

LT+25: A Language Testing Symposium in Honor of John B. Carroll & Robert Lado
Sauli Takala

Abstract

Testing of Written Communication

It is proposed that a joint presentation be prepared by the authors. Sauli Takala's presentation will give an account of developments in the testing of writing. The paper will first present a theoretical model of written communication and outline the domain of school-based writing. It will then address the following topics: setting and analyzing writing assignments, methods and issues related to the rating of compositions (validity, reliability) and present some empirical findings of a 14-nation empirical study of written composition, which is being conducted by the International Association for the Evaluation of Written Composition (IEA) and which is coordinated by Takala. Complementing the presentation, Liisa Havola will take a more integrative point of view by exploring some essential relationships between reading and writing. After a review of the topic, she will use writing of summaries as an empirical example of such a relationship and present some pilot results of her large-scale study.

Check Kelly for early history

Let me begin my presentation on a personal note. I came to work on applied linguistics quite by chance. In the mid-1960's the first steps were taken in Finland to introduce foreign languages into the elementary school. There was interest in knowing things: (1) Does the early introduction of a foreign language have a detrimental effect on the development of the pupils' first language? (2) What and how much do the pupils learn of what they are taught in L2 lessons? To get answers to these questions the National Board of

General Education commissioned a study from the recently established Institute for Educational Research, University of Jyväskylä, with a grant that covered the salary costs of one research assistant. This was the beginning of commissioned educational research in Finland.

When I was once leaving a lecture on the language of the old Swedish Bible, I was addressed by a gentleman who asked me point-blank whether I would be willing to take on a temporary job as a research assistant at the Institute for Educational Research. I was baffled by the question having English and Swedish philology as my majors. I explained that I did not have the necessary background. He told me that the job concerned the construction of language tests in English and that my professor had recommended me as a suitable candidate. I needed the money and accepted.

When I started the job, it turned out that the person from whom I was to take over had been to Ann Arbor and got acquainted with the Lado testing approach. I was given Lado's book "Language Testing" (the 1962 Longman edition) and a set of Michigan tests of foreign languages. Thus my entry into the field of applied linguistics was through language testing and Lado's book was my "Bible". I have never ceased to appreciate the contribution that Lado made through his systematic and consistent approach.

Another early influence was John B. Carroll. His impressive reviews of research literature (Carroll, 1963) his systematic outlines of research problems (Carroll, 1967), his theory-based analyses of constructs (Carroll, 1976, 19***, reading comprehension) and his taxonomies (Carroll, 1968; 1977**) and his expertise in test construction (Carroll & Sapon, 19**) and his psychometric prowess (19**) had a powerful influence on my own thinking. Both Lado and Carroll had, therefore, an undisputed place on the list of the heroes of the L2 profession that I recently presented at the FIPLV ****th world conference of language teaching in Helsinki (Takala, 1985).

In more recent times, the role of speaking and hearing was clearly

emphasized at the expense of reading and, especially writing. Thus, e.g., the syllabus for the upper secondary schools in state of Hessen stated (1957) that listening and speaking precede reading and writing. The instructions for Hamburg from the same period specify that oral exercises are central in language study and that written exercises grow from the oral ones. The influential Ankara conference (1966), sponsored by the Council of Europe, recommended that students should be able to write what they are able to say.

Finocchiaro (1977) suggested that writing should be taught and practised only to a limited extent in the teaching of foreign languages in primary grades.

Several problems have occupied those researchers who have been working on the teaching and assessment of writing. Among them are the following:

- (1) How can writing ability be defined or at least delimited?
- (2) Is writing ability one unified construct or can it be measured by measuring its different components, for instance, factors related to content and form?
- (3) If writing ability is measured by way of components, how should they be weighted, if at all?
- (4) How can a test instrument with sufficient validity and reliability be constructed?

How Has Writing Been Defined?

Lado (1962) has defined the ability to write as follows: "

We will then define writing a foreign language as the ability to use the language and its graphic representation productively in ordinary writing situations. More specifically we mean by writing a foreign language the ability to use the structures, the lexical items, and their conventional representation, in ordinary matter-of-fact writing.

Lado made a clear distinction between creative writing and ordinary

writing. Also, consistently with his habit emphasis and habit transfer, Lado believed that the testing of writing could be advanced best by listing the particular problems that a writer's particular linguistic background was expected to create.

Lado distinguished between an integrated method of testing writing by asking students to produce a connected piece of writing and a method of testing writing with separate factors such as punctuation, spelling, structure or vocabulary. This latter method would make it possible to sample the problems systematically. Lado recognized, however, that the validity of the synthetic approach was not readily conceded and he discussed ways of improving the objectivity of scoring composition tests.

Lado recommended a manysided test of writing, and suggested the following as one possible design: (1) Objective, partial production, multiple-choice items (50-80) dealing with specific problems of spelling, punctuation, grammatical structure, and vocabulary. (2) Twenty or thirty items of the objective, partial production type on a single connected passage testing chiefly matters of sequence and transition signals. (3) Three pictures with instructions to write a paragraph about each with grading based on mechanics only (= number of errors per 100 words). (4) Two short compositions on assigned topics (30 minutes each) with grading based on style, content and mechanics.

It is assumed that the student has something to say and a personal point of view. The student must observe the normal requirements of form and present his views effectively. According to Harris (19) writing is a complex skill, which must simultaneously take into account several points: 1) content, 2) form, 3) grammar, 4) style, 5) mechanics. Valette (1967) ***** Grieve (19) ***** Pilliner (19) *****

Carroll, J. B. 1961. "Fundamental considerations in testing for English

- language proficiency of foreign students." In *Testing the English proficiency of foreign students*. Washington, D.C.: Center for Applied Linguistics, 30-40.
- 1963. "Research on teaching foreign languages." In Gage, N. L. (ed) 1963. *Handbook of research on teaching*. New York: Rand McNally, 1060-1100.
 - 1976. "Psychometric tests as cognitive tasks: A new "structure" of intellect". In Resnick, L. B. (ed) 1976. *The nature of intellect*. Hillsdale, N.J.: Erlbaum, 27-56.
 - 1976. "Promoting language skills: The role of instruction." In Klarh, D. (ed) 1986. *Cognition and instruction*. Hillsdale, N.J.: Erlbaum, 3-22.
 - 1968. "The psychology of language testing." In Davies, A. (ed) 1968. *Language testing symposium: A psycholinguistic approach*. London: Oxford University Press, 46-69.
 - 1967. "Research problems concerning the teaching of foreign or second languages to younger children." In Stern, H. H. (ed) 1967. *Foreign languages in primary education*. London: Oxford University Press, 94-109.
 - 1965. "The contributions of psychological theory and educational research to the teaching of foreign languages." *Modern Language Journal* XLIX, 5: *****

First Draft

TESTING WRITING ABILITY: A REVIEW

Sauli Takala
 Institute for Educational Research
 University of Jyväskylä

1. Some Basic Issues in the Testing of Writing Ability

Several problems have occupied those researchers who have been working on the teaching and assessment of writing. Among them are the following:

- (1) How can writing ability be defined or at least delimited?
- (2) Is writing ability one unified construct or can it be measured by measuring its different components?
- (3) If writing ability is measured by way of components, how should they be weighted, if at all?
- (4) How can good writing tasks be constructed?
- (5) How can valid and reliable rating methods be developed?

2. A Brief Historical Sketch

As Kelly (1969) notes, in classical times the peak of education was the art of rhetoric, which combined artistry in word use, logical reasoning, and, usually the techniques of public speaking. In classical times, what was written was usually also read aloud and elocution was an important part of training.

Kelly also suggests that throughout the history of language teaching, four types of exercise have been used in teaching composition: transcription and consequent rote learning of models, structural variation of models, imitation of masters, and original writing.

 Paper *planned to be* presented at LI+25: A Language Testing Symposium in Honor of John B. Carroll & Robert Lado, Quiryat Anavim, Israel, May 11-13, 1986; *was not presented as I could not attend the symposium*

In medieval times the practice of verse composition held an important position in Latin and Greek, but in more recent times prose writing has totally eclipsed verse writing, which Milton would have approved. Medieval rhetoric concentrated on written composition, following the teaching of Quintillian and Cicero's *Topica*. However, in the 1800's free composition, which had been increasingly criticised, was largely replaced by text exegesis and translation. While translation had in the time of the Renaissance been advocated as a useful method of cultivating stylistic consciousness, it later became to be used to teach more elementary skills of making correct sentences and joining them together. Teachers were recommended to analyse carefully shortcomings in student writing. Translation was acquiring a basically negative stance. By the end of the 19th century, translation always preceded free composition or totally ousted it from the curriculum. A long way had been travelled from an emphasis on ideas and graceful expression (feel for language) to an emphasis on correct structure and linguistic equivalence.

In more recent times, the role of speaking and hearing was clearly emphasized at the expense of reading and, especially writing. Thus, e.g., the syllabus for the upper secondary schools in state of Hessen stated (1957) that listening and speaking precede reading and writing. The instructions for Hamburg from the same period specify that oral exercises are central in language study and that written exercises grow from the oral ones. The influential Ankara conference (1966), sponsored by the Council of Europe, recommended that students should be able to write what they are able to say.

Finocchiaro (1977) suggested that writing should be taught and practised only to a limited extent in the teaching of foreign languages in primary grades.

Most testing experts have not addressed the situation beyond the elementary or intermediate stages of language teaching and learning. Yet, hundreds of thousands of students need to write a lot in a language which is not their first language. This applies to those countries where the language of in-

struction is one of the many dialects of a country, or a created standard language, or a language of the former colonizing power. Another group affected is the students who go to study abroad and during the course of their studies need to answer written examinations, write term papers, and write theses. A third group are those who, after completing their professional education, need the ability to produce at least the first draft of letters, memoranda, contracts, papers, instructions, etc. As international contacts intensify and the language skill requirements increase, the literate bias of our own post-industrial culture tends to make the skills related to written language more and more important.

3. How Has Writing Been Defined?

Lado (1962) has defined the ability to write as follows:

We will then define writing a foreign language as the ability to use the language and its graphic representation productively in ordinary writing situations. More specifically we mean by writing a foreign language the ability to use the structures, the lexical items, and their conventional representation, in ordinary matter-of-fact writing.

Valette (1967) considers writing to be the most sophisticated of the four language skills. According to her, communication through the written word "possesses a certain degree of finality and demands real proficiency from the writer if it is to be effective" (p. 131). Valette took a developmental point of view in her recommendations concerning the testing of writing. Thus tests should be structured so that they measure the various aspects of student progress of acquiring the writing skill: the mechanics - vocabulary, spelling, grammar - have to be acquired before the student can aspire to precision of expression, fluency, and style. (**** note how correctness, rather than, communicative effectiveness seems to dominate her thinking here****) Valette lists a number of ways testing partial aspects of writing, much in the style of Lado. In discussing composition, she states that "a composition measures the student's ability to organize his thoughts, to choose his vocabulary, to formulate his sentences - in short to commit his ideas to paper" (p. 157). She notes problems related to the amount of time needed for scoring and the

objectivity of scoring. Among composition tasks she mentions "point of view" composition (physical descriptions, emotional states), letter-writing conventions, and thought-provoking essays.

Harris (1969) points out that the teaching of writing as an integrated course is normally deferred until rather advanced courses in foreign language study. He views writing as a complex skill involving the simultaneous practice of a number of very different abilities, only some of which are strictly linguistic and some of which are never fully achieved by many students, even in their native language. Harris recognizes five general components of the writing process: content, form, grammar, style, and mechanics. He reviews the defense of the essay examination (real measure of writing abilities, motivates students to actually write, easy and quick to prepare) and the criticism levelled against it (unreliability, avoidance of problems, long scoring time). Harris himself recommends a combination of the objective and free writing tests, as did Lado.

Heaton (1975) differs from most of the earlier language testing experts by having a more sophisticated view of writing. He is conversant with old - or at least rediscovered - theory of written discourse, as shown by his discussion of the purpose and audience of writing and the forms (modes) of writing.

Heaton (1975) emphasizes that it is important to distinguish between the terms composition and essay. He writes:

The writing of a composition is a task which involves the student in manipulating words in grammatically correct sentences and in linking those sentences to form a piece of continuous writing which successfully communicates the writer's thoughts and ideas on a certain topic. Moreover, since in real-life situations there is generally a specific purpose for any writing, composition writing frequently takes the form of letters, reports, extracts from diaries, etc. Essay writing, on the other hand, involves far more than the production of grammatically correct sentences: it demands creativity and originality, since it is generally intended not only to inform but also to entertain. Essays on such topics as Clouds, The Importance of Being Last, and The Countryside at Night are written to sparkle and impress, and good essayists are as rare as good poets. (p. 127)

Heaton concludes that it is generally neither reasonable nor realistic to demand creativity and originality in the form of an essay, while it is reason-

able to expect students to write accurate English for a meaningful purpose. He also stresses the communication aspect of writing in insisting that

the student should be presented with a clearly defined problem which motivates him to write. The writing task should be such that it ensures he has something to say and a purpose for saying it. He should also have an audience in mind when he writes. (p. 128)

Heaton considers the writing skills to be complex and difficult to teach, requiring the mastery of grammatical and rhetorical devices but also the mastery of conceptual and judgement elements. He lists the skills under four main areas: (1) grammatical skills: the ability to write correct sentences, (2) stylistic skills: the ability to use language effectively, (3) mechanical skills: the ability to use correctly conventions of written language, and (4) judgement skills: the ability to write in an appropriate manner for a particular purpose with a particular audience in mind, together with an ability to select, organize and order relevant information.

It would appear that the more recent book on language testing by Oller (1979) represents somewhat of a retrogressive step if we think of the way he deals with writing. He is very much occupied in trying to get the area of testing ordered neatly. He is most interested in the construct of language ability and in conceptualizing tests in line with his view. While this is interesting, it is too narrow a perspective. Language testing needs to take a broad view of human activity: it should place language activities within the broader context of general human activity and purposes. Thus language testing needs to consider what are the constants, parameters and variables of language use. Roughly speaking the constants are: sender/addressor, receiver/addressee/audience, topic, channel and text. The parameters represent the various characteristics that specify the actual characteristics of the constants (e.g., the identity of the writer and audience, purpose of writing, assumed background knowledge, the perspective from which the topic is dealt with, etc.). The variables are the modes of organization and the use of rhetorical and linguistic resources. The problem with Oller's approach, like so much of American research, is - as I have argued elsewhere (Takala 1984) - that it is

too much preoccupied with linguistically based concepts and is not sufficiently sociological, psychological and educational in terms of its research questions and units of analysis.

It is possible that Oller's claim that language ability is unitary was based on a number of assumptions, several of which have proved questionable. First, he seems to share the view that children had essentially learned most of the structure of their L1 in the early years. Second, like so many experts in L2, he has not been interested in advanced foreign language skills, and thus not in writing in L2. Also, he does not seem to have been aware of recent research in literacy. All of these would have indicated that while various language skills obviously are related, there are also clear differences. Speaking and writing, for instance, emphasize somewhat different functions of language and they prefer somewhat different structures of language (cf. Perera, 1984; Takala, 1982).

The present author (Takala 1982) has defined writing as follows:

Writing is a multilevel interactive and goal-directed process of constructing, encoding and communicating meaning by means of a conventional system of visible marks (p. 220).

To conclude the discussion on the concept of writing, it seems evident that it is useful to make a basic distinction between two forms of writing: writing without composing and writing with composing.

4. Test Types for Measuring Writing Ability

Lado made a clear distinction between creative writing and ordinary writing. Also, consistently with his habit emphasis and habit transfer, Lado believed that the testing of writing could be advanced best by listing the particular problems that a writer's particular linguistic background was expected to create.

Lado distinguished between an integrated method of testing writing by asking students to produce a connected piece of writing and a method of testing writing with separate factors such as punctuation, spelling, structure or vocabulary. This latter method would make it possible to sample the prob-

lems systematically. Lado recognized, however, that the validity of the synthetic approach was not readily conceded and he discussed ways of improving the objectivity of scoring composition tests.

Lado recommended a many-sided test of writing, and suggested the following as one possible design: (1) Objective, partial production, multiple-choice items (50-80) dealing with specific problems of spelling, punctuation, grammatical structure, and vocabulary. (2) Twenty or thirty items of the objective, partial production type on a single connected passage testing chiefly matters of sequence and transition signals. (3) Three pictures with instructions to write a paragraph about each with grading based on mechanics only (= number of errors per 100 words). (4) Two short compositions on assigned topics (30 minutes each) with grading based on style, content and mechanics.

It is assumed that the student has something to say and a personal point of view. The student must observe the normal requirements of form and present his views effectively. According to Harris (19) writing is a complex skill, which must simultaneously take into account several points: 1) content, 2) form, 3) grammar, 4) style, 5) mechanics. Valette (1967) ***** Grieve (19) ***** Pilliner (19) *****

5. Validity as an Issue in Assessing Writing Skills

The validity of writing assessment can best be addressed in terms of construct validity, content representativeness (or validity) and curricular validity. Since we do not have any clear notion of the psychological structure of writing, i.e., how general or how task-specific it is, construct validity can best be guaranteed by an analysis of the general features of writing situations and a resulting defensible specification of the domain of writing tasks. This is a functional approach to construct validity and it was used in the IEA International Study of Written Composition. In other words, since it is not easy to say directly what writing ability consists of, we chose to look at what it is for in general and in what situational contexts it occurs. This

means that we have focussed on the initial conditions of writing and on its functions. Another, complementary approach has been more genetic. In other words, we have attempted to sketch the initial conditions that lead to a final product through a sequence of events (writing processes).

Domain of Writing

Murphy (1974) notes that if we are to understand western views of communication, we must recognize the dominant didactic impulse, the laying down of precepts for techniques that allow the speaker achieve, within the situation of discourse, the desired goal. Thus rhetoric had a pragmatic orientation: to convince the interlocutor. Aristotle, in his *Rhetoric*, defines rhetoric as the faculty of discovering all the available means of persuasion. He also made a distinction between epic poetry on the one hand and tragedy and comedy on the other. Aristotle clearly preferred the tragedy over the epic as higher art form, which attains its end more perfectly. He also refers to "drama" as poems imitating persons who are acting and doing something.

Kinneavy (1971) gives a succinct review worth quoting at some length:

... in Antiquity, three main aims of language structured the training in the art of discourse: the literary, the persuasive (rhetorical), and the pursuit of truth (dialectical). The analysis of literary texts was the province of the secondary school: the other two aims were "collegiate" and university concerns. In composition, which was directed to a preparation for rhetoric, certain forms or modes were thought to be basic to all composition (narrative, description, eulogy, and definition) and structured the composition program. (p. 8).

However, Kinneavy suggests that the common classification of the modes of discourse (forms, genres, types) into narration, exposition, argumentation, and description was not fully established before the mid-1800's (Bain's *English Composition and Rhetoric*, 2nd ed, 1867).

More recently, Moffett (1968), Britton et al. (1975), D'Angelo (1975), Kinneavy (1971), Wilkinson et al. (19***) and others have attempted to define models for teaching composition. I will, however, refer to the work of Vähäpassi (1982, 1983, 1986), as it has constituted an important part of my own work on writing. Briefly stated, Vähäpassi suggests that in any writing

situation, there is a writer who writes about something with a certain purpose and audience in mind. Writing is both an act of communication and an activity of cognitive processing.

Vahäpassi systematizes the domain of writing by taking communication and cognitive processing as two main dimensions of her typology. On the communication axis he distinguishes several dominant purposes of writing and specifies main categories of audience. On the cognitive processing dimension, she distinguishes three hierarchical levels of processing and specifies main categories of content which is processed. This system produces a grid and various text types can be located within its cells (see Figure x).

6. Schemes for Rating Written Products

There are several ways of classifying methods of measuring writing ability. Wesdorp (1981) suggests the following classification: global rating, primary trait scoring, analytic scoring, scale rating, interlinear method, objective testing.

In this paper I will mainly discuss writing with composing and discuss holistic scoring, analytic scoring and primary trait scoring as the most common forms of rating written products. I shall begin with holistic scoring.

Typical of holistic scoring (e.g., Cooper 1977) is that the rater takes a script and either (1) matches it with another piece of writing in a graded set of scripts, or (2) rates it for the quality of certain features considered important to that kind of writing, or (3) assigns it a letter or number grade. The placing, rating or grading is done quickly, on the basis of the first impression, after the rater has practised the procedure together with other raters. Holistic scoring, when conducted with rigor, uses scoring guides, or rubrics, which distinguishes it from a more haphazard impressionistic scoring.

The best known analytic scoring procedure is probably that developed by Diederich (1974). The Diederich scale was developed empirically by using factor analysis. A sample of writing was scored by experts representing diff-

erent disciplines. The factors extracted were: ideas, organization, wording, flavor, and mechanics. The last category is sometimes sub-divided into usage, punctuation, spelling, and handwriting. Each factor is rated on a scale from 1 (low) to 5 (high), and ideas and organization are rated on a scale from 2 to 10 (ie., they received a double weighting). Thus the scores can vary from 10 to 50.

Another example of an analytic scoring method is given by Quellmalz (1979). She defines an expository scale consisting of general impression, essay focus/main idea (the subject and main idea are clearly indicated), essay organization (the main idea is developed according to a clearly discernible method of organization), support (generalizations and assertions are supported by specific, clear supporting statements), and mechanics (the essay is free of intrusive and mechanical errors).

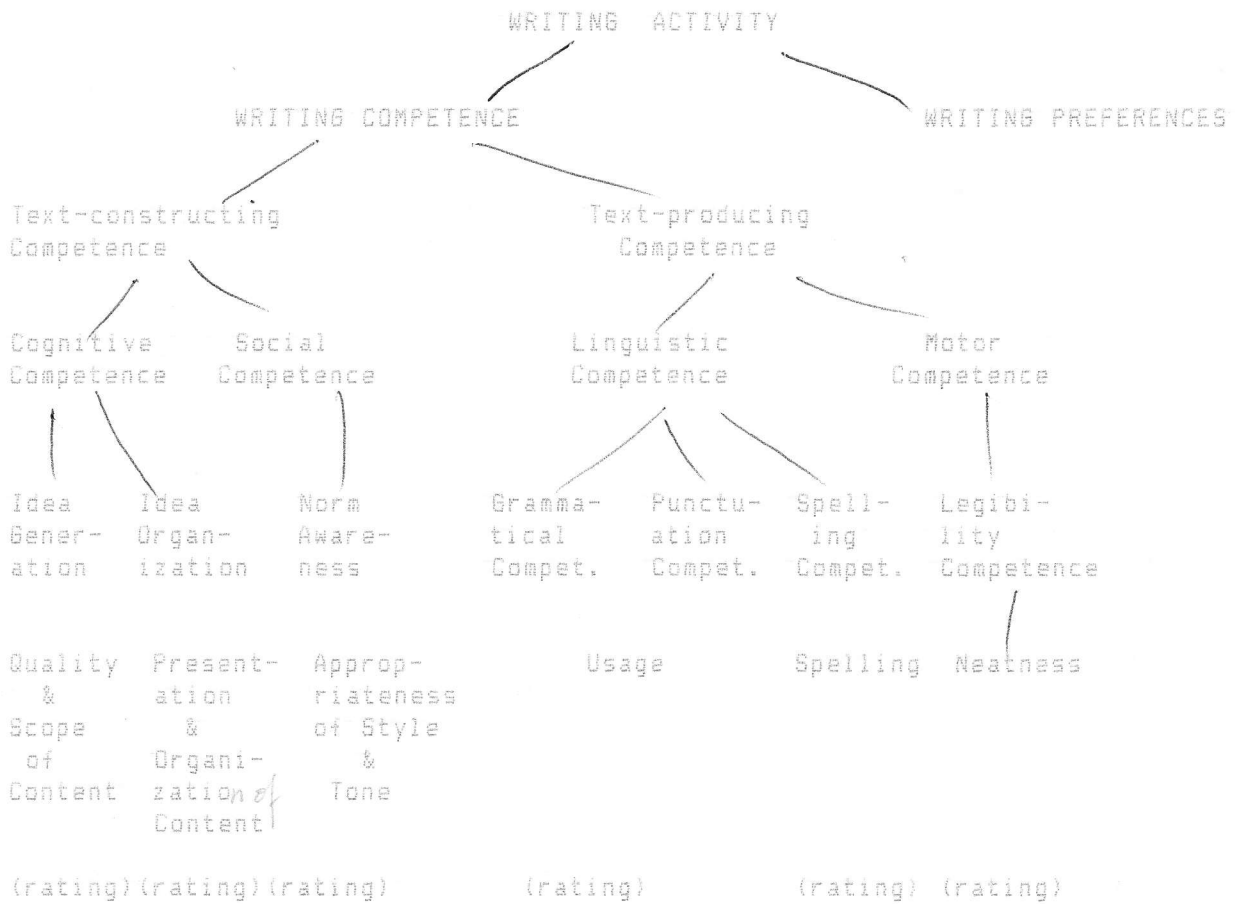
Mullis (1980) explains that the rationale of primary-trait scoring is that writing is addressed to some audience and it is judged in view of its effect on that audience. Primary-trait scoring focusses on assessing whether a piece of writing has certain characteristics or primary traits that are crucial to success with a given rhetorical task. Lloyd-Jones (1977) expresses the goal of primary-trait scoring as follows: "to define precisely what segment of discourse will be evaluated (e.g., by presenting rational persuasion between social equals in a formal situation) and to train readers to render holistic judgments accordingly. (p. 37). He states further that the main steps are to define the universe of discourse, to devise exercises which sample that universe precisely, ^{to elicit} to the writers' cooperation, to devise workable scoring guides, and to use the guides.

The universe of discourse is defined by a three-part model, which can be discourseser oriented (expressive discourse), subject oriented (explanatory discourse) or audience oriented (persuasive discourse). The scoring guide consists of (1) the task itself, (2) a statement of the primary rhetorical trait of the writing which should be elicited by the task, (3) an interpreta-

tion of the task indicating how each element of the stimulus is presumed to affect the writer, (4) an interpretation of how the situation of the task is related to the posited primary trait (a summary of 2 and 3), (5) a system for defining the shorthand which is to be used in reporting descriptions of the writing (the actual scoring guides), (6) samples of papers which have been scored (definition of the score points), and (7) discussions of why each sample paper was scored as it was (extensions of the definitions).

Lloyd-Jones (1977) suggests that primary-trait scoring has certain advantages which outweigh its difficulty. The explicitness of the scoring guide helps to establish the validity of the scoring. By focussing sharply on specific types of discourse, more information can actually be obtained from writers' strengths and weaknesses than by a more global approach.

The present author has attempted to define writing as a construct in the following way (Takala, 1983, 1985):



Task 1	Task 4	Task 9	Tasks 2&3	Task 6	Task 5	Task 8	Task 7
Mes- sages	De- scription of	Letter of Advice	Summary Para- phrase	Argu- ment	Story	Free Com- posi- tion	Essay
Appli- ca- tion	a) Ob- jects b) Pro- cess						

Pragmatic/ functional writing	Study- related writing	Mastery of persuasion & arguments	Description of events & related percep- tions & feelings	Free Expres- sion	Reflect- ive writing
-------------------------------------	------------------------------	---	--	-------------------------	----------------------------

Weisdorp (1981) has assessed the practicality of various methods of assessing writing and summarizes his conclusions in a table form as follows:

	Global Rating	Primary Trait	Analytic Scoring	Scale Scoring	Inter-linear	Objective testing
	Indiv	Jury	Jury	Jury	Jury	
Chances of obtaining high reliability?	No	Yes	Yes	Yes	Yes	Yes
Chances of obtaining reasonable content validity?	No	Yes?	Yes	Yes	Yes	No
Practicality in teaching?	No	No	Yes	Yes	No	Yes?
Feasibility in selections?	No	Yes	Yes	Yes	Yes	No
Chances of positive washback on teaching?	No	No	Yes	Yes	Yes?	No

7. What Does a Rating Depend On?

Among many interesting questions, the Swedish FRIS-Project (Lindell, 1980) has explored whether ratings can be consistently predicted by a linguistic analysis of the scripts. The answer was affirmative: above all, productivity predicted expert ratings. In other words, we can get a fairly good estimate of the quality of a script by simply checking its length. More specifically, the most important factors were the number of different words, the number of unusual (non-frequent) words, the number of punctuation marks, and word length. Again, the dominant importance of good vocabulary in contrast to syntactic competence is demonstrated.

References

- Carroll, J. B. 1961. "Fundamental considerations in testing for English language proficiency of foreign students." In Testing the English proficiency of foreign students. Washington, D.C.: Center for Applied Linguistics. 30-40.
- 1963. "Research on teaching foreign languages." In Gage, N. L. (ed)

1963. Handbook of research on teaching. New York: Rand McNally. 1050-1100.

- 1976. "Psychometric tests as cognitive tasks: A new "structure of intellect". In Resnick, L. B. (ed) 1976. The nature of intellect. Hillsdale, N.J.: Erlbaum. 27-56.

- 1976. "Promoting language skills: The role of instruction." In Klahr, D. (ed) 1986. Cognition and instruction. Hillsdale, N.J.: Erlbaum. 3-22.

- 1968. "The psychology of language testing." In Davies, A. (ed) 1968. Language testing symposium: A psycholinguistic approach. London: Oxford University Press. 46-69.

- 1967. "Research problems concerning the teaching of foreign or second languages to younger children." In Stern, H. H. (ed) 1967. Foreign languages in primary education. London: Oxford University Press. 94-109.

- 1965. "The contributions of psychological theory and educational research to the teaching of foreign languages." Modern Language Journal XLIX, 5: *****

Lindell, E. 1980. "Six reports on free writing: A summary of the FRIS project". Malmö School of Education, Didakometry No. 61.

Quellmalz, E. 1979. Defining writing domains: Effects of discourse and response mode. UCLA Graduate School of Education, Center for the Study of Evaluation. (Interim Report)