. 29-46.

Carver, R.P. 1974. "Two dimensions of tests: psychometric and edumetric". American Psychologist 28: 512-18.

Cziko, G.A. 1981. "Psychometric and edumetric approaches to language testing: implications and applications". Applied Linguistics 2: 27-44.

Dulay, H. et al. 1982. Language Two. New York: Oxford University Press.

Elashof, J.D. and R. Snow. 1971. Pygmalion reconsidered. Worthington: C.A. Jones.

Faure, E. et al. 1972. Learning to be. Paris: UNESCO.

Harding, A. et al. 1980. Graded objectives in modern languages. London: Centre for information on language teaching and research.

Kohonen, V. 1983. "Learners, teachers and graded objectives in communicative language teaching". In Nyyssönen and Mauranen (eds). AFinLA Yearbook 1983. 37-67. AFinLA Series of Publications No 36. Oulu and Helsinki.

Kohonen, Viljo. 1984. "On cognitive achievement and self-concept in various learner groups: some empirical findings". In Differentieringsfrågor i språkundervisningen. Seminarierapport nr. 11. 108-49. Nordiska språk-och informationscentret, Helsingfors.

Koululait 1983. Lakikokoelma 13. Helsinki: Valtion painatuskeskus.

van der Linden, W. J. 1982. "Criterion-referenced measurement: its main applications, problems, and findings". Evaluation in Education 5: 97-118.

Littlewood, W. 1984. Foreign and second language learning. Cambridge: CUP.

Peruskoulun englannin kielen oppimäärä ja oppimääräsuunnitelma. Kouluhallitus 1982.

Porter, D. 1983. "Assessing communicative proficiency: the search for validity". In Johnson, K. & D.Porter (eds.), 1983. Perspectives in communicative language teaching. New York: Academic Press. 189-209.

Rask, S. 1982. Arviointi oppilaan tukena. Vantaa: Kunnallispaino.

Ropo, Eero. 1984. Oppiminen ja oppimisen tyylit. Acta Universitatis Tamperensis, ser. A vol. 172. Tampere.

Rosenthal, R. and L. Jacobson. 1968. Pygmalion in the classroom. New York: Holt, Rinehart and Winston.

Rosenthal, R. and D.B. Rubin. 1978. "Interpersonal expectancy effects: the first 345 studies". The Behavioral and Brain Sciences 3: 377-415.

Strevens, P. 1980. "The paradox of individualized instruction: it takes better teachers to focus on the learner". In Altman, H. and V. James (eds) 1980. 17-29.

Takala, S. 1985. "Criterion-referenced testing in foreign-language teaching". This volume.

Warries, E. 1982. "Relative measurement and the selective philosophy in education". Evaluation in Education 5: 191-203.

Widdowson, H. 1979. Explorations in applied linguistics. Oxford: OUP.

Sample test / English / 7th Grade name_____
Autumn 1983

Test 1: Fill in the missing words from the box:

	0	1
1)		
2)		
3)		
4)		
5)		
6)		
7)		
8)		

Paul and his friend Simon 1)_____ (haluavat (want)) to buy an old boat. Their father can help them to 2)_____ (korjaamaan) _____it. (repair)
They get the 3)_____ (paikallis (local)) paper every day. Today they see an interesting 4)_____ (ilmoitus (advertisement)) for an old rowing boat. They are very 5)_____ (kiireisiä (busy)). It's 6)_____ (hauskaa (fun)) to go to see the boat. It is 7)_____ (todella) _____old, but very 8)_____ (halpa (cheap)) so they buy it. (really)

repair, mean, hurt, loud, reach, bonnet, nearest, prize, usually, busy, local, game, want, perhaps, really, interested, fun, problem, harvest, cheap, expensive, fine, advertisement

Fill in the missing words according to the cue in Finnish:

	0	1
9)		
10)		
11)		
12)		
13)		
14)		
15)		

Sheila Jackson lives in the country. The Jacksons 9)_____ there 10)_____ (viljelevät (grow)) and 11)_____ (vehnää (wheat)) (kauraa (oats)) There are cows there, too, and Sheila helps to 12)_____ (lypsämään) _____them. Because Roger is the 13)_____ (nopein (quickest)) runner in his school he is often in London in summer. He usually wins the 100 meters 14)_____ (kilpailu (dash)). His friend Michael always has a good 15)_____ (istumapaikka (seat)) when Roger runs.

Fill in words that suit to the context:

	0	1
16)		
17)		
18)		
19)		
20)		

Paul and Simon have parties in an old barn. There is no 16)_____ in the barn, but they have a 17)_____ so they can dance. Paul's 18)_____ music is old rock but Simon is a heavy rock 19)_____. They 20)_____ their friends to their parties.

score:_____/20

Test 2: Read the text and underline all verbs in the past tense, both regular and irregular.

Mr Glenn went shopping yesterday. He needed a new coat. In the shop he saw many nice coats, but they were all quite expensive. He tried on a dark blue coat, but it looked too big. The lady showed him a smaller coat and Mr Glenn liked it a lot. He bought it and drove home.

	0	1
21)		
22)		
23)		
24)		
25)		
26)		
27)		
28)		
29)		
30)		

score: _____ / 10

Look at the pictures and tell what your friend did last summer.

	0	1	2
31)			
32)			
33)			
34)			
35)			
36)			

Last summer my friend 1) _____ a good book. Then she 2) _____ some strawberries and 3) _____ the car. Then 4) _____

_____

5) _____

6) _____

score: _____ / 9

Test 3: Jill tells about herself. Ask her name, age, where she lives, etc. You can also ask her other things than those mentioned in the text. Ask at least 5 questions!

My name is Jill Harper. I'm 12. I live in Hartford. It is a town near London. I've got one brother, his name is Jack. I go to school in Hartford. I like it very much. After school i play tennis with my friend Terry. Terry usually wins, because she is better than me.

	0	1	2
37)			
38)			
39)			

1) _____

2) _____

3) _____

4) _____
5) _____
_____

	0	1	2
40)			
41)			

score: _____/ 10

Test 4: John rings Tom. Write down what John says to Tom.

Tom: Brighton 653144
John: Hello, 1) _____. 42)
täällä puhuu John (this is John speaking) It's about the match.
Tom: oh yes. When is it?
John: 2) _____. 43)
Tänä iltana seitsemältä (tonight at seven)
Tom: Do you think Bob will win?
John: 3) _____ 44)
Viime vuonna Philip oli ensimmäinen (last year Philip won)
4) _____ 45)
Teddy toinen ja Bob viides (and Teddy came second and Bob fifth.)
Tom: But I think he is much better now.
John: Yes, I hope so. 5) _____ 46)
Tavataan stadionilla (we'll see at the stadion)
_____.

	0	1	2	3
42)				
43)				
44)				
45)				
46)				

score: _____/ 13

Test 5: Tell about last summer's events: where you were and what you did. You can tell, for instance, about travelling, sports or your hobbies or about the weather. Mention at least six events.

1) _____ 47)
2) _____ 48)
3) _____ 49)
4) _____ 50)
5) _____ 51)
6) _____ 52)
_____
_____

	0	1	2	3
47)				
48)				
49)				
50)				
51)				
52)				

score: _____/ 18

Test 6: Read the text carefully and then answer the questions in Finnish.

Sport is a very important part of life in an English school. Nearly all schools have their own sports fields. Each school has its own teams, and they play against the teams of other schools. Most games between schools are on Saturdays. Many of the pupils who are not playing come to watch their school team.

In winter boys play football or rugby. Rugby is a very hard game, in which the players carry the ball in their hands. There are many famous football clubs in Britain, and even people in other countries follow their favourite English team.

If you go to England in summer, you can see men in white clothes playing on a big round field. They use a thick wooden bat to hit a small red ball. This game is cricket. Nearly every village has its own cricket team, and there are matches every weekend. To foreigners cricket seems a slow game - an international match can take five days.

0 1 2

1) Mitä kaikilla englantilaisilla kouluilla on?
(what have all English schools got?)

_____ 53)

2) Milloin ottelut yleensä pelataan?
(when are games usually played?)

_____ 54)

3) Mitä ne oppilaat tekevät, jotka eivät pelaa?
(What do those pupils do who do not play?)

_____ 55)

4) Miten palloa kuljetetaan rugbyssä?
(How is the ball carried in rugby?)

_____ 56)

5) Millainen peliasu on kriketinpelaajilla ?
(what do cricket players wear?)

_____ 57)

Your total score:_____/ 90

score:_____/ 10

Appendix 2.

Item Analysis

ITEM No.	Item corr.	Facility Value (%)
TEST N.O 1 (Vocabulary)		
1/ 1	0.67	80.77
2/ 2	0.73	81.73
3/ 3	0.68	85.58
4/ 4	0.55	84.62
5/ 5	0.69	75.00
6/ 6	0.53	72.12
7/ 7	0.64	77.88
8/ 8	0.69	74.04
9/ 9	0.73	69.23
10/ 10	0.84	60.58
11/ 11	0.78	69.23
12/ 12	0.62	77.88
13/ 13	0.63	74.04
14/ 14	0.68	74.04
15/ 15	0.72	73.08
16/ 16	0.74	59.62
17/ 17	0.72	60.58
18/ 18	0.46	36.54
19/ 19	0.51	45.19
20/ 20	0.45	31.73
Mean	0.65	68.17
s	0.10	14.73
TEST N.O 2 (Past tense)		
1/ 21	0.60	86.54
2/ 22	0.58	91.35
3/ 23	0.68	74.04
4/ 24	0.53	59.62
5/ 25	0.52	81.73
6/ 26	0.56	88.46
7/ 27	0.56	88.46
8/ 28	0.58	84.62
9/ 29	0.62	82.69
10/ 30	0.57	75.96
11/ 31	0.42	61.54
12/ 32	0.63	77.88
13/ 33	0.62	80.77
14/ 34	0.79	66.35
15/ 35	0.71	56.25
16/ 36	0.70	66.83
Mean	0.60	76.44
s	0.09	10.84
TEST N.O 3 (Cued Interview)		
1/ 37	0.73	75.96
2/ 38	0.66	78.85
3/ 39	0.71	71.15
4/ 40	0.75	65.87
5/ 41	0.67	63.94
Mean	0.70	71.15
s	0.03	5.70
TEST N.O 4 (Dialogue compl.)		
1/ 42	0.66	81.73
2/ 43	0.66	63.78
3/ 44	0.76	75.00
4/ 45	0.57	83.65
5/ 46	0.52	28.85
Mean	0.63	66.80
s	0.08	20.12
TEST N.O 5 (Past Events)		
1/ 47	0.77	61.22
2/ 48	0.80	57.37
3/ 49	0.79	60.58
4/ 50	0.79	54.17
5/ 51	0.86	51.92
6/ 52	0.77	49.36
Mean	0.80	55.77
s	0.03	4.36
TEST N.O 6 (Reading Compr.)		
1/ 53	0.32	44.23
2/ 54	0.44	84.62
3/ 55	0.50	79.81
4/ 56	0.26	91.35
5/ 57	0.53	63.46
Mean	0.41	72.69
s	0.10	16.95

Correlations Between Tests (7 = Total Scores)

	1	2	3	4	5	6
2	0.904					
3	0.844	0.851				
4	0.846	0.817	0.792			
5	0.842	0.866	0.773	0.799		
6	0.580	0.567	0.523	0.503	0.514	
7	0.958	0.955	0.894	0.895	0.925	0.651

Test	Items	Mean	s	Reliability
1	20	13.635	6.097	0.942
2	16	14.125	5.196	0.902
3	5	7.115	3.017	0.875
4	5	8.337	3.362	0.822
5	6	10.038	6.009	0.930
6	5	7.269	2.661	0.651
7		60.519	23.801	0.977

Frequency Distribution of Total Scores