

Bureau for International
Language Co-ordination

CONFERENCE
REPORT
1983

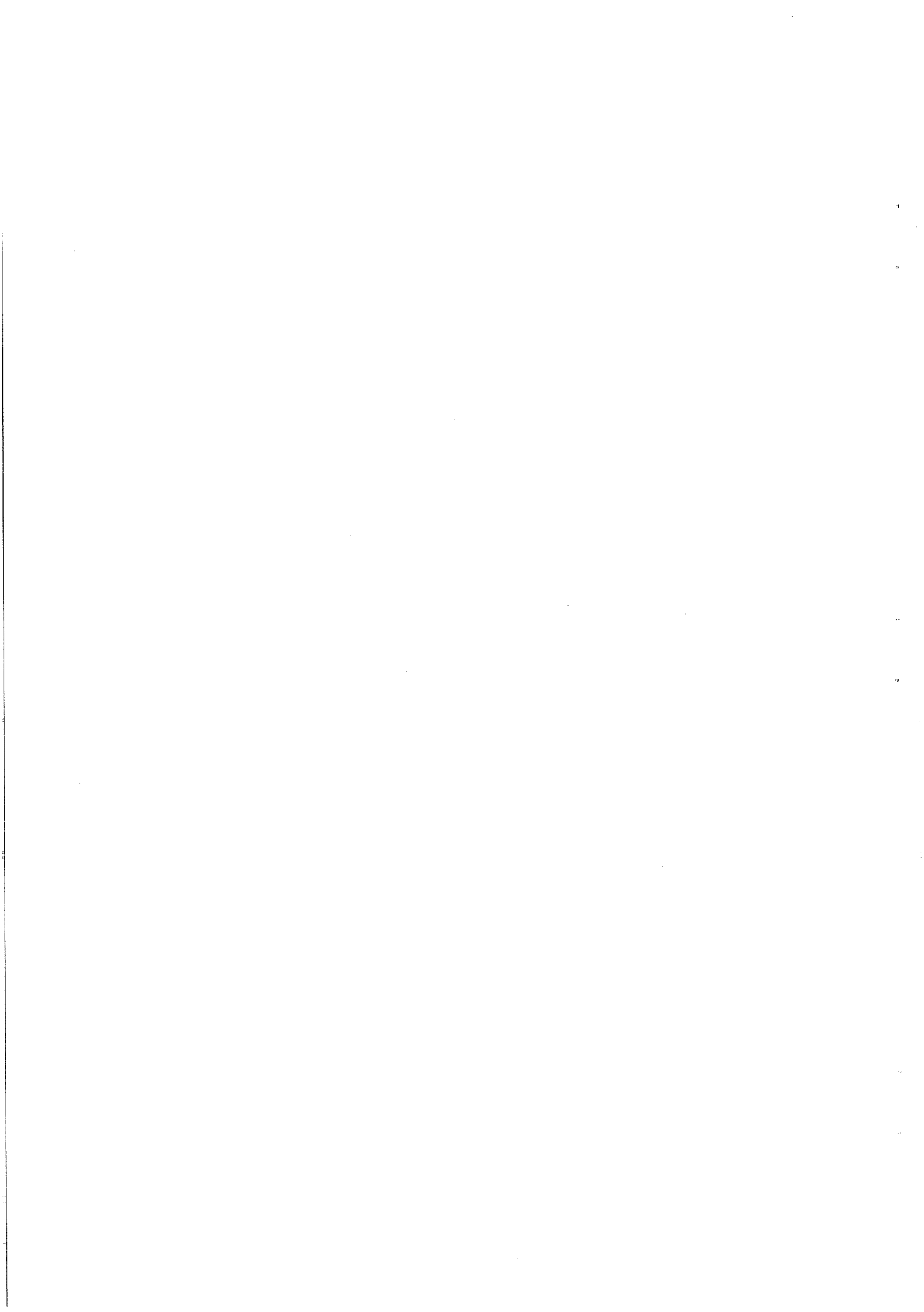
HÜRTH
FEDERAL REPUBLIC OF GERMANY

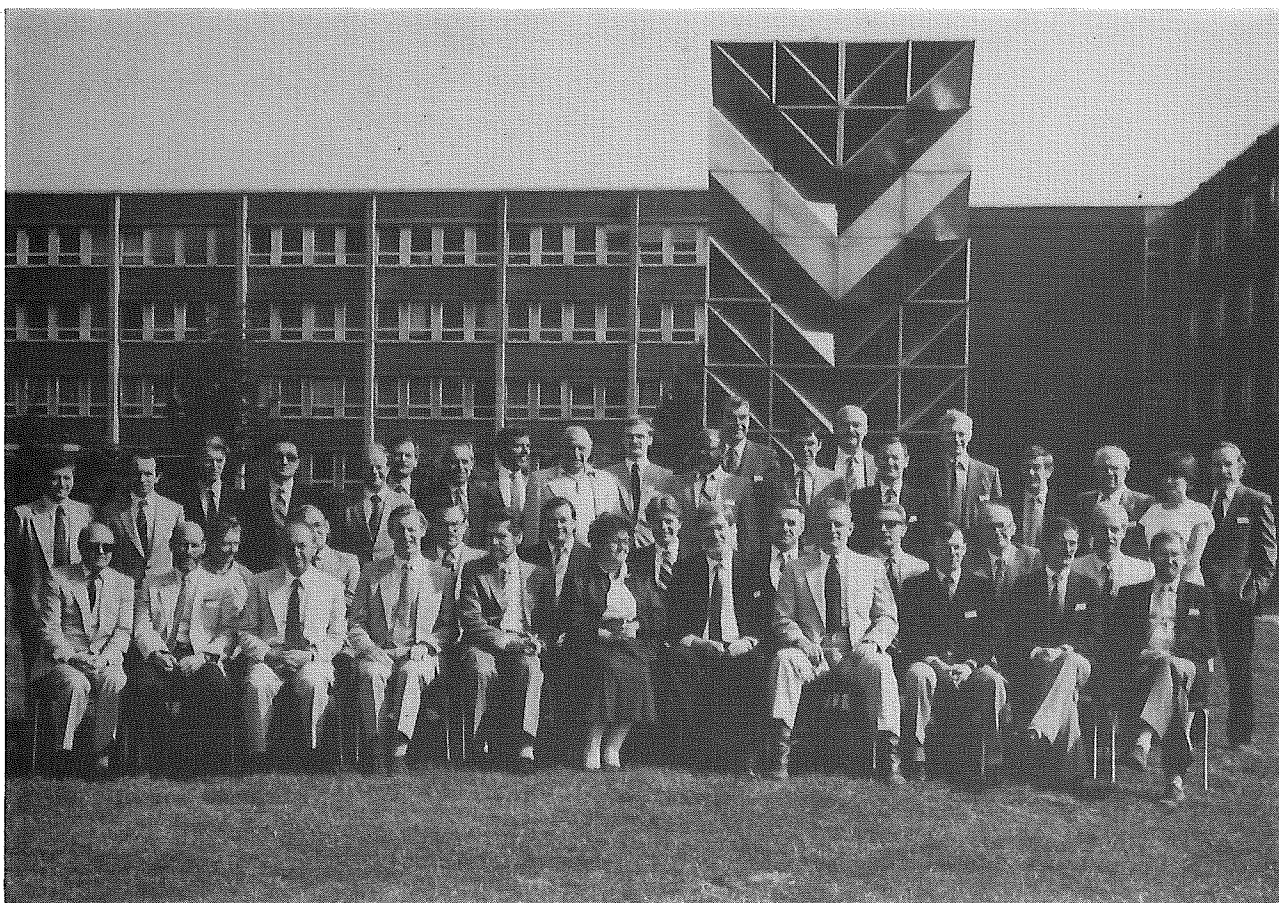
BILC SECRETARIAT

Bundessprachenamt
D 5030 HÜRTH
Federal Republic of Germany

TABLE OF CONTENTS

	<u>Page</u>
<u>I. CONFERENCE ORGANISATION</u>	
1. Conference Photograph	1
2. Conference Programme	3
3. List of Participants	5
<u>II. PRESENTATIONS</u>	
1. List of Speakers	9
2. From Aptitude to Acquisition: 20 Years of Research into Learning a Second Language - Werner Hüllen	11
3. Language Aptitude Testing in the Canadian Armed Forces - Percy Rangongo	23
4. Language Aptitude Testing in the United States - Ray T. Clifford	25
5. Oral Language Proficiency Testing at the Foreign Service Institute - John Ratliff	35
6. The Assessment of Language Proficiency and Aptitude for Specific Jobs - Brendan J. Carroll	43
7. Holistic Language Testing - Facts and Fallacies - Christine Klein-Braley	47
8. Recent Experience with an Aptitude Test for Studies in Modern Languages - Günter Trost	65
<u>III. NATIONAL REPORTS</u>	
1. Belgium	85
2. Canada	87
3. France	93
4. Germany	103
5. United Kingdom	107
6. United States	109
7. SHAPE	121
<u>IV. STUDY GROUP REPORTS</u>	
1. Study Group 1 and 2 (combined report): Language Aptitude Testing	127
2. Study Group 3: Language Proficiency Testing and the Curriculum	133
3. Study Group 4: Language Proficiency Testing and the Teacher	137

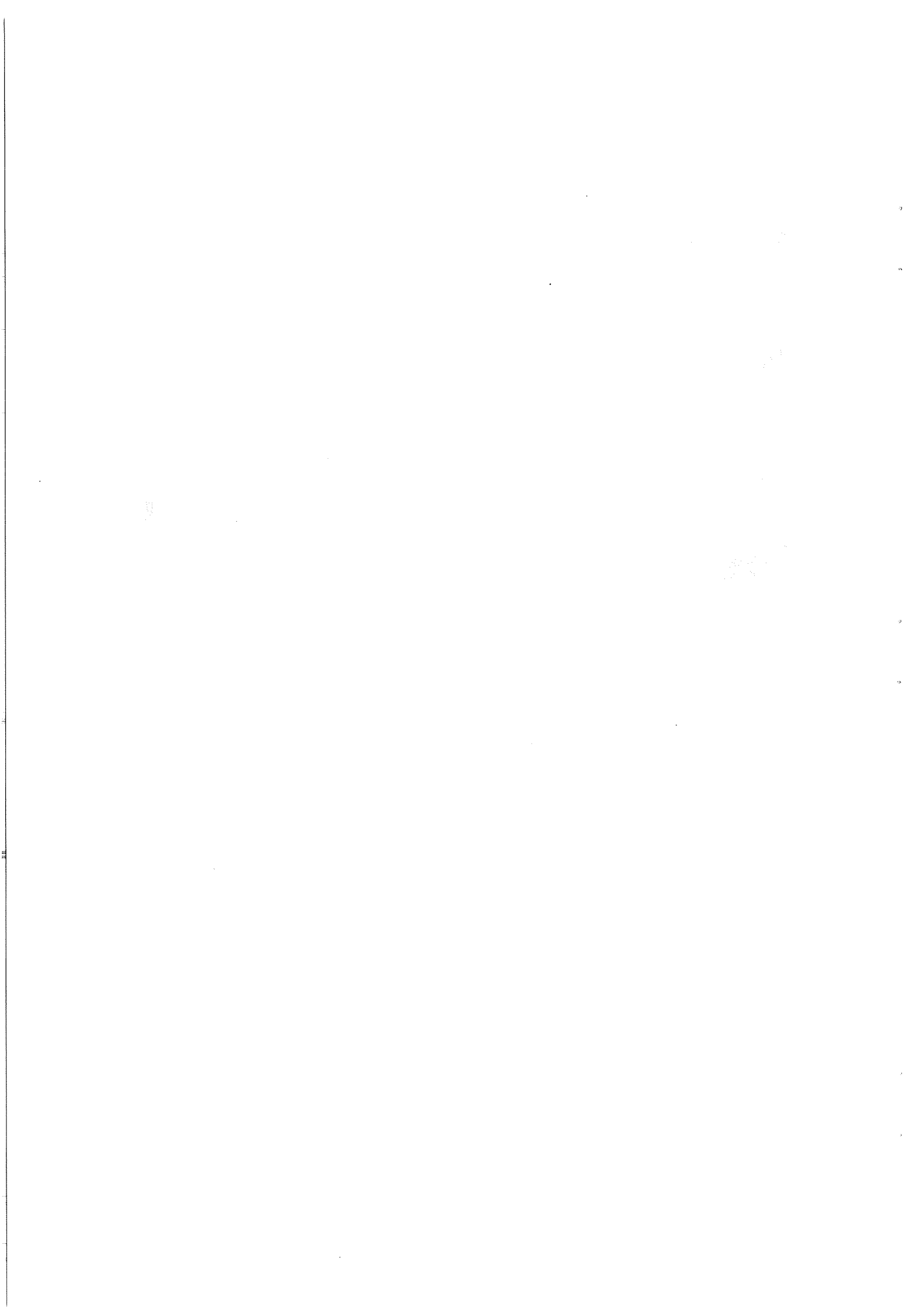




Front Row: (left to right) RDir Leben, Mr. Ellis, Col Lenci, WingCDR Thomas, Ministerialrat Dr. Tobollik, Ministerialdirigentin Buchberger, Präsident Maur, Col McNerney, Maj Noordsij, Dr. Curica, Cdt Filleul

Second Row: RDir v. Herzenberg, Mr. Gregory, LtCol Thomson, (seated) Mr. Ratliff, Dr. Clifford, LtCol Thériault, LtCol Barbeaux, LRDir Rohrer, Maj (ret) Walinsky

Rear Row: LtCol Borrelli, LtCol Stefanelis, Cdt Tancré, Col Tatal, Mr. Worrall, LtCol Coltelli, Col Perrotta, LtCol Bellillo, Mr. de Lespinois, Mr. Hüllen, Mr. Rangongo, LtCol Brace, Mr. Schwarz, RDir Wech, Cdt Barré, SqnLDR O'Hagan, Maj Pearce, ORR Jüsten, Miss Hamacher, Mr. Rutherford



PROGRAMME

Date	Time	Event	Remarks
05 June	1400 - 2200	Arrival and accomodation of delegates	
	2000	Evening meal	
06 June	0900	Conference assembles	
	0930	Opening Address	
	1000	Administrative Briefing	
	1030	Coffee and Photograph	
	1115	Prof. Dr. W. Hüllen: From Aptitude to Acquisition: 20 years of research into second language learning	
	1230	Lunch	
	1400	Tour of Bundessprachenamt facilities	
	1530	Mr. M. P. Rangongo: Language Aptitude Testing in the Canadian Forces	
	1700	Study Group Session 1	
	1830	Dinner	Evening free
07 June	0900	Dr. R. I. Clifford/ Mr. J. Ratliff: Aptitude Testing in the U.S.A.	
	1000	Coffee	
	1030	National Reports 1 to 4	
	1230	Lunch	
	From 1330	Excursion to Moselle Valley	Includes wine tasting and dinner
08 June	0900	Mr. B. J. Carroll: The Assessment of Language Proficiency and Aptitude for Specific Jobs	
	1000	Coffee	
	1030	Dr. C. Klein-Braley: Holistic Language Testing - Facts and Fallacies	
	1230	Lunch	
	1400	National Report 5	
	1500	Study Group Session 2	
	1600	Steering Committee Meeting 1	
	From 1900	BILC Dinner	

Date	Time	Event	Remarks
09 June	0900	Dr. G. Trost: Recent Experience with an Aptitude Test for Studies in Modern Languages	Tours of Cologne, Bonn or Aachen can be arranged
	1000	Coffee	
	1030	Study Group Session 3	
	1130	Steering Committee Meeting 2	
	1230	Lunch	
	Afternoon	Open	
10 June	0900	Study Group Reports	
	1000	Coffee	
	1030	Study Group Reports	
	1130		
	1230	Lunch	
	1400	Steering Committee Report	
	1530	Sum-up by Conference Chairman, open forum and closing	
	From 1630	Departure of delegates	
11 June	Morning	Remaining delegates depart	

BILC CONFERENCE 1983

LIST OF PARTICIPANTS

CONFERENCE CHAIRMAN

Präsident Hanns Maur Präsident Bundessprachenamt

DISTINGUISHED VISITOR TO BILC-83

Ministerialdirigentin Head, Vocational Training
Ingeborg Buchberger Branch in the Social Services
Division, MoD Bonn

NATIONAL DELEGATIONS

BELGIUM

Head of Commandant Head, English Department
Delegation D. Filleul (Daniel) Royal Military Academy
Brussels

Member Commandant Royal Military Academy
J. Tancre (Jacques) Brussels

CANADA

Head of Lieutenant Colonel Director Language Training
Delegation J. R. Thériault (Ray) (DLT), National Defence
Headquarters, Ottawa

Members Lieutenant Colonel Commandant, Canadian Forces
R. H. Thomson (Jock) Language School (CFLS)
Ottawa

Mr. M. P. Rangongo (Percy) Pedagogical Advisor,
DLT, National Defence
Headquarters, Ottawa

FRANCE

Head of Lieutenant Colonel Officier pédagogie du Centre
Delegation E. H. Barbeaux (Edouard) de Langues et Etudes
Etrangères Militaires
(C.L.E.E.M.), Paris

Member Commandant Chef de la Section audio-
B. F. Barré (Bernard) visuelle du C.L.E.E.M.

FEDERAL REPUBLIC OF GERMANY

Head of Delegation	Regierungsdirektor E. Leben (Erwin)	Deputy Head, Language Training and Language Services Section; MoD Bonn
Members	Regierungsdirektor I. von Herzenberg (Ingeborg)	Head, Training Materials Development
	Regierungsdirektor P. Müller (Peter)	Head, Language Section, Military District IV - Wiesbaden
	Regierungsdirektor F. Wech (Fritz)	Head, Language Section, Military District VI - München
	Oberregierungsrat G. Jüsten (Gerd)	Inspector, External Language Training
	Mr. F. Gregory (Frank)	Head, Teacher Training
	Mr. M. P. M. Schwarz (Michel) (M. A.)	MatDevlpt Sectn: Principles of LngInstr and MatDevlpt

ITALY

Head of Delegation	Colonel F. Lenci (Frederico)	Commandant, Air Force Language School
Members	Colonel A. Perrotta (Aldo)	Commandant, Army Language School
	Lieutenant Colonel F. Borrelli (Frederico)	Army Language School
	Lieutenant Colonel G. Bellillo (Gianmarco)	Ministry of Defence, Air Staff, Rome
	Lieutenant Colonel C. Coltelli (Claudio)	Ministry of Defence, Rome

NETHERLANDS

Head of Delegation	Major L. Noordsij (Leen)	School Militaire Inlichtingen- dienst, Harderwijk
-----------------------	-----------------------------	--

PORTUGAL

Head of Delegation	Dr. R. M. S. Curica (Rui)	Air Force Academy of Portugal
-----------------------	---------------------------	-------------------------------

UNITED KINGDOM

Head of Delegation	Wing Commander I. J. Thomas (Ivor)	T Ed2, Royal Air Force
Members	Lieutenant Colonel J. E. Goldsmith (John)	Commandant, Higher Education Centre, Mülheim
	Major C. C. Pearce (Carl)	Command Language Adviser, BAOR
	Mr. G. G. Worrall (George)	Language Adviser, Institute of Army Education
	Squadron Leader V. W. J. O'Hagan (Tim)	Officer Commanding CLTC
	Mr. A. Rutherford (Alex)	Language Training Adviser to the Royal Navy

UNITED STATES

Head of Delegation	Colonel D. A. McNerney (David)	Commandant, Defense Language Institute, Foreign Language Centre (DLIFLC), Presidio of Monterey, California
Members	Dr. R. Clifford (Ray)	Academic Dean, DLIFLC
	Mr. P. J. de Lespinois (Pierre)	Director of Training, DLIFLC
	Lieutenant Colonel R. C. Brace (Dick)	Dean of Academics, DLIELC
	Mr. J. Ratliff (John)	Associate Dean, Foreign Service Institute

SUPREME HEADQUARTERS ALLIED POWERS EUROPE

Head of Delegation	Mr. D. Ellis (David)	Head, Language Training SHAPE/Proficiency Testing IMS NATO
Members	Mr. A. McLean (Alistair)	English Teacher, SHAPE
	Mrs. J. Lawrence (Joyce)	Administrator, SHAPE Language Circle

OBSERVERS

Colonel	E. Tatal (Erol)	Turkish Army Attaché, Bonn
Lieutenant	A. Stefanelis (Antonius)	Greek Deputy Defence Attaché, Bonn

CONFERENCE ORGANIZATION

Präsident	Hanns Maur	Conference Chairman
Leitender Regierungsdirektor	J. Rohrer (Josef)	Chairman, BILC Secretariat Chairman, Steering Committee
Mr.	H. Walinsky (Herbert)	BILC Secretary
Frl.	M. Hamacher (Monika)	Secretariat
Herr	Ch. Hüllen (Christopher)	Secretariat

LIST OF SPEAKERS

Mr. B. J. Carroll
Newbury, Berkshire
England

Dr. R. T. Clifford
Academic Dean
DLIFLC
Presidio of Monterey

Professor Dr. W. Hüllen
Universität Essen, Fachbereich 3
Literatur- und Sprachwissenschaften
4300 Essen

Dr. C. Klein-Braley
Universität Duisburg, Fachbereich 3
Sprach- und Literaturwissenschaft
4100 Duisburg

Mr. M. P. Rangongo
Pedagogical Advisor
National Defense Headquarters
Ottawa, Canada

Mr. J. Ratliff
Associate Dean
FSI
1400 Key Blvd.
Rosslyn
Virginia 22209

Dr. G. Trost
Institut für Test- und Begabungsforschung
der Studienstiftung des Deutschen Volkes
(Institute for Test Development and Talent Research)
Koblenzer Straße 77
5300 Bonn 2



From Aptitude to Acquisition:
Twenty Years of Research Into Learning a Second Language

Werner Hüllen

What I am attempting to do is take you along the path that researchers into language acquisition and language learning have ~~done~~ during the past twenty years. According to the topic of this conference, the point of departure will be John B. Carroll's term language aptitude; the point of arrival, then, will be what his fellow researchers have made out of this term and its meaning. Of course, our main interest here is not the persons and their ideas, but the ways in which their hypotheses apply to teaching and learning. To understand and improve these should be our main concern.

followed

An attempt like this can of course only be an attempt in the true meaning of the word; it can never be exhaustive or definitive in any sense. Moreover, an attempt like this depends much on the works and publications of others which are to be quoted or mentioned, but which cannot be reported in detail.

My paper will proceed in three major sections. The first centres around Carroll's ideas and the term aptitude as defined by him. The second centres around the term acquisition, and this means that Chomsky and psycholinguistics as stimulated by Chomsky come into focus. The third section centres around Krashen's so-called *monitor-model*. The thread that runs through all three of them is the question of what foreign language aptitude is and in which way it can influence our teaching.

1. In 1962 John B. Carroll (1962, also 1972 and 1981) published a paper in which he defined five variables which were supposed to predict and to assess success in foreign language learning. The type of learning he had in mind was the intensive army course with soldiers in the United States. The variables were the following:

1. general intelligence, defined as the ability to understand verbal instructions or infer them from a context if not explicitly given;
2. amount of time a learner is willing to devote to the learning task, particularly if his aptitude forces him to spend more time than others;
3. foreign language learning aptitude, defined as actual learning time under best teaching conditions;
4. presentation of teaching material, and
5. opportunity to learn.

Carroll's model of prediction had as a practical aim the planning of foreign language courses. This is why its author was not really interested at that time in research outside

his programme, for example in the question of whether general intelligence is something inborn or something acquired. An aptitude for learning a foreign language was assumed, but the question was left open as to whether this aptitude was a gift of nature or the result of previous learning. Carroll obviously presupposed an interrelationship between the five variables; however, it proved very difficult for him and others to analyse this. As his main aim was to develop courses and teaching material for these courses, the two last variables (presentation of teaching material and opportunity to learn), obviously, are dependent on the three previous ones.

As is well-known, Carroll's model became very influential - both in theory and for many practical purposes. It was enlarged and refined (Jakobovits 1970, Wienold 1973). Aptitude always played a part in it as being something which not all (or not many) people have. The most important practical application of Carroll's model is the Modern Language Aptitude Test, MLAT, (Carroll/Sapon 1959), which has dominated techniques of assessment. It has been used as a tool to find out the special aptitude of an individual, which, together with the other variables, set conditions for the presentation of teaching material and the optimization of opportunities to learn (Carroll 1973).

It is easy to see that the theory of second language teaching and learning, as discussed nowadays, has moved from Carroll's terminology. However, the problems he described remain.

What Carroll called 'general intelligence' is today analysed under the heading 'communication in the classroom'. We look upon the teaching situation as a kind of communication in its own right and presuppose that the ability to understand and contribute to this communication is a prerequisite for learning a language (Cazden et al. 1976, Mehan 1979, Lörcher 1983). The 'amount of time', as defined by Carroll, has given rise to a wave of studies under the heading 'motivation'. It led to the definition of certain kinds of motivation and their analysis as one particularly powerful variable in the learning process (Gardner/Lambert 1972, Solmecke 1976). 'Presentation of teaching material' has become a major concern of grammarians, for instance, in their endeavour to construct a pedagogic grammar (Corder 1973, Hüllen 1976 and 1979, Bausch 1979) and to construct curriculum functions and notions which are relevant for communication (Wilkins 1976, van Ek 1976). Finally, 'opportunity to learn' has become the key word in the latest and most fashionable explanatory model of second language learning and acquisition, the so-called *monitor model* in which it figures under the heading of 'input' (Krashen 1981, 1982; Dulay et al. 1982).

But what about language aptitude? The term has certainly vanished; if you look at new publications which mirror the state of the art (e. g. Stevick 1982), including theory of testing (e. g. Stevenson 1982), you will find that the term is not mentioned any more. The problem behind it, however, may be said to have done nothing less than dominate the development of language learning (but not teaching) analysis ever since

Carroll's model appeared. Curiously enough it is an aspect of the problem that Carroll, as mentioned before, did not even attempt to elucidate which caused this question, namely, of whether language aptitude is a product of nature or nurture, whether it is an inborn faculty or the result of an educational process (Felix 1982: 187 - 227).

Carroll broke aptitude down into four sub-abilities, each of which, obviously, refers to a level of language:

1. ability to store and reproduce phonetic data after a certain time;
2. grammatical sensitivity to linguistic categories in understanding or reading as well as in speaking or writing, even if unable to define or describe them or to state explicit rules;
3. the ability to memorize lexical data, i. e. to learn words and retrieve them when necessary;
4. induction, i. e. the ability to reproduce structured speech or writing with the help of insights gained from language data previously experienced; that is basically the ability to apply rules learned to a new context.

From Carroll's writings it is not clear whether he understood the subdivision of language aptitude to mirror four independent faculties of the mind, or whether he thought of one unitary language aptitude, which only for the sake of teaching and testing was broken down into four 'departments' according to the generally acknowledged levels of linguistic description. This means that Carroll's definition of aptitude is located somewhere between psychology and linguistics, but not clearly attributed to either of them.

For a psychological description, the term aptitude is fairly vague. It combines four quite different functions of the human mind without saying how this combination is brought about. Storing and reproducing phonetic data is a sensory faculty; grammatical sensitivity is a very fuzzy concept of what we could call abstract knowledge without a clear statement as to whether meta-knowledge (which normally goes together with abstract knowledge) is attached to it; the ability to memorize and retrieve is a neuro-physiological brain capacity; and induction is a high order of logical thinking. If Carroll supposes them to be 'departments' of an all-embracing aptitude, he hypothesizes rather than describes this.

It is probably not unfair to say that these four sub-abilities actually are surface descriptions of what can be observed in a foreign language class with reference to the possibilities of testing. Carroll tried to gain access to the linguistic products of language behaviour with the help of linguistics, which had always analysed these products as a system on various levels (phonetic, grammar (= syntax), lexis, and what could approximately be called text).

Carroll's psychological terms, then, appear to be conceptual bridges which are supposed to lead from the result of linguistic production back to the processes that create them in the mind. That presupposes a direct relation between a psychic process and the linguistic product, even though the latter is described in purely linguistic terms. In this respect Carroll adhered to the structuralist tradition for which the relation between linguistics and psychology was a straight one, whereas his thinking generally had the tendency to overcome this very tradition and its linguistic levels-and-skills-approach (Lado 1961, Valette 1967).

2. We know that the discussion about how people adopt¹ a language (their first or any other one) has been dominated by Chomsky's ideas from the moment of his arrival on the psycholinguistic scene. (The term 'adopt' is used here to describe man's general ability to process linguistic data and use language, whereas the term 'acquire' is reserved for adopting a language under natural, unguided circumstances, and the term 'learn' is reserved for adopting a language in a classroom. Except in pathological cases, the first language can only be *acquired* whereas a further language can either be acquired or learnt. An acquired further language is called a 'second' language, whereas a learnt further language is called a 'foreign' one. The difference between a natural and a formal setting for adopting a language seems important enough to make this terminological distinction. This, however, does not mean that acquisition and learning are totally different procedures. As man's mind and a man-made language are involved in both cases, there are, of course, overlapping areas between them.)

Chomsky entered the linguistic scene at about the same time as Carroll published his model for predicting learning success. The central term of Chomsky's theory is language acquisition device as a name for man's innate capacity to process linguistic data and, in doing so, to construct a mental grammar (Chomsky 1965, Lenneberg 1967). The LAD is supposed to be the linguistic counterpart of language universals. Man's innate capacity to acquire or to learn a language presupposes the fact that all the languages on earth have common features. At the same time, LAD is the battering ram against all kinds of behavioristic explanations.

The question which John B. Carroll did not dare to tackle, then, was answered by Chomskian psycholinguistics and has, basically, not been asked again. Language aptitude, in a much broader sense than Carroll used the term, is today looked at as a general human property. It is part of everybody's cognitive endowment and not a distinction of the gifted vis-à-vis the non-gifted. All analyses of the matter subsequent to Chomsky's stimulating work take this as a starting point in one way or the other.

These analyses were devoted to a description of mental grammars mainly in the syntactic domain (Miller 1962, 1974; Braine 1963; Johnson 1965). They were understood to be an attempt at making the general notion of 'competence' concrete. However, the con-

¹In this discussion I will use 'adopt' to cover both 'learn' and 'acquire'.

cept of the LAD, as developed and elaborated by Chomsky and the many scholars in his wake, was not meant to explain acquisition of any other than the first language. Still, this concept has set the pattern for all subsequent attempts to explain this. From Chomsky's ideas, psycholinguistics found its way to so-called second language acquisition theory, which is an explanation of the adoption of a language that shares with first language acquisition the fact that it is not guided by teachers or teaching material and with foreign language teaching that it is not applied to a first language (Wode 1981). Nearest to the LAD in this context is perhaps Selinker's concept of a *latent psychological structure* (1972), by which he tried to explain that acquiring or learning a second or foreign language must rest on man's genetic capacity, but in a way different from acquiring the first language, because all further languages are more often than not only imperfectly mastered.

In this context the opinion prevails that acquiring a second language is much the same as acquiring a first language (Dulay/Burt 1972, 1974). There is a particularly strong resentment against the old contrastive hypothesis explained in behaviouristic terms (Bausch/Raabe 1978) according to which any language to be learned after the first language is under its strong influence. So-called 'morpheme order studies' try to show that the sequence in which some grammatical features are internalized by a child acquiring his first language is repeated by children, and even by adults, who are acquiring a second language. Obvious differences between the two processes of acquisition are looked upon as either accidental or not essential to what goes on. Though the strength of the claim varies with authors, today the so-called *identity hypothesis* dominates; first language and second language acquisition are regarded as basically the same (Felix 1982).

The areas which have been investigated in the context of this hypothesis are fairly narrow and far from comprehensive. It is mostly acquisition of syntactic phenomena that has attracted researchers, and within them it is mainly negation and interrogation which has been described (Felix 1982: 17-18, 20-32, 33-47). Using Carroll's terminology, we could say that grammatical sensitivity has been explored to some detail, but that the phonetic, the lexical and the textual aspects of the process, let alone its pragmatical relations with general communicative behaviour, have been neglected. This one-sidedness is hardly recognized by researchers, because in spite of such obviously limited research, broad theories about the mechanics of second language acquisition were set up. The gist of what they say is that the acquisitional process consists of discrete phases whose sequence is a fixed order though individual variations occur (Felix 1982). I shall not go into the details of criticism to which I personally subscribe in many respects that can be raised against this theory.

Another hypothesis about second language acquisition which follows Chomsky's ideas in a less direct way, but would not be possible without his work, has been described under the heading 'interlanguage' or with similar terms (Bausch/Kasper 1979). Its central idea is that adopting a language is not

a linear process which can be measured at its end, that is from a native speaker's competence, but that it must be looked upon as a sequence of interstages, all of which build systems with limited autonomy. What a language learner knows after, say, three months of linguistic experience and instruction, is not a native speaker's competence minus everything the learner does not know, but a 'small' language in itself which can be described with the help of a set of rules just as a full language can. The fact that these rules are not identical with the rules as accepted by a speech community does not undo their systematic character. Mistakes, therefore, are essentially not something negative, but something positive, relative to the stage of learning a learner has reached (Hüllen/Jung 1979, Knapp-Potthoff/Knapp 1982). This description of the learning process reminds everybody of first language acquisition, where psychologists have always distinguished certain phases which gradually approach normal competence (Clark and Clark 1977).

The autonomy of interlanguages is, of course, only a relative one, because its rules are dependent on the first language of the learner and the target language as used by native speakers. There is simply no other way of describing them. However, the most important point is that progress in learning a foreign language cannot be measured by its relationship to full competence, but only by comparing at least two interlanguages which a learner goes through. The linguistic product, therefore does not mirror some kind of mental process directly, but does so only with reference to other linguistic products, either of the same, or of a following, interlanguage. With respect to Carroll's ideas this means that foreign language learning aptitude can only be tested by comparing linguistic products of one learning stage in order to find out the state of an interlanguage at a given point in time, and of all linguistic products of subsequent learning stages in order to find out what progress has been made.

The development from Lado via Carroll and Chomsky to Selinker implies the important change from structuralist and behavioristic patterns of explanation to cognitivistic ones. For the problem under discussion, this means, among other things, the departure from explaining aptitude and proficiency in a foreign language with the assumption of a number of independent skills to be examined in discrete item tests. It means explaining aptitude and proficiency in a foreign language with the assumption of an overall competence which accounts for all achievements in the same way and which is tested globally, for example, with the cloze test (Spolsky 1973, 1978) or with dictation (Oller/Streiff 1975). Whereas Carroll stands between the two, Oller, for example, advocates strongly the so-called unitary competence hypothesis (Vollmer 1982). This problem will be taken up later.

It is important to stress that those scholars who have analysed second language acquisition in the way just explained are not interested in predicting its success and are not interested in finding conditions which would improve this success. This means they are not interested in Carroll's main question. For them, success is basically inherent in the procedure itself. Since adopting a second language is done with the help of the same

mental processes as adopting a first language, it is taken for granted that the outcome will be roughly the same and, if this is not the case, incidental parameters are to blame but not the theory. As language acquisition is investigated in a so-called natural context, the planning of courses for a formal teaching situation is outside the researcher's interest anyway.

Thus, foreign language learning theory has moved away from Carroll's central problems by substituting acquisition for aptitude and so-called natural conditions for the classroom; but in doing so, it has answered one of the main questions we raised. The distinction between a mental process and a linguistic product has become clearer. We are no longer interested in assessing discrete items of language achievement at a given point of time, but a construction of systematic interlanguages which follow each other in time and the comparison of which allows the researcher to infer underlying mental processes and strategies. I am not saying that all the results of interlanguage explanation are satisfactory, but at least the question of the difference between process and product now appears to have been answered.

3. There is one model which has absorbed the developments reported, but which is, like Carroll's model, devoted to the acquisition and learning of a second language in the classroom. This is Krashen's so-called *monitor theory* (Krashen 1981, 1982).

Krashen distinguishes between a 'filter', an 'organizer' and a 'monitor'. The 'filter' consists of the learner's anxiety and general motivation and controls the amount of data experienced. The 'organizer' arranges incoming linguistic data in a way which results in interim constructions (interlanguages) and finally in the commonly accepted structure of the language itself. It does its work below the level of consciousness and is responsible for what is called 'acquisition', as distinguished from 'learning'. The 'monitor', however, is man's ability to consciously formulate rules about the structure of a language. In the optimal case, the organizer and the monitor lead to the same end, which for Krashen means that the monitor becomes superfluous. Whether a person acquires or learns depends on (1) the kind of linguistic input he is exposed to, (2) the communicative situation he is in and, most of all, (3) on his personality. If he depends on conscious strategies and likes to produce formally correct language, he will use the monitoring system. If he depends on subconscious processes and trusts in natural communication, he will use his inborn organizer. According to this system, the subconscious and conscious faculties work independently of each other. Subconscious acquisition is the more effective device, whereas conscious application of rules only steps in when difficulties in understanding or in expressing oneself arise and when the temperament of the language user requires it.

One can indeed be very critical of this theory and the function it gives to the subconscious and the conscious means of adopting a second language (Knapp-Potthoff/Knapp 1982: 97-101). However, at least Krashen is aware of the fact that the process of first language acquisition cannot simply be repeated in the second/foreign language class, and that here formal teaching is useful and necessary. I am personally inclined to think that Krashen's definition of the monitor is wrong, but I acknowledge that he is aware of the two ways in which a new language can be approached and that a conscious approach is at least to some extent indispensable in scholastic surroundings.

This brings John B. Carroll's main problem into focus again. Krashen's monitor model is about the nearest to Carroll's that has been developed. Indeed, he does not speak of aptitude because this is taken care of by the LAD, but he speaks of grammatical sensitivity (under the term 'organizer') and of induction (under the terms 'organizer' and 'monitor'). He more or less ignores the phonetic and lexical aspects of language acquisition. Krashen does not speak of general intelligence either, probably because he takes it for granted. The motivational aspect is not considered in terms of time, but in terms of anxiety, which, however, could easily be converted into time. (A learner with a high anxiety filter, which narrows down input, will take longer to collect a certain amount of linguistic data than a learner with a low anxiety filter, which enlarges input.) Presentation of teaching material and opportunity to learn are made dependent variables when Krashen suggests special ways of teaching that meet the demands of the other variables and when he considers input essential. Input, however, is controlled by teaching methods and the books used. Krashen expects success in foreign language teaching when the variables work together in an optimal way, but he refrains from investigating how they are actually linked to each other.

Our walk through twenty years of research has revealed a meandering development and has led to a result which is only partially satisfactory: we have cleared up a critical point which is central to Carroll's model and to the assumption of a foreign language learning aptitude, but we have done so only at the expense of moving away from other critical points and leaving them uninvestigated. This applies to the replacement of aptitude by acquisition and the replacement of the formal teaching situation (which consists of the presentation of teaching material and of the opportunity to learn) by so-called natural conditions. Krashen's model counterbalances this development only to a certain extent, because he attributes all inherently successful processes in adopting a new language to acquisition and makes learning an artificial way out if nothing else helps. Krashen's model is obviously applicable to the situation of immigrants who live, for example, in the United States, and are exposed to a full language experience there. In addition, they can go to a school in order to make up for what they do not grasp in the natural way. But this model is obviously not applicable to a teaching and learning situation where a foreign language is solely (or mostly) to be learned through the teaching input in a class and with no help from a speech community. This, however, is the case with most foreign language teaching as it is carried out in schools or, for instance, in army courses.

When Chomsky and his successors speak of acquisition, they think of the genetic endowment of man as a language processing animal. When Carroll spoke of aptitude, he thought of a possible faculty which exists in addition to man's genetic faculties. Carroll wanted to account for the obvious experience of teachers and learners alike that some learn more quickly and with more success (whatever that means) than others. This additional gift may be genetic, too, or it may be acquired. This means we either have to presume that people have differing faculties for language acquisition, irrespective of the conditions in which they live and of the experiences they have had, or they have these faculties as a result of such conditions and experiences. In this respect, the question of whether there is a unitary competence or a multiple competence for learning a foreign language becomes important. Intuitively speaking, the first assumption would speak in favour of a genetic explanation, the second assumption in favour of a social and experiential explanation. The issue is not settled. The last empirical project that I know of (Vollmer 1982) finds foreign language proficiency to be strongly dominated by first language proficiency and, as far as this is not the case, mainly expressed in the high order skills of listening and reading, followed by the elementary skill of pronunciation. This means that if there is something like an inherited aptitude, it obviously applies to first and second language alike; what is not explained by this is obviously taken care of by a proficiency in integrated skills of understanding which may be acquired.

From all this it follows that we should go on trying to investigate what Carroll understood as aptitude with the tools which have been found during the last twenty years. Basically, this means that interlanguage analysis should be applied to language processes in the classroom with the aim of researching differences between individuals (not similarities as up to now) and relating them to parameters which might be responsible for such differences. As the theory stands now, these will be parameters of early language experience and motivation. We should set up a general postulate that investigations into the language learning process must take as much of the whole learning personality as possible into consideration - a postulate which current research certainly does not fulfil (Koordinierungsgremium 1983).

Bibliography

- Bausch, Karl-Richard (ed.). 1979: Beiträge zur Didaktischen Grammatik. Königstein.
- Bausch, Karl-Richard and Horst Raabe. 1978. "Zur Frage der Relevanz von kontrastiver Analyse, Fehleranalyse und Interimssprachenanalyse für den Fremdsprachenunterricht", Jahrbuch Deutsch als Fremdsprache 4, 56-75.
- Bausch, Karl-Richard and Gabriele Kasper. 1979. "Der Zweitspracherwerb: Möglichkeiten und Grenzen der 'großen' Hypothesen", Linguistische Berichte 64, 3-35.
- Braine, M. D. S. 1963. "The Ontogeny of English phrase structure", Language 39, 1-13.
- Carroll, John B. 1962. "The prediction of success in intensive foreign language training", in: R. Glaser (ed.): Training Research and Education. Pittsburg (reprinted 1965)
- Carroll, John B. 1972. "Fundamental considerations in testing for English language proficiency of foreign students", in: H. B. Allen and R. K. Campbell (eds.): Teaching Differences and Universals in Language Learning Aptitude. Rowley, Mass.
- Carroll, John B. 1981. "Twenty-five years of research on foreign language aptitude", in: K. D. Diller (ed.): Individual Differences and Universals in Language Learning Aptitude. Rowley, Mass.
- Carroll, John B. and S. M. Sapon. 1959. MLAT Manual. New York. (frequently reprinted).
- Cazden, Courtney B. et al. 1976: Functions of Language in the Classroom. New York (4. edition)
- Chomsky, Noam. 1965: Aspects of the Theory of Syntax. Cambridge, Mass.
- Clark, Herbert H. and Eve V. Clark. 1977: Psychology and Language. New York.
- Corder, S. Pit. 1973: Introducing Applied Linguistics. Harmondsworth.
- Dulay, Heidi et al. 1972: Language Two. New York/Oxford.
- Dulay, Heidi and Marina Burt. 1972. "Goofing: An indicator of children's second language learning strategies", Language Learning 22, 235-252.
- Dulay, Heidi and Marina Burt. 1974. "Natural sequences in child second language acquisition", Language Learning 24, 37-53.
- Ek, Jan A. van. 1976: The Threshold Level for Modern Language Learning in Schools. London.
- Felix, Sascha W. 1982: Psycholinguistische Aspekte des Zweitspracherwerbs. Tübingen.

- Gardner, Robert C. and Wallace E. Lambert. 1982: Attitudes and Motivation in Second Language Learning. Rowley, Mass.
- Hüllen, Werner. 1976: Linguistik und Englischunterricht 1 und 2. Heidelberg.
- Hüllen, Werner. 1979. "Sprachfunktionen in einer didaktischen Grammatik", in: K. R. Bausch (ed.): Beiträge zur Didaktischen Grammatik. Königstein.
- Hüllen, Werner und Lothar Jung. 1979: Sprachstruktur und Spracherwerb. Düsseldorf/Bern.
- Jakobovits, Leon A. 1970: Foreign Language Learning. A psycholinguistic analysis of the issues. Rowley, Mass.
- Johnson, N. F. 1965. "The psychological reality of phrase structure rules", Journal of Verbal Learning and Verbal Behaviour 4, 469-475.
- Knapp-Potthoff, Annelie and Karlfried Knapp. 1982: Fremdsprachenlernen und -lehren. Stuttgart.
- Koordinierungsgremium im DFG-Schwerpunkt 'Sprachlehrforschung' (ed.). 1983: Sprachlehr- und Sprachlernforschung: Begründung einer Disziplin. Tübingen.
- Krashen, Stephen D. 1981: Second Language Acquisition and Second Language Learning. Oxford.
- Krashen, Stephen D. 1982: Principles and Practice in Second Language Acquisition. Oxford.
- Lado, Robert. 1961: Language Testing. New York.
- Lenneberg, Eric H. 1967: Biological Foundations of Language. New York.
- Lörscher, Wolfgang. 1983: Linguistische Beschreibung und Analyse von Fremdsprachenunterricht als Diskurs. Tübingen.
- Mehan, Hugh. 1979: Learning Lessons. Cambridge, Mass.
- Miller, George A. 1970. "Some psychological studies of grammar", American Psychologist 17, 748,762.
- Miller, George A. 1970. "Four psychological problems in psycholinguistics", Philosophy of Science, 183-199.
- Oller, John W. and V. Streiff. 1975. "Dictation: A test of grammar based expectancies", English Language Teaching 30, 25-35.
- Selinker, Larry. 1972. "Interlanguage", IRAL 10, 209-231.
- Solmecke, Gert (ed.). 1976: Motivation im Fremdsprachenunterricht. Paderborn.

Spolsky, Bernard. 1973. "What does it mean to know a language; or how do you get someone to perform his competence?", in: J. W. Oller and J. C. Richards (eds.): Focus on the Learner. Rowley, Mass.

Stevenson, Douglas K. 1983. "Foreign Language Testing: All of the Above", in: C. J. James (ed.): Foreign Language Teaching. Lincolnwood, Ill.

Stevick, Earl W. 1982: Teaching and Learning Languages. Cambridge.

Valette, Rebecca M. 1967: Modern Language Testing. New York.

Vollmer, Helmut J. 1982: Spracherwerb und Sprachbeherrschung. Tübingen.

Wienold, Götz. 1973: Die Erlernbarkeit der Sprachen. München.

Wilkins, David A. 1976: Notional Syllabuses. London.

Wode, Henning. 1981: Learning a Second Language. Tübingen.

Language Aptitude Testing
in the Canadian Armed Forces

Percy Rangongo

When Canada agreed to provide a speaker on language aptitude, it was with the expectation that one would be available from the Public Service Commission of Canada. Unfortunately, budgetary considerations have precluded this. In Canada, it is the PSC that has conducted practical research relating to the multi-factor construct that John B. Carroll has identified as language aptitude. The bulk of the work I am going to outline for you was done at the Federal PSC of Canada, in the official languages programme namely, the teaching of French and English as second languages.

In the wrap-up, I shall make a brief allusion to aptitude testing within the Canadian Forces Language training programme. In its earliest days, the PSC's Language Training Bureau was interested in the MLAT as an instrument for grouping students for teaching purposes. This use was soon followed by the exploitation of the sub-tests as diagnostic tools. With the inception of intensive language instruction, pressure was brought to bear upon the Language Training Bureau to improve its output and accelerate student progression. Most of the research referred to here is published in articles by Wesche 1981, and Wesche, Edwards und Wells 1982. Two references in this regard will be provided with the typed copy of this talk. The balance of the references are in internal PSC reports. From the prognostic analyses of the MLAT it was found that the MLAT total score correlated between .36 and .69 with student attainment of their Language Knowledge Examination Standards, and .32 to .63 with achievement test scores during training.

PSC review of remedial training data led to the institution of a streaming programme that not only took a student's aptitude test total score into account, but also his or her learning style. This led to the development of two alternate teaching programmes in addition to the strictly audio-lingual teaching method entitled, *Dialogue Canada*, an inductive package patterned on Credif's *Voix et images de France*. These new programmes were *L'Approche Traditionnelle* and *The Functional Approach*. The former is analytical and the latter is patterned along the lines of *language for specific purposes*. Moderate and high aptitude students evidencing inductive and analytical learning styles were found to progress equally well when they followed the approach that is appropriate to them. In seeking to improve success predictions, counsellors at the client centre, have to link performance on the individual subtests of the MLAT with differentiated treatment recommendations. Teachers are not only provided with an indication of the approach consistent with a student's learning style, but also the kinds of treatment indicated by performance on MLAT sub-tests, e. g. an individual who is high on the 'Phonetic Script' subtest is considered a suitable candidate for the audio-visual course, and students who are low on grammatical sensitivity as evidenced by poor scores on 'Words in Sentences' or 'Spelling Clues' are recommended to

follow the *functional approach*. Students with high *grammatical sensitivity* have been found to do well using analytical approaches. 'Number learning' and 'paired-associations' on MLAT have been linked with memory ability and success with audio-visual methodology.

The question may be asked as to what validity this diagnostic use of the sub-tests enjoys: It is reported that performance prediction using this diagnosis in conjunction with student pre-course interviews permits 85-90 per cent predictions. Internal reports using information from formal attitude scales also indicate high student instructor satisfaction. Besides research with the MLAT and parts of the 'Pimsleur' Battery, the PSC has also produced an adaptation of the MLAT for use with Francophone Learners of English. Data from its standardisation indicates that it has as high a predictive value as the MLAT.

On a theoretical plane - Wesche, Edwards and Wells (1982) have conducted a factor analytic study relating aptitude to intelligence. This study interrelates subtests of the MLAT and the subtests of Thurstone's Primary Mental Abilities (PMA). The main finding is that the 'Reasoning' sub-test of the PMA correlates highest with the MLAT total score - the correlation coefficient is over .60. This preliminary work would seem to negate, or indicate a need to temper, Carroll's assertion of a separation between general intelligence and his construct of general language aptitude.

Within the Canadian Forces (CF), aptitude testing has been limited to use of the MLAT as a selection tool. In practice, the prognosis of non-success provided by the MLAT has not deterred nomination of poor aptitude candidates to language training, anyway. Internal studies have revealed clear differential distribution for officers and other ranks. Some promising avenues of research in this area within the CF are indicated in internal work that has linked overall MLAT performance and initial listening comprehension. It is hoped that these directions will soon be pursued.

This brief talk has barely skimmed over the detailed work that has been done at the PSC. It is recommended that members obtain and read the two published references.

Bibliography:

- Diller, Karl C. Editor, Individual Differences and Universals in Language Learning Aptitude Chp VII - Carroll, 1981
- Diller ibid Chp VIII - Marjorie B. Wesche, 1981
- Wesche M.,
Edwards, H.,
Wells W. Foreign Language and Intelligence, Applied Psycholinguistics 1982

Language Aptitude Testing in the United States

Ray T. Clifford and John W. Thain

Language testing, that is aptitude testing, in the United States is motivated by practical concerns rather than theoretical concerns. Professor Hüllen mentioned yesterday that there didn't seem to be much recent research in the United States about aptitude testing. There are good reasons for this. The academic community hasn't done much aptitude testing over the past decade. Faced with ever-dwindling enrollments in foreign languages, they had no need for aptitude tests which might further reduce the number of students in their classes. In the government, on the other hand, we were already following an ongoing program of language aptitude testing which was working very well.

The rationale for aptitude testing in the government is quite straightforward. We are constantly striving to make our teaching as efficient as possible, and aptitude testing helps us accomplish that goal. Assuming a correlation of 1.0, or a perfect correlation between aptitude test results and student achievement, it is a very simple process to establish a minimum passing score for a course and then from that score determine what the minimum aptitude score should be for students coming into the course. (See Figure 1.) Unfortunately the correlations between aptitude and foreign language learning are not 1.0, but range instead from 0.2 to maybe 0.6, depending on the aptitude test and the particular course in question.

A typical correlation of 0.4 might yield a score dispersion something like that in Figure 2. Thus when we define success in terms of a minimum class score under real-world conditions, that score cuts across a very broad band of aptitude scores. If we choose to set the minimum aptitude score in the middle of that band, we make two kinds of mistakes. Some of the people who exceed the minimum aptitude score will fail the course and some who fail the aptitude test could have successfully completed the course. In government programs where we are looking for maximum efficiency the first kind of mistake is much more serious than the second. To minimize that problem we simply raise the minimum aptitude score. When the minimum aptitude score falls at a higher level as in Figure 3, we have minimized the "false positives," the number of people predicted to succeed who will not. Unfortunately there are a great number of people who are rejected from our classes who could have learned the language. Still, as long as we have enough people coming into the services, we can afford to be selective and by setting very high aptitude scores our programs are quite successful. In fact, we have reduced our academic attrition to between ten and 15 percent, or in other words, we have 85 and 90 percent academic success because of this initial screening for aptitude.

The aptitude tests which are most commonly used in the United States are the Modern Language Aptitude Test (MLAT); The Pimsleur Language Aptitude Battery (PLAB); and the Defense Language Aptitude Battery (DLAB). These tests have sections which test first language ability, memory, listening comprehension, one's ability to work with new grammatical structures, and the ability to see patterns in a foreign language. Subtests on some batteries are sometimes very similar in

format to subtests in other test batteries; however, different test authors do not necessarily call such similar subtests by the same names. Some typical language aptitude test items are included in Table 1.

Examples of subtests which measure auditory ability are the Number Learning and Phonetic Script subtests of the MLAT, the Sound Discrimination and Sound Symbol Association subtests of the PLAB, and the Recognition of Stress Patterns subtest on the DLAB. These subtests have been found to have correlation coefficients in the range of .15 to .40 with various criterion measures in different studies. The amount of preparation given the examinee varies from use of a practice item in the Stress Patterns Subtest from DLAB to a mini-lesson in the number system of an artificial language found in the Number Learning Subtest of MLAT.

The names of the various subtests give a partial clue as to the types of items included. For example in the Phonetic Script Subtest of MLAT, the examinee is given some practice in matching short syllables with phonetic transcriptions; then he hears similar syllables on tape and has to choose which of the four phonetic transcriptions correctly represents the syllable he has heard. The Sound-Symbol Association Subtest of the PLAB is similar to the MLAT Phonetic Script Subtest, except that the item stimuli heard on tape are polysyllabic nonsense words and the multiple choice transcriptions are in English. In the Sound Discrimination Subtest on the PLAB, the examinee receives practice in discriminating between pitch, orality, and nasality in a West African language. He then must discriminate similar sounding words in this language in context. In the Recognition of Stress Patterns Subtest on DLAB the subject must choose which of four polysyllabic words in a foreign language has a different stress pattern than the other three options.

The subtests of auditory ability mentioned above use nonsense words and syllables, words from artificial languages, and words from foreign languages not likely to be known to the student. The common underlying assumption on these subtests is that if an examinee can show auditory ability by mastering novel short utterances in the short testing time available, he will have a similar ability to perform the more complex task of learning a foreign language over a longer period of time.

Examples of subtests which measure verbal ability include the PLAB Vocabulary Subtest, the PLAB Language Analysis Subtest, the DLAB Foreign Language Concept Formation Subtest, the MLAT Spelling Clues Subtest, the MLAT Paired Associates Test.

Items in the PLAB Vocabular Subtest consist of a single English vocabulary item as a lead followed by four other English words as English multiple-choice options, only one of which has approximately the same meaning as the English lead. This kind of item format is widely used to test knowledge of first language vocabulary in other tests which are not used as foreign language aptitude test.

The MLAT Spelling Clues Subtest is similar to the PLAB Vocabulary Test, except that the English lead is spelled approximately the way it is pronounced in English rather than by its correct spelling.

In the MLAT Paired Associate Subtest, the examinees memorize English equivalents of a Kurdish vocabulary list. The test items consist of a Kurdish word stimulus followed by four multiple choice options each consisting of a single English word.

In the PLAB Language Analysis Subtest, the examinee is given a few words from a foreign language along with their English translations. He is also given simple grammatical rules in the foreign language. His task is to translate English sentences into the foreign language using the vocabulary and grammar rules he has been taught.

In the DLAB Foreign Language Concept Formation Subtest, the examinee is given a set of phrases in an artificial language, each of which is accompanied by a set of pictures rather than English equivalents. Part of the examinee's task is to deduce the grammatical rules of the language from the pictures by matching pictures with phrases in the artificial language.

The MLAT Words in Sentences Test measures the ability to recognize pairs of words or phrases in context that serve similar grammatical functions.

Like the auditory ability tests mentioned above, some verbal ability subtests present the examinee with a very small corpus of foreign or artificial language material to learn. But other verbal subtests have both English leads and English multiple choice options. Instead of giving the student non-English material to learn, these tests sample the knowledge of English vocabulary and grammar the students already have. The underlying assumption is that if the examinee is able to make such subtle distinctions in his own language, then he will be able to achieve similar facility in a foreign language.

A subtest that cuts across the distinction between auditory and verbal ability is the DLAB Foreign Language Grammar Subtest. In this subtest the examinee reads English phrases which constitute the item leads. His task is to choose which of four phrases he hears is the accurate translation of the lead in a pseudo-English artificial language which consists of the same English lexical items but with radically different morphology, syntax, and grammar. This subtest forms the core of DLAB, which has proven to be a better predictor of student achievement at the Defense Language Institute than either MLAT or PLAB. Since 1978, the U.S. Department of Defense has used DLAB for all its aptitude testing.

Another aptitude testing approach is the development of language specific aptitude tests. Recently, the United States Air Force attempted to produce language specific aptitude tests for Russian, Arabic, and Chinese. These tests are to be used as aids in selecting students for training in Slavic, Middle Eastern, and Far Eastern languages respectively. Sketchy preliminary data indicate that these tests have been moderately successful in predicting attrition in DLIFLC foreign language courses. The subtests of these language specific aptitude lists are much like the subtests on the MLAT, PLAB, and DLAB, but they start out by giving the students training in a limited corpus of the target foreign language rather than using an artificial language. The amount of training given the examinee is also somewhat longer than that given on the major commercial

aptitude tests. In fact these tests might be considered to be a beginning language lesson followed by an achievement test and the results are simply used as an aptitude indicator for the whole course which is to follow.

There is another whole category of aptitude testing where little research has been done--and that's in the area of predicting chances for the continued success of students with prior language learning experience. All of the available aptitude tests have been developed to test students who have never learned a second language.

Success in a previous language course should be an ideal predictor of future success, but because of the diversity of language programs in American high schools and colleges and the lack of universally accepted measures of proficiency, the completion of a given number of course credit units and grades received are very poor indicators of the degree of foreign language proficiency attained. We therefore test these prospective students over again for functional language proficiency to place them into the advanced courses and to predict whether they will be successful or not.

Unfortunately we found that students of equal functional ability in a given language did not have equal chances for future success. That is, they did not all develop at the same rate beyond their present proficiency level. In fact, some of them seemed to be on a plateau from which they never advanced. In identifying these low potential individuals it has been useful to consider not only global language ability expressed in terms of our standard zero to five proficiency ratings, but also the various performance elements that contribute to the overall ability rating.

For the speaking skill interagency agreement identifies these performance elements as grammar, pronunciation, vocabulary, socio-linguistic/cultural, and fluency/integrative factors. It is important to note that a combination of individual elements scores cannot be used to predict proficiency by means of regression equations. That is, performance elements do not contribute linearly to prediction at all points on the language proficiency scale. One particularly important example of nonlinearity is illustrated by the performance element of persons who have been identified as a "terminal 1+" or "terminal 2+." A typical profile of a "terminal 2+" is shown in Figure 4. Such terminal, or low potential, cases are so called because experience has shown that they have very little chance to increase their overall language proficiency. In terms of time on task required to progress to the next level, their aptitude is very low indeed. The single most important feature which identifies these cases is a low subscore on grammatical accuracy accompanied by higher subscores on other performance elements such as vocabulary.

Such terminal cases have typically learned inaccurate grammatical structures which they find difficult to unlearn, and which prevent them from accomplishing higher level communicative tasks. Frequently these persons have studied in programs or been in situations which placed initial emphasis on rapid acquisition of coping skills for lower level communicative tasks without a corresponding emphasis on grammatical accuracy. Their deficiencies in grammar cannot be compensated by higher scores on other elements, because their grammar difficulties prevent them from achieving success when they move on to higher level communicative tasks.

In contrast we find that a flat profile, that is a case where an individual has scores approximately equal across all elements, indicates a background that has great promise for the future, that is, the individual will continue to make progress in global language ability. This finding has implications not only for aptitude assessment for advanced students, but also for curricular design and methodology in beginning language courses.

In summary, we do use language aptitude tests in the United States. In the academic community they are not now being used, simply because they are not needed. In the government we have to be more selective because we have to be more efficient, and obviously we currently use aptitude testing for the good of the government, not for the good of the individual. We are also beginning to explore the possibility of aptitude testing for specific languages and have just begun to look at the problem of predicting continued success for students with prior language experience who want to obtain a truly professional capability in the language.

Figure 1

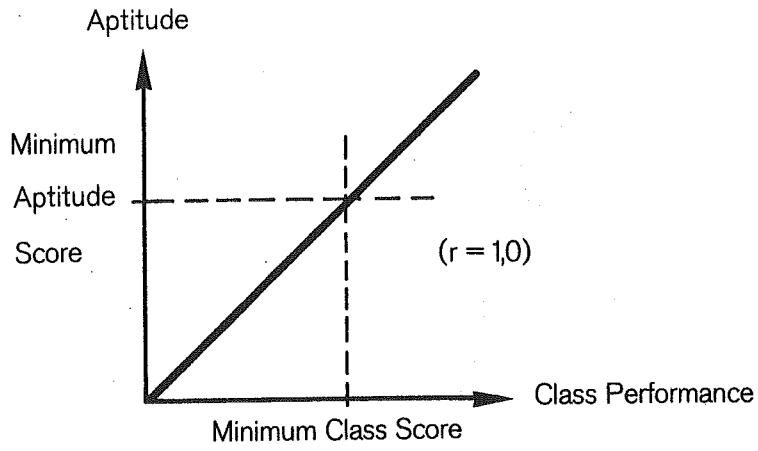


Figure 2

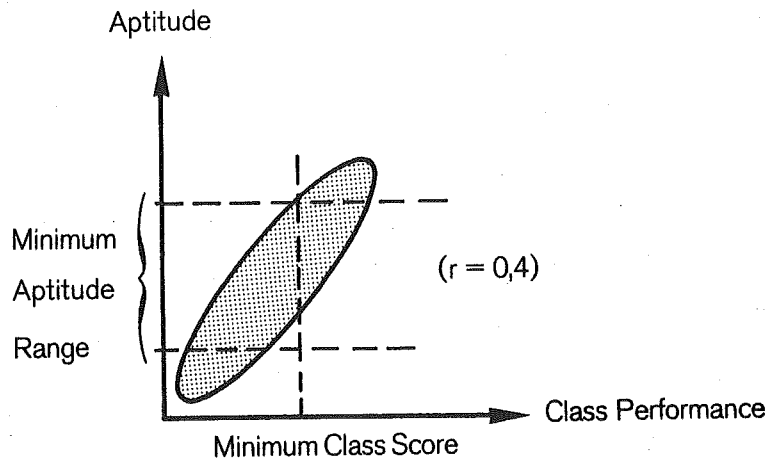


Figure 3

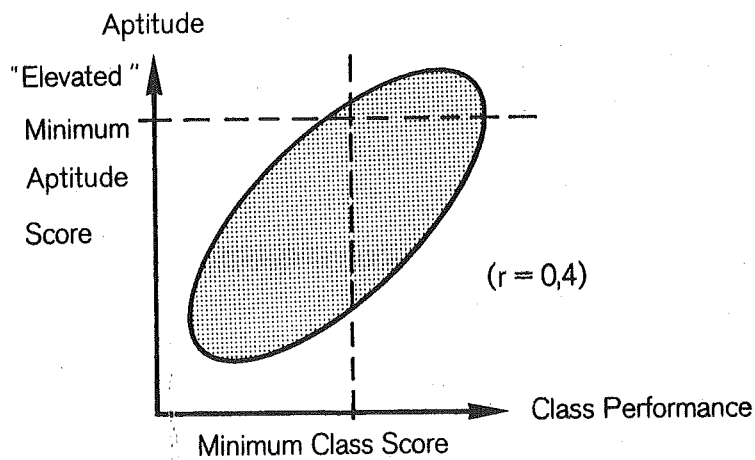


Figure 4

TYPICAL PERFORMANCE PROFILE FOR A TERMINAL 2+

SPEAKING PERFORMANCE PROFILE					EXAMINER	DATE
					TEST NUMBER	
Pronunciation	Fluency/Integrative	Sociolinguistic/Culture	Grammar	Vocabulary	Tasks	
P =	Fl =	SC =	G =	V =	T =	
5 FULLY ACCEPTED BY ENS	5 FULLY ACCEPTABLE TO ENS ON ALL SUBJECTS.	5 USE OF REGISTER, CULTURAL REFERENCES, AND COLLOQUIALS EQUIVALENT TO AN ENL	5 EQUIVALENT TO ENS.	5 EQUAL TO ENS IN BREADTH AND DEPTH ON ALL SUBJECTS.	5 ALL FACTORS INTEGRATED INTO PERFORMANCE EQUIVALENT TO THAT OF AN ENS	
4 BARELY MISPRONOUNCES.	4 HIGH DEGREE OF FLUENCY, EFFORTLESS, SMOOTH, NORMALLY WITHIN RANGE ACCEPTABLE TO NS.	4 RESPONDS APPROPRIATELY ON ALL LEVELS NORMALLY PERTINENT TO PROFESSIONAL NEEDS. TL CULTURE DOMINATES.	4 ONLY OCCASIONAL ERRORS; NO PATTERN OF DEFICIENCY. MAKES USE OF HIGH-LEVEL DIS-COURSE STRUCTURES.	4 EXTENSIVE, PRECISE, AND APPROPRIATE TO EVERY OCCASION.	4 ABLE TO TAILOR LANGUAGE TO FIT AUDIENCE, COUNSEL, PERSUADE, REPRESENT A POINT OF VIEW, NEGOTIATE, AND INTERPRET FOR OTHERS.	
3 ACCENT MAY BE FOREIGN, NEVER INTERFERES. BARELY DISTURBS ENS	3 SPEAKS WITH FACILITY; BARELY HAS TO GROPSE FOR WORDS. FLUENCY AND COMPREHENSIBILITY ACCEPTABILITY HAS SOME PROMINENT DEFICIENCY TO USE PHRASE AND CIRCUMLOCUTION AND REPEATER PHRASES, ALTHOUGH INTERLANGUAGE MAY SURFACE IN IDIOMATIC EXPRESSIONS. CANDIDATE ACTUALLY THINKING IN TL.	3 MAKES FREQUENT APPROPRIATE USE OF TL CULTURAL REFERENCES AND EXPRESSIONS. SOCIOLINGUISTIC INACCURACIES MAY EXIST, BUT DO NOT RESULT IN MISUNDERSTANDING.	3 ONLY SPORADIC ERRORS IN BASIC STRUCTURES. OCCASIONAL ERRORS IN LOW FREQUENCY STRUCTURES, AND MORE FREQUENT ERRORS IN LESS COMMON, COMPLETE STRUCTURES.	3 BROAD ENOUGH TO CONVERSE AND EXPRESS OPINIONS IN FORMAL AND INFORMAL CONVERSATIONS ABOUT FACTUAL, SOCIAL, PROFESSIONAL, AND ABSTRACT TOPICS.	3 CAN CONVERSE IN FORMAL AND INFORMAL SITUATIONS, RESOLVE PROBLEM SITUATIONS, DEAL WITH UNFAMILIAR TOPICS, PROVIDE EXPLANATIONS, DESCRIBE IN DETAIL, OFFER SUPPORTED OPINIONS, AND HYPOTHESIZE.	
2 OFTEN FACILITY, BUT INTELLIGIBLE	2 SPEAKS WITH CONFIDENCE, BUT NOT WITH FACILITY. INFREQUENTLY USES PARAPHRASE AND REPEATER PHRASES. INTERLANGUAGE IS APPARENT, BUT NOT DOMINANT.	2 TL CULTURE MAY PREDOMINATE, BUT SOCIOLINGUISTIC AND CULTURAL CONTRASTS DO NOT OFFEND NS.	2 CAN HANDLE ELEMENTARY CONSTRUCTIONS IN LIMITED DISCLOSURE.	2 SUFFICIENT TO SPEAK SIMPLY WITH SOME CIRCUMLOCUTIONS IN CASUAL CONVERSATIONS ABOUT CONCRETE TOPICS SUCH AS OWN BACKGROUND, FAMILY, AND INTERESTS, WORK, TRAVEL, AND CURRENT EVENTS.	2 ABLE TO FULLY PARTICIPATE IN CASUAL CONVERSATIONS. CAN EXPRESS FACTS, GIVE INSTRUCTIONS, DESCRIBE, REPORT ON, AND PROVIDE NARRATION ABOUT CURRENT, PAST, AND FUTURE ACTIVITIES.	
1 ERRORS FREQUENT; INTELLIGIBLE TO NS USED TO DEALING WITH FOREIGNERS.	1 SLOW, STRAINED, EXCEPT FOR ROUTINE EXPRESSIONS. AVOIDS CERTAIN CONSTRUCTIONS AND VOCABULARY. INTERLANGUAGE MAY DOMINATE.	1 SUFFICIENT CULTURAL/SOCIOLINGUISTIC KNOWLEDGE TO DEAL WITH NS USED TO DEALING WITH FOREIGNERS.	1 ERRORS FREQUENT, BUT INTELLIGIBLE TO NS USED TO DEALING WITH FOREIGNERS.	1 INCLUDES BOTH CONTENT AND FUNCTION WORDS, BUT LIMITED TO EVERYDAY SURVIVAL AND COURTESY REQUIREMENTS.	1 CAN CREATE WITH THE LANGUAGE AND ANSWER QUESTIONS, PARTICIPATE IN SHORT CONVERSATIONS.	
0 UNINTELLIGIBLE.	0 SO HALTING THAT CONVERSATION IS IMPOSSIBLE.	0 NO EVIDENCE OF SOCIOLINGUISTIC OR CULTURAL AWARENESS.	0 TOTALLY WRONG OR INCONSISTENT.	0 INADEQUATE FOR EVEN SIMPLE CONVERSATION.	0 NO FUNCTIONAL ABILITY.	

ENS = educated native speaker

NS = native speaker

TL = target language

NL = native language

TABLE 1

EXAMPLES OF ITEM TYPE FORMATS
FROM MAJOR APTITUDE TESTS USED IN
THE UNITED STATES

1. Phonetic Script Item Format.

Lead utterance on tape: /pirp/

Written options: 1. piyrp
2. pirp CORRECT
3. peyrp
4. piyp

2. DLAB Recognition of Stress Patterns

Lead utterance: None

Tape recorded options: 1. /wokpúdin/ CORRECT
2. /tárpufin/
3. /ládbukim/
4. /zádgúbim/

3. Words in Sentences.

VIENNA is the capital of Austria.

He wanted to go dining in California.
1 2 3 4 5

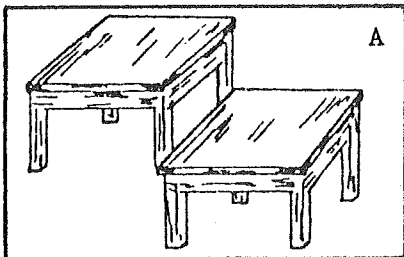
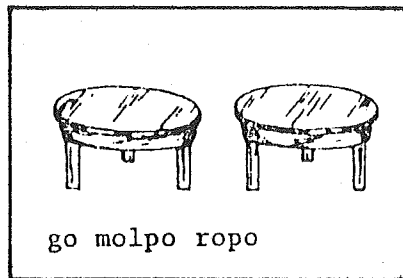
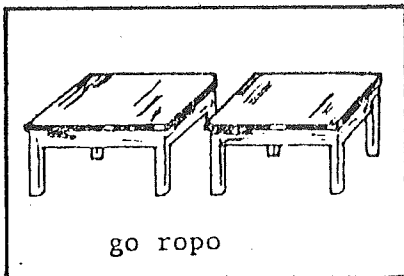
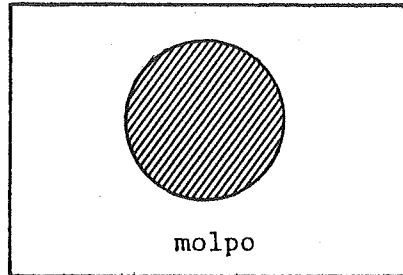
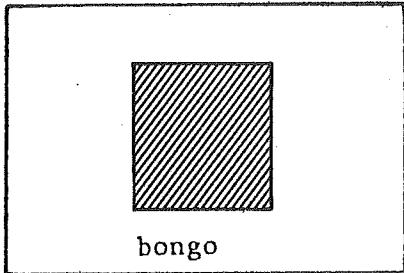
No. 1 option is correct, because he, like Vienna to the first sentence is the subject of the sentence.

4. DLAB Foreign Language Grammar.

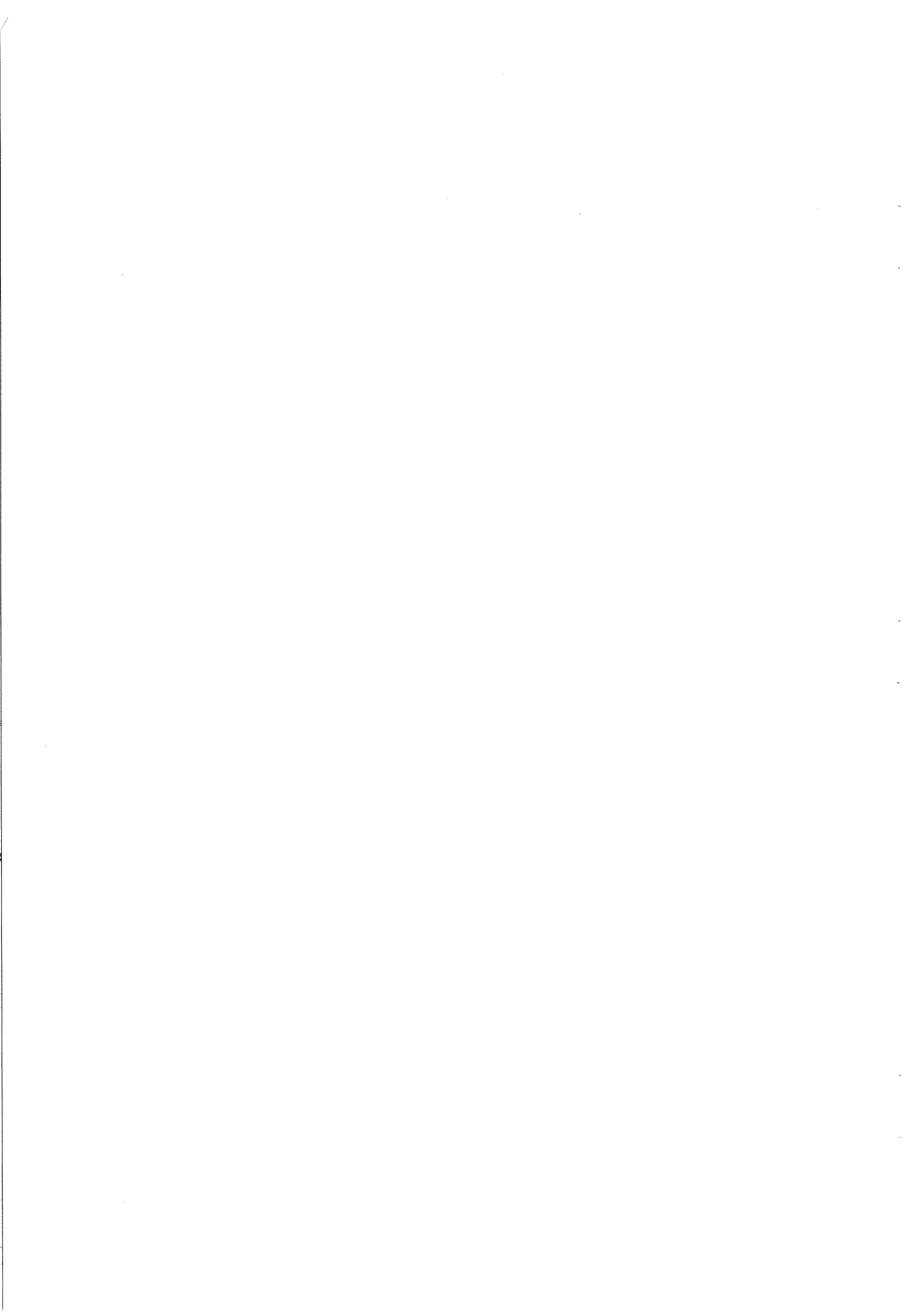
Printed lead: The house is large.

Options on tape: 1. /hawsu larju/ CORRECT
2. /hawsu larju/
3. /larji hawsu/
4. /larju hawsu/

5. DLAB Foreign Language Concept Formation



- A. 1. molpo go ropo
2. go bongo ropo CORRECT
3. go ropo molpo
4. ropo go bongo



Oral Language Proficiency Testing
at the Foreign Service Institute

John Ratliff

For many years, the Foreign Service Institute's Language School has played a major role in developing the theory and practice of oral language proficiency testing. The oral interview proficiency test and the S-scale definitions have traditionally been known as the *FSI system*. Although much of the early work was done by FSI personnel, the *FSI test* has gone well beyond the walls of our Institute. With various refinements, the system has been adopted by other U.S. Government training institutions, by educators and testers throughout the U.S., and by the military training institute of the NATO countries.

The system has remained remarkably intact over the years. It consists of four main components:

1. **TEST FORMAT:** A face-to-face 30-60 minute encounter between the examinee and a native speaker (the 'interviewer'), observed by a testing specialist (the 'examiner') certified to award scores. Test events are not rigidly controlled, consisting largely of conversation at the direction of the interviewer.
2. **EVALUATION CRITERIA:** The features of the examinee's performance that are considered indicative of proficiency, and from which a score is derived. These features have been grouped under five 'factors': Accent, Grammar, Fluency, Vocabulary, and Comprehension.
3. **PERFORMANCE STANDARDS:** A scale of excellence for each factor, used to describe a performance in order to arrive at a global S-score.
4. **LEVEL DEFINITIONS:** A scale of overall ('global') proficiency in functional terms, from S-0 (no proficiency) to S-5 (proficiency equivalent to that of an educated native speaker).

The strength of this system has been its flexibility. The nature of the interview can be adjusted to fit the circumstances, and experienced testers agree on global S-scores despite disagreements on the significance of specific points. For many years this system has been the best option for assessing communicative competence, when that - rather than course achievement - is what is wanted.

The Concerns

All tests are inferential. The only truly direct assessment is on-the-job observation which, in our case, is out of the question. By definition, a 'test' is an attempt to make valid inferences from a short-term demonstration. The validity of the score depends on the strength of the inferences:

"I observed the examinee do A, B and C well, but have difficulty with D and E. I infer from these observations that in real life he can handle X-type situations fairly successfully, but not Y-type situations".

FSI testing specialists took a hard look at the FSI test. They concluded that it was too inferential. That is, judgements are sometimes based on too little evidence or too shaky evidence. There is a tendency to make 'inferential short-cuts': "People who make errors in noun plurals can't be above an S-2".

Conversely, there was sometimes evidence of communicative competence that could not be scored in terms of the five traditional factors. This was particularly apparent in special cases such as 'rusty' native speakers, whose ability to communicate was impressive despite error-laden speech. Less dramatic but similar differences also were noted between more typical examinees.

CONCERN #1: STRENGTHEN THE INFERENCES THAT LEAD TO S-SCORES

A second concern was that of inter-language comparability. Could an S-3 in Chinese do what an S-3 in Spanish could? More? Less? How can we be sure? The S-Scale is defined in functional terms - what people can do at each level - but these functional scores are derived from a non-functional assessment. Testers tend to notice 'correctness' (errors or lack of errors) rather than functional 'effectiveness'.

To ascertain effectiveness it is necessary to pose communicative tasks of some sort and observe the examinee's attempts to perform them. Moreover, to ensure comparability between languages, all examinees should confront the same tasks.

CONCERN #2: DEVELOP A STANDARDIZED TASK-BASED TEST FORMAT

The Revisions

For our assessments to include functional effectiveness as well as linguistic correctness, we had to revise the evaluation criteria and the performance standards associated with them. The test format had to be changed for two reasons: (1) to ensure that examinees had the opportunity to demonstrate their functional capabilities, and (2) to ensure that examinees in different languages had the same opportunities.

The level definitions, already expressed in terms of functional effectiveness, were left unchanged. Indeed, all revisions were designed to be consistent with the well-established S-Scale and

the thousands of S-Scores awarded over the years. Our goal was to improve the validity of our test scores, not to change test itself.

The three revisions are described below.

1. EVALUATION CRITERIA

Testers listen to everything, but what should they listen for? We wanted our testers to include in their judgements everything that contributes to or detracts from the effectiveness of communication. Here is no dearth of possible evaluation factors. On the contrary, there are too many, and they tend to overlap or subsume each other. Since there is a finite limit to the number of factors that can be noticed simultaneously, we resolved to stay with five. The following outline indicates how and why the traditional five factors were altered.

- We wanted to observe features of the examinee's performance such as 'organization', 'turn-taking', 'conversational repair', 'linking/cohesion', and other characteristics of skillful discourse.

NEW FACTOR: Discourse Competence

- We wanted to put all rule-governed matters of correctness together under one factor. Dropping 'accent' and 'grammar', we established a composite factor that includes both morphology and phonology, as well as some aspects of word-ordering.

RE-CONSTITUTED FACTOR: Structural Precision

- We wanted to assess aptness of expression and appropriateness of usage, as well as breadth of vocabulary. We expanded 'vocabulary' to include these concepts.

EXPANDED FACTOR: Lexicalization

- We wanted to observe how readily language comes to the examinee's mind, but not to count off for native-like pauses. We refined the definition of 'fluency' along these lines.

RE-DEFINED FACTOR: Fluency

- We wanted to make it clear that the 'comprehension' factor refers only to comprehending speech in face-to-face situations - where the listener affects the speaker. Non-interactive comprehension, such as listening to speeches or the broadcast media, is not measured by this test.

DE-LIMITED FACTOR: Comprehension

To summarize, the revised model of communicative competence is as follows.

Communicative competence consists (primarily) of five inter-related factors:

DISCOURSE COMPETENCE

STRUCTURAL PRECISION

LEXICALIZATION

FLUENCY

COMPREHENSION

2. PERFORMANCE STANDARDS

The revised factors serve to identify aspects of an examinee's performance that contribute to (or detract from) communicative effectiveness. To score a test, however, it is not enough for the examiner to say, "I now know the examinee's strengths and weaknesses." the examiner must also determine the extent of the examinee's abilities and disabilities.

To record the degree of effectiveness, examiners must refer to some set of performance standards. The nature of this scale is critical. Direct reference to the S-scale level definitions begs the question; seat-of-the-pants judgements are not stronger inferences. On the other hand, if the factors were scaled in abstract terms (say, from 1 to 10 or along a semantic differential), or expressed in terms of an error count, then our revision would amount to little more than a change in terminology.

It was essential to develop a scale of effectiveness. Our revised performance standards set forth for each factor degrees of 'impact' (pro or con) upon the success of the communication during the test.

There is no lessened concern at FSI for good grammar or good pronunciation, however. Any pattern of error that interferes with the success of communication lowers the score, as always. The revised standards do allow examiners to lower scores for reasons other than incorrectness, and to raise scores for reasons other than correctness.

Specifically, the revised performance standards are as follows:

BLOCKING: On a given factor, an examinee is rated 'blocking' if the imperfections in his performance are such that they prevent the interaction from proceeding.

DYSFUNCTIONAL: This rating is given when the imperfections force a temporary interruption in the interaction until significant 'repairing' is accomplished.

INTRUSIVE: The imperfections are clearly noticeable and disconcerting but not otherwise problematic.

ACCEPTABLE: The examinee's performance is such, on this factor, that the examiner observes no significant impediment or enhancement of effectiveness.

SUCCESSFUL: The examinee's demonstrated facility on this factor is a significant contribution to the success of the communication.

SUPERIOR: Control of this factor stands out as a major attribute of the examinee's success as a communication.

These general definitions constitute a six-point scale of increasingly positive impact. For each factor, a very specific performance standard statement has been written for each degree of successfulness. In all the result is a 6 by 5 grid of thirty performance standard statements.

These statements relieve examiners of trying to describe the examinee's performance in their own words. Instead, they listen for evidence which will allow them to select the statement which best describes the degree of effectiveness they heard during the test.

The scoring grid contains only abbreviated statements. The full texts are contained elsewhere in our testing literature and form the basis of examiner-training. During a test, the tester checks the appropriate boxes on the grid. These checkmarks are then transferred to a scoring table to derive a raw score. A conversation chart converts the raw scores to S-scores.

Note that the scale of statements for a factor is not 'smooth', with equal increments between the six points. The 'distance' between intrusive and acceptable lexicalization, for example, is not necessarily the same as the distance between acceptable and satisfactory lexicalization. The statements are, therefore, already 'weighted' in consonance with existing level definitions.

3. FORMAT

The traditional format of the test - the 'oral interview' - had to be better structured to ensure comparability between languages. Moreover, assessing communicative effectiveness requires richer samples of what examinees can and cannot do. Chit-chatting and question-answering alone show little about how well the examinee can manage a conversation or organize an oral presentation.

The ability of the examinee to do these and other tasks common to the work of the foreign services is just what we want the test to reveal. Clearly then, the test should involve several very representative tasks which will permit testers to observe what we want them to assess.

Accordingly, a standardized test format was developed consisting of three representative tasks. They are:

1. a conversation
2. an interview
3. a briefing

The conversation is a warm-up period where tester and examinee exchange socio-personal information. This gives an early indication of fluency, appropriateness of usage, among other impressions.

The interview has the examinee interviewing the tester to obtain his views on a subject of mutual interest. This task reveals how well the examinee can ask question, make his agenda known, and understand what he is being told. The examinee provides the examiner with a point-by-point translation of the tester's views.

The briefing requires the examinee to deliver a 5-minute oral presentation on a professional topic and then respond to questions and comments from the tester. This task allows the examinee to demonstrate his ability to perform 'linking', 'transitions', 'backtracking', 'verbal underlining', 're-stating', and many other critical skills.

All tests in all languages follow this format. Test security is maintained by providing testers with numerous 'kits' containing sets of topics from which the examinee may choose. These testing kits are periodically retired and re-combined.

CONCLUSION

All FSI testers have now been re-trained to administer tests according to the new format. The change has been well received by both testers and students. The general reaction is that the revised format provides for a much better display of communication skills. The standardized format is also fairly easy to learn. Gone is the necessity for surgical probing to determine whether or not the examinee has mastered the subjunctive. The tester interacts with the examinee in a reasonably natural way. All parties play themselves - there is no role-playing.

One spin-off benefit is that many more instructors can share the testing load at the Institute. Another is that examinees seem to accept and understand the test results better.

The re-training of examiners is a more ambitious and challenging effort, involving much tape-listening and discussion with our testing specialists. Examiner re-training will continue throughout 1983. Our Spanish and French examiners who have completed the process have responded enthusiastically. They report that the new factors legitimize some of their observations which, in the past, could not influence the scoring.

We expect the average scores to remain very much the same - at least until our improved curricula take hold. Many individual

scores, however, will come out higher or lower by the new scoring system. An extensive statistical analysis during the 84-85 school year will measure the effect of the revisions. We will make the results available when the study is completed.

Our efforts to strengthen the existing testing system were undertaken to provide our client agencies with better predictions about what their employees are capable of doing in the foreign language. If features of our revised system would help meet the testing obligations of other institutions, we would be happy to make more information available. In any event, you would of course want to develop your own testing kits with tasks drawn from the professions of your students. Please feel free to contact me at the Foreign Service Institute, or write directly to Mr. Gary Crawford, Chairman, Department of Program and Staff Development, Foreign Service Institute, 1400 Key Blvd., Rosslyn, Virginia 22209.

DEPARTMENT OF STATE
FOREIGN SERVICE INSTITUTE
LANGUAGE PROFICIENCY REPORT

NAME		For Office Use Only	
AGENCY	GRADE OR RANK		
SOCIAL SEC. NO.	DATE OF BIRTH	Examiner: _____	
LANGUAGE	LANGUAGE CODE	Tester: _____	
TEST DATE	TESTING PLACE	If State, have you been employed less than 30 days?	
TEST RESULT RAW SCORE _____ S- /R-			

RATING SCALE

←-----|-----→
COMPREHENSION

BLOCKING	DYSFUNCTIONAL	INTRUSIVE	ACCEPTABLE	SUCCESSFUL	SUPERIOR
Misunderstandings of very simple language	En recognizes non-understandings, which are frequent	Difficulties in comprehension slow down the interaction	Wide range of comprehension. Little need for repeats/paraphrase	Understanding is limited only by cultural background	Understanding is essentially complete
DISCOURSE COMPETENCE					
BLOCKING	DYSFUNCTIONAL	INTRUSIVE	ACCEPTABLE	SUCCESSFUL	SUPERIOR
Unable to produce continuous discourse	Minimally elaborate performance at task	Performance shows lack of cohesion	Discourse is cohesive	Discourse is well-organized	Broad flexible control of functional rhetorical speech devices
STRUCTURAL PRECISION					
BLOCKING	DYSFUNCTIONAL	INTRUSIVE	ACCEPTABLE	SUCCESSFUL	SUPERIOR
Structural accuracy is largely random	Frequent miscommunication because of lack of structural control	Inconsistent accuracy even in high frequency usages	Structural inaccuracy is rarely the major cause of misunderstandings	Range and control of structure is relatively great	Structural accuracy and flexibility approaches bi-lingual level
LEXICALIZATION					
BLOCKING	DYSFUNCTIONAL	INTRUSIVE	ACCEPTABLE	SUCCESSFUL	SUPERIOR
Lexicalization is inaccurate and very narrow	Lexicalization is narrow and highly unnatural	Lexicalization is unusual and/or imprecise	Lexicalization is clear and relatively natural	Lexicalization is very rarely imprecise	Lexicalization is precise and appropriate
FLUENCY					
BLOCKING	DYSFUNCTIONAL	INTRUSIVE	ACCEPTABLE	SUCCESSFUL	SUPERIOR
Speaks with great difficulty; pauses exceed speech	Pauses frequently and irregularly; pauses not filled	Speaks readily but not smoothly. Frequent false starts and incompletions	Speaks readily & fills pauses suitably	Speaks readily during even unusual speech tasks; occasionally searches for words	Speaks effortlessly and smoothly; hesitations do not have linguistic causes.

VERSION USED: W- _____ W - _____ READINGS: _____
 I- _____ I- _____
 B- _____ B- _____

SIGNATURES: _____

RATERS: _____

REVIEWED: _____

The Assessment of Language Proficiency and Aptitude
for Specific Jobs

Brendan J. Carroll

Introduction

The aim of this paper is ^{to} put forward certain ideas about specific purpose language tests both of proficiency and aptitude, and to present practical ways of developing and using such tests.

A. Basic Concepts

There are several conceptual areas in which conventional, general purpose tests differ from communicative, specific purpose tests. We list 9 such areas:

Conventional, general

Communicative, specific

- | | |
|---------------------------------------|------------------------------------|
| 1. Concerned with general performance | Related to specific contexts |
| 2. Based on linguistic analysis | Based on communicative function |
| 3. Separate 4-skill components | Multi-mode events |
| 4. Artificial, 'language-like' | Realistic, authentic |
| 5. Mainly objective measurement | Considerable subjective assessment |
| 6. Norm-reference, correlational | Critereon-reference, operational |
| 7. Focus on analysis | Focus on synthesis |
| 8. Tests of knowing | Tests of acquisition |
| 9. Selecting micro-features | Range from macro-meta- to micro- |

B. The Test Product

The different conceptual emphasis lead to different processes of test development and, given practical constraints, a palpably different test product. The communicative tests we have developed have a number of characteristic features:

1. They form a conscious continuum in the context of given jobs and relevant learning programmes showing a harmony of needs, programmes, tests and work contexts.
2. In the construction phases of Design, Development, Operation and Monitoring (D - D - O - M), considerable emphasis is placed on pre-specification of test form according to functions specified and on operational effectiveness, including formal validation.

3. Test performance, either of groups or individuals, is shown in fully described and exemplified levels - for my tests, a 1 to 9 Scale is used.
4. Band performances are expressed both as a profile and as an overall rating with an individual's Performance Profile matched against the Target Profile of the relevant job.
5. Progress monitoring is carried out by establishing successive profiles and calculating rate of progress in hours per band (h.p.b.); progress at the rate of 200 h.p.b., depending on circumstances, is an all-round rule of thumb estimate.
6. Statistical treatment is applied so as to clarify the internal structure of items/sub-tests, but greater emphasis is given to matching test features with external performance criteria and to the congruence of the test with the comprehensive pre-specification of needs.

C. Communicative Aptitude

Earlier work on language aptitude has not shown a very impressive track record, although the prediction correlations quoted are often statistically significant. My diagnosis is that such aptitude studies of linguistic performance are insufficiently comprehensive in a communicative context. Specifically, the manipulation of linguistic patterns at the micro-level (usage) bear only a contingent relationship to communicative operations at the macro-level (use).

Similarly, procedures for measuring later, predicted performance and indices of aptitude obtained from bio-data. For example, I have some evidence to indicate correlations between later speed of progress in English and age, educational level, languages spoken, countries visited and interest in England. A weighted multiple correlation of such indices could well improve progress prognosis.

At present, I am tracking progress on a 1000 hour English course and, expressing 'aptitude' in terms of hours per band progress (more speed, more 'aptitude') am trying to collect discrimination indices for:

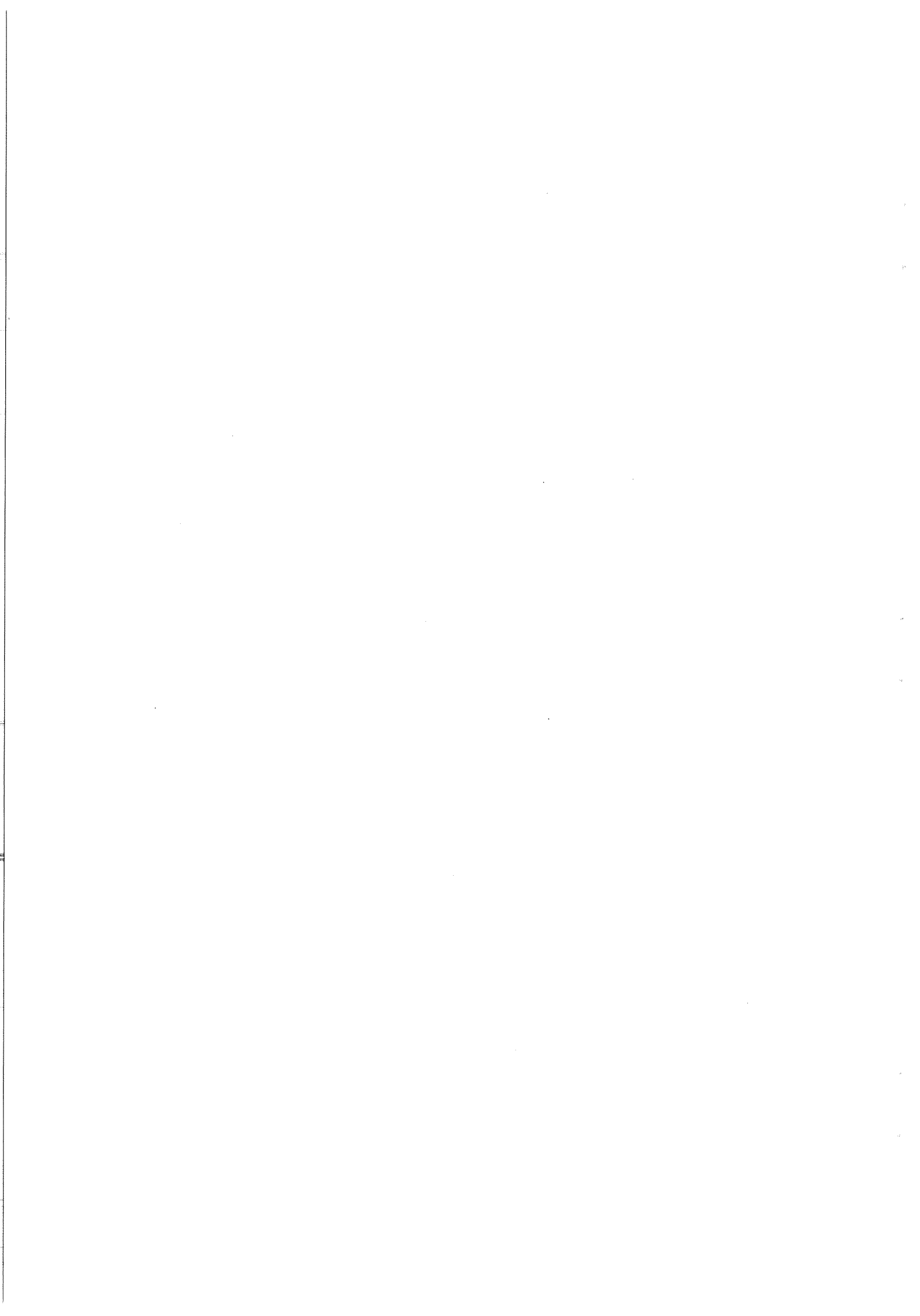
- | | |
|----------------------------|---|
| 1. Age | 7. Job history |
| 2. Educational level | 8. Personal interest |
| 3. Language background | 9. L. learning motivation |
| 4. Travel | 10. Marital status |
| 5. Personality (extr/intr) | 11. Communicative aptitude (strategic) test performance |
| 6. L/Verbal intelligence | 12. Linguistic aptitude. |

D. Tests and Examinations referred to

1. The English Language Testing Service (U.K.)
2. The Royal Society of Arts exam in the Communicative use of English
3. The new SHAPE placement test system
4. The Pergamon English Test Services
5. The new Associated Board Test in English for Academic Purposes
6. The ARAMCO Professional English tests

In Conclusion

Improved test efficiency and communicative scope are allowing us to investigate further the concept of communicative aptitude so as to produce more effective and predictive assessments. It was suggested that BILC co-operate in this research for the establishment of a workable Communicative Aptitude Scale (C.A.S.)



Holistic Language Testing - Facts and Fallacies

Christine Klein-Braley

In 1975 Bernard Spolsky presented a classification of language testing which has since been generally accepted. His three trends follow each other historically, although all of them can still be found alive and kicking today. His first trend was the pre-scientific or traditional. Here tests are developed and scored by the teacher on the basis of his authority. His right to administer any sort of test is unquestioned, both by himself and by the pupils.

In Spolsky's second trend, the psychometric-structuralist or modern trend, testing is the domain of the specialist. Psychometrists and linguists cooperate to develop testing instruments which conform to all the demands of psychometric theory with regard to their measurement properties. The experts tend to view the classroom teacher as an amateur in the field. The teacher, on the other hand, feels that the professional tests offered to him are often not really tailored to the specific needs of his teaching and his pupils.

The third trend in Spolsky's classification is the psycholinguistic-sociolinguistic. Here once again the experts are at work, but this time it is the linguists rather than the psychometrists who determine test format and content. As the name suggests, it was primarily developments in linguistic theory which paved the way for the new tests.

Examination of the similarities and differences between the trends reveals that the first and the third trend share a preference for holistic testing, whereas the second and the third trend share a concern for measurement. The second trend has a distinct preference for discrete-point items, if possible clothed in multiple-choice form. In the second trend only the experts have the time and the money to engage in large-scale test construction. In the first and the third trends the teacher can construct his own tests, but in the third trend his tests are only likely to be satisfactory testing instruments if he follows the advice of the experts. They do the basic research. However, if he takes their advice seriously, then he can be fairly certain that his testing procedures are reliable and valid.

One of the major questions, of course, is what led to the transitions between the trends. Prescientific language testing had primarily made use of holistic test forms: essays, precis, translation, summary, Nacherzählung, dictation. With the developments in linguistics in the 60s in the USA came new approaches to language in general. Examples for this are the move from prescriptive to descriptive grammar; the concern for the corpus as a source - indeed as the only legitimate source - for a description of a language; the behavioristic attitude to language learning. It was accompanied in language teaching by pattern practice drills; the language laboratory; contrastive analysis as an approach to curriculum construction and, last but not least, by the discrete-point item approach to language testing.

It is probably not entirely by accident that Robert Lado is pre-eminent in three of these areas. A central figure in the applied linguistics of the 60s he developed the foundations for language teaching using the behavioural approach; he advocated contrastive analyses as an approach to curriculum development and he wrote the major work on language testing incorporating the suggestions from the other two areas. I should like to emphasise at this point that Lado's works seem recently to have been more attacked than read. In fact he is very much more sophisticated and sensible in his writings than his detractors have often been.

That Lado should have interested himself in language testing was not surprising in view of the fact that testing in general was 'in' in the 60s. The arrival of the first computers and efficient mechanical calculators had made the statistics involved in classical test theory rather less laborious than they had been hitherto, and the publication of basic text books had made test theory accessible to a much larger public. So there was an interest in testing generally. Lado was the scholar who made these findings available to the language specialists.

Now, the interesting fact about classical test theory is that it is at bottom a theory of error. The basic idea behind all the statistical fireworks is that if error can be minimised then what is left is reliable variance. There are two major concepts for error elimination. The first is the concept of multiple measurement. If something is measured many times, then in the long run the error-free measurement is likely to lie very near the average over the total number of measurements. For this reason classical test theory equates reliability with multiple measurement. This is shown, for instance, in the Spearman-Brown prophecy formula, which enables us to calculate the reliability a test would gain if its length and therefore the number of measurements involved were increased.

The second concept is that of unidimensionality. Any scale in a test should be developed in such a way that it measures only one aspect or trait. And once again statistical methods are available to examine the question.

The demands for multiple measurement from the psychometrists coincided with an approach to language which either worked down from the corpus level to the smallest possible unit (in general the phoneme), or which started with the phoneme and worked upwards to the sentence. The sentence - however difficult it was to define - was the largest unit in the grammar. This attitude to language determined the language teaching - pattern practice, which also operates at the sentence level - and fitted very well into the test format required. The tests which were then constructed on the basis of structural linguistics, using classical psychometric techniques repeatedly showed themselves to be highly reliable and acceptably valid.

Thus the change from the first to the second trend is primarily concerned with making language measurable in an objective and reliable manner. But in addition the concept of validity changed. The holistic tests used in the first trend were viewed as obviously valid by their users, since they consisted of job samples. But the tests of the second trend were not highly face valid, and

so their validity had to be demonstrated in empirical investigations. In fact the high correlations obtained between second trend language tests and other available measures including other second trend language tests were sufficiently high to satisfy the psychometrists, although the users, as Stansfield showed in 1976, were not really convinced.

The change from the second to the third trend again depended largely on a change in linguistic theory. The use of language in action came to be viewed as predominant, rather than knowledge of the rules of language. We might characterise this as a change of emphasis from competence to performance - as Spolsky (1975) pointed out, second trend language tests were ignoring two important aspects of language, namely the fact that it was creative and the fact that it was redundant. According to him, the first problem was that of getting someone 'to perform his competence', the second then of interpreting this performance in such a way that generalisations could be made from it to other performances of the same competence.

However, once the Pandora's box of measurement had been opened, it was no longer possible simply to ignore measurement and measurement concerns. This is why even when tests from the third trend look identical to tests from the first trend, they are not, in fact, the same, since the innocent assumption that the tests are *of course* objective, reliable and valid is no longer possible.

In addition to resuscitating tests from the first trend, the third trend attempted to develop new types of tests which would be more appropriate to language use than those of the second trend. At the same time they had to be equally satisfactory measuring instruments. Again, it was the concept of multiple measurements which seemed to offer a viable approach. Rather than just one essay, the examinee was now expected to produce a number of short texts, each of which was evaluated by at least two trained raters. Even better, however, were tests which despite their holistic approach consisted of a large number of items and which could therefore be evaluated by techniques offered by classical test theory.

There was however one snag: the problem of unidimensional scales. Classical test theory could only be used if these third trend tests themselves were unidimensional, and this meant that now language itself had to be viewed as unitary rather than divisible. In view of the changes in linguistic theory, however, this seemed not unreasonable, and in fact Oller and Hinofotis proposed the theory of unitary competence in 1976.

The idea of language as a unitary trait was not new - it was, in fact, the implicit basis for the first trend tests - but even the results from the second trend tests supported the idea, since all the allegedly independent subtests of the test batteries had such high variance which could be assigned to such dimensions as SYNTAX or LEXIS or even READING COMPREHENSION. For approximately the last decade the argument has been raging about this question of unitary or divisible language competence. The early arguments were based on the high correlations between subtests in trend two tests and also on the

high correlations between trend two subtests and the new integrated tests. Oller, for instance, argued in 1971 that since dictation correlated higher with any of the discrete-point item tests in the ELSPE battery, it was the test which could be substituted for all the others if only one test could be administered. The legitimation for the dictation as a test was derived from a theory of language processing: the examinee had to segment an incoming sequence of sound into subsequences which corresponded to words of the language being tested and write these down. Later factor analytic studies were advanced which showed that all subtests including the integrative tests had high loadings on the first general factor. When, as was often the case, the integrative test showed the highest loading on this general factor, then this was interpreted as evidence for the construct validity of the integrative test.

Recently, however, these assumptions have been called into question. Both John B. Carroll and Ulrich Raatz have shown that factor analysis is not a suitable technique for resolving this question. Raatz (1981), for instance, was able to demonstrate empirically that the factor structure of a matrix depends on the population to whom a test is administered. If this is heterogeneous, then a general factor is more likely to emerge than if it is homogeneous. He showed that if a separate factor was to emerge for any type of trait or dimension, then at least two such tests must be entered into the factor analysis.

Further he was able to show that both the method chosen for analysis and the type of rotation selected would affect the results obtained. In fact he reaches the conclusion that a researcher can easily obtain any solution he finds convincing by judicious manipulation of the data and the techniques followed. Carroll (in press) makes the same points, but in addition draws attention to the importance of the teaching and learning history of the sample examined. Since it is normal practice to teach all the four skills in an integrated fashion, one would expect to find a general factor with loadings for all four skills when the results are put through a factor analysis. In fact, as Helmut Vollmer concluded in 1981, the evidence on this question is inconclusive.

As early as 1976 Upshur had pointed out that there probably was no such thing as a genuine discrete-point item test. Any discrete-point item would have to contain vocabulary and grammar and syntactical and morphological constraints. Farhady (1983) drew attention to the logical fallacy involved in the one-sided interpretation of correlation coefficients in favour of integrative tests. A correlation of .8 between a so-called grammar test and a dictation does not necessarily mean that the dictation is a better test than the grammar test. It could quite simply mean that both are good tests of what they are intended to measure. Indeed, given that the grammar test is highly reliable, a correlation of this magnitude could be good indirect evidence that the dictation, too, is reliable, and since this is something which is very difficult to measure using the techniques of classical test theory, the interpretation might even be that the dictation is at least as good as the grammar test.

Perhaps the most important point has again been made by our psychometrist colleagues. Both Carroll and Raatz have considered the implications of a unitary theory of language, and both come to

the conclusion that, in fact, neither teaching nor testing would have to change even if this theory could be shown to be correct. Even such an ultimately holistic skill as driving a car must necessarily be taught as a discrete-point activity, and it is therefore legitimate to test these points individually as well as testing the integrated performance as a whole. While a holistic test of language may well be a legitimate device for proficiency testing, we need discrete-point tests in achievement and diagnostic testing, and we would be foolish to discard these test techniques in favour of the much less precise information offered by integrative testing.

We can therefore conclude that whether language competence is unitary or divisible, the teacher and tester will continue to need both discrete-point item and holistic tests - a point made by Carroll in 1961 when he said:

(discrete-point testing) makes for highly reliable and valid testing. (...) I do not think, however, that language testing (or the specification of language proficiency) is complete without the use of the approach ... requiring an integrated facile performance on the part of the examinee. It is conceivable that knowledge could exist without facility. (...) For this reason I recommended tests in which there is less attention paid to specific structure points or lexicon than to the total communicative effects of an utterance.
(1961:318)

In any case the argument is not yet over. In perhaps the most sophisticated attempt so far made to decide which of the two hypotheses is correct, Bachman and Palmer come to the conclusion:

This study ... has provided some support for the hypothesis that there are independently measurable speaking and reading traits - enough support to warrant continuing with phase 2 of the investment.
(1981:162)

* * *

One of the major psycholinguistic test forms to emerge from the third trend has been cloze tests. Indeed cloze procedure has probably received more attention from researchers in the last ten years than any other type of language test. Again, it was Oller who suggested the use of cloze procedure for L2 testing in 1971. At the same time he recommended that the scoring techniques used in L1 testing should not be followed. He also modified the theoretical assumptions. Cloze procedures is considered to be a technique for constructing reading tests in L1, or, alternatively, a technique for measuring text difficulty (cf. Klein-Braley 1981, 1982 a, 1982 b). But Oller suggested that in the L2 setting cloze procedure could be used as a technique for measuring pragmatic expectancy grammar - Oller's term for the mechanism involved in processing language.

Cloze procedure conformed to Spolsky's demand for tests measuring creativity and redundancy in language. Since the text is unknown, the examinee is confronted with a novel piece of language. Since every nth word is missing, the examinee must use both his own language and the clues that the redundancy of language provides to restore the text. The scoring technique which accepts as correct any word which fits the text syntactically and semantically ensures that the L2 examinee is not penalised for his inability to use the words of the original author.

As I have stated already, an immense amount of research has been performed using cloze tests. Unfortunately much of this research is useless, mainly because cloze tests were far too quickly accepted as valid language tests, so that research into cloze tests themselves and the assumptions underlying these tests was neglected. It quickly became the custom to include a cloze test in a research design as the 'anchor measure', the test which was ipso-facto valid, and which could therefore be used as the central measure against which all the other tests could be evaluated. It also became normal to recommend cloze tests to teachers, since they were quick and easy to produce, and had received the legitimation of the testing experts.

One of the reasons for the rapid popularity of cloze tests was the underlying theory.

It is difficult to find a coherent account of the theory underlying cloze tests. Many attempts seem more an ex post facto description of processes which might take place inside the examinee when s/he is completing a cloze test than a genuine theoretical foundation for test construction. And in a sense, this was the order things happened in, although there is no doubt that some type of test of reduced redundancy was something people were looking for at the moment when cloze tests appeared on the scene. Oller himself, incidentally, gave the credit to Christine Conrad, his co-author in the 1971 paper.

Since, as will become obvious later, the C-test is a development of the cloze procedure, I have tried to assemble the underlying theoretical assumptions so that they look more like a theory and less like an unrelated collection of flashes of genius. This is necessary if one is to be able to test the theory, and in empirical terms the only good theory is one that can be tested, even if it turns out in the end to be wrong.

We start with Chomsky's concepts of *competence* and *performance*. Competence is to be viewed as an abstract system of rules internalised in the language user; performance is the rule system in action used for communication. Language use is controlled by the user's competence, and if the competence is fully developed and nature, then the user will perform linguistically as a native speaker of the language in question. This is difficult to define, but obviously easy to discern, and explains, for instance, the frequent compliments I receive on my allegedly native-like performance in German. No-one ever compliments my German husband on his ability to speak German, so obviously Germans have, in fact, a very fine ear for the distinctions between native and non-native performance.

It seems that in both L1 and L2 the learner begins with a competence of zero. As learning proceeds, so the level of competence is expressed in better performance (here using 'better' in the sense of more adult-native-speaker-like). Recently scholars such as Selinker (1972), Corder (1967), Krashen (1981) and Nemser (1971) have discussed the development of interim or approximate systems of the learner. Analysis of the learner's performance permits assumptions to be made about the state of his internalised rule system or competence at the time of testing.

Competence, therefore, is not testable or, indeed, observable in any way. Performance, on the other hand, is both observable and testable. Therefore the only way of assessing competence is to put the learner in a controlled situation where a specific objectively-defined performance is required. The level of performance will reveal the level of competence. This can be either absolute, as suggested for instance by Carroll, Carton and Wilds in terms of the performance of the adult educated L1 speaker, or it can be relative, for instance in terms of the examinee's peer group (better/worse or top/bottom or 1.3 standard deviations above the mean).

One technique for forcing the user to activate his/her competence is to present the learner with a piece of mutilated language. If the user's competence is fully developed s/he will be able to use all levels of language to restore/reconstruct the text - there will be grammatical, syntactical, lexical, semantic, collocational, contextual, pragmatic, logical, situational clues - and no doubt many others. It is not even necessary to know how the examinee arrives at the solution; it is sufficient in the context of the test that s/he finds an acceptable one.

The final score obtained by the examinee in a language processing test is a numerical estimate of his competence. Naturally, this is only an indirect estimate, but in testing situations many estimates can only be indirect. One of the main problems with cloze procedure however is that the necessary reference point on the scale is missing. Whereas with other L2 tests the performance of the native speaker adult is taken as the reference point - and the assumption is made that his score would be 100% - there is no guarantee at all that the adult educated native speaker will obtain 100% on a cloze test. This assumption is, of course, made when the adult educated native speaker is the *scorer*: in this case he *is* able to judge whether the solutions offered are acceptable, but research in the L1 area shows that this just won't hold water (cf. in this context Taylor 1953 and Klein-Braley 1982 a) (Arnhem).

Cloze tests are therefore based on the concept of competence as a language processing mechanism. Competence governs all linguistic expression (language use) and is always actively involved, even in such apparently passive activities as reading or listening. Other theoretical concepts which have been used in connection with cloze tests include information theory and redundancy - which is another way of saying that the examinee has a large number of clues available to help him in restoring the text. A useful contribution from Oller ties in the concept of short-term memory (1973). The higher level learner can chunk more segments simultaneously into higher units, and thus hold longer sections

of text in memory at any one time. Thus he can make use of information that is not available to the lower-level learner.

We still have to explain why we can speak of 'cloze' tests in the abstract rather than having to refer to a published test battery such as the ELSPE, the TOEFL or the Cambridge Proficiency. This is because cloze procedure is a technique for test construction, rather than a test form in itself. Cloze tests, as Stevenson pointed out in 1978, are tests 'without a well-defined content'. The text used for testing is, of itself, irrelevant. What cloze tests aim at doing is obtaining a random sample of the examinee's performance when confronted with potentially *any* mutilated text, and they do this by using a random deletion technique. The elements deleted in the text are considered to be a random sample of all the elements in the text, the text itself is considered to be a sample taken randomly from all the possible texts in the language. And random sampling, as any statistician will tell us, is always a Good Thing.

It is, in fact, the statistical nature of test construction which forms the legitimation for generalising from one single performance to all other possible performances. The fact that actually random deletion is not used, but that instead words are deleted systematically from the text using a given number n (which is usually between 5 and 10) does not necessarily conflict with this idea, provided we can show that random deletion and systematic deletion are equivalent.

However, only if we are willing to accept these theoretical assumptions is there any justification at all for the claims so often made that 'cloze tests are reliable and valid tests of general language proficiency'. Only if the cloze procedure is in some way special and different from other types of test construction procedures, can there be any rationale behind the automatic reliability and validity so often claimed. You can test this for yourself by trying the statement 'I have found, after having constructed, administered and scored over a thousand *multiple-choice* tests to ESL students in the last three years, that the *multiple-choice* test is an extremely simple, yet valid language proficiency test.' (adapted from Aitken 1977:66 who, of course, was referring not to multiple-choice tests but to cloze tests.)

We can now set up a paradigm for testing the assumptions of cloze tests, and this was, in fact, the paradigm is used in my own study (Klein-Braley 1981).

Assumptions

1. Mechanical n th word deletion produces a random sample of all possible elements of the language.
2. All cloze tests using mechanical n th word deletion set the mental process of pragmatic expectancy grammar into action and test its efficiency.
3. The actual text chosen for test construction is irrelevant (within very broad limitations).

4. Tests may vary in difficulty but they will still rank examinees in the same order.
5. The actual deletion rate chosen for test construction is irrelevant
6. The starting point for deletions in the text is irrelevant.
7. Therefore all cloze tests are parallel or equivalent tests.
8. Therefore if one cloze test can be shown to be reliable, all cloze tests are reliable.
9. Therefore if one cloze test can be shown to be valid, all cloze tests are valid.

The results of the two major studies performed on cloze tests in the last few years (Alderson 1979 and Klein-Braley 1981) show that in fact none of these assumptions stands up to empirical testing. It is not my aim to present the detailed results here, although I am very willing to make them available to anyone who is interested. They can be summarized as follows:

1. The cloze construction principle (deletion of every nth word) does not ensure random deletion of the elements in the text.
2. Each of the factors text, deletion rate, starting point affects the quality of the test as assessed by reliability and validity coefficients.
3. The ratio of structure/content words deleted affects the difficulty of the test.
4. The test constructor cannot tell in advance how his text, his starting point and his deletion rate will affect the test.
5. Tests using the higher deletion rates (8, 9, 10) need to be extremely long if 50 items are to be ensured.
6. Use of only one very long test can result in unfairness for the examinees as a result of test/text *content*.
7. Exact scoring, while objective, is inappropriate for L2 examinees since the tests are then too difficult.
8. Acceptable scoring is not objective.
9. The scoring method used affects the reliability and validity of the test.
10. The two methods of scoring do not necessarily produce equivalent results (at least for advanced L2 students).
11. Reliability coefficients for homogeneous groups are low (homogeneous = either classroom groups or monolingual groups).
12. Cloze tests have little relationship to activities in the classroom since the scores are difficult to interpret.

13. Empirical studies show weak or zero relationships between cloze scores and teacher judgements.
14. The reference point (e. g. native speaker performance) is missing.
15. Validity could not be demonstrated for German L2 English speakers (Klein-Braley); while for Alderson's validity coefficients seemed to vary more or less at random.

This summary includes the findings of both researchers, and I think you will agree that it would be difficult to be more negative about a test form than this.

The odd thing is that in spite of their negative findings, both Alderson and Klein-Braley recommended that the theory should be retained and that a more satisfactory test form might solve the technical problems. In other words, it was not the theory, as such, which was at fault; the problem lay in the operationalisation of the theory in classical cloze procedure.

In 1981 I finished the work on my cloze study, and at that point Ulrich Raatz, who had been involved in the cloze research in the capacity of one doctoral adviser, moved into the project directly. We decided that any improvement over cloze tests should have the following qualities:

1. It should use much shorter texts but produce at least 100 items.
2. There ought to be no problems involved in the choice of deletion rate and starting point.
3. The deletions ought to be an absolutely representative sample of the text.
4. It should not favour examinees with specialised knowledge.
5. It should use only exact scoring, and therefore be entirely objective.
6. Adult educated native speakers should obtain virtually perfect scores.

In addition, the new procedure should provide reliable and valid tests and be relatively easy to develop.

Our solution to the problem is the *C-test*.

The C-test is a cloze test in which every second word is deleted. However, in order to ensure that solution is possible at all, we leave the first half of the word standing. If the word has an odd number of letters we delete the larger half. If the word has only one letter, like the English words *a* and *I* we ignore it in the counting. The examinee restores the missing halves of the words. Only entirely correct restorations are counted as right.

Example C-Test

Once upon a time, there was a little girl who lived with her mother, who was a widow. They were _____ so poor _____ that one day the _____ had not _____ left to eat. The _____ little girl went _____ into the _____ woods to _____ play. She _____ was so hungry that _____ she began _____ to cry. An old _____ woman came _____ up to _____ her. "Why _____ are you _____ crying, my _____ child?" she _____ asked. "Because _____ I am so _____ hungry," said the little girl. "Then you shall be hungry no more," said the old woman.

We began with a number of pilot studies. Our first question was, of course, whether the technique would work at all. We also needed to know whether the native speaker would achieve a virtually perfect score. The third question was whether the non-natives would have a sufficiently wide range of scores to make the results interpretable.

The results of these pilot studies were positive. So we then started to look into the linguistic assumptions. Two of our students were persuaded to collect 40 English and 40 German texts of 200 words each from a wide variety of sources. They analyzed these into nouns, adjectives, verbs, adverbs, prepositions, articles and remainder. Then they deleted every second word and compared the distribution of the deleted elements to the distribution of the deleted elements to the distribution of the elements in the original texts. The results showed that for every one of the 80 texts the deleted samples are representative for the entire text.

One final problem remained: that of text length and possible unfairness as a result of text content. We therefore decided that any C-test would contain at least 5 independent texts. This had two advantages. The first was that the test would be fairer. The second was that each of these texts could be viewed as one test item. This avoided the knotty problem of statistically interdependent items, and enabled us to use Cronbach's Alpha to determine reliability.

In the meantime we have performed a large number of empirical investigations into various aspects of the C-test. Inevitably German groups, or groups of subjects available in Germany, constitute our largest sample groups. We realize that in some cases the numbers are rather small. However this is a classical example for home-brewed research: the total manpower (womanpower) available consists of Raatz, myself and our student assistant, Margit Backhaus - to whom we are incidentally very grateful! Six months ago we were joined by Edgar Suessmilch, who was responsible for the immigrant worker groups. But so far we have been working without outside support or financial assistance.

We would therefore describe our results as promising, but not at the moment, as conclusive.

The first groups to be examined were schoolchildren in England and Germany learning their own language. Here we were confronted with the problem of text selection, which we solved pragmatically by using texts taken from text or story books specifically aimed at the age groups concerned.

Table 1 shows the results for the English group.

RELIABILITY AND VALIDITY OF C-TESTS FOR ENGLISH L1 GROUPS

#	AGE	N	REL	VAL
1	10	44	.91	.73
2	11	15	.75	.86
3	12	30	.66	-
4	10	26	.88	.88
5	15	19	.77	.10
6	13	26	.92	.77
0				

In two cases the teacher involved had only been teaching the class for a month, and while one of the teachers allowed herself to be persuaded into giving an assessment of the children's standing in English, the validity coefficient is low. However in those cases where the teacher knew the class, the validity coefficients are high, as are the reliability coefficients in almost all cases.

The German groups were tested much later in the school year, and here in all cases the teachers knew their pupils well. In this

RELIABILITY AND VALIDITY OF C-TESTS FOR GERMAN L1 GROUPS

#	GRADE	N	REL	VAL
7	3	130	.86	.80
8	4	92	.88	.72
9	6	63	.80	.47
10	7	49	.76	.47
11	8	62	.85	.60
12	3	110	.88	.52
	3	110	.87	.65

table all the validity coefficients are satisfactory, and in only one case does the reliability coefficient fall below .80. We view these results as an indication that the C-test performs well for the L1 learner provided s/he has not yet approached the final adult native speaker competence too closely.

The second group for whom we developed C-tests consists of LF learners. Our term LF covers all learners of a language who are not resident in the country where the language is used as a national language. We need this distinction because, as you will see shortly, we also have L2 groups in the study.

The first LF group is the students with whom I am concerned in my everyday work. They are all students of Anglistik at the University of Duisburg, and they are either aiming at becoming teachers

RELIABILITY AND VALIDITY OF C-TESTS FOR GERMAN LF GROUPS
(University students of English)

#	N	REL	VAL
13	27	.81	.72
14	64	.80	.62
15	19	.91	.79
16	14	.77	.90
17	67	.91	.71
18	67	.90	.69
19	67	.84	.65
19a	43	.91	.88

or at taking an M.A. degree with English as a main subject. These are advanced students of English, and constitute an extremely homogeneous group. They are, incidentally, the group for which the cloze tests failed to work because of their low reliability.

The validation criterion here is the DELTA test - a psychometric-structuralist test in the best tradition, which has been very carefully examined and shown to be both highly reliable and valid (cf. Klein-Braley/Lück 19 and Schwibbe/Schwibbe, 1983 unpublished MS). Although in all cases the validity coefficients are satisfactory, the reliability coefficients reflect a problem which has been very difficult to solve. It is very hard to find texts which are sufficiently challenging for these students. With all the other groups we can proceed in the usual way, starting with an easy text and working our way up to a difficult text, but here this approach is not possible. Not until experiment 19a did I succeed in finding five texts which approximated to the desired

difficulty level of around 50%. The low reliability coefficients were obtained for those trials where the test was too easy overall.

With the German LF group, students at a language school in Cologne, these problems did not appear since in this group we have the full spread of language learners from complete beginners to highly proficient students. Again a carefully constructed and validated test was available as a criterion, and both reliability and validity coefficients are most satisfactory. In experiment 20 we used the same German C-test as that for the German children in the third grade. Although the results were satisfactory the teachers reported that the students found the texts too childish. We therefore asked the teachers to select suitable texts themselves. The test constructed from these texts was used in experiment 21.

RELIABILITY AND VALIDITY OF C-TESTS FOR LF LEARNERS OF GERMAN
(Language school)

#	N	REL	VAL
20	99	.92	.82
21	153	.93	.85

Finally we have the L2 groups. As you probably know, Germany has a large number of immigrant workers who were recruited during the prosperous years. Many of these workers have de facto if not de jure made their homes in Germany, and one of the burning educational questions in Germany at the moment is that of teaching German to their children. In some schools there are only 20% German children as against 80% children of immigrant workers. This reflects the differential fertility of the two groups.

#	NAT	GRADE	N	REL	VAL
22	L2tot	3	231	.86	.72
	Turks	3	92	.85	.61
	Greeks	3	107	.85	.74
	other	3	32	.83	-
	German	3	109	.92	.80
23	L2tot	4	245	.92	.69
	Turks	4	94	.87	.41
	Greeks	4	96	.92	.82
	other	4	55	.88	.73
	German	4	88	.88	.82

In a large number of schools there are therefore special classes provided for these children, in which they are taught some subjects in their own language, but where the main emphasis is certainly on teaching them German as fast as possible so that they can move into the normal classes.

If one talks to the teachers involved in these schools, the first demand one hears is for an efficient and accurate technique for deciding just how much German a child knows. However such tests are not, at present, available.

Edgar Suessmilch, a colleague working at Düsseldorf University, decided therefore to examine whether the C-test could be used in this area. Table 5 shows his results. Again, as you can see, they are extremely encouraging and it seems worthwhile continuing to experiment in this area. Here we have, however, exactly the opposite problem to that with the German students of English - the tests are both too long and too difficult. We have been able to reduce the number of test parts without losing reliability or validity, indeed in some cases the coefficients increase. However a lot more work is needed in this area.

I think you will agree that our results so far are promising. It looks as though the C-test may indeed be one solution, possibly one among many, to the problems with classical cloze tests. Since this is so, you may wonder why in the first part of my talk I was so insistent on the need for both types of test, holistic and discrete-point item tests.

One of the things which disturbs me about the work in the general area of applied linguistics is our overenthusiasm for everything which is new and different. We seem to follow the precept of 'off with the old and on with the new' without ever pausing to consider whether we aren't in fact throwing a cap which still has a fair amount of use in it over an unnecessary windmill. Why, in fact, can't we have *both* pattern practice *and* communicative activities in the classroom? Why can't we have cognitive code learning *and* the language laboratory? Why can't we have both grammar *and* functions?

One of the things we are afraid of is that the C-test might become a successor to the cloze test in a way that neither of its inventors could accept. Let us - and any of you who would like to join in - do the spade work properly before we begin to propagate it for use by the non-expert. Teachers have far too much to do in the course of their everyday work to be able to replicate all the research carried out by the test expert. At the same time they have a right to demand that test forms propagated by the test expert have been properly investigated before they start to be recommended for general use.

We feel that the C-test has its rightful place in placement and proficiency testing, probably also in achievement testing by helping the teacher to gain a general overview of the pupils in his class when he meets them for the first time. But we are fairly certain that C-tests are not diagnostic instruments (except maybe in certain specific areas such as dyslexia), and also that they are unlikely to be useful in normal teaching-based achievement testing.

We would be delighted if the C-test ultimately turned out to be a new and useful tool for both testers and teachers. We would, however, find it extremely regrettable if it - like the cloze - were to be prematurely propagated as the philosopher's stone or the test to end all tests, before it has been very thoroughly investigated.

Bibliography

Aitken, K. G.

Using cloze procedure as an overall language proficiency test.
TESOL Quarterly 2: 240-226
1977

Alderson J. C.

The cloze procedure and proficiency in English as a foreign language.
TESOL Quarterly 13: 219-226
1979

Bachman, L. F. and A. S. Palmer

A multitrait-multimethod investigation into the construct validity of six tests of speaking and reading.

In: Palmer, A. S., P. J. M. Groot and G. A. Propser (eds): The construct validation of tests of communicative competence.

Washington, D. C., TESOL, 149-165
1981

Carroll, J. B.

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.
1961

Carroll, J. B.

Psychometric theory and language testing.

In: Oller J. W. Jr. (ed), Issues in language testing research.

Rowley, Mass., Newbury House, 80-107.
1983

Carroll, J. B., A. S. Carton and C. Wilds

An investigation of cloze items in the measurement of achievement in foreign languages.

College Entrance Examination Board Research and Development Report.

Cambridge, Mass., Laboratory for Research in Instruction. Harvard University.

1959

Corder, S. P.

The significance of learners' errors.

IRAL 5, 161-170

1967

Klein-Braley, C.

Empirical investigations of cloze tests. An examination of the validity of cloze tests as tests of general language proficiency in English for German university students.

University of Duisburg. Doctoral Dissertation.

1981

Klein-Braley, C.

On the suitability of cloze tests as measures of reading comprehension.

In: Lezen in Onderwijs en Onderzoek. Toegepaste taalwetenschap in artikelen 13, Jaargang 1982:2

Amsterdam, VU Boekhandel/Uitgeverij, 49-61

1982a

Klein-Braley, C.

Textverständlichkeitsforschung.

Paper presented at the Erstes Colloquium zur Quantitativen Linguistik,

Essen

1982b

Klein-Braley, C. and H. E. Lück

The development of the Duisburg English Language Test for Advanced Students (DELTA)

In: Grotjahn, R. and A. Hopkins (eds), Research on language learning and teaching. Quantitative Linguistics 6.

Bochum, Verlag Dr. N. Brockmeyer, 85-126

1980

Krashen, S. D.

Second language acquisition and second language learning.

Oxford, Pergamon Press.

1981

Lado, R.

Linguistics across cultures. Applied linguistics for language teachers,

Ann Arbor, University of Michigan Press.

1957

Lado, R.

Language Testing

Longon, Longmans

1961

Nemser, W.

Approximative systems of foreign language learners.

IRAL 9, 115-123

1971

Oller, J. W. Jr.

Cloze tests of second language proficiency and what they measure.

Language Learning 23, 105-118

1973

Oller, J. W. Jr. and C. Conrad

The cloze technique and ESL proficiency.

Language Learning 21, 183-195

1971

Oller, J. W. Jr. and F. Hinofotis

Two mutually exclusive hypotheses about second language ability: factor analytic studies of a variety of language subtests.

Paper presented at the Annual Meeting of the Linguistic Society of America

1976

Raatz, U.

Are oral tests tests?

In: Klein-Braley, C. and D. K. Stevenson (eds): Practice and Problems
in language testing 1

Frankfurt etc., Verlag Peter D. Lang, 197-212

1981

Schwibbe, G. and M. Schwibbe,

Validierungsuntersuchungen zum Duisburg English Language Test for
Advanced Students (DELTA)

Unpublished MS

1983

Selinker, L.

Interlanguage

IRAL 10: 209-231

1972

Spolsky, B.

Language testing: art or science?

Paper presented at the AILA World Congress, Stuttgart

1975

Stansfield, C.

Teachers' attitudes towards publishers' tests.

In: Benseler, D. P. (ed), Second language teaching 76. Proceedings
of the Pacific Northwest Council on Foreign Languages. Corvallis,
Oregon, P.N.C.F.L.

1976

Stevenson, D. K.

Face validity and loss of faith - some effects of recent cloze re-
search on traditional views of language proficiency.

Paper presented at the 9th Jahrestagung of the GAL, Mainz.

1978

Taylor, W. L.

Cloze Procedure - a new total for measuring readability.

Journalism Quarterly 30, 414-438.

1953

Upshur, J. A.

Discussion of 'A program for language testing research.'

In: Brown, H. D. (ed) papers in second language acquisition. Language
Learning Special Issue Number 4, 167-174

1976

Recent Experiences with an Aptitude Test
for Studies in Modern Languages

Günter Trost

1. Preliminary Remarks

Two psychometricians meet in the street. One says: "How is your wife?" Replies the other: "Compared to what?"

There is no doubt that people working in the field of educational assessment are handicapped by a "déformation professionnelle" which makes them incapable of saying simple things such as "My wife is well". It is true as well, though, that even in every day life most qualifying statements of persons on persons are implicitly based on comparison. In the testing business the process of comparison is made explicit by a plain definition of a particular reference group or of a particular criterion of competence or proficiency and usually by the quantification of the relation of an individual's test result and these reference norms.

I would like to talk about an Aptitude Test for Studies in Modern Languages which presently allows for a comparison of individual test results with the results of a large group of persons who are interested in studying modern languages and which, after the completion of longitudinal validity studies, will allow for a comparison with the test scores of those who actually perform well in the respective study courses at university.

Four years ago, at this very place, I presented the preliminary version of that test, and I reported on test construction and on the results of studies with this test (TROST/BICKEL 1981). Today I would like to give you an account of the revision of the test which has occurred in the meantime, present the test in its actual form and report on recent experiences with the application of this test - among other aptitude tests - in the framework of a counselling programme. Before I do so, let me briefly resume the concept on which the development of this aptitude test is based, and the test's genesis.

2. Concept of Test Development

Since 1975, the West German Institut für Test- und Begabungsforschung (Institute for Test Development and Talent Research) has been developing seven 'study-field oriented tests' (Studienfeldbezogene Tests) aiming at the assessment of the specific cognitive aptitudes which are relevant for success in the following university subject areas or 'study fields':

Business and Economics	Modern Languages
Engineering Science	Natural Sciences
Law	Psychology
Mathematics	

(TROST/BICKEL/BLUM/CHRISTIAN 1979)

The tests are conceived to be used exclusively for counselling purposes. They are devised for students who have completed secondary education ("Abiturienten") and wish to obtain an orientation on their individual chances to succeed at university in the particular study courses they are interested in. This information, along with the results of an interest questionnaire, shall help them in their decision about whether to go to university, respectively in which courses to enrol.

I have elsewhere furnished evidence for the assumption that there is a strong need for better counselling instruments in our country (TROST 1980: S. 74-83; TROST 1982: S. 173). The introduction of aptitude tests can fill an obvious gap in the present counselling programmes.

The basic idea underlying the development of several "study-field oriented tests" starts from the general experience that, in order to get a good estimate of a person's cognitive aptitudes for several given areas of study, it is not satisfactory to use one and the same test battery yielding a profile of the individual's test scores which can be compared to the respective "ideal" profiles. Usually, this procedure merely leads to rather general recommendations. One cannot do justice to the special requirements of a large variety of subject areas by using a single test. Furthermore, the majority of the school leavers do have certain preferences with regard to their future studies and wish to get information about their aptitudes for these study areas. The other extreme, to construct a special test for every single subject offered at university, though, would be totally uneconomical.

Instead we tried to construct a small number of separate tests for those university subject areas which can be treated as units because in careful analyses of their typical requirements they proved to be similar. Yet for those pupils who do not have any preferences we intend to offer a more general "broad-band" test (BLUM/HENSGEN/TROST 1982).

The following report will be focused on one of the seven tests: the "Aptitude Test for Studies in Modern Languages" (Beratungstest für philologische Studiengänge).

3. Development of the "Aptitude Test for Studies in Modern Languages"

The need for counselling based on information about the individual prerequisites to meet the requirements of particular university courses is especially strong among those who are considering studying modern languages. Frequently these areas are chosen from indecision or ignorance about alternatives, or as "second-rate solutions" when applicants have not been admitted to their first choices.

Consequently, the ratio of students who drop out from university or change their subject area after a few semesters is exceptionally high in modern languages; the average duration of study till graduation is longer than it is in most of the other sub-

subject areas; and students enrolled in modern language courses are among those who are least satisfied with the subject area they have chosen (BICKEL 1980: Chapter 3).

Therefore it seemed especially important to provide an instrument for counselling those who are thinking of studying modern languages. Our intention was to assess aptitudes which are needed for studies in all modern languages, German included.

After a review of the literature and a thorough analysis of the study requirements in English, French and German, a first form of the test was developed in close cooperation with members of modern language faculties. This form, which was laid out in two versions, was tried out on a sample of 244 "Abiturienten" in Northrhine-Westfalia who were interested in studying modern languages at university. On the basis of the results of the test analyses, the test was revised in terms of structural modification and selection of the best items out of the two preliminary versions.

The second test form emerging from this revision was applied to a sample of 872 Abiturienten in four states of the Federal Republic of Germany. The results of the test analyses are reported elsewhere in detail (BICKEL 1980: Chapter 7). Let me only mention the outcome of factor analyses.

Two factors showed up. One was called "competence in usage of verbal material", the second was named "aptitude to detect systems in verbal material". Only 39 per cent of the total variance in the test scores were explained by the two factors, i. e. a considerable amount of specific variance - beside the error variance - was left with each subtest.

Seven months after the 1977 tryout of the test, the same test was administered again to a sample of 238 Abiturienten in two states who had taken it before. The stability of the test or its retest reliability proved to be high: the correlation of the total test scores earned in the first and in the second session was $r_{tt} = .93$ (BICKEL 1980: Chapter 8).

The third experimental study also shed light on the construct validity of the test. Specific hypotheses as to the relation of the test to a variety of other diagnostic measures (Gymnasium marks, intelligence tests, a general scholastic aptitude test, a test of critical thinking, a test of verbal memory, a knowledge test, a biographical inventory and interest questionnaire) were tested. Table 1 shows some of the findings. In general, the results of the analyses corresponded with most of the hypotheses on which the test construction had been based, so that there is some evidence for the construct validity of the test.

The outcome of the second and third empirical tryouts of the test also yielded indications of how the test could further be improved. So, when the continuation of the research project was granted in 1981, the first thing to do was a comprehensive revision of the test:

- The least reliable, least "construct-valid" and most time-consuming subtest, "Paragraph Organisation", was removed.
- In all of the six remaining aptitude subtests the number of items was increased from 18 to 24 items.
- The foreign language part of the test was revised most thoroughly: The number of items was enlarged from 24 to 32; most items of the French version and all items of the English version in the foreign language part were replaced. With the kind assistance of Professor Helmut Bonheim and Dr. Barbara Kreifelts of the University of Cologne we could introduce new items which had already been tried out and had attained favourable item statistics.

The test version emerging from this revision was checked once again by a group of experts (members of faculties of modern languages).

4. Present Structure of the "Test for Studies in Modern Languages"

Table 2 shows the structure of the "Test for Studies in Modern Languages" in its present form. The test consists of seven subtests; five of them are presented in German, one contains verbal material in artificial languages, and one measures foreign language competence (the testees can choose between items in English or French). The administration of the total number of 176 items takes about 3 1/2 hours. Sample items of each of the subtests are presented in tables 3-5 in the appendix (cf. BICKEL 1980: S. 109-121).

5. Application of the Test in a Counselling Programme

In one of the "Länder" of the Federal Republic of Germany, Niedersachsen, an "integrated counselling system" has been established during the last four years. All institutions in the "model region" who are involved in counselling activities (schools, universities, regional counselling centres for pupils, federal agencies for vocational counselling) cooperate in joint counselling programmes, in joint training courses for counsellors etc. One of the focuses of the project is counselling at the transition into higher education. In this context, special "project courses" are offered in grades 12 and 13 which are to help pupils in making their decisions when leaving school, and guided visits to universities of the region are arranged. So far, standardised diagnostic instruments have not been used in the counselling programme.

Last winter, our aptitude tests including the test for modern languages were introduced into this system in terms of a pilot programme. Over a time span of two years, the tests will be offered to approximately 6,000 pupils in the following way: All pupils of the "model region" in the 12th and 13th school year are informed about the test programme via the schools. They are asked to indicate the test(s) they are interested in on a form. The schools collect the forms and send them back to our institute. According to the pupils' preferences, we invite them individually to the respective test term(s) in their own school or in a neighbouring school. Along with the invitation

they receive a questionnaire inquiring into their study interests and vocational preferences, their major subjects and achievements in school and their extracurricular interests and activities. The pupils are required to fill the inventory in at home and return it on the test term.

Different tests are administered on different days in a given local region so that the testees are not forced to choose between two or more tests. All tests take place in the afternoon when there are usually no lessons.

As soon as the scoring of all tests is finished, we mail the individual test results, printed out verbally and in graphs, to the private addresses of the testees. The subtests and what they assess are explained. In addition, the testees are given the average results of certain reference groups and their standing relative to these group statistics. (After the completion of the validity studies we can hopefully provide an estimate of the candidates' future success in a given field of university studies). A sample of the feedback print-outs for the aptitude test for modern languages is given in tables 6-8 in the appendix. For the interest questionnaire the individual results are displayed in a similar way. Along with this material goes the suggestion to discuss the test and questionnaire results with a counsellor (the schools and counsellors in the "model region" receive an information booklet on the test programme, its background and its purposes).

This autumn and winter we will be able to apply a sequential model of testing: In the first stage a "broad-band" test is offered especially to those who do not yet have clear notions of their preferred university subjects or who have not decided whether to go to university at all. This general test renders a screening of the candidates' structure of abilities. On the basis of these results, the testees can be referred to the more specialised tests mentioned above, if they want to receive more precise information on their academic aptitudes. The "study-field oriented tests" will then be offered in the second stage.

The application of the tests in the present project aims at two targets: a practical one and a scientific one.

With respect to the practical aim we regard the project as a feasibility study which will answer the following questions:

- To what extent do the pupils respond to the invitation to take one or several of the tests?
- What relevance do the test results have for the pupils' career planning?
- Can the organisational and technical problems be solved: cooperation with the schools in announcing and administering the tests, central rapid test scoring in our institute, printing out and mailing the test results as well as reference values and interpretations to the testees etc.?

Observation during test administration, a follow-up questionnaire sent to all participants along with the test results, and interviews with teachers and counsellors give us information on the questions listed above.

The main scientific purposes are to carry out test analyses with the revised test forms and to start a longitudinal validity study on a sufficiently large sample of persons who take the test in a real counselling setting and not in a merely experimental situation. All testees will be followed up three to four years after the test term: By a questionnaire we will assess the career choices they have made after graduating from school, the academic or nonacademic success, the satisfaction with the chosen career etc.

6. First Results of the Application of the Test in the Counselling Programme

Last winter, all secondary school students in the "model region" were invited to take one or several of six "study-field oriented tests" and/or a general scholastic aptitude test. 3,700 students responded, and 2,600 students actually participated in the testing programme.

308 of those participating took the Aptitude Test for Studies in Modern Languages; in the optional part of this test measuring proficiency in a foreign language, 249 testees chose the English version, and only 59 pupils chose the French version.

As to the "feasibility" of the programme, we can state that the technical and administrative problems related to the implementation of the testing programme, the scoring procedures and the printing out of the results along with some interpretations could be solved. Furthermore, it has become obvious that the students' interest in the tests is high and that the testees widely accept the present layout, organisation and administration of the tests as well as the present way of giving feedback on the test results. What relevance the test results have for the pupils' career planning cannot yet be determined.

Out of the various results of the analyses with the Aptitude Test for Studies in Modern Languages, I selected three issues on which I will briefly report:

- general test characteristics,
- relation between performance in the test and school marks,
- difference in the average test results of certain groups of testees.

Table 9 shows some of the results of the conventional test and subtest analyses. The average difficulty index of the total test (without the foreign language part) is $P = .52$ indicating that the test is fairly adequate to the aptitude level of the examinees. The foreign language subtests are somewhat easier ($P = .62$ and $.58$). The split-half reliability coefficient for the total aptitude test ($r_{tt} = .91$) is satisfactory. The reliabil-

ity coefficients for the subtests range from .56 to .87; their median is .74. (The reliability values as determined by the KUDER-RICHARDSON formula 20 are similar).

The coefficients for the correlations between test results and school marks are given in table 10. In general, we find the highest correlations with performance in the subjects German and English but also in Mathematics - mainly for those subtests which require the testee to detect structures and systems in verbal material. The coefficients for the relation between the particular subtests and the average mark in school are fairly low; the highest values stand for the correlation of the test for competence in French and of the subtest "Artificial Languages" with this criterion. The correlation between the total aptitude test score and the average mark is $r = .36$; it means that the test only measures to a small degree aptitudes which are already attested in the average mark in school.

Table 11 allows for some comparisons of the average test performance of certain groups of testees. Male pupils did slightly better in the total test than female. This is perhaps due to a much stricter self-selection among male test participants (76 per cent of all those who took the Aptitude Test for Studies in Modern Languages were girls). Girls, however, obtained somewhat better average results in the foreign language part of the test.

The mean test scores of those who took the test in the 12th school year are a little lower than the mean results of those who were tested in their 13th school year. It may be concluded that training in school (or mere aging?) over a time span of one year has a slight influence on test performance.

Finally we examined the average test score of pupils who, before taking the test, had indicated different subject areas as their first choices for future studies at university. Table 11 shows that, on an average, the highest test results were obtained by those who named German as their favourite area of study; the test performance of the rest of the groups - favouring English, French or "other languages" - did not vary significantly.

So much for the outcome of the first application of the test in the framework of a counselling programme.

Much more important than these findings, of course, are the results of the longitudinal validity study which is going on and which will prove the test's predictive power - or its lack of predictive power.

"The trouble with longitudinal studies is", as William OWENS of Georgia University put it last year on the Edinburgh Congress of Applied Psychology, "that subject and investigators age at the same rate!" So we still need some aging and a lot of patience until we will dispose of the hard data which justify the introduction on the new test battery in Germany on a general basis as an instrument for counselling.

References:

- BICKEL, H. 1980: Diagnose der Studierfähigkeit für die Philologien.
München: Minerva Publikation.
- BLUM, F., HENSGEN, A. & TROST, G. 1982: Informationsbroschüre zur Erprobung der Beratungstests für Oberstufenschüler und Abiturienten.
Bonn: Institut für Test- und Begabungsforschung.
- TROST, G. 1980: Tests zur Beratung von Abiturienten bei der Wahl des Studienfaches. In: GÖBEL, U. (Hg.): Tests - Wegweiser für die Bildungs- und Berufsplanung, S. 71-123. Köln: Deutscher Instituts-Verlag.
- TROST, G. 1982: The Role of Aptitude Tests in Counseling. In: VALISNO, M.D. (Ed.): Broadening Opportunities Through Assessment in Education, S. 171-193.
Princeton: International Association for Educational Assessment.
- TROST, G. & BICKEL, H. 1981: An Aptitude Test for Studies in Modern Languages as Part of a Test Battery for Counselling Applicants for Higher Education. In: KLEIN-BRALEY, C. & STEVENSON, D.K. (Eds.): Practice and Problems in Language Testing. Volume 1: Proceedings of the First International Language Testing Symposium of the Interuniversitäre Sprachtestgruppe held at the Bundessprachenamt, Hürth, 29-31 July 1979, S. 124-144. Frankfurt/Bern: Peter D. Lang.
- TROST, G., BICKEL, H., BLUM, F. & CHRISTIAN, H. 1979: Studienfeldbezogene Tests zur Abiturientenberatung. Bonn: Bericht Nr. 6 des Instituts für Test- und Begabungsforschung.

A p p e n d i x :

Tables 1-11

Diagnostic Instruments	Correlation Coefficient r	Number of Subjects
general scholastic aptitude test (ATS) -- verbal section --	.69	104
intelligence test (LPS)	.54	86
average mark in school (12th school year)	-.42	860
Watson-Glaser Critical Thinking Appraisal (2 subtests)	.36 / .32	79
test of verbal memory (L-G-T)	.34	86
knowledge test (D-W-T) (scale of literature)	.31	86
biographical inventory (MBI) (performance scales)	-.02 to .19 Median: .18	79
interest test (D-I-T)	-.12 to .12 Median: .02	86

Table 1 : Correlations between the total score in the "Aptitude Test for Studies in Modern Languages" and Scores in other diagnostic instruments

Name of Subtest	Cognitive Functions Required	Number of Items	Time
Sentence Completion	recognition of functionally and formally adequate expressions; differentiated use of language	24	20'
Artificial Languages	recognition of principles in patterns of unknown languages	24	50'
Word Composition	discrimination of different formal and functional principles of the composition of words	24	20'
Sound and Rhythm	identification of similarities or differences in the sound and rhythm of words and phrases	24	20'
B r e a k			
Analysis of Mistakes	identification of mistakes in grammar, style or logic	24	20'
Reading Comprehension	comprehension, analysis and interpretation of literary texts	24	35'
Competence in a Foreign Language: English or French	reading comprehension and proficiency in (alternatively) English or French	32	30'
Total Test		176	195'

Table 2 : Structure of the "Aptitude Test for Studies in Modern Languages"

Satzergänzung

Bei den folgenden Aufgaben sind jeweils in einem Satz ein oder mehrere Wörter durch einen Strich ersetzt. Unter jedem Satz sind fünf mit (A) bis (E) bezeichnete Ergänzungsvorschläge angegeben. Suchen Sie denjenigen Ergänzungsvorschlag heraus, der sich am besten in den sprachlichen und gedanklichen Zusammenhang des Satzes einfügt.

Wer gleich zu Beginn von Verhandlungen alle seine Positionen als "unverrückbar" kennzeichnet und Maximalforderungen stellt,
_____ Möglichkeit einer flexiblen Verhandlungsführung.

- (A) verschenkt keine
- (B) strebt nach der
- (C) übersieht die
- (D) begibt sich der
- (E) gewinnt die

Sprachensysteme

In den folgenden Aufgaben werden Ihnen jeweils einige Ausdrücke einer fremden Sprache und deren deutsche Übersetzung vorgestellt. Zur Bearbeitung der daran anschließenden Aufgaben sollen Sie versuchen, die Bedeutung der Wörter sowie einige grammatikalische Regeln der fremden Sprache herauszufinden.

Die folgende Aufgabe bezieht sich auf die Ausdrücke:

kataro	=	er hofft immer
kama	=	er denkt nach
ukama	=	ich denke nach
ukata	=	wir hoffen
usaka	=	wir schlafen

"Ich schlafe immer" heißt in der fremden Sprache:

- (A) usakata
- (B) ukataro
- (C) sakarotu
- (D) usakaro
- (E) kamasaka

Table 3:

Sample items: "Sentence Completion" and "Artificial Languages"

Wortbildung

- (A) Gesundheit
- (B) Zufriedenheit
- (C) Kindheit
- (D) Frechheit
- (E) Sturheit

(Abweichung im
Ableitungsprinzip)

- (A) Geschirrspüler
- (B) Zitronenfalter
- (C) Seifenspender
- (D) Landvermesser
- (E) Sargtischler

(Abweichung in der inhalt-
lichen Beziehung der Wort-
glieder)

Klang und Rhythmus

- (A) kurzgefaßt
- (B) weitgereist
- (C) verärgert
- (D) tiefbetrübt
- (E) durchgemacht

(Abweichung im
rhythmischen Prinzip)

- (A) Mit lachender List dem Lug gelauscht.
- (B) Tote Tauben täuschen den Tapferen.
- (C) Wie Walter weiteren Wegweisern wich.
- (D) Der Unhold unkte unter der Ulme.
- (E) Die müde Süße einer Rübenblüte.

(Abweichung von der Alliteration)

Eugen, der Juwelendieb,

- (A) Nahm, was er nur kriegen konnte,
Ohne daß ihm ein Befremden
- (B) Über sich zurücke blieb.
Eines Tages aber stahl
- (C) Er (man wirds nicht glauben wollen)
- (D) Einen ganzen wundervollen
- (E) Grade nicht benutzten Saal.

(Abweichung vom Reimschema)

Fehleranalyse

Bei den folgenden Aufgaben ist jeweils ein Satz in vier mit (A) bis (D) bezeichnete Abschnitte gegliedert. Einer der vier Abschnitte kann einen oder mehrere Fehler sprachlicher, stilistischer oder gedanklicher Art enthalten. Dieser Abschnitt ist herauszufinden. Enthält der ganze Satz keinerlei Fehler oder Mängel, so ist die Antwort (E) zu markieren.

- (A) Verschärft wurden die Bestimmungen,
- (B) die die Befugnis zum Tragen von Waffen betreffen;
- (C) sie besagt unter anderem, daß nur der Waffen tra-
gen darf,
- (D) der besondere hoheitliche Aufgaben erfüllt.
- (E) (kein Fehler)

Table 4:

Sample items: "Word Composition", "Sound and Rhythm" and "Analysis of Mistakes"

Textverständnis

Lesen Sie die folgenden Texte bitte aufmerksam durch. Es schließen sich jeweils einige Fragen an, die sich auf die Texte beziehen. Suchen Sie unter den mit (A) bis (E) bezeichneten Antwortvorschlägen die beste Antwort auf jede der gestellten Fragen.

Tatsächlich ist der Fortschritt der Naturwissenschaften wie ein alter Wüstenpfad übersät mit den ausgebleichten Gerippen fallengelassener Theorien, die einstmals ewiges Leben zu besitzen schienen. Der Fortschritt der Kunst bringt ebenso qualvolle Neuurteilungen etablierter Werte. Der Dichter, der Maler, der Wissenschaftler: jeder überlagert das Universum mit seiner mehr oder weniger kurzlebigen Vision, jeder konstruiert sein eigenes parteiisches Realitätsmodell, indem er alle jene Aspekte der Erfahrung hervorhebt, die ihm bedeutsam erscheinen, und die anderen wegläßt. Die gleiche Technik der Abstraktion findet sich in der Zeichnung des Cartoonisten, dem Diagramm des Physikers, dem Atlas des Geographen, dem stilisierten Landschaftsbild oder Porträt. Die Technik ist die gleiche; nur die Medien sind verschieden und die Kriterien dafür, was relevant ist.

Mit welchem Mittel sucht der Autor hauptsächlich Überzeugungskraft zu gewinnen?

- (A) Wiederholung von Argumenten
- (B) Anhäufung von Fakten
- (C) Pointen
- (D) Bildhafter Ausdruck
- (E) Verwendung von Ironie

Welche Intention verbindet der Autor mit seinem Text?

- (A) Es geht ihm um die Erhaltung etablierter Werte.
- (B) Es geht ihm um das Lob des Fortschritts.
- (C) Es geht ihm um die Betonung der Nutzlosigkeit von Kunst und Wissenschaft.
- (D) Es geht ihm um die Vereinheitlichung des Weltbildes.
- (E) Es geht ihm um eine erkenntnistheoretische Grund-satzfrage.

Ergebnisse im Beratungstest für das Studienfeld Philologien

1. Übersicht über die einzelnen Untertests des Beratungstests für das Studienfeld Philologien und über Ihre Ergebnisse

Der Untertest **SATZERGÄNZUNG** prüft syntaktische, semantische und pragmatische Aspekte der Sprachkompetenz. Er erfordert das Erkennen der logischen und der sprachlichen Struktur eines Satzes, die Verarbeitung der inhaltlichen Informationen und der sprachlichen Signale sowie die Beurteilung der Angemessenheit eines ergänzenden Ausdrucks. Im Vordergrund stehen dabei Sprachverständnis und Sprachbeherrschung. Ihr Ergebnis in diesem Untertest ist, bezogen auf die Gesamtgruppe der Teilnehmer an diesem Test, durchschnittlich.

Im Untertest **SPRACHENSYSTEME** sind aus vorgegebenen Ausdrücken einer fiktiven Sprache Wortbedeutungen und grammatikalische Regeln abzuleiten. Dies erfordert die Fähigkeit zum Erkennen von Gesetzmäßigkeiten und Strukturen in Sprachmustern.

In dieser Aufgabengruppe ist Ihr Testwert überdurchschnittlich.

Beim Untertest **WORTBILDUNG** wird die Fähigkeit zum Erkennen von Besonderheiten in kontextfreiem Sprachmaterial erfaßt. Hierbei liegt das Gewicht auf den Fähigkeiten zur sprachlichen Abstraktion und zur Differenzierung.

Das von Ihnen in diesem Untertest erzielte Resultat ist überdurchschnittlich.

Der Untertest **KLANG & RHYTHMUS** prüft die klanglich-rhythmische Diskriminationsfähigkeit anhand von Wörtern, Wortpaaren, Sätzen und Gedichtzeilen. Die Aufgaben erfordern die Lösung von der Sprache als Bedeutungsträger, die Beachtung von phonetischen Unterschieden und das Erkennen von metrischen Abweichungen.

In diesem Untertest ist Ihr Punktwert durchschnittlich.

Im **GESAMTTEST** (ohne den Untertest "Fremdsprachenkenntnis") haben Sie von 144 Aufgaben 92 richtig gelöst. 78 Prozent aller Teilnehmer an diesem Test haben ein niedrigeres oder ein gleich gutes Ergebnis erzielt.

Im Vergleich mit den Teilnehmern mit dem Fachwunsch **ANGLISTIK** ist dieses Ergebnis weit überdurchschnittlich. Setzt man Ihr Resultat in Beziehung zu den Ergebnissen der Interessenten für ein Studium der **GERMANISTIK**, so haben Sie durchschnittlich abgesehen.

Verglichen mit der Teilnehmergruppe, die das Studium der **ROMANISTIK** in Erwägung zieht, ist Ihr Punktwert überdurchschnittlich.

Im Untertest **FEHLERANALYSE** sind logische, grammatikalische und stilistische Fehler zu entdecken. Breit gefächert wird damit die sichere Sprachbeherrschung und die Fähigkeit zur sprachkritischen Beurteilung von Texten geprüft.

Ihr Ergebnis in diesem Untertest ist überdurchschnittlich.

Der Untertest **TEXTVERSTÄNDNIS** erfaßt in thematischer Streuung die Fähigkeit zum Umgang mit Texten. Über das Sprachverständnis in bezug auf den Inhalt des Gelesenen hinaus prüft er insbesondere das Erkennen von Intentionen des Autors, von rhetorischen Figuren, von Situationseinflüssen und von Stilvarianten sowie die Interpretationsfähigkeit.

In dieser Aufgabengruppe ist Ihr Ergebnis weit überdurchschnittlich.

Der Untertest **FREMDSPRACHENKENNTNIS** mißt fremdsprachliche Kompetenz in, wahlweise, Englisch oder Französisch. Neben Kenntnissen in bezug auf Vokabular, Stil, Grammatik und Idiomatik setzt er ein hohes Maß an Leseverständnis voraus.

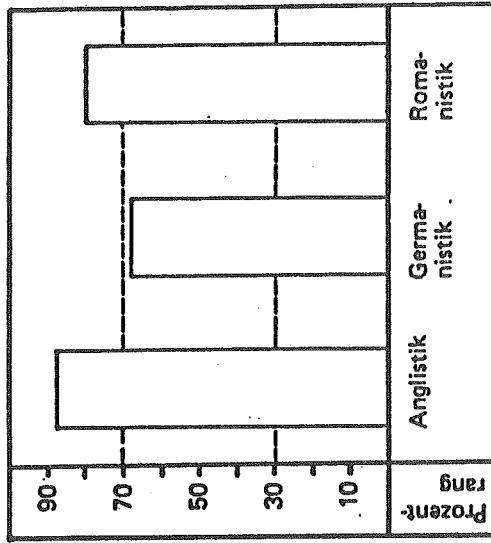
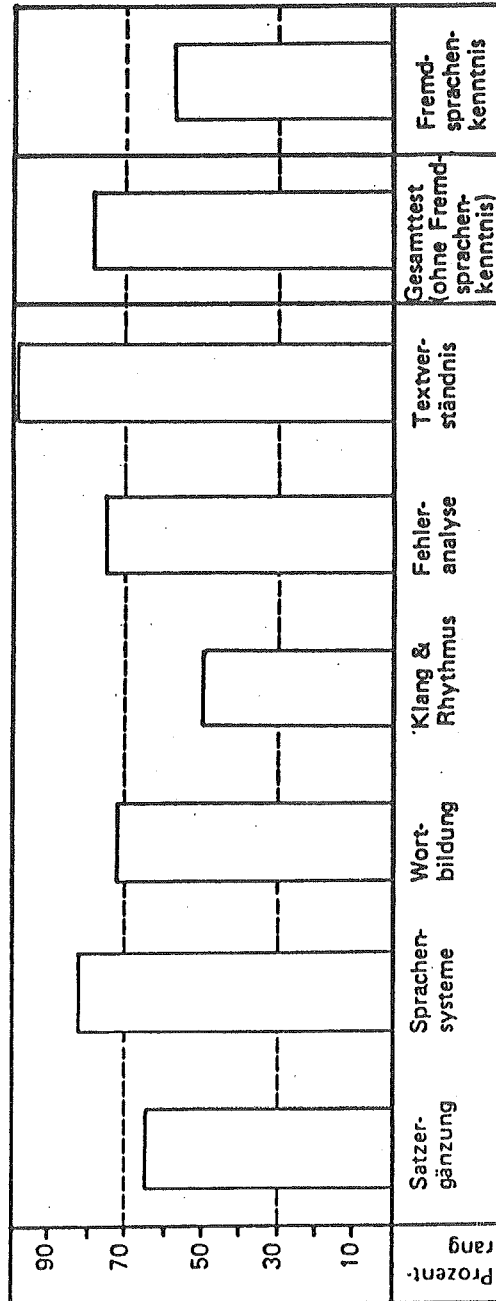
Das von Ihnen in diesem Untertest erzielte Resultat ist, bezogen auf die Gruppe aller Teilnehmer, die denselben Fremdsprachentest bearbeitet haben wie Sie, durchschnittlich.

Table 6:

Verbal description of the "Aptitude Test for Studies in Modern Languages" and of the individual results of a test participant

Ergebnisse im Beratungstest für das Studienfeld Philologien

2. Ihr Abschneiden in den einzelnen Untertests und im Gesamttest, verglichen mit den Ergebnissen aller Teilnehmer an diesem Test (Prozentrang-Angaben*)
3. Ihr Abschneiden im Gesamttest (ohne den Untertest "Fremdsprachenkenntnis"), verglichen mit den Ergebnissen der Teilnehmer mit den Fachwünschen Anglistik, Germanistik und Romanistik (Prozentrang-Angaben*)



* Ein Prozentrang von 70 in einem Test bedeutet beispielsweise, daß 70 Prozent der Testteilnehmer in der jeweiligen Bezugsgruppe eine niedrigere oder allenfalls eine gleich hohe Punktzahl erzielt haben. 30 Prozent der Testteilnehmer haben demnach einen höheren Testwert erreicht. Bei einem Prozentrang zwischen 30 und 70 wird das Testergebnis als durchschnittlich eingestuft. Die Prozentrang-Angabe beim Untertest "Fremdsprachenkenntnis" ist auf die Gruppe aller Teilnehmer bezogen, die denselben Fremdsprachentest bearbeitet haben wie Sie.

Table 7: Information on the percentile rank of a participant as compared to certain reference groups

Ergebnisse im Beratungstest für das Studienfeld Philologien

4. Ihre Testergebnisse im einzelnen und einige Vergleichswerte

Bezeichnung des Untertests	Anzahl der Aufgaben	Mittelwert der Teilnehmer			maximal erreichte Punktzahl*	Ihre Punktzahl*	Ihr Prozentrang** bezüglich der Teilnehmer				
		insgesamt	mit dem Fachwunsch Anglistik Germanistik	Romanistik			insgesamt	Anglistik	Germanistik	Romanistik	
Satzergänzung	24	12,3	11,4	13,0	12,1	24	13	65	76	55	65
Sprachensysteme	24	11,9	11,2	12,6	11,7	22	14	81	87	74	85
Wortbildung	24	12,6	11,7	13,1	12,7	24	14	73	80	67	73
Klang & Rhythmus	24	12,6	11,4	13,7	12,6	24	11	49	61	37	51
Fehleranalyse	24	15,0	14,1	15,8	14,8	23	17	76	83	67	78
Textverständnis	24	15,4	15,0	15,7	15,5	24	23	98	98	97	98
Gesamttest (ohne Fremdsprachenkenntnis)	144	79,8	74,8	83,9	79,4	119	92	78	87	69	81
Fremdsprachenkenntnis	32	14,8	16,9	12,7	-	30	17	62	50	77	-

* Die "Punktzahl" gibt die Anzahl der jeweils richtig gelösten Aufgaben an. Jede richtig gelöste Aufgabe wird also mit einem Punkt bewertet. Falsch gelöste und nicht bearbeitete Aufgaben werden nicht bewertet. Die "maximal erreichte Punktzahl" ist die höchste Punktzahl, die von einem Testteilnehmer in dem betreffenden Untertest oder im Gesamttest erzielt worden ist.

**Siehe Erläuterung auf Blatt 2.

Table 8: Printout of the participant's individual test scores and some reference data

Subtest	Number of Items	Mean	Standard Deviation	Average Difficulty	Split-half Reliability (SP.-BROWN)
Sentence Completion	24	12,3	3,4	.51	.56
Artificial Languages	24	13,1	4,8	.54	.87
Word Composition	24	13,0	3,8	.54	.74
Sound and Rhythm	24	13,5	4,5	.56	.83
Analysis of Mistakes	24	11,7	3,4	.49	.59
Reading Comprehension	24	12,3	3,7	.51	.64
Total Test (without For. Language Part)	144	75,5	15,8	.52	.91
Competence in:					
English (N = 249)	32	19,9	5,4	.62	.79
French (N = 59)	32	18,6	5,2	.58	.74

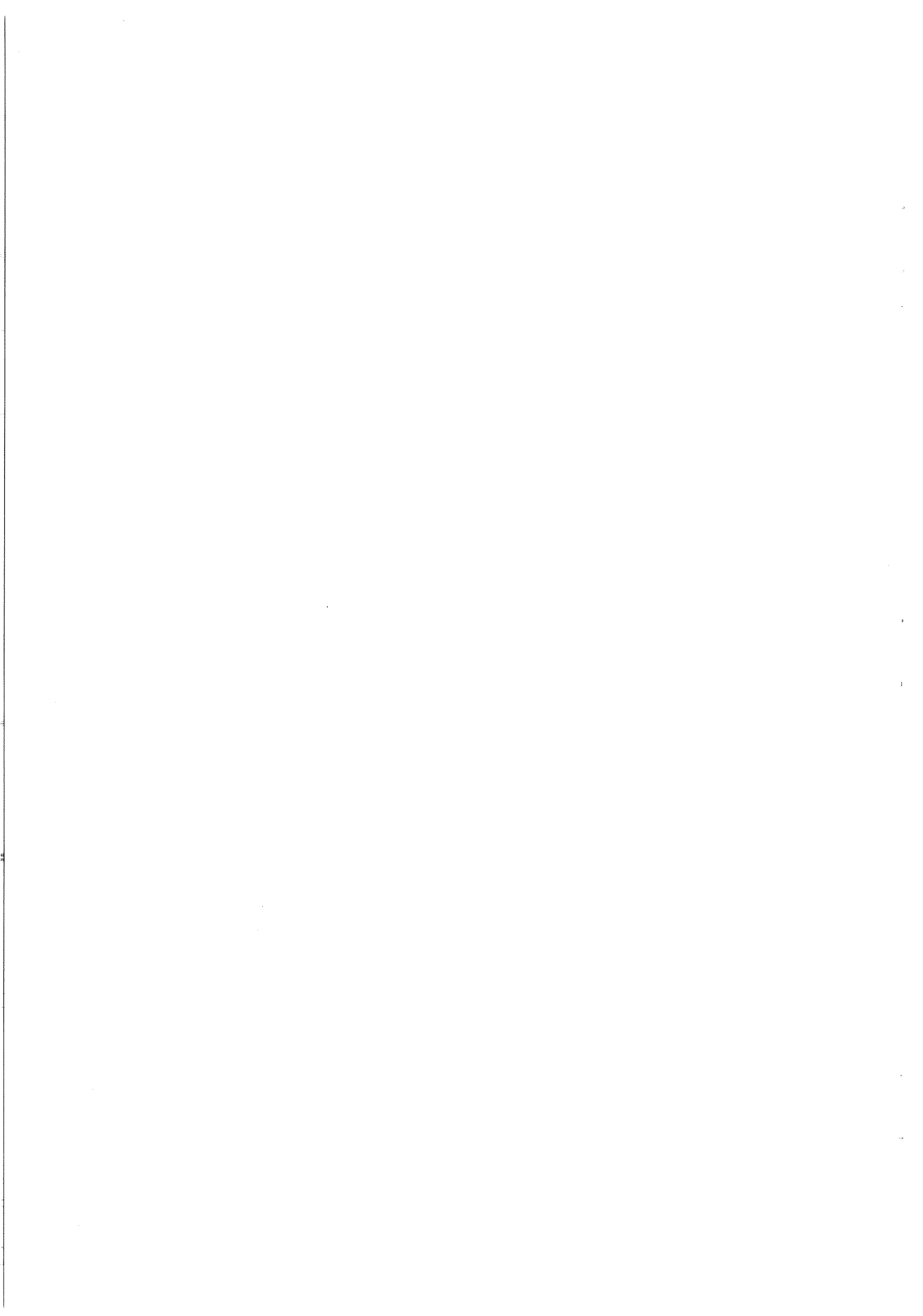
Table 9: Results of test analyses on the 1983 sample (N = 308)

Subject Subtest	German	English	French	Latin	Math.	Biology	Chem.	Physics	History	Fine Arts	Music	Average Mark
Sentence Completion	.28	.23	.02	.04	.18	.18	.15	.10	.26	-.04	.04	.26
Artificial Languages	.22	.19	.17	.19	.33	.18	.23	.19	.08	.12	.13	.33
Word Composition	.17	.19	.24	.18	.25	.11	.12	.15	.09	.08	.11	.26
Sound and Rhythm	.20	.23	.09	.14	.24	.02	.10	.09	.06	.07	.17	.22
Analysis of Mistakes	.19	.16	.07	.07	.16	.04	.07	.06	.21	.07	.00	.19
Reading Comprehension	.22	.14	.04	.06	.03	.05	.03	.06	.19	.02	-.02	.16
Total Test (without For. Language Part)	.32	.29	.17	.18	.30	.15	.18	.16	.21	.08	.12	.36
Competence in: English (N = 163-248)	.08	.53	.27	.16	.11	.02	.00	.03	.05	.11	.03	.18
French (N = 39-58)	.43	.34	.43	.45	.18	.19	.40	.25	.40	.02	.02	.40

Table 10: Correlations between test scores and school marks (N = 202-306)

Subtest	Male (N=73)	Female (N=235)	12th School Year (N=209)	13th School Year (N=99)	English (N=86)	German (N=122)	French (N=59)	Other Studies (N=41)
Sentence Completion	13,2 (3,0)	12,0 (3,4)	12,0 (3,3)	12,8 (3,3)	12,1 (3,3)	12,8 (3,3)	11,6 (3,3)	11,8 (3,7)
Artificial Languages	13,2 (4,6)	13,0 (4,9)	12,9 (4,8)	13,3 (4,9)	12,6 (5,1)	13,1 (4,7)	13,4 (4,4)	13,4 (5,1)
Word Composition	12,8 (3,9)	13,1 (3,8)	13,1 (3,7)	12,8 (4,0)	13,0 (3,7)	12,6 (3,7)	13,5 (4,2)	13,6 (3,7)
Sound and Rhythm	13,6 (5,3)	13,5 (4,3)	13,4 (4,7)	13,8 (4,2)	13,8 (4,4)	13,7 (4,4)	12,9 (4,3)	13,4 (5,5)
Analysis of Mistakes	12,4 (4,0)	11,5 (3,1)	11,5 (3,2)	12,0 (3,5)	11,8 (3,2)	12,0 (3,6)	11,3 (3,1)	10,9 (2,9)
Reading Comprehension	12,9 (3,7)	12,2 (3,7)	12,1 (3,7)	12,9 (3,6)	11,7 (3,6)	13,0 (3,6)	12,1 (3,8)	12,0 (3,8)
Total Test (without For. Language Part)	78,1 (18,2)	75,2 (15,0)	75,0 (16,0)	77,7 (15,4)	75,1 (15,6)	77,2 (16,4)	74,8 (15,2)	75,0 (15,6)
Competence in English or French	18,8 (6,2)	19,8 (5,3)	19,1 (5,6)	20,5 (5,2)	22,1 (4,1)	18,1 (5,6)	19,6 (5,0)	18,6 (6,4)

Table 11: Means and (in parentheses) standard deviations of test scores of different groups of testees



NATIONAL REPORT - BELGIUM

The report revolved around a presentation of a videotape produced by the English Department at the Royal Military Academy Brussels in collaboration with the Closed-Circuit Television Centre.

Designed as a working document illustrating this year's Conference theme, that is to say Language Aptitude and Proficiency Testing, the videotape meant to show how CCTV can assist language teachers in trying to assess the oral proficiency of their students. Using CCTV this way also shows how language instructors can train officer-cadets to deliver a speech, make a briefing or respond on social occasions.

About ten video cassettes (i. E. 10 hours film) have actually been necessary to permit a significant selection covering a wide range of proficiency levels from 1st year cadets to 4th year ones.

A questionnaire had been made available for the interview which was meant to make cadets come up with pat answers. They were free to consider the questions as starting points for an offhand discussion when their proficiency allowed creativity.

Conversation had to be as natural as possible even though the topics had been selected by the teacher in the hope of putting some life into the discussion and of obtaining participation from the audience.

Reading was also an unprepared performance that was to be followed by a summary without having another go. This exercise was meant to show their reading and reading ability as well as their speaking ability.

The speeches had been prepared beforehand by the cadets, who had been given ample time to choose and prepare their subject.

* * *

Having seen this film now, it may be said - and the videotape is conclusive on this - that depending on the first language of the learner, the results in oral proficiency differ in a dramatic way, especially when the same amount of time is devoted to language training irrespective of the mother tongue.

It is obvious that some language communities need to spend more hours if they want to achieve some or the same proficiency; and lack of motivation would not do as an explanation for poor achievement.

Continuing seeking new ways to tackle this problem is an essential concern we cannot depart from.

Appropriate aptitude tests may lead to new techniques so as to more effectively train the less gifted cadets to the required standards.



NATIONAL REPORT - CANADA

1. The Directorate of Language Training at National Defence Headquarters (NDHQ) in Ottawa has seen a level of activity in the past year almost unparalleled since its foundation. This is not likely to cease for another two years at least, when in the Fall of 1985, we should have in place a completely new curriculum oriented towards communicative competence, using military functions, military situations and audio-visual support exploiting military scenarios; as well as a brand new set of instruments to test students, again emphasizing communicative competence.
2. This high level of activity originates from a study of the Official Languages (French & English) Command & Control, Training and Infrastructure, done by BGen McLaws of NDHQ. The study had been commissioned to find the reasons for the rather low rate of progress made towards bilingual objectives (numbers of bilingual personnel) which had been set some 5 years earlier.
3. This activity in the Directorate then, is that which relates to the implementation of the recommendations of this study. There were 42 in all and since the Fall of 1982, a couple more have been added.
4. During the next few minutes, I propose to review for you some of these recommendations and how we propose to tackle their implementation.
5. Generally speaking, the study found that our slow rate of progress could be attributed to factors grouped into two main categories:
 - a. those factors which affect the success rate in training of personnel to a maintainable level of bilingualism:
 - (1) motivation and selection of personnel to undertake second language training, and, subsequent employment, in order to re-inforce and improve expensively acquired skills;
 - (2) curriculum content, teaching staff, facilities and testing; and
 - (3) administration and Command and Control.
 - b. in the second category, we find those factors which are deemed to have affected the overall progress. They are:
 - (1) Policy - guidelines to co-ordinate achievement of objectives were unclear; and
 - (2) Resources were insufficient and badly used.

6. The decision to implement these recommendations led to the creation of three working groups, each tasked with specific items. The first of these, under the chairmanship of a full Colonel of the Personnel division at NDHQ works on those items that relate directly to the Personnel who become our students on the French course. They are:
 - a. to greatly enhance the motivation of Canadian Forces (CF) Anglophone members to become bilingual; (combination of exhortation from the Chief of the Defence Staff (CDS) to possible career enhancement)
 - b. insert the Continuous French Course (CFC) in the list of manning priorities for the Canadian Forces with regards to student course loading, in order to visibly enhance its importance; (no longer will managers, Branch Heads etc. be able to use the excuse of too much work for not wanting to let their personnel go off on courses since all who go on course get the person-year replacement)
 - c. develop and adopt a dynamic manpower model to enhance the effectiveness and efficiency of second language training and bilingual manning functions;
 - d. select a better average quality of candidate for the French course than had been the case heretofore;
 - e. to meet quantitative CFC quotas overall, to the maximum extent possible; and
 - f. immediately or soon after graduation, post those who attain a "Functional" level or better exclusively to positions requiring at least Functional bilingual capability or to positions in FLUs; and with respect to those who come close to, but fall short of the Functional level, post them to Functional French (FF) or French Language Units (FLU) positions to the extent that they can be expected, in the judgement of the Canadian Forces Training Systems (CFTS), to be capable of handling the language requirements and eventually reaching Functional level.
7. The second working group, under the chairmanship of the senior Policy Officer (Director General Official Languages (DGOL) Branch) is looking into those recommendations of the study which deal with overall CF policy. This group is responsible to:
 - a. examine and revise CF objectives for LT in terms of target numbers and target dates in relation to resource availability and affordability;
 - b. conduct critical and stringent review throughout the CF to determine which positions need to be identified as bilingual and what the language level for each should be; and
 - c. conduct a review of all language training programs to establish priority for resource allocation.

8. The third working group, under my chairmanship must implement a total of 27 recommendations, the majority of which emanate from the McLaws study, and a few others, such as the revamping of all language testing in the CF, which have been added since Dec 82. These are grouped into six different categories:

a. First Category - Command and Control

- (1) the responsibility for overall co-ordination of Language Training Plans and Programs has been passed from DGOL to the Director General Recruiting, Organization and Training (DGRET); and
- (2) DLT is moved organizationally to DGRET.

b. Second Category - Language Training Policy clarified

- (1) currently, Language Training (LT) policy is contained in 4 different instructions. All of these will be amalgamated into one instruction to include the basic principles of the McLaws study recommendations; and
- (2) the terms "bilingual" and "functional" are being re-defined in "operational" terms - i. e., in terms of what the member is required to know to perform CF functions. (Big job - responsibility of Percy Rangongo, whom you all know).

c. Third Category - Improve Language Training

- (1) formation of a Standards Cell under the direction of Canadian Forces Training Systems Headquarters (CFTSHQ) to monitor progress in LT and the standards of teaching of second languages;
- (2) provision of teacher supervision at Phase I centres; and
- (3) application of the very successful CF Individual Training System to language training: (Done by CFTS but under the watchful eye of Pedagogical Advisor - Percy Rangongo)

d. Fourth Category - Improve success rate

- (1) no student with low aptitude will be placed on language training courses unless he is known to be well motivated (MLAT "D" students); (One of CFTS' demands, easier said than done given the possible career implications of LT and the human rights factor)
- (2) implementation of a widespread Modern Language Aptitude Tests (MLAT) and "A" tests competence (understanding) to facilitate the selection process for Career Managers;

- (3) attempt to consolidate Phase I centres in order to gain flexibility by enabling better grouping of students (at present near to impossible with only two classes per centre);
- (4) increase homework (with a purpose) and increase classroom time (by eliminating unnecessary breaks); and
- (5) permit early graduation when students are ready (to further enable grouping of students and thereby allow the teachers to give more time to the less talented).

e. New Departures being tried out

- (1) development of a new curriculum oriented towards communicative competence, using military functions and CF situations and audio/visual support exploiting military scenarios;
- (2) improved training of teaching staff; and
- (3) in March of this year, 24 Anglophone recruits started studying French in St. Jean. This pilot project if successful, could radically change the traditional method of teaching French to Anglophones when they are well into their career. The project is being closely monitored by my office.

9. In 1982, the CF commissioned a group of PhDs in measurement, educational psychology, administration and linguistics from the University of Ottawa to conduct an in-depth study of our testing system: While the study team found that there did not exist at present, better, more accurate instruments to test linguistic competence (French and English), our battery of tests were found to be behind the state of the art.
10. The implementation of the recommendations of this study has already started and will take until Summer of 85 to complete. NDHQ/DLT is going to write testing specifications and these will serve to construct new English and French proficiency batteries of tests aimed at testing "Communicative Competence" as opposed to testing pure knowledge of language structure and syntax.
11. In conclusion, the aim of this tremendous amount of work currently being done in NDHQ is to increase the amount of bilingual anglophones in the CF. We are banking on the measures to improve our teaching which in turn should improve our rate of success.
12. I should mention here that no increase in resources is forthcoming and that training in the two official languages must continue.
13. During the year, we have had 934 students in 4 different programs studying French and 1211 students in 2 programs studying English. These numbers do not include the Official Languages programs of the three Military Colleges nor the Summer programs

of the various officer candidates, nor do they include our foreign language students and programs. For this, I'd ask LtCol Jock Thomson, the Commandant of the Canadian Forces Language School in Ottawa to conclude this presentation by providing you with an update on Foreign Languages programs.

(By LtCol Thomson)

14. In essence, there have been no major changes, since BILC 82, in Canadian Forces Foreign language training programs. However, the purpose of this report is to give a short overview of our main programs, and to highlight some new aspects within them.

Formal Foreign Language Training Programs

15. All formal, full-time foreign language training for Canadian Forces is conducted by Canadian Forces Languages School (CFLS) Ottawa, Foreign Language Department. Currently, the school has the capability to teach twelve languages, namely:
- a. Arabic; (commencing Fall 83)
 - b. Chinese (Mandarin);
 - c. Czechoslovakian;
 - d. German;
 - e. Italian;
 - f. Japanese;
 - g. Norwegian;
 - h. Polish;
 - j. Russian;
 - k. Serbo-Croatian; and
 - m. Spanish.
16. The courses are normally between 205-230 training days in length, and the average attendance is about 35 students, ranging from Colonel to Private.
17. Also this year the Directorate of Language Training (DLT) is trying out a different approach to language training by having one student to take a combination of university and commercial language courses, mainly to see how well such training compares to our own programs, and if it can be used as a viable alternative when our own training resources are fully committed.
18. In addition to the above, each year approximately 7-10 students, who for service reasons are unable to take CFLS formal courses, are given tailored "crash" courses through

commercial languages schools. Although this method is not desirable, it has the side benefit of keeping DLT up-to-date on commercial school's capabilities and sources of teachers.

Self-Study Packages

19. Over the past year there has been marked increase in requests for foreign language self-study packages, particularly from personnel being posted to Germany, especially those who will have some official dealings with German officials, but not sufficient to justify the formal CFLS course. To meet this requirement a package similar to the US "HEADSTART" will be developed, using the new audio/video production facilities which were recently inaugurated at CFLS, St-Jean, and which many BILC delegates had an opportunity to visit during last year's BILC conference in Canada. In this regard, Canada would be interested to learn about any other countries experiences with self-study packages and if possible obtain some samples for evaluation.

English As a Foreign Language

20. Although Canada has a large English language training program, as a result of the Government's Official Languages initiatives, teaching of English as a foreign language to foreign students is a new experience. In 1982, through a bilateral agreement with Italy, ten Air Force officers were given 51 training days of English instruction, prior to pilot training. A further 10 students will be trained this year. The object was to ensure that all candidates had the minimum English language proficiency profile for pilot training, i. e., 3322. Since most candidates had a good proficiency on arrival in Canada, a portion of the course could be dedicated to aircrew technical and environmental language training.

21. This concludes Canada's National Report.

RAPPORT NATIONAL - FRANCE

Ce rapport qui traite des modalités d'application de la politique des langues étrangères, adoptée en 1981 par le Chef d'Etat-Major de l'Armée de Terre et dont l'essentiel a fait l'objet du rapport national 1982, se décompose en quatre parties:

- I - Présentation des postes réclamant une qualification linguistique.
- II - Définition des degrés de langue et des emplois correspondants.
- III - Description des épreuves des examens.
- IV - Répartition des responsabilités dans l'enseignement des langues.

* * *

I. POSTES RECLAMANT UNE QUALIFICATION LINGUISTIQUE.

L'importance des langues se traduit, en temps de paix comme en temps de guerre, par l'existence de très nombreux postes d'officiers, sous-officiers et militaires du rang pour lesquels la qualification linguistique, qu'elle soit prioritaire ou non, est nécessaire.

A. Certains postes correspondent à une aptitude langue écrite.

Il s'agit de fonctions dans lesquelles l'officier ou le sous-officier traitant doit être capable de saisir le sens général d'un document étranger, se rapportant à son domaine de travail, pour pouvoir l'exploiter sans, pour autant, en effectuer une traduction rigoureuse.

Il s'agit également des postes de traducteurs, ouverts, suivant les documents à traduire, à toutes les catégories de personnels ou réservés à certaines d'entre elles.

Il existe enfin des postes d'analystes de presse étrangère spécialisée (militaire ou non) qui exigent des connaissances particulières et sont donc réservés aux officiers et aux sous-officiers supérieurs. Ces postes ne sont pas l'apanage de la branche renseignement, même s'ils y sont particulièrement nombreux, et se situent souvent à un niveau élevé.

B. Certains postes correspondent à une aptitude langue parlée.

Ils intéressent des personnels de toutes catégories, capables de comprendre une langue étrangère parlée, et, en règle générale, de s'exprimer oralement dans celle-ci avec aisance.

C'est le cas des personnels employés pour accompagner un visiteur étranger, pour participer à des réunions internationales de travail ou d'état-major, pour interroger des étrangers, pour assister à des manoeuvres ou à des exercices d'armées alliées, pour bénéficier d'échanges au pair.

C'est aussi le cas des personnels appartenant à des unités susceptibles d'être engagées dans des actions extérieures ou stationnées en pays étranger (Europe ou outre-mer).

Il s'agit également des interprètes en consécutive travaillant au profit de personnalités militaires de rang élevé.

C. Pour d'autres emplois, la maîtrise de la langue écrite et parlée est nécessaire:

dans le domaine des relations internationales: attachés des forces armées (A.F.A.), attachés des forces terrestres (A.F.T.), officiers de liaison instructeurs (O.L.I.), détachements de liaison, missions militaires françaises de liaison (M.M.F.L.), stagiaires dans des écoles ou des états-majors à l'étranger, cadres des forces internationales (ONUST), secrétaires de postes A.F.A., membres des missions de l'assistance militaire technique;

dans le domaine de l'enseignement: professeurs, instructeurs ou répétiteurs de langues vivantes au centre de langues et études étrangères militaires (C.L.E.E.M.), dans les écoles et les garnisons.

Les connaissances nécessaires pour tenir ces différents postes n'étant pas identiques, les niveaux de qualification sont matérialisés par un système de certificats militaires à 3 degrés.

II. DEFINITION DES DEGRES DE LANGUE ET DES EMPLOIS CORRESPONDANTS.

A. Certificat Militaire de Langue écrite.

1. C.M.L.E. 1.

Le premier degré sanctionne l'aptitude à traduire un texte à caractère général incluant le vocabulaire militaire de la presse non spécialisée. Le candidat devra donc posséder une connaissance de la langue courante permettant la traduction d'un texte extrait d'un journal, d'une revue civile ou d'une revue militaire non spécialisée traitant d'un sujet d'actualité ou d'ordre général, ne comportant que les termes militaires appartenant au vocabulaire courant.

La possession du C.M.L.E. 1. permet à tous les personnels de remplir une fonction de traducteur de documents non spécialisés et, aux cadres, d'exploiter des documents de travail.

2. C.M.L.E. 2.

Le 2ème degré sanctionne l'aptitude à traduire tout texte militaire se rapportant à l'armée de terre du pays considéré et traitant de problèmes généraux, de l'organisation, de la tactique, des personnels, des matériels de cette armée. Le candidat devra donc posséder une bonne connaissance du vocabulaire militaire de l'armée de terre.

Le C.M.L.E. 2. correspond à des emplois de traducteurs spécialisés officiers ou sous-officiers, en unité ou en état-major.

3. C.M.L.E. 3.

Le 3ème degré sanctionne l'aptitude à traduire et analyser tout texte militaire se rapportant à l'armée de terre ainsi que tout texte d'ordre général se rapportant aux autres armées, aux services communs ou aux formations paramilitaires. Le candidat devra donc avoir acquis des connaissances détaillées sur l'armée de terre, connaître le travail en état-major aux échelons élevés, posséder des connaissances générales sur l'organisation des autres armées, tant françaises que du pays considéré, et posséder une bonne connaissance du vocabulaire militaire interarmées.

Le C.M.L.E. 3. correspond à des postes d'officier traitant d'organismes spécialisés ou d'états-majors de rang élevé, chargé de traduire ou d'analyser des documents importants; il permet également la participation à des jurys d'examens écrits ou à la préparation des C.M.L.E.

B. Certificat Militaire de Langue Parlée.

1. C.M.L.P. 1.

Le premier degré sanctionne l'aptitude à participer à une conversation orientée sur un sujet d'ordre général comportant le vocabulaire militaire courant. Le candidat devra donc être capable de s'exprimer couramment dans la langue considérée tant sur un sujet de la vie quotidienne ou d'actualité que sur un sujet militaire à caractère général.

Le C.M.L.P. 1. permet de faciliter les contacts avec des personnels étrangers, notamment lors de visites, de réunions interalliées, de manoeuvres ou d'exercices, et d'être affecté préférentiellement à des unités stationnées ou susceptibles d'être employées hors de France.

2. C.M.L.P. 2.

Le 2ème degré sanctionne l'aptitude à s'exprimer couramment sur un sujet se rapportant à l'armée de terre soit d'ordre général soit d'ordre particulier jusqu'au niveau de la division. Le candidat devra donc posséder la maîtrise de la langue étrangère parlée, avoir acquis des connaissances générales sur l'organisation, l'emploi, les matériels et connaître la technique d'état-major jusqu'au niveau division inclus.

Le C.M.L.P. 2. permet de tenir des postes à l'étranger, de mener des réunions de travail avec les alliées, d'être affecté dans des postes d'officiers de liaison ou de remplir des fonctions d'instructeur ou de répétiteur de langue.

3. C.M.L.P. 3.

Le 3ème degré sanctionne l'aptitude à s'exprimer couramment sur tout sujet d'ordre général interarmées ou d'ordre particulier se rapportant à l'armée de terre, à assurer une mission d'accompagnement d'une personnalité militaire de rang élevé et à interpréter en consécutive une conversation non spécialisée. Le candidat devra donc posséder une bonne culture générale et des connaissances militaires développées tant sur les forces armées françaises que sur celles du pays considéré.

Le C.M.L.P. 3. permet l'accès:

- à des postes de professeur ou d'officier chargé des problèmes de langues dans des organismes spécialisés;
- à des emplois d'officier traitant dans des états-majors de haut niveau ou en relations suivies avec les étrangers;

- aux fonctions d'interprètes en consécutive;
- aux missions d'accompagnement de personnalités militaires de rang élevé;
- aux fonctions de membre de jurys d'examens de langues.

4. C.M.A.I.

Il sanctionne l'aptitude à interpréter en consécutive une discussion portant sur tout sujet militaire se rapportant à l'armée de terre ou d'ordre général interarmées.

Le C.M.A.I. correspond aux emplois d'interprétariat en consécutive pour les missions de haut niveau, soit pour des interlocuteurs de rang très élevé, soit sur des sujets hautement spécialisés.

III. EPREUVES DES CERTIFICATS MILITAIRES DE LANGUE

A. Nature, durée et coefficient des épreuves des C.M.L.E.

Epreuve.	Nature.	Durée.	Coefficient.
Examen du premier degré.			
Traduction de la presse quotidienne non spécialisée.			
Version.	Traduction d'un extrait de la presse non spécialisée, comportant du vocabulaire courant militaire.	1 h 30	1 (1)
Thème.	Traduction d'un extrait de la presse non spécialisée, comportant du vocabulaire courant militaire.	1 h 30	1 (2)
Examen du deuxième degré.			
Traduction de textes militaires généraux ou spécialisés armée de terre.			
Version n° 1.	Traduction d'un texte militaire de caractère général.	1 h 30	1
Version n° 2.	Traduction d'un texte militaire à caractère technique armée de terre.	1 h 30	1
Version n° 3.	Traduction d'un texte militaire à caractère tactique.	1 h 30	1
Thème.	Traduction ou rédaction d'une lettre à caractère officiel ou protocolaire, adressée à une autorité militaire.	1 h 30	1
Examen du troisième degré.			
Toutes traductions.			
Version n° 1.	Traduction d'un texte militaire à caractère interarmées ou se rapportant aux autres armées ou aux services communs.	2 heures	1
Version n° 2.	Traduction d'un texte à caractère tactique, scientifique ou technique de l'armée de terre.	1 h 30	1
Thème.	Traduction d'un texte d'état-major (ordre, décision, compte rendu...).	1 h 30	1
Analyse.	Analyse en langue française d'un texte militaire en langue étrangère.	2 h 30	2
(1) Note minimum pour être admis : 12/20. (2) Note minimum pour être admis : 8/20.			

.../...

B. Nature, durée et coefficient des épreuves des C.M.L.P.
(y compris C.M.A.I.)

Epreuve.	Nature.	Observations.	Coefficient.
Examen du premier degré.			
Lecture et conversation militaire.			
A	Lecture et traduction à vue d'un article de la presse étrangère de caractère général pouvant comporter du vocabulaire militaire courant.	Texte d'une centaine de mots (1) : 15 mn + 5 mn de préparation.	1
B	Conversation courante dans la langue à base de questions et de réponses sur un sujet d'ordre général comportant un aspect militaire.	25 minutes.	1
Examen du deuxième degré.			
Lecture, écoute et conversation militaire (armée de terre).			
A	Lecture et traduction à vue d'un texte militaire étranger (presse étrangère militaire spécialisée, règlement, ordre...).	Texte d'une centaine de mots (1) : 15 mn + 5 mn de préparation.	1
B	Analyse en français d'un texte militaire étranger enregistré.	Audition : 5 mn. Analyse : 15 mn.	1
C	Exposé sur carte, en langue étrangère, d'une situation tactique du niveau régiment ou division du pays considéré, à partir d'un thème en langue française remis au candidat, suivi de questions sur les unités mises en œuvre.	20 mn + 15 mn de préparation.	2
Examen du troisième degré. (2)			
Interprétation et connaissance du pays.			
A	Interprétation en consécutive d'une conversation sur un sujet militaire spécialisé relatif à l'armée de terre du pays considéré.	20 mn.	1,2
B	Interprétation en consécutive d'une conversation sur un sujet militaire non spécialisé à un niveau interarmées.	20 mn.	1,2
C	Interrogation dans la langue sur le pays considéré : connaissances générales d'histoire et de géographie, situation politique, économique et sociale actuelle.	10 mn + 5 mn de préparation.	0,8
D	Interrogation dans la langue sur l'armée de terre du pays considéré.	15 mn + 5 mn de préparation.	1
E	Interrogation dans la langue sur les autres armées, les forces paramilitaires ou les services communs (connaissances générales).	10 mn + 5 mn de préparation.	0,8
<p>(1) Ou équivalent en caractères.</p> <p>(2) La réussite à l'examen de langue parlée du 3^o degré, avec une moyenne égale ou supérieure à 14/20 permet d'obtenir le C.M.A.I.</p>			

RENOUVELLEMENT DU C.M.A.I.

Epreuve.	Nature.	Durée.	Coefficient.
A	Epreuve A du C.M.L.P. 3.	0 h 45	1
B	Epreuve B du C.M.L.P. 3.	0 h 45	1

C. Conditions d'inscription au 2ème degré écrit ou parlé.

Nul ne peut faire acte de candidature à un deuxième degré écrit ou parlé s'il n'a pas obtenu au préalable le C.M.L.E. 1. et le C.M.L.P. 1., ce qui constitue la garantie d'une bonne connaissance de base de la langue.

IV. RESPONSABILITES DANS L'ENSEIGNEMENT DES LANGUES.

A. Au niveau des Régions.

Les commandants de Région Militaire et de territoire, le Commandant en Chef des forces françaises en Allemagne ont la charge d'organiser les cours de préparation aux examens. Ils sont aidés dans cette tâche par la présence d'officiers "langues" désignés à tous les niveaux de la Région, de la Division militaire territoriale, de la garnison.

B. Rôle du C.L.E.E.M.

Le Centre de Langues et Etudes Etrangères Militaires est chargé:

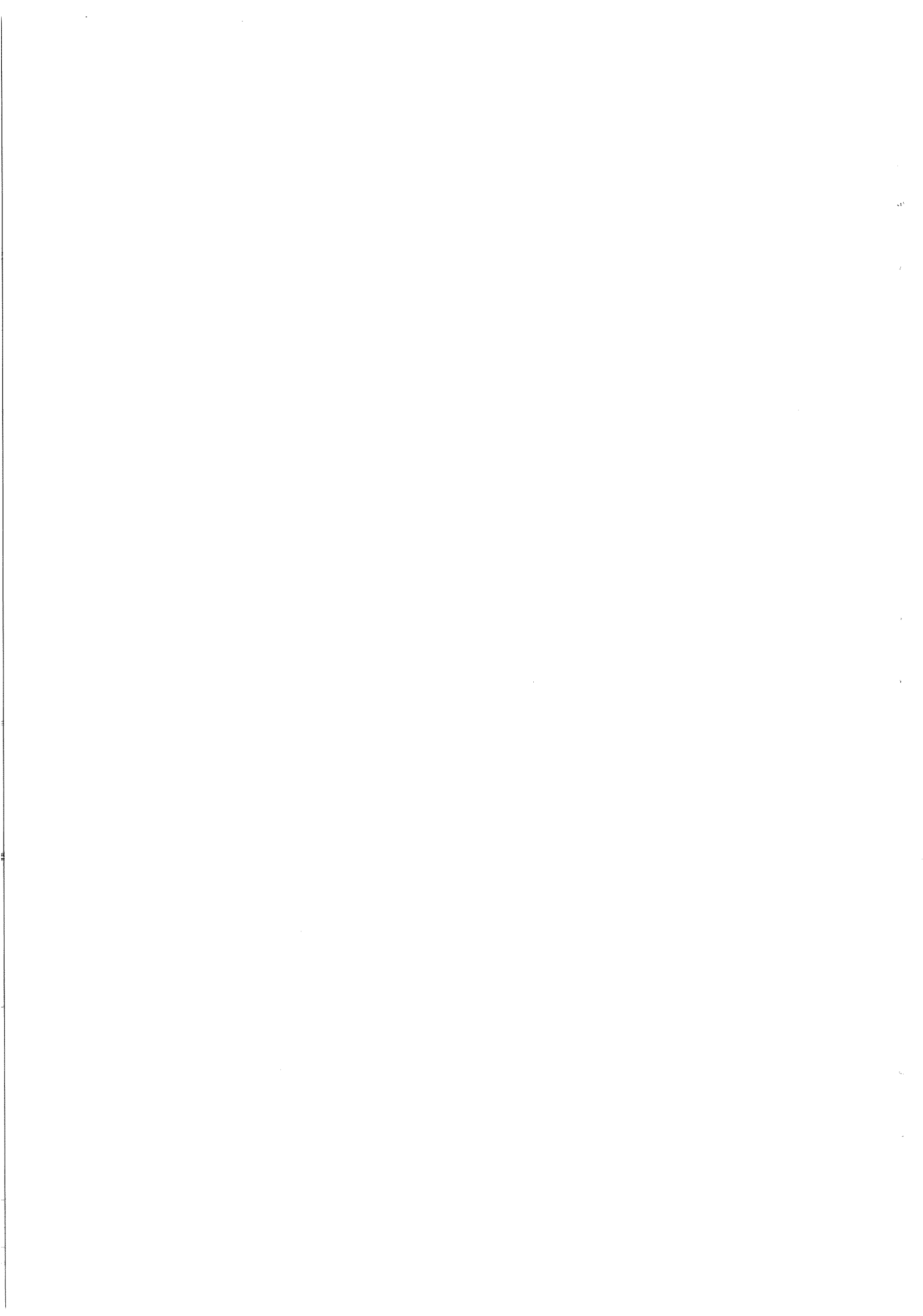
- de définir des programmes d'enseignement adaptés aux différents niveaux de connaissances à acquérir,
- d'élaborer et de diffuser toute la documentation correspondante, qu'elle soit écrite ou enregistrée sur cassettes ou vidéo-cassettes,
- de prendre à sa charge la préparation des candidats dans certaines langues soit par des cours par correspondance, soit par des cours et stages particuliers effectués dans ses locaux,
- de choisir les sujets d'examen et pour une grande partie des candidats, de corriger leurs épreuves écrites ou de leur faire passer les épreuves orales.

Les certificats militaires de langues étrangères restent le moyen privilégié de sanctionner les connaissances linguistiques des personnels de l'armée de terre, qu'il s'agisse de l'ensemble des cadres ou des spécialistes de haut niveau.

Leur préparation et le choix de la langue restent pour la majorité des personnels militaires une affaire d'appréciation personnelle, à base de volontariat même si le Commandement peut dans certains cas imposer la langue et le niveau à acquérir. De même la préparation de ces certificats reste le plus souvent à base de travail individuel, en dehors des heures de service.

Trois mille à trois mille cinq cents candidats se présentent cependant chaque année à ces certificats dans une trentaine de langues (plus de 50 % en Anglais, de 20 à 25 % en Allemand).

La meilleure adaptation des certificats militaires actuels aux emplois à tenir devrait permettre de disposer désormais, à partir de cette ressource importante en candidats, de personnels qualifiés en nombre suffisant.



NATIONAL REPORT - GERMANY

In our last year's National Report we gave an account of the way in which the Bundessprachenamt, in particular the Language Training Division, was to be reorganized. This reorganization became effective on 1 February last year and we are glad to say that the optimism with which we then viewed the operation was not unwarranted.

The SLP testing system which was also mentioned in last year's report and which kept us very busy for a considerable amount of time is now firmly established.

Our main concern now is a project which we launched last September to adapt our teaching methodology to the training system, both at this Institute and at the external language training establishments for which the Bundessprachenamt is technically responsible. In this report I would like to discuss the major problem which arose during this adaptation process - the problem of defining the nature and the role of the new generation of course material to be developed.

In the past we had generally used uniform course material in our classes but we now had to find an alternative to the uniform textbooks and peripheral teaching and learning aids to which we had become accustomed.

You may ask: What's wrong with uniform course material? Well, in our context, there are a lot of things that are wrong with it. I will touch upon only a few.

The teaching objectives that can be defined for many of our training establishments are very diverse; in the Bundessprachenamt they may, depending on the language being taught, differ drastically from one class to the next. The thematic interests of our learners are objectively, and subjectively, anything but homogeneous.

Uniform course materials must, by their very nature, be thematically largely indifferent and, therefore, of little interest to our type of learner. Apart from this point, which is sufficiently important by itself, uniform course material can have negative psychological consequences. It can easily assume a dominating role in the classroom because the teacher may be tempted to use it as a prescriptive manual that dictates what and how he teaches. The teacher's own knowledge of the subject matter and the world, and his own creativity can hardly come into play. Under these circumstances he's not the master; at worst, he may be reduced to the role of subservient mediator between the textbook and the learners. The learner may be negatively affected because he is tempted to believe that the prescribed uniform textbook contains everything (all the linguistic and communicative devices) that he needs to learn to use the target language. We can easily find students who have been led to believe that a particular French textbook contains the French language. They probably also believe that the linguistic and thematic content of that textbook represents what

they are supposed to master for the final examination. Such students may never become aware that in learning other languages we are not learning examples of hypothetical language use but that we should be learning and can only learn the language through examples. In other words: the semantic content of a given uniform textbook, consisting of pre-formulated semantic givens that cannot stimulate the learner's own imagination and creativity are frequently learned for the sole purpose of being reproduced in examinations. Unfortunately this is what many examinees expect and are satisfied with, because they are spared the demanding task of verbalizing spontaneously in communicatively relevant situations.

For the reasons I have cited, and for reasons of a more technical nature that I cannot go into this report, we decided to discontinue the development of uniform course material for all possible applications and have replaced it by a system in which the individual teacher (or a small team of teachers of a teaching unit) with clearly defined communicative objectives, are given a greater chance to make full use of their own professional expertise and creativity.

To accomplish this significant change in our approach we have developed what we call a SYLLABUS: a voluminous compendium of guidelines for the construction of communicative skill-oriented curricula complete with detailed, minutely elaborated examples of instructional units.

Guided by this syllabus, a team of teachers from each teaching unit or establishment develop a curriculum for each level and for each language skill. In this task they are aided by experts from the Bundessprachenamt, and in certain cases, when the department is not adequately staffed to do the job, the Bundessprachenamt may decide to develop the curriculum itself.

Each CURRICULUM developed in this way will be a fairly rigid framework that determines and describes the format, the content and possibly the chronological order in which the instructional units are to be used with any group of learners. The coordinator for the various skill-oriented curricula is the respective head of the teaching unit.

The INSTRUCTIONAL units or lessons are developed within this framework by the individual teacher or a small group of teachers. Experts from the Bundessprachenamt are always available to conduct relevant training seminars for teachers or to give advice on the construction of a particular instructional unit.

By giving the teacher, and eventually also the learner, more freedom in the learning process we are following our conviction that only purposeful language use can make another language come to life in the classroom. In our opinion, the use of artificial language violates an essential predisposition of man's memory, i. e. its constant search for meaningful information. In this connection: authenticity of the learning material is one thing; authenticity in the learner's handling of the material is another. We are convinced that by giving the teachers and, in the sense

that I have explained this before, also the learners, individual responsibility for instructional units of a course, the goal of serious use of language in the classroom and authenticity in the learner's handling of course material can be attained.

Further details and relevant documentation on this project will be made available as it is completed.



NATIONAL REPORT - UNITED KINGDOM

The Head of the UK Delegation did not present a formal report but instead made a short statement on recent LT (language training) activity in the UK.

He referred to the fact the organization of LT in the UK was being reviewed and that the formation of a possible tri-service Defence School of Languages was the subject of a feasibility study. The total LT task was also being reviewed and lists of approved annotated posts for LT drawn up. National training times for SLP levels in most common languages were also being agreed.

Over all, LT activity in the UK was on the increase. The Army, for example, despite the effects of planning for major organizational changes, continued to meet in full its LT commitment. German LT for personnel serving in BAOR continues to increase. There are plans to include some majors among a pool of officers and soldiers who receive formal training in German up to SLP 2200 (Army Colloquial Test). This will bring the annual total up to 750 trained personnel of whom a significant number go on to higher levels. Several thousand other soldiers have received basic training in German of one kind or another; amongst these were personnel serving in the Falkland Islands who would shortly move to Germany. The increase in LT as a whole is well evidenced by the fact that there has been a 50 % increase in the past year in BAOR in the number of loans of self study packages.

Within the RAF LT was also dynamic and there had been a complete reequipment of learning packages for self study and a number of organizational measures were in train which would be implemented in the next year or so.

Finally, the UK delegate from BRNC Dartmouth reported on work that was being done on the production of an Anglo/German Naval technical glossary. He paid tribute to the help and cooperation received from the Bundessprachenamt.



NATIONAL REPORT - USA

DEFENSE LANGUAGE INSTITUTE; FOREIGN LANGUAGE CENTER; (DLIFLC)
MONTEREY, CALIFORNIA

1. COMPREHENSIVE TEST DEVELOPMENT PACKAGE

The Defense Language Proficiency Test III (DLPT III) will replace DLPT Is and IIs, which now exist in over 40 languages. The earlier tests have serious deficiencies, as almost half of them are more than 20 years old and the language in some is obsolete. Most tests have only one form so there is a chance of compromise. Also, the current DLPTs test only listening and reading and only test the lower part of the language proficiency scale (Levels 0-3). Thus, the particular norm-reference scoring system devised for the DLPTs has led to inflated scores over the years. The new DLPT IIIs will test the full proficiency range (Levels 0-5) in three skills - listening, reading and speaking. Writing tests can be added later if users determine a requirement. The DLPT III will serve as an end-of-course test at DLIFLC. However, like the present DLPTs, it will test the language at large and not any particular course of instruction. The DLPT III will be a two-tiered system. Only those who score level 3 on a lower range test will be eligible to take the upper range test in that particular skill. The DLPT III design and initial development effort was accomplished as a joint endeavor between DLIFLC, NSA, CIA, FSI, US Office of Education and other agencies from the private sector. A nationwide conference was held at the Presidio of Monterey in November 1981 for this purpose. DLIFLC also initiated a dialog among the Inter-agency Language Roundtable members to standardize the government language skill level descriptions. These cooperative efforts culminated in a newly developed set of functional skill level descriptions. The primary aim of the new skill level standards is to equate level attainment across language across various government agencies. They will also serve as the basis for the terminal level objectives (TLOs) for all DLIFLC courses. A prototype DLPT III is being developed in Russian. The lower range test was completed at the end of October 1982, and the upper range test will be complete in October 1983. Work was begun in eight other languages in FY83. Four other languages will follow in FY84. To eliminate the problem of inflated scores, DLIFLC developed a plan to recalibrate the current DLPTs using the new skill level standards. Six high enrollment languages have new scoring systems at this time. All other languages tested at DLI will be recalibrated by the end of FY83. This brings DLPT results into line with results of tests used by other government agencies. A new graduation policy has been established concerning all classes beginning instruction in calendar year 1983. The minimum graduation standards will be achieving level 1 in each of the two skills (Listening and Reading) on the DLPT II, or level 1 in two of the three skills (Listening, Reading and Speaking) on the DLPT III. Any student not achieving these standards will be awarded a certificate of attendance in lieu of a graduation diploma. The minimum standards may be raised at a later date. To accompany the new courses being developed by DLIFLC, Proficiency Advancement Test I and II will be developed. They will be administered after

the first and second phases of the course. These tests will measure course objectives and also measure the student's ability to handle unfamiliar material. These tests contain a diagnostic feature that will provide information about students' weaknesses to help place them in remedial programs, both in-house and in the field.

2. INPUT AND LOAD

The student input picture at DLIFLC shows a consistently rising trend from FY79 through the mid 80's. The increase in foreign language input is attributed to greater awareness of the significance of foreign language proficiency both in the intelligence services as well as in other activities targeted for overseas assignments. Since FY79 student loads have increasingly taxed barracks and classroom facilities at the Presidio to the point where, in 1980, we had to garner the BOQ's for use as enlisted billets and move one Directorate of over 200 people off-post to provide additional classrooms. By FY81, even this additional capacity was exceeded. To ease the burden at the Presidio, we opened a branch facility at Lackland AFB in November 1980, teaching Russian to approximately 400 Air Force and Army students and opened a branch at a former Public Health Service Hospital at the Presidio of San Francisco on 1 October 1982, teaching Korean, Spanish and German to approximately 400 single Army Enlisted students. By the end of FY83, we will have experienced a 70 % increase in student load over that of 1979, a rise of over 14 % per year. The present classroom shortages at the Presidio is being alleviated by using staggered scheduling and, by contracting for the use of a local, recently vacated elementary school. These solutions are expected to relieve the immediate demand for classroom space until 1985 when the first of two new General Instructional facilities should be ready for use. The medium/long range solution to our over-loading at the Presidio is the new military construction now in progress.

3. RUSSIAN CULTURAL HOUSE

DLIFLC has established a Russian Cultural House (RCH) for total immersion language training for a one year trial basis. Current Terminal Learning Objectives for Army Cryptologic Training Service students at DLIFLC are beyond the levels of linguistic competence attainable within the 47 weeks allocated for the Russian Basic Course. In order to meet National Security Agency requirements for highly specialized and competent linguists, DLIFLC implemented the LEFOX Program, a 27 week program immediately following the basic course and designed to improve CTS Linguists' competence attained before arrival on station for initial utilization tours. In another effort to improve the quality of linguists it was decided that intensive language instruction would best be conducted in an atmosphere of total immersion in the linguistic and cultural milieu of the target country. Our total immersion concept requires the student to use the target language exclusively in daily living. The House fa-

cility, is designed for six to eight students to reside and train in the facility, rotating students every eight weeks. The facility will be operated by DLIFLC Russian instructors controlling the out of class environment, supervising student preparation of Russian style meals, and aiding students with their spoken Russian. Students would continue to attend scheduled classes. Russian television programs and radio broadcasts will be provided by use of videocassette recordings. The Russian Cultural House is scheduled to start operation on 15 June 83.

4. EDUCATIONAL TECHNOLOGY DIVISION

The Educational Technology Division is DLI's most recent effort to bring technology to DLI. The purpose of the division is to identify relevant opportunities for application of technology to language training; convert these opportunities into scheduled, funded, formal projects; and plan and manage milestones until a project becomes self-sustaining. The Educational Technology Division has already created two foreign language videodiscs and has hosted a major instructional technology conference. Current projects include a joint NSA/DLIFLC/FORSCOM research agreement; production of 13 videodiscs to support the existing German Gateway course; evaluation and training in computer literacy for DLI faculty and staff.

5. DEFENSE FOREIGN LANGUAGE PROFESSIONAL DEVELOPMENT PROGRAM

The Defense Foreign Language Professional Development Program was developed by DLIFLC, and approved by the General Officer Steering Committee during August 1981. The purposes of the program is to further develop the linguistic skills of the service NCOs and Petty Officers (POs) that are assigned to DLIFLC as language trainers and to enable DLIFLC to better meet terminal learning objectives of user activities by employment of these NCOs and POs with field experiences. At present the NCOs and POs are assigned to the following 8 languages: Arabic, Chinese, Czech, German, Korean, Polish, Russian and Spanish. The TDA recognizes 65 NCO and PO positions (34 Army; 22 USAF; 7 USN; 2 USMC). Sixty-one NCOs and POs are presently assigned. The participating NCOs and POs are used in the same way as civilian universities use graduate assistants. Dependent on their language capabilities, they participate in the instruction process by instructing in grammar, translation, transcription and language drills, by handling language laboratories, and by administering tests. The NCOs and POs are also extensively involved in student tutoring, counseling and evaluation, and in course maintenance and development functions. The result of the Professional Development Program is an improvement in the linguistic skills of the participating NCOs and POs and is contributing to better qualified linguists for the services.

6. DLIFLC ATTRITION

Student attrition at DLIFLC is continuing to be studied with increased emphasis. Students are eliminated for either administrative or academic reasons; administrative attrition is beyond the control of DLIFLC and includes elimination for medical reasons, loss or non-award of security clearance, change requested by user agencies and failure to adapt to military life. Academic attrition consists of two categories, lack of effort (LOE), and lack of aptitude (LOA). Academic attrition rates are affected by a variety of factors such as scores on the Defense Language Aptitude Battery (DLAB), years of education, prior language learning, age, difficulty of the language, and individual motivation. No single factor as yet stands out as a useful predictor for students whose DLAB scores exceed the minimum cut-off of 89. We continue to collect and analyze the data, e. g., for possible combinations of factors. Our recent emphasis on earlier identification of attritees has significantly moved academic attrition toward the early weeks of the course. The overall attrition figures for DLIFLC and for the top ten languages, including dependents, (about 4 % of all DLIFLC students) is about 25 %.

7. DLIFLC CONSTRUCTION PROGRAM

Long awaited construction programs at DLIFLC have begun. While DLIFLC has excellent language training facilities, significant increases (approximately 500 each year) in student load has resulted in severe overcrowding with attendant degradation of instruction and the soldierization process. The construction program is centered in three general areas: Barracks, Classroom and Morale, Welfare and Recreational Facilities. With the latest Army stationing and installation plan figures indicating a requirement for 2000 additional enlisted barracks spaces to accommodate current and projected student loads, the FY84 Barracks project has been accelerated to FY83. Simultaneously, the burgeoning student load has had a tremendous impact upon our classroom, requiring extensive use of our 1905 vintage converted Horse Cavalry Barracks. DLIFLC has leased one nearby public school facility for administrative space to open up additional classrooms, is negotiating for another public school facility for additional classrooms, has opened up two Branch Facilities at Lackland AFB and Presidio of San Francisco and has implemented a staggered shift schedule of classes to cope with programmed increases in student load. As the FY83 Instructional Facility is completed we will start closing out the Branch Facilities. The available morale, welfare and recreational facilities are aged and inadequate for student population. An extensive facility upgrade as well as the entire listing of proposed projects is as follows:

FY	PROJECT	SQ FT	PN	CAPACITY	CWE (\$000)
83	Enlisted Person Dormitory (barracks)	167,865	T049	704 Per	23,200
83	General Instruction Facility	76,958	047	80 CR	8,100
84	Utilities Upgrade		T096		5,635
84	Dining Fac Modernization (Bldg 625)	10,145	050	600 Per	1,473
84	Enlisted Person Dormitory (barracks)	182,172	055	603 Per	14,600
84	General Instruction Facility	77,895	T083	80 CR	9,219
85	Dining Facility	9,500	T097	500 Per	1,900
85	Physical Fitness Ctr (w/outdoor complex) Includes ball fields.				
85	Academic Library	13,200	036		1,500
85	Child Care Center (152 Children)	11,400	T020		1,700
85	Recreation Center (Enlisted)	19,800	029		3,000
85	Logistical Support/Print Plant	46,000	042		2,500
86	Academic Auditorium	25,000	030	1500 Per	4,200
86	Military Personnel Center	25,000	T103		2,400
86	Post Exchange Facility (AAFES)	15,800			
86	Bowling Center (NAF)	14,400	T102	16 Lanes	2,400
87	Company HQS, Admin Complex	48,000	T100	3 Bldgs	4,800
87	Instructional Media Center	11,000	T106		1,200
87	Training Development/Eval Fac	41,000	T095		4,900
87	Skill Development Center (NAF)	11,000	T032		1,500
88	BOQ				
88	Post Exchange Auto Serv Ctr (AAFES)	2,500			4,600
89	International Cultural Ctr	3,600	043		445
89	Post Library	8,000	T056		1,200
89	Barracks Modernization (3 bldgs)		T107		
89	Enlisted Pers Club Annex (NAF)				
89	Youth Center (NAF)		T057		
90	General Instruction Facility	77,000	T104	80 CR	12,000
90	Barracks (w/dining facility)	161,700	T105	700 Per	21,000
90	Officers/Faculty Club				
90	Modernization (NAF)				
90	Exchange Snack (NAF)				
90	General Purpose Warehouse				
90	Post Chapel				

8. ACADEMIC PROGRAM HIGHLIGHTS

An Academic Master Plan was developed having six parts: instruction, course development and revision, a testing component, research and evaluation, faculty training, and professional development. DLIFLC has developed and is implementing a new unified teaching methodology. This is called Progressive Skills Integration (PSI), a progression of classroom activities that allows students to gain confidence to work with a language in real life situations. PSI methodology is defined enough to tell whether a teacher is following the method, yet a professional teacher can still: choose among options to meet specific class needs. PSI was developed with the assistance of several visiting professors and was published as DLIFLC PAM 350-10, Academic Policy and Standards. By the end of FY83 the entire professional workforce at DLIFLC will have been trained in PSI. Additionally, DLIFLC also has developed a supervisory course for managers who supervise foreign language teachers. All of DLIFLC's department chairpersons and supervisors have been training, so that they can assist in the major faculty training effort. Lastly, a two-week performance-oriented Basic Instructor Training Workshop trained 166 instructors in 1982.

9. BASIC COURSES

Development of DLI's new Basic Courses replaces dated materials used in the classroom today, and serves as instructional cores for future Intermediate and Advanced Language Courses, as well as certain non-resident courses. A five-year development plan, established the development priorities based on Terminal Learning Objectives furnished by the National Security Agency and performance requirements as provided by our DOD-wide user agencies. Using the systems approach - a first in the development of foreign language course materials - we perform a job and task analysis and then design the courses emphasizing communicative skills. Our comprehensive language training program is directed toward the development of four primary skills: listening comprehension, reading, speaking, and writing. The following basic courses (Phases I-III) are presently under development: Korean (47 weeks); Arabic (MSA) (47 weeks); Iraqi (16 weeks); Chinese (47 weeks); Polish (47 weeks); Czech (47 weeks); German (32 weeks); French (28 weeks). Development of a Spanish Intermediate Course (Phase IV) (8 weeks) has also begun. The Russian Basic Course (Phases I-III) (47 weeks) and the Arabic Egyptian (16 weeks) and Syrian (16 weeks) courses were completed in 1982 and are used in the respective departments. We are presently conducting Analysis and Design efforts for new basic courses in: Greek (47 weeks), Italian (28 weeks), and Japanese (47 weeks). In addition to projects within the Directorate of Training Development, several language departments in the Directorate of Training are accomplishing related tasks such as updating vocabulary and usage, revising tests, improving instructional techniques, etc.

10. RESIDENT TRAINING

In FY82 the average student load at Monterey was 2487 and 360 at Lackland AFB. At the Foreign Service Institute we trained an average load of 104 students. Training was conducted in 35 languages and dialects. Load projection for FY83 is 2724 in Monterey, 400 at Lackland, 268 at Presidio of San Francisco and 100 at FSI. This year DLIFLC commenced training in Dari and the Vietnamese Basic Course was extended from 42 to 47 weeks to add the Saigon dialect.

FOREIGN SERVICE INSTITUTE

In language training, FSI stresses the development of communication skills which personnel from all the foreign affairs agencies need to perform effectively on the job. About half of FSI's approximately 8,000 language students per year here and abroad are State employees or dependents; the others come from such agencies as DOD, USIA, AID, Commerce, Treasury and USDA.

During the past year, the Institute has consolidated and expanded the curriculum initiatives begun in 1981 which were designed to upgrade the quality and relevance of language training. One part of this effort has been the intergration of language and area studies. This program has succeeded in better preparing students for working effectively overseas. In carrying out this program, particular emphasis is being placed on developing substantive reading skills that officers will use on the job.

Another successful new activity which was expanded in 1982 is the series of "Bridges" exercises in which students use foreign languages to practice such functional skills as interviewing for information, negotiation, and defending U.S. policy. These exercises are now an established part of the language curriculum and FSI believes that - as a result - students are arriving in country better able than ever to use foreign languages on the job.

FSI has been field-testing new Arabic and Russian course materials, needed for these two high priority languages. These materials will be completed in the near future. In addition, FSI is introducing improvements in its testing of spoken language proficiency which it believes will result in more information about language competence.

The Familiarization and Short Term (FAST) program of 6-10 week language courses continues to show excellent results since its inception in 1981. FSI is working on expanding the number of such courses beyond the 16 now being offered. These courses are primarily designed for staff support personnel and dependents.

The Department and other foreign affairs agencies are working closely to implement the Model Foreign Language Competence Posts programs established by the Foreign Service Act of 1980. Under this program, all personnel assigned to the Model Posts (Dakar and Montevideo) will have an appropriate level of competence in the language of the country by October 1, 1983.

DLI, ENGLISH LANGUAGE CENTER
ANNUAL REPORT

1. INTRODUCTION

- a. Each fiscal year, the Military Departments provide DLIELC with the number of foreign military trainees (FMT) programmed to attend DLIELC prior to their entry into US technical/professional training programs. In FY 82, 2,531 FMT entered DLIELC. Approximately the same number are programmed to attend in FY83. The average daily student load of 600 has FMT personnel from approximately 50 different countries.
- b. The English language proficiency skill level required for entry into a technical/professional program is determined by each Military Department and is expressed in terms of an English Comprehension Level (ECL) test score on a scale of 0-100. The majority of the programs which are highly technical or hazardous in nature require an ECL of 80. Prerequisites for less technical courses vary from 65 to 80 ECL. The FMT is given an ECL screening test in-country prior to departure for CONUS. If the FMT does not meet the English language proficiency requirements for direct entry into the technical or professional program, or if the FMT requires Specialized English terminology training as a course prerequisite, the individual is programmed for additional language training at DLIELC.

2. ENGLISH LANGUAGE TRAINING COURSES

- a. The American Language Course (ALC) is a proficiency-based course and is variable in duration. Upon entry at DLIELC, an FMT is placed at the appropriate proficiency level in the American Language Course and receives six hours of instruction daily. During the last seven weeks of scheduled training at DLIELC, providing the minimum ECL score has been achieved, the FMT studies specialized technical terminology appropriate for the scheduled follow-on training program.
- b. DLIELC conducts three courses for selected FMT who are involved with the teaching of English in their homelands.
 - (1) The Basic English Language Instructor Course is a 27-week course. During this time, the trainees study the structure and phonology of English, and the DLI methodology of teaching English as a Foreign Language (TEFL). Emphasis is placed on TEFL techniques and peer teaching.

teaching. Six classes of this course are scheduled annually.

- (2) The Advanced English Language Instructor Course is a 13-week course. It is intended for experienced TEFL instructors who need to be updated on the TEFL "state-of-the-art" and on teaching techniques. This course is conducted quarterly.
- (3) The Advanced Programs in English Language Training Management Systems Course is an eight-week course. It is conducted twice a year for FMT who are managers, administrators, and/or supervisors in host country ELTP.

c. Three additional courses are also conducted by DLIELC as required for FMT.

- (1) Language Laboratory Maintenance Training provides instruction and practice in the installation, maintenance, and operation of language laboratories. Course duration varies from three to eight weeks depending upon previous experience of the trainees in electronics.
- (2) OJT Professional/Specialized, CONUS consists of on-the-job training in general laboratory procedures, i. e., operational and preventive maintenance procedures.
- (3) Observer Professional/Specialized, CONUS is tailored to cover those areas in the operation and administration of an English Language Training Program (ELTP) which are most appropriate to the observer(s) as defined by the host country.

3. OVERVIEW OF DLIELC TRAINING ACTIVITIES

The following special projects highlighted the academic training program during the past fiscal year:

- a. Egyptian Programs: The Egyptian Army, Navy, Air Force programs continued strongly with the large majority of the students receiving from eight to 12 weeks language training.
- b. US Army Pre-Basic ELT Pilot Program: The second increment consisting of 151 pre-basic US Army enlistees completed training during the fiscal year. Entry ECL scores varied from 0-60, and after completing 13 weeks of ELT, their ECL scores ranged from 0-100 with a mean score of 56.4. The soldiers also received US Army military familiarization training concurrently with the ELT.
- c. General English Language Training Program for US Army and USAF ROTC Students: The US Army students started to arrive at DLIELC in January 1982. A total of 50 students per year were programmed. The USAF students began entering in June 1982. The DLIELC mission is to train them in the General English Section for a period of up to 16 weeks for improvement of their speaking and comprehension abilities.

- d. EURO-NATO Joint Jet Pilot Training Program: At the request of managers of the EURO-NATO Joint Jet Pilot Training Program (ENJJPTP) at Sheppard AFB, Texas, DLIELC set up a nine-week Graduate Instructor Pilot Language Refresher Course stressing live radio communications and instructor pilot language development. Deficiency areas covered were live radio communication vocabulary and comprehension, and radio speech development.
- e. DLIELC continued to monitor all approved US Army Nonresident English Language Programs (NRELP) in CONUS and overseas, and to provide American Language Course (ALC) textbooks and other related training materials. Fifty-eight US Army Programs are located in CONUS; 151 are located overseas. US Army NRELP support also included on-site staff assistance visits and instructor orientation workshops.
- f. DLIELC continued to furnish ALC material support to the NRELP for Filipino personnel employed at the US Naval Ship Repair Facilities at Subic Bay, Philippines. Student load is approximately 200 per year. DLIELC was requested by Chief of Naval Operations to assist the US Navy in assessing the need for English language training for its limited English-speaking enlisted personnel above the rank of E-3. DLIELC staff, in conjunction with a USN representative, designed a field survey plan to administer the official English Comprehension Level (ECL) test and to interview these individuals.
- g. A DLIELC LTD continued to assist in the upgrading of the USAF NRELP for Portuguese nationals employed at Lajes Field, Azores.
- h. DLIELC deployed a one-member Language Training Detachment (LTD) to Mogadishu, Somalia. There were no changes in LTD assignments to Morocco, Saudi Arabia, Thailand, North Yemen, and Zaire. Two positions in Egypt were carried over from the previous year. Additional personnel were deployed as Mobile Training Teams (MTT) to meet requirements generated by the Security Assistance Training Program (SATP) in Egypt, Costa Rica, Honduras, Japan, and Sudan.

4. CURRICULUM DEVELOPMENT ACTIVITIES

- a. DLIELC uses a systems approach to the planning and development of English language instructional programs which ensures that personnel are taught the language skills necessary for the successful completion of follow-on technical training. Priorities in the curriculum development effort are established through analysis of student input, service needs, and the systematic updating of requirements (course review). Currently, all established priorities for development or revision of the American Language Course (ALC) materials are based on the Resident ELTP requirements.

- b. DLIELC continued a major overhaul of the ALC General English materials. Books 1 and 2 (formerly 1110 and 1210) of the Elementary materials went for operational tryout in Morocco, Yemen, and the General English Section of DLIELC. Books 3 to 6 are now in the process of being keyboarded and edited. Intermediate materials have been planned, titled, and charted in detail and three units of Book II (the first of the Intermediate series) have been written and keyboarded. Periodic course maintenance was performed on current materials.
- c. The implementation of a modular approach to language skills acquisition in technical areas has resulted in the revision of the Specialized English Training (SET) curriculum. DLIELC now offers a more diversified language-oriented curriculum in technical areas. SET curriculum was reduced from 50 technical areas to 28 to place greater emphasis on language skills. Two special language and study skills modules were integrated into the SET curriculum as prerequisites for all of the curriculum, making the SET curriculum a nine-week course of study. These modules, plus the language and study skills incorporated into the 28 technical curriculum areas, greatly enhance the students' comprehension of materials utilized in their technical follow-on training.
- d. In the Instructor Development area, new courses have been designed to meet changing needs in the fields of faculty development and foreign instructor training. The Advanced English Language Instructor Course was developed and is now in Training Material Operational Tryout (TMOT). This course replaced the English Language Instructor Refresher Course and is of the same length, 13 weeks. A 27-week course, Basic English Language Instructor Course, was designed, and the first module, DLIELC Methodology, is being written. This course will replace the English Instructor Course, of the same length, which has been the principal instructor training course at DLIELC.
- e. Approximately 1500 new English Comprehension Level (ECL) test items have been validated for entry into the Computer Generated Test (CGT) item bank. Computer software problems have precluded the use of these items. Present plans are to justify and purchase stand-alone computer hardware to support the CGT system. Extensive software revisions will have to be made in order to switch to this new hardware. We do not anticipate that the new system will become operational before FY84. We will, therefore, be limited to using our present inventory of operational and reserve ECL tests for at least the next 12 months.



SHAPE LANGUAGE CENTRE (SLC) - REPORT

1. Language Training

The year (1982-1983) has seen some re-structuring of class schedules and programmes to meet internal requirements; self-study versions of the basic courses are being developed and a revised English programme is planned for September 1983 (essentially time-tabling).

A major review of the English programme has started with Dr. Patricia McEldowney, University of Manchester, England (author of a language course and articles and Chief Examiner for the Northern University Joint Matriculation Board (JMB) for which she wrote the tests).

On 23rd June, at the NATO Headquarters, Janine Courtillon (CREDIF) will be presenting the French course "ARCHIPEL" - the most recent publication by CREDIF/Hatier - of which she is co-author.

Both video support and computer assisted learning have been followed up after last year's conference; the hardware and software are being researched. An experimental class has been run using the BBC course "Ensemble" and can be reported on. A major use of the video equipment has been the filming of seminars and tests (reported on elsewhere).

The SLC continues to provide technical assistance to the IMS NATO, Brussels, where it manages the language programmes for several NATO agencies.

2. Language Testing

The test battery described in last year's report is almost complete and alternative forms (for listening and reading) are being developed. (A separate presentation on this can be made at the conference to include video recordings of the oral test in English.)

3. Seminars

The lists of seminars attended or organized by the SLC Staff in 1982 and 1983 are enclosed; copies of individual reports can be obtained on request. One seminar is to be highlighted: the presentations by Dr. McEldowney and Dr. Trocme on December 17 1982 (LTC(82)SR/20) on which a video recording is to be made available at the conference.

4. A major project this year has been the recording for the Belgian Ministry of Education of a self-study course in English to be used by adults following further education. The material was recorded in September 1982 and is now in final form.

5. Language Circle

To the list of languages taught, Danish, Greek and Turkish have been added this year (1982-83). The Circle continues to offer language classes to both adults and youth, with special intensive courses in the summer.

6. Other matters (tour of ACE agencies/commands by the Head of Language Training, distribution of tests through the BILC Secretariat, exchange of materials as suggested by the Laboratory Working Group at the 1982 Conference etc.) will be reported on in the Steering Committee.

LANGUAGE CENTRE SHAPE/NATO

SEMINAR REPORTS 1982

SERIAL N° LTC(82) SR/ (N/REF N°)	DATE	PLACE	SPEAKER	SUBJECT	LTC REP.
1 (196)	27 Jan	British Council (B.C.) Brussels (Bx1s)	W.S. FOWLER	The compatibility of communicative language acquisition and grammati- cal accuracy in EFL	CW
2 (206)	22 Feb	B.C. Paris	Keith MORROW (Bell School, Norwich, U.K.)	The use of authentic materials in ELT	DHE (+)
3 (209)	24 March	B.C. Bx1s	Louis ALEXANDER	The use of authentic and semi- authentic materials in LT	NH/CW
4 (205)	15-19 Feb	CREPA Paris	Eve OGONOWSKI	"All's Well 2"	CW (+)
5 (276)	1 April	B.C. Bx1s	Christopher J. BRUMFIT	Pedagogic issues in the teaching of literature at secondary level	DHE/CW

(+) SHAPE LTC has supplementary documents for these reports.

SERIAL N° LTC(82) SR/ (N/REF N°)	DATE	PLACE	SPEAKER	SUBJECT	LTC REP.
6 (275)	22 April	Teacher's Association Bx1s	Michael SWAN	Practising grammar : communicative methods	CW
7 (226)	28 April	B.C. Bx1s	Hamish NORBROOK (BBC)	Video materials in the language classroom	CW (+)
8 (277)	6 May	B.C. Bx1s	Adrian PILBEAM Mark ELLIS (LTS London)	Systems one (A video-based English Language Training package)	DHE/CW (+)
9 (257)	13 May	B.C. Bx1s	David CARVER (Moray House, Edinburgh)	Self-access learning in ELT	MGA/FB
10 (258)	17-21 May	R.M.C. St Jean Montréal Canada	BILC 82 (see report for list of speakers)	The impact of advances in instr. tech. upon second L. L.	DHE (+)

SERIAL N° LTC(82) SR/ (N/REF N°)	DATE	PLACE	SPEAKER	SUBJECT	LTC REP.
11 (259)	19 May	B.C. Bx1s	John P. LOWERCAN Robert KENDREW (Nelson)	Filmscan	CW (+)
12 (336)	26 May	LTC SH	Henri BESSE (CREDIT France)	"Interlignes"	LTC SH LTC IMS NATO
13 (333)	24 May	FSI Washington USA	Staff of PSD SLS FSI	"Bridges"	DHE (+)
14 (334)	10 June	LTC SH	Maire O'REILLY	Threshold courses	LTC Staff
15 (335)	25 June	NATO HQ Bx1s	Varicus (LTC)	Resource centre / activity days	LTC Staff (SH + NATO)

SERIAL N° LTC(82) SR/ (N/REF N°)	DATE	PLACE	SPEAKER	SUBJECT	LTC REP
16 (344)	24 Sep	LTC SH	- Film	Using magazine pictures in the language classroom	LTC + LC (+) teachers
17 (376)	29 Oct	LTC SH	- Film	Use of recorded texts (Sound tape-recordings)	LTC + LC (+) teachers
18 (371)	10 Nov	Connaught Hall University of London	Paul Mason Cicero School	Language Teaching by Telephone	CW (+)
19 (374)	26 Nov	LTC SH	Maire O'Reilly	Challenges/problems of adults learners of languages - "Threshold" courses	LTC + LC (+) teachers
20 (375)	17 Dec	LTC SH-- Bldg 306	P. McEldowney H. Trocmé	(a) Progressions in an English language course (b) Applications to language teaching of brain and memory research	LTC, LC (+) IMS + various

Report of Study Groups 1 + 2

Language Aptitude Testing

Study Group 1

Language aptitude tests, placement and counseling

What connections can be seen between aptitude tests (content and results), the placement of student in classes and the counseling/guidance while learning? What are we doing? What should be done?

Chairman: Mr. D. Ellis

- | | |
|------------------------|-----------------------|
| - LtCol E. H. Barbeaux | - Col D. A. McNerney |
| - Cdt B. F. Barré | - Mr. A. Rutherford |
| - Dr. R. M. S. Curica | - Cdt J. Tancre |
| - RDir E. Leben | - LtCol R. H. Thomson |
| - Col F. Lenci | |

1. Inputs to this Study Group came from:

- a. Addresses 1, 2, 3, 4 and 6 (Plenary Session)
- b. Articles identified during the conference (see bibliography at Annex)

Mr. B. J. Carroll (address 4) and Dr. Günter Trost (address 6) each sat with the group for one of its three one-hour sessions.

Language aptitude (LA) tests, placement and counseling.

2. What are we doing?

From the group's discussion it appeared that LA tests are being used in North American countries, primarily for selection and placement but not for counseling. Most European countries represented (with the exception of Italy, where a test is now being validated at the Air Force School, and the UK, where parts of the MLAT are being given to personnel prior to 'long' courses) are not using aptitude tests. This is explained by the different situation applying over entry to language programmes, the European countries usually having no choice in the selection of personnel.

North American countries are generally satisfied with the LA tests being used but it was suggested that language-specific aptitude tests would be useful and should be explored.

It was noted that there is a growing tendency to make all test results available to candidates but that this is by no means universally applied.

3. What should be done?

Areas which merit research and experiment and which would lead in the long-term to significant improvements in cost effectiveness of language programmes are:

- a. language-specific aptitude tests (cf 2 above)
- b. use of bio-data obtained from candidates prior to testing and training (for details see Mr. Carroll's address to the conference).
- c. use of aptitude tests for
 - grouping of students
 - indications on length and pace of instruction
 - counseling during training
 - determination of learning strategies
 - researching differences between learners of a first as against a second (or more) foreign language

A heterogeneous test population would be needed and BILC would provide an ideal channel for organizing such a research project.

Study Group 2

Language aptitude testing, needs analysis and selection

Do we need aptitude tests to secure information about the existence and type of an individual's general language abilities and skills or do we need aptitude tests in order to predict more accurately whether someone should or should *not* learn languages? Can language aptitude in our respective contexts be an instrument of selection?

Chairman: WingCDR I. J. Thomas

- | | |
|--------------------------|-------------------------|
| - LtCol F. Borrelli | - LtCol J. R. Thériault |
| - RDir I. von Herzenberg | - Col E. Tutal |
| - Mr. P. J. de Lespinois | - LtCol A. Stefanelis |
| - Maj L. Noordsij | - Mr. G. G. Worrall |

4. In Group 2 we all agreed that when students were going to be trained only to the Elementary Level 1.1.1.0. of STANAG 6001 aptitude testing should not be required.

We further agreed that there is misunderstanding among users about the term aptitude.

Managers think of aptitude as an instrument to be used in the selection of students. Teachers think of it as an instrument that can give a specific profile on a student so that teaching strategies may be adapted to his specific learning habits. We decided that priority should be given to the development of an aptitude test that would assist the manager in the selection of students, predict learning time to achieve a pre-determined level of proficiency and permit adequate grouping of students.

5. Conclusions

Despite use of aptitude tests by several national authorities (and in particular the substantial research projects at the PSC Canada and the systematic testing in the USA) little is yet known about the predictive power of aptitude tests within the NATO context. Furthermore, limitation of their use to a selection instrument may be leaving many areas of potential importance untapped.

Research and experimentation outlined in paragraph 3 would bring to light much of the information at present not available and would thereby lead to economies in training programmes.

In a joint session of the groups we also discussed at length a proposal that would permit the collection of data on language aptitude with reference to language training in the NATO context.

This proposal was accepted by the Steering Committee Policy Subgroup.

Details of this proposal are included in the minutes of the Steering Committee.

Bibliography

1. Wesche M., Edwards H. and Wells W.

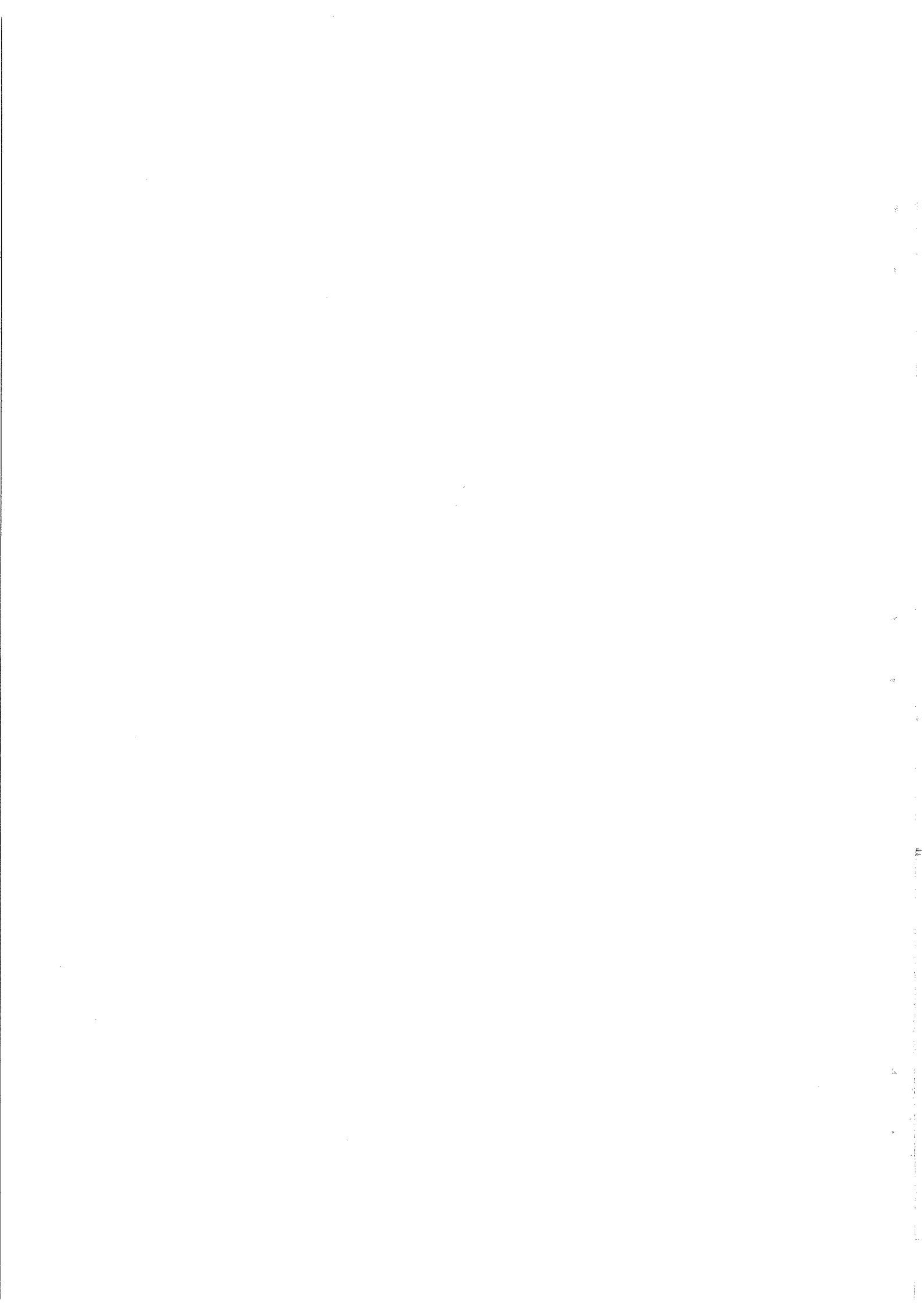
Foreign Language Aptitude and Intelligence Applied Psycholinguistics (1982) 3. 127-140.

2. Diller K. C. (Editor)

Individual Differences and Universals in Learning Aptitude
Newbury House 1981

3. CILT/CRS (British Council)

Foreign Language Testing Specialised Bibliography B 8,
August 1981



Rapport des groupes d'études 1 et 2

Groupe d'étude 1

1. Le groupe s'est basé sur des informations tirées
 - a. des conférences 1, 2, 3, 4 et 6 (Session Plénière)
 - b. des articles signalés durant cette session du BILC (cfr la bibliographie en annexe)

Mr. B. J. Carroll (conférence 4) et Dr. Günter Trost (conférence 6) ont pris part à l'une des sessions de travail du groupe.

Tests d'aptitude à l'apprentissage d'une langue (LA), évaluation et assistance.

2. Que faisons-nous actuellement?

Suite aux discussions, il apparaît que les 'LA tests' sont utilisés aux Etats Unis et au Canada, principalement pour la sélection et l'évaluation et NON pour l'assistance.

La plupart des pays européens représentés, n'utilisent PAS ces tests, à l'EXCEPTION de l'Ecole Militaire de l'Air d'Italie où ils sont en cours de validation, et du Royaume Uni où certaines parties du 'MLAT' sont utilisées pour des candidats à des cours de longue durée.

Ceci s'explique par les différentes modalités de désignation de candidats à un cours de langue (les centres de langues de ces pays n'ayant PAS la possibilité de choisir leurs futurs étudiants).

Les Etats Unis et le Canada sont satisfaits de l'emploi des 'LA tests' mais on pourrait suggérer que des tests d'aptitude spécifique à la langue choisie, seraient utiles et que des recherches dans ce domaine devraient être entreprises.

L'on a noté qu'il existe une tendance croissante à rendre les résultats des tests accessibles aux candidats mais que ceci n'est pas universellement appliqué.

3. Que doit-on faire?

Les domaines où recherches et expériences seraient utiles et pourraient conduire à long terme, à une amélioration significative de la rentabilité d'un programme de langue sont:

- a. tests d'aptitude spécifique à chaque langue (cfr § 2)
- b. emploi de données biographiques fournies par le candidat avant de subir des tests ou de suivre le cours (cfr conférences de Mr. Carroll et Mr. Trost)

c. emploi de tests d'aptitude pour

- regrouper les étudiants par catégorie
- préciser la durée et le rythme du cours
- assister suivre en cours d'instruction
- déterminer la stratégie pédagogique
- rechercher des différences entre les 'apprenants' d'une première langue étrangère par opposition à une seconde ou une troisième langue

Une population hétérogène est nécessaire pour l'organisation d'un pareil test; le BILC est idéalement placé pour organiser un projet de recherches.

4. Groupe d'étude 2

Le groupe estime que le test ne doit pas être utilisé pour des candidats qui ne vont pas au delà du niveau élémentaire (niveau 1.1.1.0. du STANAG 6001).

Le groupe estime d'autre part qu'il y a une certaine confusion à propos du terme 'aptitude'. Pour les directeurs de cours il s'agit d'un outil permettant le choix des étudiants. Pour les enseignants il s'agit d'un outil indiquant un profil déterminé permettant d'adapter une stratégie pédagogique aux habitudes d'apprentissage spécifiques.

Le groupe a décidé que la priorité devrait être donnée à l'élaboration d'un test d'aptitude qui aiderait le directeur dans le choix des étudiants, permettrait de prévoir la durée d'un cours pour atteindre un niveau de compétence pré-établi, permettrait un regroupement adéquat des étudiants.

5. Conclusions

Malgré l'emploi de tests d'aptitude par différentes nations (en particulier le programme de recherche du 'CSP' du Canada et l'emploi systématique de tests aux Etats Unis) très peu d'éléments sont connus pour déterminer la valeur prévisionnelle des tests d'aptitude dans le contexte de l'OTAN.

D'autre part ramener leur emploi à un outil de sélection pourrait laisser plusieurs domaines d'importance potentielle, inexplorés.

Les recherches et les expériences citées dans le § 3 pourraient faire ressortir une bonne partie de renseignements (indisponibles aujourd'hui) et conduire ainsi à des économies supplémentaires dans le programme d'instruction.

Lors d'une session commune, une proposition de 'collectes' de données d'aptitude a été discutée, en se référant à l'enseignement des langues dans le cadre de l'OTAN. Cette proposition a été acceptée par le sous-groupe de travail du comité directeur, chargé d'étudier la politique à suivre. Les détails de cette proposition sont repris dans les minutes de ce comité.

Report of Study Group 3

Language Proficiency Testing and the Curriculum

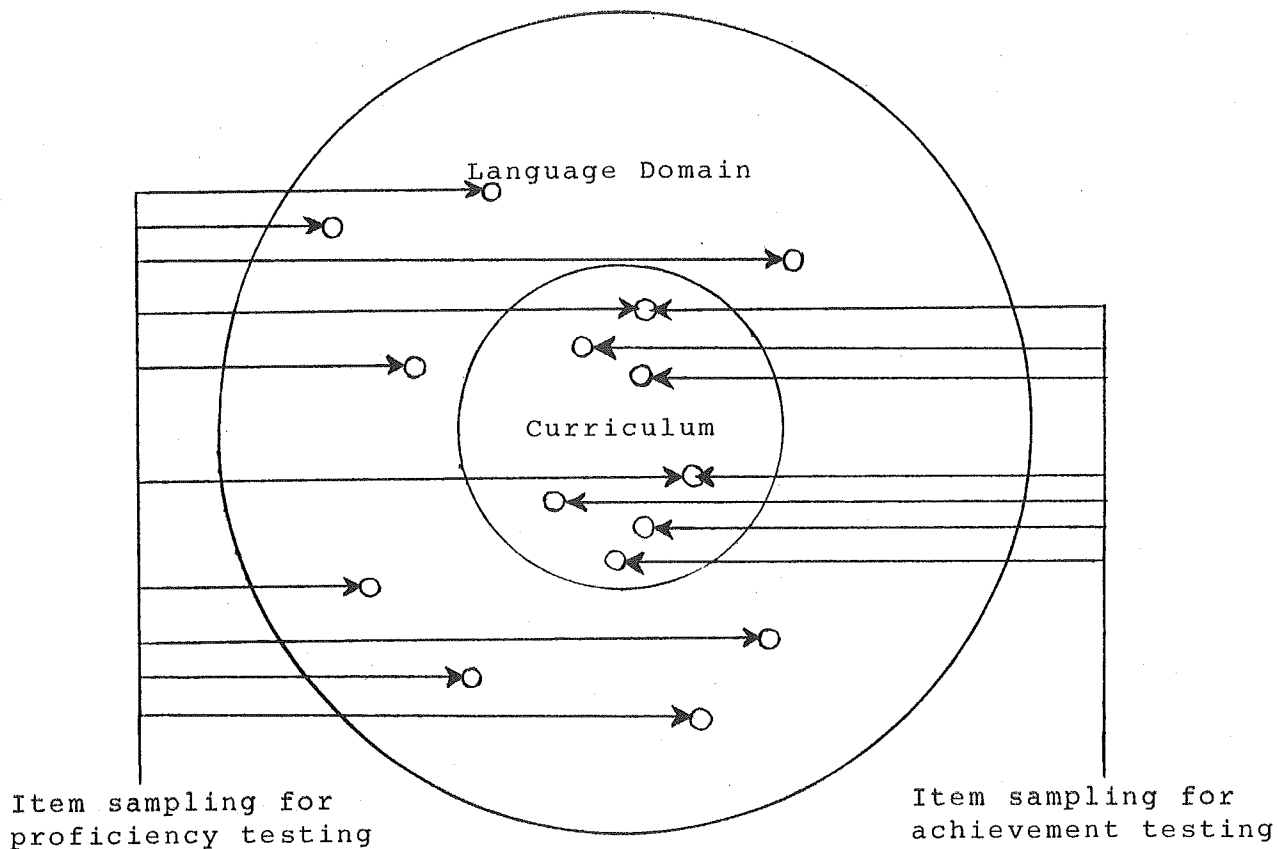
What is influencing what? Do we have to teach what we are testing? In what sense?

Chairman: Dr. R. Clifford

- LtCol G. Bellillo
- Cdt D. Filleul
- LtCol J. E. Goldsmith
- ORR G. Jüsten
- Mr. A. McLean
- SquLDR VWJ O'Hagan
- RDir P. Müller
- M. M. P. M. Schwarz

Statement of the problem

Normally one links an achievement test to its corresponding curriculum. Such a test exactly reflects the contents of that curriculum and as such it reflects a defined subset of the required language domain. In communicative or functional foreign language teaching, oriented towards the teaching of the four basic skills (listening, speaking, reading and writing) according to the STANAG 6001, the question of the relationship between curriculum and proficiency testing is a crucial one. While proficiency tests may include elements of the curriculum, they go beyond the curriculum because they go with the total required language domain of which the curriculum is only a representative subset.



Commentary

If there is a change in orientation in a system of teaching which includes needs analysis, syllabi, and curricula, what should be dealt with first? The teaching or the evaluation of performance? In other words: what influences what? Is it necessary on a basis of a needs analysis to first change the curriculum or is it better to first change the final exams?

It is probably more effective to first consider the concept of testing, because obviously as a consequence of changing the testing procedures the teachers will adapt and revise the contents of their courses as well as their methodology. To ask teachers to define learning objectives for the curriculum, without having given them a concrete idea of the training requirement through test examples, will only lead to the development of a great number of objectives - which are bound to be of variable value if not even contradictory - and to endless discussions among policy makers. On the other hand if the final examinations are created as proficiency tests dealing with the total required language domain rather than as achievement tests referring exclusively to a specific curriculum then teaching and learning will be directed towards the total required language domain. Students will not learn examples but will learn through examples.

Recommendations

1. If the precise content of a test is known it becomes the instructional goal, and thus a sub-subset of the language domain for the teachers and the total language domain for the students. To prevent this it may be appropriate to provide examination candidates with sample examinations showing text content and format rather than the actual test which is to be administered. (See for example the 'Prüfmittelbeispiele' of the Bundessprachenamt.)
2. In light of what is been said above it is essential to maintain the confidentiality of tests so that tests are not compromised. If teachers participate in the production of tests or know of their contents because they administer them, it is imperative to produce several forms of each test or at least to produce a large number of test items, so that these tests cannot be taught in the class and will remain genuine proficiency tests.
3. Whereas the slogan: "Teach what you test and test what you teach" is appropriate for achievement testing, and the slogan: "Teach in the same way you test and test in the same way you teach" applies to proficiency testing; the ideal would be to integrate the concepts of proficiency and achievement testing and accomplish both goals.

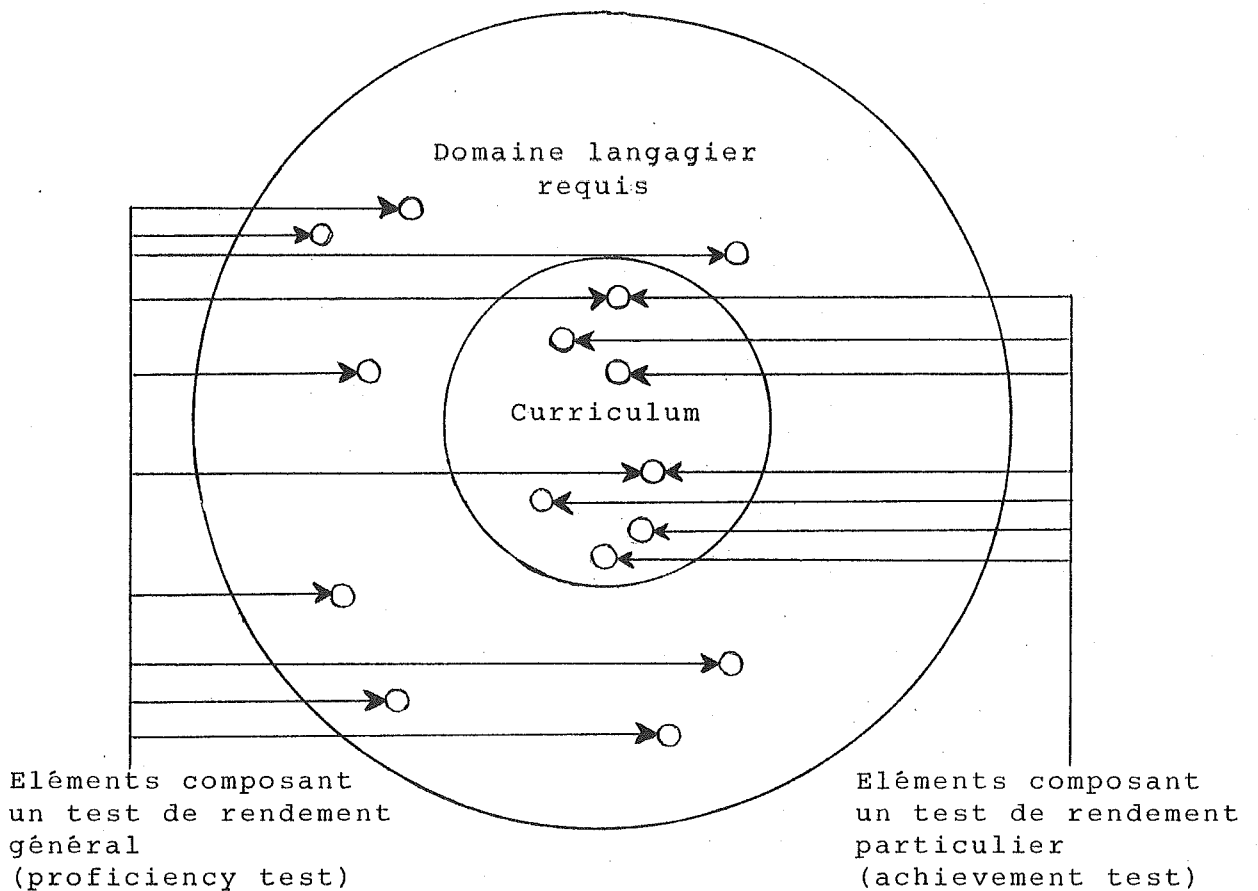
This is possible only if the curriculum exactly represents the required language domain as defined by a needs analysis. Such an approach would be desirable as it would completely subsume the fallacious dichotomy between teaching and testing. However, this has only been accomplished in specialized courses where it has been feasible to develop well defined objectives for the entire language domain.

Rapport du groupe d'étude 3

Le curriculum et les tests de rendement général

Le problème

Il est courant de combiner curriculum/syllabus et test de rendement particulier (achievement test): ces tests reflètent exactement le curriculum en tant que tel, c'est-à-dire en tant que substrat du domaine langagier requis. En langues étrangères, dans le cadre d'un enseignement fonctionnel ou d'un enseignement de la communication orienté aux 4 habilités langagières fondamentales (écouter - parler - lire - écrire) selon STANAG 6001, la question de la relation du curriculum aux tests de rendement général (proficiency tests) s'impose: ces tests incluent en tant que tels des éléments du curriculum mais le dépassent toutefois parce qu'ils se situent au niveau du domaine langagier requis, niveau que le curriculum ne fait que représenter.



Commentaires

Dans un système d'enseignement comprenant l'analyse des besoins, le syllabus et les curricula que faut-il, en cas de changement d'orientation, considérer d'abord? L'enseignement ou l'évaluation des performances langagières? En d'autres termes, qu'est-ce qui influence quoi?

- Faut-il sur la base d'une analyse des besoins d'abord changer le curriculum en revisant les objectifs d'enseignement, c'est-à-dire en 'collectionnant' les objectifs jugés nécessaires par les professeurs?

ou bien

- Vaut-il mieux changer d'abord les épreuves terminales, les tests?

Il semblerait qu'il est plus efficace de repenser d'abord la conception du testing parce qu'immanquablement, par après, les professeurs s'adapteront, c'est-à-dire reconsidéreront les contenus de leur enseignement ainsi que leur méthodologie.

Demander aux professeurs de définir les objectifs d'enseignement qui feront le curriculum, sans avoir donné par les épreuves terminales la conception voulue par l'institution concernée, conduit les planificateurs à des discussions sans fin, tant les objectifs proposés par les enseignants sont nombreux et de valeur inégale, voire même contradictoires. Par ailleurs, si les épreuves langagières terminales sont considérées comme des épreuves du type rendement général, et non du type rendement particulier (achievement test), donc se rapportant uniquement à un curriculum précis, l'enseignement et l'apprentissage seront eux aussi axés directement sur le domaine langagier requis. L'apprenant apprendra non les exemples mais par les exemples.

Recommandations

Si le contenu d'un test est connu dans ses détails, ce test risque de devenir le but de l'enseignement et le substrat du domaine langagier qu'il est en fait risquera de devenir d'une part le curriculum complet de l'enseignant et d'autre part, pour l'apprenant, le domaine langagier total et exclusif. Pour remédier à cet état de fait, il n'est pas inopportun de donner aux candidats aux examens des exemples de tests (contenu et format). (Voir à ce propos les divers 'Prüfmittelbeispiele' du Bundessprachenamt.)

Dans l'optique de ce qui a été dit plus haut, il faut veiller à conserver aux tests leur caractères confidentiel. Si les professeurs participent à l'élaboration de tests ou les connaissent parce qu'ils les font passer, il est impératif de préparer plusieurs variantes de tests ou du moins beaucoup d'items pour que ces tests restent des tests de rendement général qu'il est impossible de faire apprendre en tant que tels en classe.

Si le slogan didactique "Enseigne ce que tu testes et teste ce que tu as enseigné" s'applique bien aux tests de rendement particulier (achievement tests) et si le slogan "Enseigne de la façon dont tu testes et teste de la même façon que tu enseignes" s'applique parfaitement aux tests de rendement général (proficiency tests), l'idéal n'en serait pas moins d'intégrer les concepts de test de rendement général et de test de rendement particulier pour atteindre les deux objectifs à la fois. Ceci n'est toutefois possible que si le curriculum représente exactement le domaine langagier requis et défini par une analyse des besoins. Une telle approche permet de sublimer la dichotomie fallacieuse entre enseignement et examen mais ne semble, compte tenu du point de départ qu'est l'analyse des besoins, ne pouvoir s'appliquer qu'à des cours de spécialité.

Report of Study Group 4

Language Proficiency Testing and the Teacher

Do the teachers have to be involved in designing, defining the contents and administering and evaluating language proficiency tests? Is it desirable? Is it feasible?

Chairman: Mr. M. P. Rangongo

- | | |
|---------------------|--------------------|
| - LtCol R. C. Brace | - Col A. Perrotta |
| - LtCol C. Coltelli | - Maj C. C. Pearce |
| - Mr. F. Gregory | - Mr. J. Ratliff |
| - Mrs. J. Lawrence | - RDir F. Wech |

In the opinion of Study Group 4, involvement of the Teacher in various aspects of Language Proficiency Testing may be required, desirable, or feasible as indicated below:

<u>Type of Involvement</u>	<u>Required</u>	<u>Desirable</u>	<u>Feasible</u>
Designing the test	No	Yes	Depends*
Defining the contents	Yes, for special language No, for test content	Yes	Depends*
Administering the test	Some instances**	Yes	Depends*
Evaluating the test	No	Yes	Depends*
Evaluating student Performance	No	Yes	Depends*

We also examined the various advantages and disadvantages of teacher involvement in the test. Our findings are as follows:

Advantages

1. Gives teachers credibility in the eyes of the students
2. Makes them better teachers
3. Motivates teachers
4. Contributes to the self-respect of teachers
5. Increases teacher accountability
6. May contribute to increased mutual understanding between teachers and test developers

* In some countries, the existence of a standard test which must be used by all may preclude involvement of the teacher in some or all aspects.

** In some small or isolated programs, there may be no alternatives to having the teacher administer the test.

Disadvantages

1. Test may lose credibility in eyes of examinee's employer
2. May increase danger of test compromise
3. May not always result in a better test

The group concluded that, assuming the existence of clearly-stated course objectives and well-trained professional instructors, the involvement of instructors in testing may ensure a better test, particularly through feedback and suggestions to the test preparers. At the same time, teacher involvement in all aspects of testing is rarely essential to the development of a good measuring instrument. However, involvement of instructors in various aspects of the test contributes to the effectiveness of the program, through better instructor morale, understanding of course objectives, and a more obvious linkage between activities in the classroom and the language proficiency test.