NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

SECOND LANGUAGE TESTING:
THE RELATIONSHIP BETWEEN COMPETENCE AND PERFORMANCE

by

Hedy M. McGarrell

A Thesis submitted to

The School of Graduate Studies and Research

In Partial Fulfillment of the

Requirements for the Degree of

Master of Arts

Department of Linguistics

University of Ottawa

December 1980

A B S T R A C T

The purpose of this thesis is to investigate the difference between competence and performance in the context of communicative second/foreign language testing. It is hypothesized that performance has a more important role in language testing than past research would seem to indicate. It is generally accepted by researchers that language is primarily for the purpose of communication. Accordingly, second/foreign language proficiency tests should reflect the most characteristic features of communicative situations. It is shown that cloze, dictation, TOEFL and Michigan, i.e. four tests frequently used for the purpose of measuring overall proficiency do not, in fact, evaluate communicative ability. It is concluded that, in spite of current interest and research in testing communicative ability, fundamental changes in both theory and practice are imperative before this ability can be adequately measured.

ACKNOWLEDGEMENTS

I would like to thank my thesis supervisor Dr. P.C. Hauptman for his guidance and advice throughout. I would also like to thank Dr. M. Wesche and Dr. P. G. Patel for their comments and helpful suggestions. Special thanks also to my husband for his patience and support.

CONTENTS

LIST OF FIGURES

INTRODUCTION

The field of second/foreign language testing has attracted an increasingly large amount of attention in recent years. Language testing seems to provoke disproportionately more interest and controversy than other traditional subject matters (Jones 1977). No doubt, this is at least partially due to the complex processes and the large number of variables involved in language learning. There is still considerable uncertainty and disagreement amongst researchers as to what it means to know a language. The processes involved in first and subsequent language acquisition are very complex and as yet not fully understood, as a survey of the appropriate literature will confirm.

The lack of agreement over the meaning of "knowing a language" has resulted in a variety of views concerning the components to be tested in the second/foreign language learner. In this thesis the relationship between competence (i.e. knowledge) and performance (i.e. the ability to use this knowledge) is examined in the context of second/foreign language proficiency testing. It is hypothesized that the concept of communicative performance (defined as the ability to communicate in real-life situations and referred to as communicative ability) has a more central role than the literature seems to indicate. Furthermore, it is hypothesized that several frequently used tests do not evaluate communicative ability.

The first chapter offers a brief historical overview of language testing during the twentieth century. Spolsky (1975) identified three major trends in language testing. The discussion shows that the labels he used to describe these trends characterize not only the prevailing attitudes and biases towards language testing but also the general theories of language

learning and language teaching. Each one of the three trends uses different testing instruments, reflecting a different view of what it means to know a language and how this knowledge is acquired. During the third (and current) trend the concepts of competence and performance developed.

The second chapter examines this competence/performance dichotomy as proposed by Chomsky. A discussion of the views of some of the most influential theorists reveals that the dichotomy was considered to be too restrictive. The concept of communicative competence was subsequently developed to deal with linguistic as well as with sociolinguistic aspects of language.

Chapter Three discusses further theories of communicative competence that have been proposed. The theories discussed in Chapter Three represent a more integrative approach towards the concept than those in Chapter Two. Furthermore, researchers discussed in Chapter Three tend to be more concerned with second/foreign language learners' competence than with native speakers' hypothetical knowledge.

In Chapter Four the role of performance in second/foreign language learning is discussed and the importance of communicative ability (performance) for the purpose of language testing is shown. Although it is generally accepted that communicative competence can only be measured through performance, little research is devoted to that aspect of the dichotomy. As language is normally used for the purpose of communication, it is proposed that communicative tests should evaluate learners' communicative ability in terms of a number of features found to be characteristic of communicative situations. Such features are proposed based on Morrow (1977).

Chapter Five evaluates four widely used tests claimed to measure learners' communicative competence (overall proficiency)

against the criteria set forth in Chapter Four. An experimental test, designed specifically for the purpose of evaluating communicative ability in second/foreign language learners, is also evaluated.

1. HISTORICAL PERSPECTIVE

Throughout the history of education, testing has been closely linked to teaching. It would be difficult to consider second/foreign language testing apart from the teaching or learning process. The extent of the students' knowledge is of considerable interest to learner and teacher, and, as Ingram (1977) has pointed out, to those who pay the teachers, as well as potential employers of the students.

The 20th century alone has seen many rival approaches to teaching and testing come and go, and

> "...the dogma of one movement has become the heresy of the next" (Morrow, 1977)

Upshur (1969) noted that trends in language testing have tended to follow trends in language teaching, which in turn have tended to follow trends in linguistics. Thus there is a lag between the development of a new paradigm in linguistic theory and its subsequent application in language teaching and its exploitation in testing. Some teaching methods which were developed under the influence of a supposedly new approach, are tested by the methods from the previous doctrine.

Spolsky (1975) identified three stages in the development of language testing in the 20th century:

- the pre-scientific (mainly prior to world war II)
- the psychometric-structuralist (beginning in the late 1940's)
- the integrative-sociolinguistic (recent)

While these stages follow in the order given, they overlap both in time and approach. The integrative-sociolinguistic trend picks up several elements of the first, and the second and third trends co-exist and compete to some extent. These trends have not replaced each other, the effect has been cummulative. Ele-

ments of all three trends can be found in language testing material today. This chapter offers a brief discussion of the three major trends identified by Spolsky (1969). It shows the prevailing views of language and learning, language teaching and language testing for each of the three trends. A short outline of the major strengths and weaknesses of each trend concludes each of the three sections.

1.1. The pre-scientific trend

The pre-scientific trend in language testing (as described by Spolsky) is most apparent within the framework of the grammar-translation approach to language teaching.

1.1.1. Views of language and learning

This trend is based on the traditional humanistic orientation, which places primary importance on the literature of a country. It satisfies the predominant mental faculties school of thought that regards language learning as a way to train the mind. No relationship between a language and the environment is seen to exist. For this reason Davies (1968) has labelled the theory evident in the grammar-translation approach the "strict-separatist" view. Language is seen primarily as an object of study in itself and as such it is strictly separated from its use in society.

1.1.2. Language teaching and the learner

According to Chastain (1976) the primary aims of language teaching within the grammar-translation approach are:

- to prepare the learners to undertake exhaustive studies of the literature of the target language
- to encourage learners to have a deeper understanding of their native language
- to develop the learners' mind through practice in dealing with difficult learning situations and materials.

To attain these objectives, learners in the grammar-translation program typically memorize vocabulary and grammatical rules of the target language. Grammatical rules are taught deductively by means of long explanations. Grammatical rules and exceptions, regularities and irregularities are discussed at length in the students' native language. Kelly (1969:55) has observed that the grammar-translation approach expects the learner

> "...to be able to sort out from his grammar the rules like those of his own language and distinguish them from others."

Course content is based on a grammatical progression, but the progression chosen as well as the vocabulary included, are arbitrarily selected by the course designer. Some items are considered more difficult. Consequently, French verbs ending "-er" for example are taught before those ending "-ir". Kelly (1969) has noted that such a gradation in the presentation of material can be traced back to the early 16th century scholar Comenius. There are still teaching materials using such gradation available today.

1.1.3. Testing

Testing under the pre-scientific trend is characterized by a total lack of concern for statistical information. There is no accounting for validity, reliability or objectivity. Testing is considered to be the responsibility of the individual teacher. He sets his own standards and criteria.

Tests tend to require translation, composition or sentence completion type exercises. They are usually written and of an open-ended format. Oral examinations of any type are rarely used.

1.1.4. Strengths of the pre-scientific trend

Testing techniques advocated under the pre-scientific trend

allow a global evaluation of the learners' ability in the target language. Essay writing and similar activities require the learners to synthesize their knowledge and to produce utterances in the target language.

1.1.5. Weaknesses of the pre-scientific trend

From the short outline presented in the above, a number of problems become apparent:

- Subjectivity
  Scoring, evaluation and grading systems are subjective (Pilliner 1952) and it is difficult to assess students' ability in a systematic and objective manner. Examiners' personal prejudices and interference from stylistic considerations are likely to affect marking.
- Limited objectives
  The narrow view of language predominant under the pre-scientific trend in testing entails limited objectives for language testing and teaching.
- Examiner Training
  Language testing is entirely the teachers' responsibility. No specific training is required for language testing. Accordingly, the criteria used vary considerably. Some teachers put the emphasis on grammatical precision (according to the examples set by well-known literary authors), while others are more concerned with style or creativity.

1.2. The psychometric-structuralist trend

The psychometric-structuralist trend in language testing is most apparent within the audio-lingual approach to language teaching.

1.2.1. Views of language and learning

Language and language learning are viewed as basically simple

processes that lend themselves easily to empirical investigation and description. Views of language and language learning are greatly influenced by behaviourism in psychology and structuralist theories in linguistics. Under these influences, learning theory (e.g. Skinner 1957, Quine 1960) maintains that learning is shaped and consists of conditioned responses to stimuli from the environment. Applied to language learning, this view results in what Davies (1968) has called the "separate but equal" view of language, where the correct language response has to be associated with the correct stimulus from the environment. Mastery of a language is expected to be best achieved through mastery of what structuralism considers its constituent components.

Moulton (1961) summed up the then prevailing attitudes to language in the form of slogans:

- language is a set of habits
- language is speech not writing
- language is what people say, not what they ought to say
- languages are different

1.2.2. Language teaching and the learner

The implications of the ideas expressed in Moulton's slogans provided ideal conditions for what came to be known as the audio-lingual approach to language teaching and learning. Behaviour modification to achieve automaticity of response and structural pattern practice became the most important classroom activities. Politzer (1964) confirmed

> "...behaviourism and formal analysis of language were the chief features of the linguistic impact on language instruction." (Politzer 1964:149)

and Valette concurred

> "The new curriculum materials have been devised on the assumption that foreign-language learning is basically a mechanical process of habit formation." (Valette 1966:132)

Characteristics of audio-lingual methods include:[1]

- Rules are taught inductively. The meaning of utterances is secondary to a precise formulation of every pattern. Brooks (1960) confirmed:

  "The single paramount fact about language learning is that it concerns, not problem solving, but the formation and performance of habits." (Brooks 1960:40)

- Learning problems are anticipated and identified through contrastive analyses between the learners' mother tongue and the target language.
- The learners' mother tongue is not used in the classroom. The observation that languages are different (Moulton 1961) leads to the assumption that nothing but interference can be expected from the learners' mother tongue.
- Memorization of dialogues, pattern drills and structural pattern practice form the basis of language teaching influenced by theories such as those advocated by Lado and Fries. Learners are expected to gain automaticity of response to a given stimulus.

1.2.3. Testing

Language testing takes on a new direction under the influence of two sets of experts:

a) The psychologists; whose concern for the development of "objective" measures in educational measurement led to the application of statistical techniques in an effort to achieve reliability and validity.

b) The linguists; whose concern with the lack of linguistic principles in language testing led to "the structural linguists" approach to testing" (Spolsky 1978:vii)

The contribution of psychology towards the psychometric-structural trend in testing includes the development and painstaking standardization of short item multiple choice tests.

Such tests met the demands for objective, reliable and valid measuring instruments. The paramount importance of these criteria is evident in statements such as

> "...the shape of all tests, whether predictive or non-predictive, language or non-language, is primarily determined by the need to test the tests for reliability and validity. That is why, for instance, the multiple choice technique of answering is so common." (Ingram 1968:74)

The new developments in language testing based on a psychometric approach appeared to have at least two shortcomings:

- the new tests tended to use a written format, which limited their field of application somewhat;
- the selection of items was not based on any new theory of language learning and teaching, nor did it seem to have any foundation in theoretical linguistics.

Agard and Dunkel (1948) were among the first to draw attention to the potential practical deficiencies of the new testing techniques. They pointed out that, despite increased emphasis on aural-oral skills, there were no suitable tests to evaluate these skills. Similar criticisms were voiced by Carroll (1952, 1953) in reviews of developments in the field of language testing.

The linguistic side of the theoretical foundation for the psychometric-structural trend in language testing developed out of Lado's (1950) doctoral dissertation. In his conclusion he stated that

- a great lag exists in measurement in English as a foreign language (i.e. new developments in testing and testing theories were not supported by new developments in linguistic theory)
- the lag is connected with unscientific views of language
- the science of language should be used in defining what

to teach.

Lado further specified procedures to be followed to apply linguistic theory to the development of foreign language tests. Close to ten years later, Lado's (1961) refined ideas concerning language testing were published in what Spolsky (1978) has called "a classic exposition of the structural linguist's approach to testing". Constructing language tests could now be considered a science. Strict procedures had to be followed and intuition and proficiency in the target language were no longer a sufficient qualification for test designers.

The linguists' understanding of the nature of language became paramount in setting the specifications for tests, which theoretically were to be based on the contrastive analysis hypothesis.

Lado's approach to testing received support from current linguistics and psychology on a theoretical as well as on a practical level. On a theoretical level the majority of linguists[2] and psychologists agreed that language learning consisted of the formation of a new set of habits in the learner. On a practical level, test designers needed long lists of short items to be sampled and tested objectively. Structural linguists were able to deliver such items readily. This practice resulted in "discrete structure point" items (Carroll 1961). Carroll, one of Lado's early supporters, explained:

> "The work of Lado and other language testing specialists has correctly pointed to the desirability of testing for very specific items of language knowledge and skill, judiciously sampled from the usually enormous pool of possible items. This makes for highly reliable and valid testing." (Carroll 1961, quoted in Spolsky 1978:vii)

Thus short-item, multiple choice, discrete-point tests became the most common form of evaluation within the psychometric-struc-

turalist trend. Such tests typically consist of a selection (to be made by the examinee) of correct forms or items among a number of possible replies (i.e. the distractor items). Discrete-point testing items usually aim at testing one point of grammar (e.g. phonology, syntax) at the time. Students do not normally need to understand whole sentences to answer discrete-point items correctly. Not all multiple choice tests are necessarily discrete-point tests, but only discrete-point tests can tell exactly where the individual learner went wrong.

1.2.4. Strengths of the psychometric-structuralist trend

The type of testing techniques advocated by Lado and his supporters allow the development of tests that conform to high standards of reliability and validity. Furthermore, the tests are relatively economical and easy to administer. Large numbers of students can be tested simultaneously. Scoring and evaluation of tests is mechanical and can be left to machines. These characteristics encouraged the development of several standardized tests for the evaluation of student progress and program success. For example:

Test of English as a Foreign Language (TOEFL)
MLA Foreign Language Test for Teachers and Advanced Students
Graduate Record Examinations Advanced Tests in various languages
Michigan Proficiency Test Battery

Several of these tests have been in use for many years, although some of them have undergone modifications to allow, at least in part, for changing views about language and learning.

Lado's views on language testing (1961, 1962) have profoundly influenced subsequent testing techniques. The concepts of statistical validity and reliability have become prerequisite to language testing and directly quantifiable modes of assessment are sought even by Lado's opponents.

1.2.5. Weaknesses of the psychometric-structuralist trend

Tests developed to conform to theories of the psychometric-structuralist trend have been critisized for a number of reasons. General theorists (for example Hamp 1968, Di Pietro 1971) as well as researchers involved in the development of second language tests (Upshur 1962) have attacked Lado's premise that a contrastive analysis is central to language testing. The psychometric-structuralist approach to test design relies on the assumption that knowledge of the individual elements of a language is equivalent to knowledge of the language. In recent years it has become increasingly obvious that knowledge of the elements of language is insufficient. Learners are often unable to synthesize their knowledge in order to meet the linguistic demands of language use.

Robinson (1973) identified at least three areas, where "objectively" designed tests seem inferior to tests from the pre-scientific trend in the sense that they limit the students' role in the test.

- Students produce little or no language, often they are limited to the selection of alternative answers.
- Tests from the pre-scientific trend usually test the learners' ability to produce language. Objective tests usually assess the students' ability to recognize correct forms.
- In a subjective test, the learner is able to use his own norms of language. In an objective test, the learner has to base his replies on the language norms of the test designer, giving no indication of how he would evaluate a particular situation with regard to contextual factors.

The preceding discussion indicates that test designers have adopted an over-restrictive view of the type of language to be evaluated in order to increase reliability of the tests. Davies

(1978) called this the "reliability-validity tension". Furthermore the over-restrictive view has resulted in decontextualized language tests that are generally far removed from life-like language.

1.3. The integrative-sociolinguistic trend

This trend in language testing is most apparent in cognitive and especially functional approaches to language teaching and testing. It developed out of increasingly strong and numerous attacks on the principles of the contrastive analysis underlying the psychometric-structuralist trend in language testing. Carroll (1961), after having described the discrete-point approach, also described its short-comings. He then argued in favour of what he called an "integrative" approach, where the "total communicative effect of an utterance" would be taken into consideration. He provided a clear statement of an "integrated" position a few years later:

> "...since the use of language in ordinary situations calls upon all these aspects [of language], we must further recognize that linguistic performance also involves the individual's capability of mobilizing his linguistic competences and performance abilities in an integrated way, i.e. in the understanding, speaking, reading or writing of connected discourse." (Carroll 1968:58)

This position is, in fact inconsistent with an atomistic approach to language as a basis for testing. Integrative testing requires a global view of language that runs counter to Lado's fundamental assumptions mentioned above. As far as Carroll was concerned, there was no conflict. He felt that global performance should be tested in addition to individual elements, thus mobilizing all the elements of a particular language.

1.3.1. Views of language and learning

The integrative-sociolinguistic trend is influenced mainly

by developments in psycholinguistics and in sociolinguistics.

1.3.1.1. Psycholinguistic Aspect

The psycholinguistic aspect is based on Chomskyan linguistic theory. Chomsky's view of language is fundamentally contrary to that underlying the psychometric-structuralist trend. Chomsky pointed out the creative element in language. The infinite number of possible sentences in a language makes it impossible for any one native speaker to produce all the feasible and grammatical sentences of a language within his lifespan. Applied to the second language learning situation, Chomsky's position on first language acquisition called for a reconsideration of the principles supporting the psychometric-structuralist trend in language testing. Accordingly, second language learning is unlikely to reach the goal of native speaker ability on the basis of the behaviouristic habit formation theory and the structuralists' atomistic perception of language. Chomsky postulated the existence of "innate ideas" in every human being and an underlying rule system for every language (in addition to the existence of linguistic universals). The internalization of underlying rule systems is seen as the key to successful language learning by proponents of Chomsky's theory.

The fact that a considerable number of psychologists abandoned behaviourism in favour of cognitive learning theories has further influenced the integrative-sociolinguistic trend in language testing. Cognitive psychologists do not view cognitive learning as external, mechanistic processes but as a meaningful mentalistic process in which the learners play a central role. Ausubel (1968) for example, strongly rejected behaviourism, criticizing

> "...the prevailing tendency...for educational-psychologists to extrapolate findings from animal, rote, and perceptual-motor learning experiments..." (Ausubel 1968:ix)

He claimed that

> "The acquisition of large bodies of knowledge is simply impossible in the absence of meaningful learning." (Ausubel 1968:61)

According to Ausubel, cognitive theory regards the human nervous system as a data-processing and storing mechanism, where new ideas and information can be meaningfully learned and retained. This, however, occurs

> "...only to the extent that they are relatable to already available concepts or propositions which provide ideational anchorage." (Ausubel 1967:11)

Carroll (1961) referred to the influence of these developments in psychological theory on language learning theory as "cognitive-code learning theory" as opposed to the "audiolingual habit theory".

> "The cognitive-code learning theory may be thought of as a modified, up-to-date grammar translation theory...Learning a language is a process of acquiring conscious control of the phonological, grammatical and lexical patterns of a second language, largely through study and analysis of these patterns as a body of knowledge. The theory attaches more importance to the learner's understanding of the structure of the foreign language than to his facility in using that structure, since it is believed that provided the student has a proper degree of cognitive control over the structures of the language, facility will develop automatically with use of the language in meaningful situations." (Carroll 1966: 101-102)

A more cognitive approach to language learning theory has led to the perception of language learning as a progressive evolution of transitional stages of knowledge toward a model of competence. Language learning is seen as a process of gradual in-

ternalization of rules.

1.3.1.2. Sociolinguistic Aspect

While the first aspect of the integrative-sociolinguistic trend owes much to the influence of Chomsky and of cognitive psychology, the second part is, at least in part, a reaction to the failure of Chomskyan linguistic theory to deal with the complexity of language behaviour much more satisfactorily than the theories it replaced. The sociolinguistic aspect is centered around the argument that knowing a language involves more than the knowledge of a code. It involves the language user's ability to use the language in context. A language learner should be tested for his ability to communicate, to use linguistic elements appropriately. Chomsky's theory, however, does not allow for an account of sociolinguistic factors. Hymes proposed to broaden the scope of Chomsky's competence/performance dichotomy (see next chapter). He expained:

> "We have then to account for the fact that a normal child acquires knowledge of sentences, not only as grammatical, but also as appropriate. He or she acquires competence as to when to speak, when not, and as to what to talk about with whom, when, where, in what manner. In short, a child becomes able to accomplish a repertoire of speech acts, to take part in speech events, and to evaluate their accomplishment by others." (Hymes 1972:277)

Hymes has further pointed out that the social context of a message is as important as its linguistic context. Missing the cultural cues of a message may, in many cases, be synonymous with missing the message entirely.

1.3.2. Language teaching and the learner

Whereas audiolingual habit theory tries to discourage the

learner from comparing the target language with his mother tongue, the cognitive-code theory favours careful explanations of differences between the two languages involved. Under the influence of the cognitive approach to language teaching, the learner is encouraged to acquire conscious control of the target language. The learner is expected to reach a point at which he is able to formulate his own replies to previously unheard utterances. One of the most important tenets of the cognitive approach to language teaching is that learning should be meaningful to the learner. Chastain (1976) has stated:

> "The students should understand at all times what they are being asked to do. They should understand what they are saying, writing, reading, and hearing. New material should be organized to relate to knowledge that the students already have about their own language, the second language learned to that point in the course, and their concepts about the world about them." (Chastain 1976:147)

Paulston (1974) has related Hymes' arguments to the actual second language teaching situation. She has shown that communicative competence, and with it the sociolinguistic trend in language testing, reflect important aspects of knowing a language. Paulston's arguments imply that teachers need to do more than simply describe the target language. They should teach learners what else native speakers do when they communicate. To this effect, Kettering (1974) has described a number of cultural and social tasks (for example hiding/showing feelings through excuses, compliments, complaints) and how they can be incorporated into teaching situations.

The integrative-sociolinguistic trend in language testing is characterized by teaching methods which are no longer organized in terms of grammatical features. There are few methods that could be considered truly "cognitive" (e.g. Dialogue Canada II,

Part II).

A desire to render language learning more efficient has resulted in teaching methods organized in terms of features derived from language use. The communicative needs of the learner can be determined by predicting the type of situations in which he will most likely need to function. This allows the syllabus designer and the materials writer to present the learner with language appropriate to his own particular needs. A considerable amount of work is in progress in this area. Wilkins (1976) has outlined the categories for a proposed "notional syllabus". Munby (1978) has outlined a scheme that allows an elaborate specification of the learners' target communicative needs. The Council of Europe's Modern Languages Project has stimulated a number of specifications of a "threshold" level of language proficiency, i.e. the learners' needs (Coste et al 1976, Council of Europe 1973 and 1979, Richterich and Chancerel 1978, Roulet 1976, Trim 1980, Van Ek 1975, 1976, 1980 to name but a few). The above developments in psycholinguistics and sociolinguistics have also led to renewed interest in experiential and instrumental approaches to second language teaching such as immersion programs (e.g. Lambert and Tucker 1972, Macnamara 1966 and 1973, Newmark 1966 and 1968, Savignon 1972, Swain 1972, and others). Several syllabuses for "specific purposes" have been designed (e.g. Widdowson 1975). Language teaching has thus become learner-centered in its aim to meet learners' communicative needs.

1.3.3. Testing

The integrative-sociolinguistic trend in language testing marks a shift of emphasis from linguistic accuracy to functional ability. Tests reflect the problem-solving approach towards language teaching advocated (e.g. Chastain 1976). They are expected to reveal what underlying rules the learner has internalized and how they deviate, at a given point in time, from the way

they are used by native speakers. Corder (1967) called this stage the learners' "transitional competence", which enables them to comprehend and produce utterances within the scope of their "interlanguage" (Selinker 1972). Corder (1967) advocated techniques that give learners an opportunity to form and test their own hypotheses about the target language and to receive positive feedback from their errors. In addition, the concepts of reliability and validity, central to test construction under the influence of the psychometric-structuralist trend (Valette 1967, Harris 1969, Clark 1972) are still of considerable importance.

The integrative-sociolinguistic trend in language testing calls for testing instruments that allow an assessment of the learners' communicative competence. Prevailing theory requires an evaluation of the learners' ability to comprehend and produce utterances that are both situationally and contextually appropriate. To satisfy this condition, a number of different tests are proposed, e.g.

- oral interview
- cloze
- dictation
- essay writing

The most frequently used testing techniques seem to be cloze and dictation and updated versions of TOEFL and Michigan tests.

1.3.4. Strengths of the integrative-sociolinguistic trend

Testing techniques advocated under the integrative-sociolinguistic trend show some similarity with tests used under the pre-scientific trend. The more recent trend, however, requires tests to be objective, valid and reliable (although some concessions have to be made for tests to be practicable). The integrative-sociolinguistic trend in language testing encourages tests that are life-like and meaningful to the learner.

1.3.5. Weaknesses of the integrative-sociolinguistic trend

There are several problems involved in the design and administration of tests that conform to the objectives outlined above. Some of these problems are:

- The target of the tests, communicative competence, is still not clearly defined. Some researchers take it as synonymous to "native speaker competence", while others use it to indicate a "minimum" ability in the second language, others use it in yet another context.
- Tests evaluating global skills in a life-like situation are likely to be time consuming and expensive if a worthwhile level of objectivity, reliability and validity is to be achieved.
- Any testing situation is likely to be artificial. It is not clear, how learners' ability to handle a testing situation is related to real-life performance.
- Another difficulty is testing learners' linguistic reaction to disappointment, surprise, anger etc., as the motivation for communicating these feelings is artificial in testing situations.

The few problems outlined above serve to indicate that large-scale standardized tests (as developed under the influence of the psychometric-structuralist trend) are difficult if not impossible to develop within the framework of the integrative-sociolinguistic trend. In fact, the tenet that tests be based on learners' communicative needs indicates, that a large number of different tests has to be developed to serve a small number of learners (rather than a few tests to evaluate a large number of students, e.g. TOEFL, Michigan).

1.3.6. Diversification of research

The integrative-sociolinguistic trend in language testing has encouraged a diversification of testing research. The scope of

research extends to other areas of interest that concern not only language per se, but the wider concept of language acquisition. Oller (1976b) outlines three areas that are important in this context:

> "First, we may examine language tests qua tests; second, we may investigate learner characteristics using language tests as elicitation devices; third, we may focus attention on hypotheses about psycholinguistic and sociolinguistic constraints of various sorts on verbal sequences using language tests as research tools." (Oller 1976b:145, his emphasis)

Recent research on the nature of language proficiency in foreign/second languages indicates, that all the traditionally posited language skills are fundamentally related.

> "It seems that all human beings rather naturally attend to meaning in comprehending and producing discourse, and they are either incapable of or at least not good at attending to too much of anything else." (Oller and Perkins 1980:9)

Such statements provide further support for the integrative-sociolinguistic trend in language testing and its shift away from an atomistic approach towards language, that concentrates on form, towards an integrative or global one, that concentrates on meaning.

2. THE BASIC DICHOTOMY AND ITS SUBSEQUENT DEVELOPMENT

The purpose of this chapter is to show the evolution of the competence/performance dichotomy as postulated by Chomsky. A discussion of Chomsky's dichotomy is followed by a look at a number of reactions to his postulated theory. All but one of these reactions are critical of Chomsky's conception of competence. The theorists discussed in this chapter are primarily concerned with competence in terms of ideal or native speakers' knowledge. Except for Halliday, they all accept the distinction between competence and performance and they see these components as composed of various sub-components.

Linguistics has always been concerned with giving an account of language, but the emphasis tended to change under the influence of prevailing theories of language. From traditional, prescriptive works on what speakers of a given language should say, Bloomfieldian descriptivism turned to accounts of what people actually say. A further shift of emphasis occurred under the influence of cognitivism and the awareness that language is creative. The awareness that no individual language user would ever be able to use all the possible grammatical sentences in his language, stimulated an interest in underlying knowledge, i.e. the language user's potential to produce and understand these sentences. One of the problems with this view was to decide whose underlying knowledge a grammar should deal with, as language ability appears to be idiosyncratic even among native speakers of the same language.

2.1. Chomsky's competence/performance theory

Chomsky (1965) formally introduced the dichotomy to satisfy the methodological necessity of his approach of studying language in terms of idealized abstractions. He has stated:

"Linguistic theory is concerned primarily with an ideal speaker-listener..." (Chomsky 1965:3)

In Chomsky's view, all details of language behaviour are irrelevant and, accordingly, he ignores them. Although he considers his dichotomy related to Saussure's "langue-parole" distinction, he rejects the concept of "langue" as merely a systematic inventory of items. Instead, he proposes the "Humboldtian conception of underlying competence" (Chomsky 1964)[3]. Chomsky's dichotomy resembles that of Saussure in the sense that they both propose a distinction between what might be called shared vs. idiosyncratic knowledge of language users. Shared knowledge (competence in Chomsky's terms) takes precedence over idosyncratic knowledge in both theorists' views.

"The classical Saussurian assumption of the logical priority of the study of langue...seems quite inescapable". (Chomsky 1964:52, his emphasis)

The basic principle of the dichotomy is clearly stated in Aspects of the Theory of Syntax

"We thus make a fundamental distinction between competence (the speaker-hearer's knowledge of his language) and performance (the actual use of language in concrete situations." (Chomsky 1965:4, his emphasis)

Such a separation of language and language user operates on several distinct levels as will be seen below.

According to Chomsky, competence refers to the deep structure, i.e. the language user's knowledge of the grammatical rules underlying a given language. The study of competence concerns the study of psychologically real grammars. The language users' competence can be inferred from his performance and in turn be approximated to and compared with the ideal speaker-listener's competence which constitutes the competence model.

Performance can only be a direct reflection of competence under the idealizations outlined by Chomsky

> "...an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance." (Chomsky 1965:3)

It seems impossible then, that performance could directly reflect underlying competence. Natural speech is almost certain to show deviations from rules, memory lapses or similar interference, and an exact correspondence between linguistic rules of a language and the speaker's output is unlikely. The linguist trying to define competence has actual and potential utterances to work with. Although the basic formulation of the dichotomy given by Chomsky seems clear, some of the details are not specified and the researcher is left to interpret them. For example Chomsky has stated

> "...the actual observed language - actual performance - does not simply reflect the intrinsic sound-making connections established by the system of linguistic rules. Performance involves many other factors as well." (Chomsky 1967:397)

and

> "...in the study of grammar we abstract away from the many other factors...that interact with underlying competence to produce actual performance." (Chomsky 1966:9)

Such performance factors include perceptual and memory limitations, but the difficulty of differentiating between data that indicates the language user's competence and performance deviations is still unsolved. The distinction that competence is the concern of linguistics whereas performance should be considered the domain of psychology, does not clarify the problem at issue. A further distinction, that of considering "sentences" as lin-

guistic units, while relegating "utterances" to psychology as proposed by some researchers (e.g. Greene 1972) is of little practical help, In fact, Greene makes yet another distinction, i.e. between

- the abstract ability to produce a sentence, and
- the choice of a particular sentence on a particular occasion, usually due to a wide variety of situational variables.

This last distinction leads again to the domain of psychology, as no consideration is given to sociological factors by Chomsky at this point in time. Any investigation into the choice of utterances is likely to involve the motivation of human behaviour rather than actual language.

Greene (1972) pointed out that both linguists and psychologists are interested in explaining a speaker's ability to convey meanings that can be understood by other speaker-listeners of the same language. The extent to which linguistic and psychological accounts of competence coincide is difficult to assess, especially due to the fact that Chomsky's writings define competence in two different ways, one of these having more significant implications for psychology than the other. Again, interpretations vary. The most widely accepted interpretation maintained by linguists and psychologists (Greene 1972, Fodor and Garrett 1966, Campbell and Wales 1970, Canale and Swain 1979 amongst others) is to classify Chomsky's two definitions of competence into a) the weak version and b) the strong version.

a) Weak version

The weak version (according to Greene) defines competence as the language user's knowledge of the rules of grammar, syntax and phonology of a given language. Such competence is identified with mental ability or capacity. Performance here is the actual use (including false starts, memory lapses etc.) the language

user makes of his competence. This version is confined to the description of linguistic usage, i.e. of a generative grammar. It does not touch upon the internalization of this system of rules or operations, nor its output. The description would entail

- a set of rules capable of generating all grammatically correct sentences of a language, and
- structural descriptions that reflect the native speaker's intuitions about grammatical relationships.

The implication for a model of performance within the framework of Chomsky's weak version is that any output from a psychological model of the language user's behaviour could be put to the test by evaluating it against the linguistic model as pointed out by Greene (1972). In addition, the weak version allows for the theory that the set of rules providing the best descriptive account of a language user's intuitions is not necessarily linked with the set of operations whereby he himself arrives at these intuitions. Most linguists tend to accept this weaker version and its implications. Halliday is one exception and his point of view will be discussed in 2.2.

b) Strong version

The strong version equates a theory of competence with the linguistic rules that generate the grammatical sentences of a language. A theory of performance is equated with the interaction of the theory of grammar and the set of non-grammatical psychological factors relevant to language use. The characterization of linguistic abilities is simplified considerably by omitting any account of the role of memory or various low-level sensori-motor capacities involved in the production as well as the perception of speech. The focus is on the internalized rules of grammar in the cognitive system of the speaker-hearer. They are expected to provide the basis for his understanding of linguistic relations. Chomsky has stated:

"A person who has learned a language has acquired a system

> of rules that relate sound and meaning in a certain specific way. He has, in other words, acquired a certain competence that he puts to use in producing and understanding speech." (Chomsky 1970:184)

He has further claimed that since the ability to learn to speak a natural language is basic to human intelligence, the rules underlying language learning must be characteristic of the way the human mind works.

> "At the level of universal grammar, the linguist is trying to establish certain general properties of human intelligence. Linguistics so characterized, is simply the sub-field of psychology that deals with these aspects of mind." (Chomsky 1972: 28)

In summary, the weak version is limited to a description of what the linguistic usage of the native speaker-hearer consists of. The strong version, however, hypothesizes as to how the native speaker-hearer operates when using language. According to the strong version, the rules of transformational grammar are representative of the way in which the human mind functions. If this claim is accepted, any contrary findings are likely to result in a radical change in the form and organization of the competence hypothesized to underly language use. Most linguists reject this strong version as it is thought to be too restrictive (for a discussion of criticisms see 2.2.)

Kempson (1977), interpreting and elaborating Chomsky's strong version of competence, has equated the study of the linguistic regularities of a language with a theory of competence. Pragmatic theory, in Kempson's framework, is part of a theory of performance (Kempson 1975:220). She has further claimed that

> "...a theory which characterises the interaction between that linguistic characterisation and all the other factors which determine the full gamut of regularities of communication is

> a theory of performance...a theory characterising a speaker's ability to use his language appropriately in context, a theory of communicative competence, is, simply, a performance theory." (Kempson 1977:54-55)

Kempson's use of the term "communicative competence" differs therefore considerably from the way Hymes uses it (discussion in 2.2.)

2.2. From competence to communicative competence

The competence/performance distinction Chomsky proposed has had considerable impact not only in linguistics but also in the fields of psychology, philosophy and sociology. Most of the critics feel that Chomsky's dichotomy is either too restricting or does not consider some or other aspect it should include.

Fodor and Garrett (1966), writing on the distinction between competence and performance, accepted the weak version of the dichotomy. However, they criticized what they called Chomsky's "competence plus" view, i.e. the definition of performance being simply the theory of competence without the idealization. They stated that

> "...in the sense in which distinguishing between performance and competence is distinguishing between behaviour and the mechanisms underlying it, both linguistic and psychological models are models of competence." (Fodor and Garrett 1966: 138)

In their opinion, a satisfactory explanation of the relationship between competence and performance will need to represent that relationship as abstract,

> "...the degree of abstractness being proportional to the failure of formal features of derivations to correspond to performance variables." (Fodor and Garrett 1966:152)

Contextual information is however not included in the type of grammar motivating the empirical studies they discuss. This aspect is very important to other researchers, such as Campbell and Wales. They are primarily concerned with the psychological aspect of the language acquisition process. They accept the methodological distinction proposed by Chomsky. However, they feel that Chomsky's framework fails to allow for what they consider

> "...the most important linguistic ability...the ability to produce or understand utterances which are not so much grammatical but, more important, appropriate to the context in which they are made. (Campbell and Wales 1970:247, their emphasis)

Accordingly, the language user's ability to produce and understand all grammatical sentences and the operations involved in the selection of a particular sentence are not the only aspects to be considered. His ability to make that selection, taking into consideration contextual factors, is equally or more important. Campbell and Wales' definition of "context" includes both the situational and the verbal context of utterances. They propose a broader notion of competence, intended to include linguistic competence (i.e. knowledge of the rules of grammar) as well as knowledge of the rules of language use. Although Campbell and Wales do not explicitly confirm this, they seem to accept the concept of performance as the actualization of their definition of competence.

Habermas (1970), a social theorist, has developed the concept of communicative competence from a philosopher-semantician's point of view. His theory of communicative competence is highly idealized and abstract. He has equated a theory of communicative competence with "the mastery of an ideal speech situation". While accepting Chomsky's basic distinction between competence and performance, he has criticized Chomsky's theory on a semantic as well as on a sociolinguistic level. On a semantic level, he has questioned the theory's claim that universal meaning is rooted at a deep

structure level, therefore precluding all experience. He has further postulated the existence of "a priori" and "a posteriori" semantic universals. He has argued that some semantic universals are monologic, others are intersubjective (as opposed to Chomsky's view of language as a self-contained system). Figure 2.1., based on Habermas' (1970) representation of the four resultant classes, summarizes his views of the organization of semantic universals.

On a sociolinguistic level Habermas has rejected Chomsky's theory for its narrow perception of competence. It fails to take account of the essential dimension of communication (in a highly idealized sense). The intersubjectivity of language entails the admission of sociological content into the notion of communicative competence. He has further argued that a communicative situation depends on a structure of intersubjectivity which is linguistic. Such a structure is generated neither by linguistic rules in a monological system nor by the extralinguistic factors of performance. Accordingly, communicative competence should be defined in terms of "the basic qualifications of speech and of symbolic interaction" (Habermas 1970) the language user must have at his disposal in addition to his linguistic competence in order to partake in normal discourse.

Habermas' conception of communicative competence requires an ideal speaker, who has complete mastery of the dialogue-constitutive universals without taking into consideration actual restrictions under empirical conditions. Although Habermas' emphasis on the contextualized intersubjective nature of language should not be ignored, his concept of communicative competence is extremely idealized and unlikely to find application in language testing.

Jakobovits (1970) has rejected Chomsky's dichotomy because it only accounts for language usage and ignores factors associated with language use. He has proposed a framework for communicative competence that contains four different components:

Figure 2.1. Habermas' (1970) organization of semantic universals

Semantic universals	...which process experiences "a priori"	...which make this processing possible "a posteriori"
...which are linked to the condition of potential socialization "intersubjective"	dialogue-constitutive universals (e.g. personal pronouns, imperative, negative)	cultural universals (e.g. the system of kinship words)
...which precede all socialization "monologic"	universal cognitive schemes of interpretation (e.g. causality, substance, space, time)	universals of perceptive and motivational constitution (e.g. the system of colour words)

(based on Habermas 1970)

- a paralinguistic component
- a kinesic component
- a sociolinguistic component
- a psycholinguistic component

to account for the variables involved in language use. The fact that Jakobovits has excluded grammatical competence from communicative competence necessitates the retention of linguistic competence and performance in his framework, resulting in a double dichotomy:

linguistic competence	-	linguistic performance
communicative competence	-	communicative performance

From this it could be concluded that grammatical and communicative competence are two separate factors of "knowing a language" that exist independently. However, Jakobovits maintained in his discussions on language teaching that linguistic and communicative competence should not be taught separately, nor necessarily in that order.

Hymes, primarily an anthropologist and sociologist, was interested in the interaction of language and social setting (1967). Later (1972) he discussed language in the context of a heterogeneous speech community. In Hymes' theory, communicative competence seems to be equated with native speaker competence, especially with native speakers' ability to produce and understand utterances appropriate to the context. He has pointed out

> "...there are rules of use without which the rules of grammar would be useless" (Hymes 1972:278)

As an example Hymes has discussed a hypothetical child who has the ability to produce and understand any of the grammatical sentences of his native language. Such a child, producing any sentence (and possibly speech and silence too) at random, would likely be institutionalized. He has concluded that these factors must conform to specific rules, rules that should be accounted for in a theory of communicative competence. This is further

supported by the fact that in normal first language acquisition, grammatical accuracy and situational appropriateness are acquired together. It is therefore not sufficient to consider language use as a part of performance only and Chomsky's dichotomy should, according to Hymes, be revised accordingly. Hymes has redefined the original notion of competence and developed communicative competence to allow for the above factors. In his framework, communicative competence includes four grammatical and behavioural components reflected in the judgements and abilities of language users:

Grammatical

Whether and to what degree something is formally possible (something possible within a formal system is grammatical, cultural and on occasion, communicative).

Psycho-linguistic

Whether and to what degree something is feasible (within the means of implementation available)

Socio-cultural

Whether and to what degree something is appropriate (in relation to the context in which it is used and evaluated)

Probabilistic

Whether and to what degree something is in fact done, (i.e. something might be possible, feasible and appropriate and not occur and, if it does occur, what its occurrence entails).

This fourth component is very important as native speakers are able to anticipate and predict with remarkable accuracy what will be said in a given context.

Hymes' framework of communicative competence, then, refers to the interaction between knowledge of the rules of grammar (grammatical competence) and knowledge of the rules of language use. He has proposed to treat general psycholinguistic factors (such as memory, perceptual strategies) as aspects of communicative

competence (as did Jakobovits 1970). Performance is the language user's actual use of this communicative competence in the course of actual events.

Halliday represents the ideology of the British School (born out of Firthian philosophy and the influence of the linguists of The School of Prague). He is concerned about the way in which language is used in social interaction and how it varies in accordance with social function. Halliday (1970, 1973) has rejected Chomsky's competence/performance distinction. He has, however, agreed that there is a difference between what a language user knows in a language and what he can do with that knowledge. His rejection of Chomsky's dichotomy is twofold:

- He considers it unnecessary if it is intended to differentiate between what we can and what we cannot describe in a grammar.
- He considers it misleading if it serves to restrict the data considered.

Chomsky (1976a) has conceded that theoretical assumptions regarding relevant vs. not relevant data can be dangerous and should not be accepted as dogma to the exclusion of other research lines.

The approach Halliday has advocated (1973, 1978) in place of a competence/performance model centres around the notion of "meaning potential". He has defined this as sets of options that are available to the language user. He has proposed a system relating meaning potential to behaviour potential consisting of three stages:

- what a speaker can do (behavioural realization)
- ↓
- what a speaker can mean (semantic-meaning options)
- ↓
- what a speaker can say (grammatical options)

The direction of influence from level to level indicates that in Halliday's theory grammatical options are the direct realization of semantic options. As Canale and Swain (1979) have pointed out, this concurs with one of the axioms of modern linguistics, i.e. any natural language can express any meaning in some way or other. It may however be difficult to maintain this in the context of learners of a second or foreign language. At the beginning of language learning, meanings and behavioural realizations are likely restricted by what the language user can say. Many applied linguists (for example Canale and Swain 1979, Morrow 1977, Johnson 1977, Wilkins 1976) seem to maintain this view.

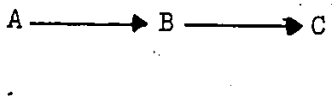
Halliday's claim that semantic options are the realization of behavioural options is somewhat less clear as it could hardly be maintained that one is only able to express semantically what social conventions permit. It would seem that social conventions permit to some degree what can be expressed semantically (for example, the advent of modern technology entails new terminology), but examples for the situation where the opposite holds can also be found. Halliday's linear progression is further put in doubt if one considers the possibility of using language with little or no reference to the social context, such as organizing one's thoughts and in creative language use (Chomsky 1975, Canale and Swain 1979). A possible solution would seem to be in a circular approach, where each option is allowed to draw on and enhance the other, as illustrated in Figure 2.2.

Whereas Hymes (and other sociolinguists and anthropologists, for example Gumperz) is primarily concerned with the incorporation of the social meaning in a framework of communicative competence, Halliday is equally concerned with the importance of the referential meaning of language, the ability

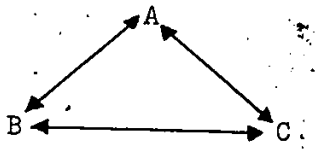
"...to communicate something...to represent our experience

of the processes, persons, objects, abstractions, qualities, states and relations of the world around us and inside us." (Halliday 1970:145-146)

Figure 2.2. Halliday's meaning potential relationship - proposed change



Halliday (1973)



Proposed change

2.3. Summary and implications for testing

The theories of communicative competence that have evolved from Chomsky's dichotomy up to this point have tended to focus mostly on one or the other area, i.e. linguistic, psycholinguistic semantic, depending on their proponents' main interest. Figure 2.3. illustrates this point.

Chomsky's dichotomy was not designed to and cannot handle the sociocultural dimension that emerges as necessary to any study concerned with the communicative aspect of language. The type of theory Chomsky has advocated is devoid of any attempts to explain the use of language in sociocultural contexts. As a result, it only serves to support the kind of testing advocated under the influence of structuralism, i.e. discrete-point testing of forms. Even modifications to Chomsky's early thinking left the grammar as a kind of mechanism, most closely related to structural theory, and far removed from communication and the social context in which it takes place. Diller (1971) has

Figure 2.3. Summary of theorists and their main concern

	linguistic	psycholinguistic	sociolinguistic	socio-philosophical	sociosemantic
Chomsky (1965)	X				
Kempson (1975 & 1977)	X				
Fodor and Garrett (1966)	x	x			
Campbell and Wales (1970)		X			
Habermas (1970)				X	
Jakobovits (1970)		x			x
Hymes (1972)			X		
Halliday (1973)					X

reported an increased trend towards pattern drill and memorization based tests which he has considered to be a direct result of Chomsky's theory.

Campbell and Wales have maintained that communicative competence is central to the process of language acquisition and as such its definition is the goal of the psychology of language. This emphasis on the psycholinguistic features of environmental factors has led them to neglect the sociolinguistic features to a great extent. Habermas, on the other hand, is concerned primarily with ideal speech situations. Like Chomsky, he has avoided all constraints imposed by the real world. This results in a decontextualized view of communicative competence.

Jakobovits' theory is in some ways similar to that of Hymes. However, he has stressed the contrast between the various aspects of a language user's communicative competence, when, in fact, other researchers seem to be interested in their relationship. Munby (1978) has pointed out that it is more important to stress relationships than pedagogically less relevant contrasts between grammatical and the sociocultural and psycholinguistic aspects of a speaker/listener's competence.

Hymes is mainly concerned with the language user's competence with regard to judgements and abilities related to or interdependent with sociocultural features. Hymes includes discourse within the framework to be covered by a theory of communicative competence, although he does not develop this area within the model he proposes.

At a lower level of idealization, Halliday's theory is similar to Hymes' in its concern with language use. At the highest level of idealization, however, Halliday's language functions are concerned with similar issues as Habermas' theory of communicative

competence. Halliday seems to cover a smaller area of data in greater detail than Hymes, but his analysis of language in use is restricted to the level of the clause.

As the preceding discussion shows, the dichotomy proposed by Chomsky has undergone considerable modifications. A great deal has been written about communicative competence and it is not always clear, what precisely is included in a definition of the term. Furthermore, the same term is likely used to describe very different phenomena. Rivers (1973:26), for example, considers communicative competence as a synonym for "spontaneous speech", while other researchers equate it with native speaker fluency. The definition of the basic framework is of considerable importance since it determines the scope of subsequent tests. In River's (1973) sense of the term, a test would concentrate on students' speaking ability only. In a framework such as postulated by Hymes, a test would concentrate on appropriateness of utterances.

3. TOWARDS A COMPREHENSIVE THEORY OF COMMUNICATIVE COMPETENCE

The purpose of this chapter is to further trace the development of Chomsky's initial competence/performance dichotomy and to investigate what it signifies in the work of different researchers. This chapter is separated from the previous chapter for two reasons:

- researchers discussed in chapter 3 tend to be concerned with how the dichotomy relates to actual speakers of a language rather than with ideal speakers
- actual speakers are usually thought of in terms of second or foreign language learners.

The group of researchers discussed in this chapter also includes a brief discussion of Oller's theory, which does not use the competence/performance distinction as such.

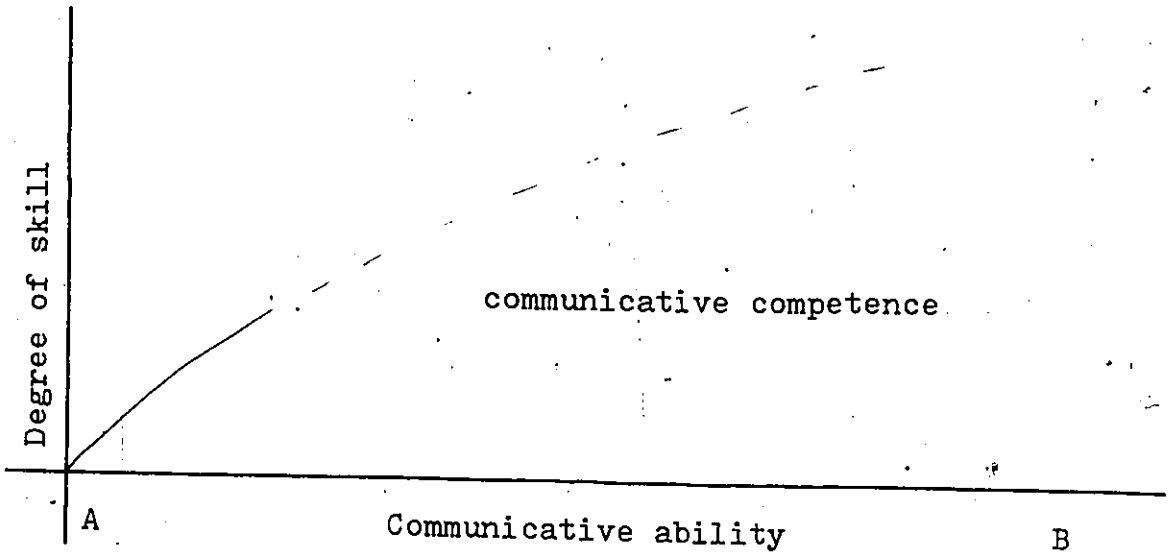
3.1. The minimum theory

Minimum theories, such as proposed by Savignon (1972), Van Ek (1976) amongst others, characterize communicative competence as the minimum knowledge required of a learner to enable him to interact successfully with native speakers of the target language. "Minimum", however, is subjective and difficult to quantify. It could consist of a few expressions found in a bilingual dictionary dealing with the two languages in question. Such interaction would, undoubtedly, be unsatisfactory for both parties involved and might be referred to as groping, rather than coping. Figure 3.1. illustrates the relatively uncertain position of communicative competence in relation to degree of skill and ability to communicate.

At point A, two speakers of different languages, whose conversation attempts are mutually unintelligible, might be able to exchange a few expressions thanks to a bilingual dictionary. One

of the two speakers, G, starts to learn speaker Y's language. As G progresses in his studies, he becomes increasingly able to cope in Y's language. The minimum ability theory tries to define what G needs to express his thoughts and what G needs to understand that which Y says to him, without breakdowns in conversation caused by G's incompetence. In this type of theory, communicative competence is not equated with native speaker competence.

Figure 3.1. Levels of communicative competence



Savignon (1972) described a study that is often referred to as a minimum theory. She defined communicative competence as

> "...the ability to function in a truly communicative setting - this is in a dynamic exchange in which linguistic competence - must adapt itself to the total information input, both linguistic and paralinguistic, of one or more interlocutors." (Savignon 1972:80, S's emphasis)

Apart from this definition, Savignon did not define an operational theory of communicative competence, nor did she include a description of communication in the study. The emphasis was on skills needed to get the speaker's meaning across, to enable

the speaker to say what he really wanted to say. Savignon did not elaborate a model of communicative competence. What she considered "the most common language situations a learner is likely to encounter" were chosen on the basis of intuition and need (i.e. depending on how the conversation during classtime developed and what appeared to interest her students). Some of the language situations chosen for their usefulness were: greetings, leave taking, information getting and giving, accepting, refusing and extending invitations, describing and reporting. Savignon considered linguistic competence less important than the ability to understand what was being said and the ability of the students to make themselves understood. (At the end of the experiment students expressed some frustration about their linguistic shortcomings, i.e. they felt lack of knowledge of structures and vocabulary reduced their ability to express themselves.) Accordingly, the criteria for evaluation included only comprehensibility and semantic accuracy in the "code" area, indicating a concern for pronounciation and vocabulary rather than structure. Savignon's tests took into consideration suitability of utterances as well as kinesic and paralinguistic features such as facial expressions and gestures.

Savignon's (1972) definition of communicative competence seems to exclude the written aspect of language. This might be due to a conception of communication as being only spoken or it might be due to time constraints imposed on her study. Savignon's study disregarded how individual utterances are linked at the level of discourse. Due to the lack of theoretical foundations, Savignon's study cannot be considered a theory of communicative competence. Furthermore, the lack of systematically defined selection criteria for content to be taught precludes it from being a model for communicative competence. It is, as Savignon indicated in the title of her book, an experiment in foreign language teaching.

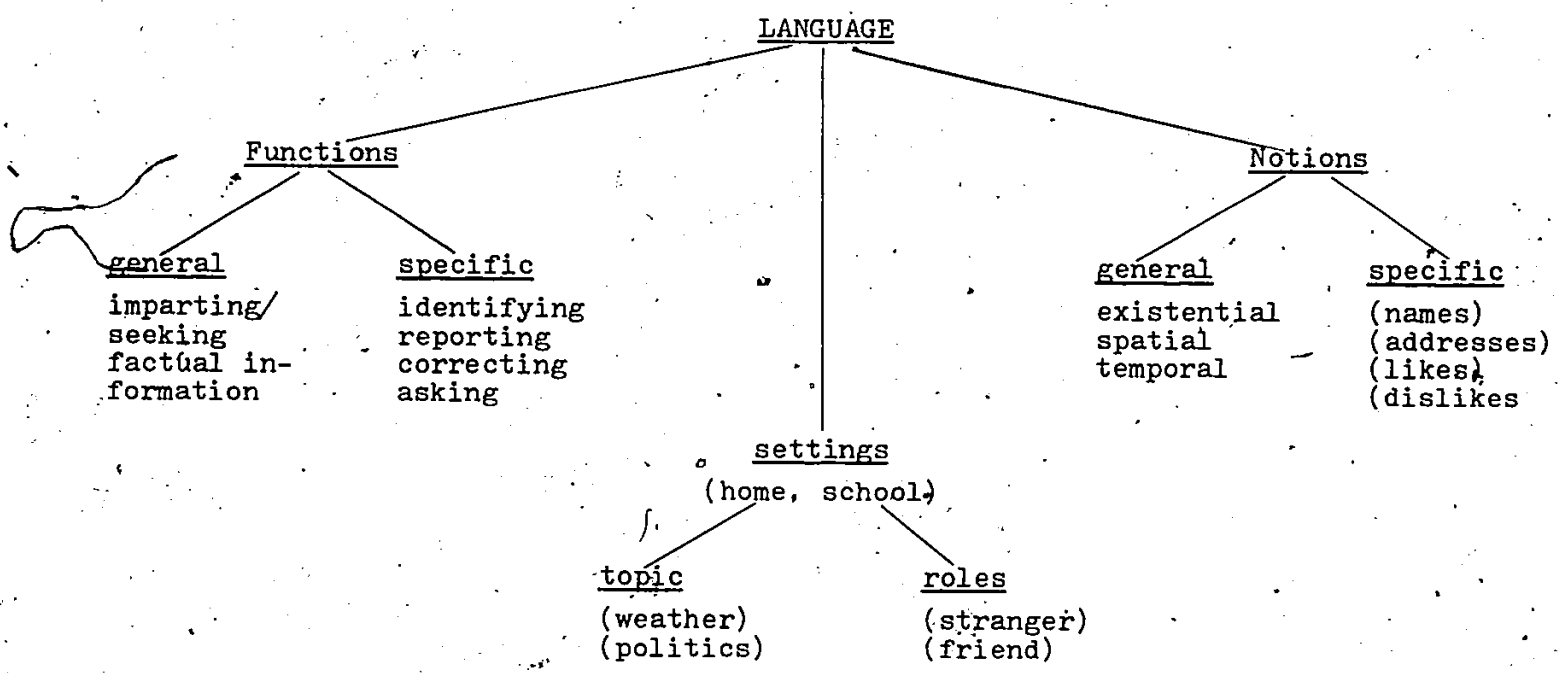
Van Ek (1976) attempted to define a "threshold level" that would enable learners to interact with native speakers of the target language. Van Ek's model is not meant to be language specific, although he offered a model for learners of English as a foreign language. Van Ek (1976) expected his "minimum objective for the teaching of (mainly oral) foreign language communication" to be only part of a more comprehensive foreign language curriculum. As such, minimum ability is considered a skill, not knowledge.

Van Ek's model stresses the importance of "language functions" (such as imparting or seeking factual knowledge, expressing and finding out emotional or intellectual attitudes) and "notions" (such as existential, spatial, temporal). His model, accordingly, includes extensive lists of such content-specifications. Grammatical forms required to express these functions and notions are of secondary importance. As in Savignon (1972), the emphasis is on getting meaning across. Van Ek further specified individual situations in terms of learner's roles, settings and topics as well as in language activities (speaking, writing, listening, reading), functions and notions. His multi-dimensional semantic-pragmatic inventory is intended for syllabus development with adult learners in a European context in mind. It could, however, be adapted to suit other requirements.

Figure 3.2. summarizes Van Ek's (1976) model by showing the main factors involved in the specifications he proposed. The particular inventories of vocabulary, structures and grammatical categories are based on which of these factors are involved in a specification of the learner's needs.

The main criterion for the purpose of evaluation in Van Ek's model is whether communication takes place. The listener who is unable to understand what is said to him and the speaker who is

Figure 3.2. Summary of Van Ek's (1976) model



unable to make himself understood have not reached the objective of communicating. A second criterion in the above framework is the degree of efficiency with which communication takes place. A speaker's efficiency in communicating can be determined by judging speaking speed, precision and linguistic accuracy (including grammatical, lexical and phonetical considerations) according to Van Ek (1976). A listener may also be judged on his ability to understand text spoken with standard or near-standard pronunciation, at normal speed and without repetitions. A language user who has to ask for constant repetition of what is being said cannot be considered a successful listener.

Van Ek's (1976) model does not include rules of language use and how they relate to the appropriateness of utterances, nor does it consider how individual utterances are linked at the level of discourse. It is simply an inventory for needs specifications, although it implies that certain sociolinguistic factors are important in determining language use. Van Ek's inventory is more structured and presented in a more formal way than Savignon's (1972) study, but their main concern is the same, namely getting meaning across.

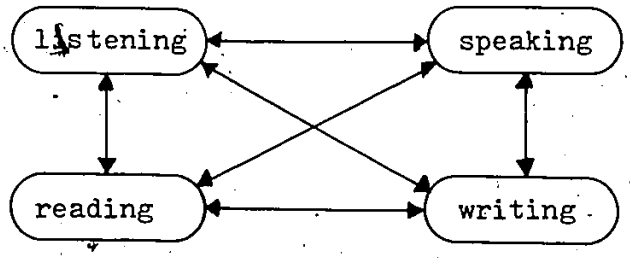
Evaluation criteria depend on a comprehensive definition of the known factors involved in communicative competence. A definition of communicative competence should, therefore, be as comprehensive as possible. A comprehensive or integrative (Canale and Swain 1979) theory of communicative competence would, ideally, explain the relationship between linguistic rules and rules of language use as well as how and to what extent context determines the selection of vocabulary, utterances and structure.

3.2. Comprehensive or integrative theories

Recently, several more comprehensive theoretical frameworks

of communicative competence have been proposed. They take into consideration the learner's ability to synthesize his knowledge of grammatical rules and how this knowledge is used in social contexts. Furthermore, they take into account his knowledge of rules of discourse underlying communication, i.e. how messages are to be combined to form units of text (written and spoken). These theories can further be considered integrative in that they do not subdivide listening, speaking, reading and writing into individual skills. They treat these skills as interrelated, the level of competency of each skill being dependent, to a greater of lesser degree, on the level of competence reached in the others (Figure 3.3.)

Figure 3.3. Interrelationship of individual skills



Munby (1978) proposed a framework of communicative competence that consists of three major components:

- a sociocultural component
- a sociosemantic view of linguistic knowledge
- rules of discourse

The sociocultural component is based on the importance of contextual appropriacy as pointed out by Hymes (1972), Cooper (1968) and Widdowson (1978) amongst others. The sociocultural orientation

focuses on the social function of language which in turn requires a learner-centered approach to language teaching, which would be based on the language needs of the learner. But, in order to know what to teach, the learner's needs have to be specified carefully. Speech functions and communicative acts can be selected accordingly which, in Munby's framework, will justify the selection of particular skills to be emphasized and linguistic forms to be taught.

The sociosemantic view of linguistic knowledge is based on Halliday's concept of meaning potential, that draws attention to the sociosemantic basis of linguistics. In Munby's theoretical framework the concept of meaning potential also indicates the central role of the semantic options available to a language user in that they translate behavioural options into options in linguistic form.

Munby postulated the component of rules of discourse to account for the language user's ability

- to use linguistic forms to perform communicative acts, and
- to understand the communicative purpose of utterances and their relationships to other utterances.

The theoretical framework for this component is based on Sinclair and Coulthard (1975) who postulated the discourse level of operation as the mediator between the levels of grammar (syntax, lexis, phonetic) and the rest of a speech event or situation as it implicates linguistic form as well as social norms. Munby's rules of discourse include:

- Widdowson's (1975) knowledge of rhetorical rules of use which govern patterning of communicative acts
- Candlin's (1976) contextual meaning of an utterance, i.e. how its value is affected by what precedes it and what follows it
- Widdowson's (1978) interpretative strategies of language users, i.e. the psychological processes that underly -

the overt language activities such as speaking.

Munby's (1978) model includes extra-linguistic knowledge and abilities of a language user in communicative competence. However, except for a brief discussion of the non-verbal aspect of communication, the area of extra-linguistic knowledge is not discussed.

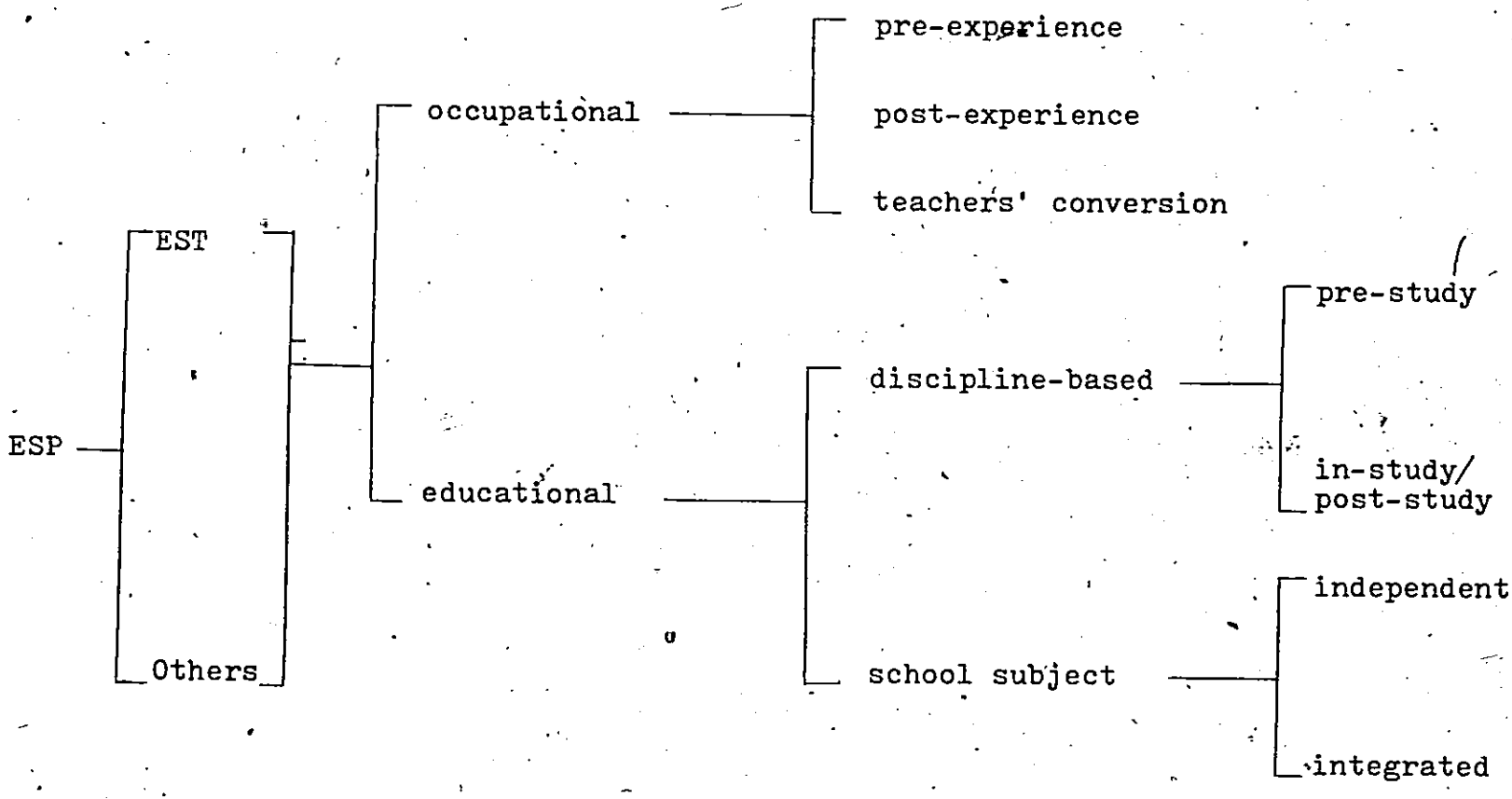
After having specified a theoretical framework for communicative competence, Munby (1978) elaborated a very detailed and exhaustive model for identifying communicative needs of learners. In his proposed "Communication Needs Processor" account is taken of eight variables he isolated as affecting communicative needs. There is, for example, the purposive domain (i.e. the purpose for which the second/foreign language is needed), broken down into micro-components as Figure 3.4. illustrates. The figure shows how the type of ESP (English for Special Purposes) required is determined, i.e. depending on whether "purpose" is occupationally or educationally motivated. The next step in Munby's model establishes precise occupational (or educational) purposes, for example:

- specific occupation
- central duty of that occupation
- other known duties
- occupational classification (specifies type of worker within larger field of work).

Similar detailed specifications are established for the remaining seven variables in the Communication Needs Processor. The results for each of the eight parameters are written out as the profile of communication needs for that specific learner. The profile of communication needs is further processed to generate the communicative competence specification for the learner or group of learners. It seems, however, questionable that the learner's language needs can (and indeed need to) be specified to such a degree. Language needs are affected by a large number of variables, i.e. the learner's needs might change at

Figure 3.4. Classification of ESP (based on Strevens 1977, quoted in Munby 1978:56)

(used to determine type of ESP required by learner)



short notice and as a result, the projected language needs would have to be adjusted. It seems questionable, that a highly detailed model would be flexible enough to adapt quickly to accommodate the necessary changes. Another point to be considered is that Munby's model takes into consideration future language needs a learner is expected to have. While this seems to be a valid theoretical goal, it does at the same time make it difficult for the learner to exploit potential opportunities to practice communication while learning. For example: many non-English speaking students attend English language programs at various institutions in preparation of a career or further education in English speaking countries. For many of them, specifications of "needs" would not include basic socializing skills required to interact with peers who speak the target language. As a result, they are not taught and thus maybe unable to make contact with speakers of the target language. They meet with members of their own linguistic group and, consequently, they miss a valuable opportunity to experience and practice their target language while in the process of learning it. Munby's language processor, where the learner's needs are entered at one end in order to obtain the user's needs (the specifications of communicative competence) might have to be adjusted to be able to account for learner-incurred needs.

Furthermore, Munby's theory ignores what is generalizable from situation to situation and thus of use to most learners (except those with highly idiosyncratic language needs), i.e. the linguistic code.

Widdowson (1978), whose view of communicative competence is based essentially on Hymes' theory with the addition of rules of discourse, considered two main components necessary for the realization of language use:

- linguistic skills (listening, reading, speaking, writing, used in reciprocal or nonreciprocal activities), i.e.

those activities which operate only on what is verbally manifested and which require only a knowledge of correct grammatical usage,

- communicative abilities, which "operate on everything that is communicative in the discourse as a whole" (Widdowson 1978:73), i.e. purely linguistic elements, non-verbal elements, illocutionary acts, rhetorical conventions and other pragmatic factors.

Communicative abilities embrace linguistic skills in Widdowson's proposed framework, but the opposite does not apply. Of major importance in Widdowson's theory is the learner's ability to interpret discourse, as mentioned in the discussion of Munby's discourse rules. Widdowson argues that discourse is not determined on the basis of linguistic cohesion, but on coherent interpretation of propositional and illocutionary features of discourse. He uses the following example of spoken discourse to clarify this point, but the principle applies equally to written discourse:

A : That's the telephone (request: Can you answer it, please?)

B : I'm in the bath (excuse: No I can't answer it because..)

A : O.K. (acceptance of excuse: I'll answer it)

The example does not include any cohesive devices such as conjunctions, to show direct links between the sentences. For satisfactory interpretation, knowledge of coherence is necessary. The setting of this dialogue is most likely the home environment (or hotel room etc.) of two people who know each other very well. A change of setting or a change of role of the participants in the above example would affect its subsequent interpretation considerably and might, in some cases, make it impossible.

It is still unclear, how rules of discourse need to be formalized in order to help the language learner. The study of discourse is considered to be in an embryonic state and at least some of the many different areas currently under investigation are likely to help clarify this aspect. A detailed study of the

various aspects involved in discourse analysis is beyond the scope of this paper. It should however be pointed out that a collection of research papers presented by Freedle (1977) indicates a large variety of research perspectives assumed and pursued in the area of discourse analysis. New insights in this area can therefore be expected and they are likely to influence the course of future developments in the area of language acquisition studies.

Canale and Swain (1979) propose a very comprehensive framework for communicative competence. It seems that they do not view communicative competence as equal to native speaker competence, although it appears to be somewhat higher on the skill/ability curve (Figure 3.1.) than in Savignon's (1972) and Van Ek's (1976) proposals. (However, Canale and Swain do not clarify this point in their study.) They propose a framework consisting of three competencies:

- grammatical competence
- sociolinguistic competence
- strategic competence

The grammatical component includes

> "...knowledge of lexical items and of rules of morphology, syntax, sentence-grammar semantics, and phonology." (Canale and Swain 1979:60)

The above component represents the language user's knowledge of the code. It includes rules of phonology, syntax, morphology, vocabulary and sentence-grammar semantics. This component represents the language user's knowledge that allows him to identify and express literal meaning of utterances. It is however not clear from Canale and Swain's discussion if their grammatical component is expected to describe or explain the language user's tacit knowledge of the grammatical rules of a language.

The sociolinguistic component includes both sociocultural rules of use and rules of discourse. The sociocultural rules

refer to appropriateness of utterances within a given context as well as to appropriateness of a chosen grammatical form to express register and attitude in a particular sociocultural setting. Rules of discourse are seen as the links that render groups of utterances both cohesive and coherent. Canale and Swain (1979) point out that rules of discourse are not likely to differ substantially from grammatical rules as far as cohesion is concerned and from sociocultural rules in the sense of coherence. They contend that, within their framework, the emphasis of rules of discourse is on

> "...the combination of utterances and communicative functions and not the grammatical well-formedness of a single utterance nor the sociocultural appropriateness of a set of propositions and communicative functions in a given context." (Canale and Swain 1979:62)

Classifying rules of discourse together with sociocultural rules might prove to be most efficient, although it could be argued that those rules of discourse relating to cohesion ought to be part of the rules of grammar. A more satisfactory solution might be found as the study of discourse progresses.

Canale and Swain (1979) add the component of strategic competence to their framework for a theory of communicative competence. Strategic competence allows for systematic verbal and non-verbal communication strategies speakers of a language (both native speakers and learners of a given language) call on to compensate for breakdowns in communication. Such breakdowns, Canale and Swain argue, may occur due to limited grammatical or sociocultural competence or performance variables. A language user's strategic competence relates to either grammatical competence (in the sense that the language user knows how to paraphrase structures or lexical items not mastered or momentarily unavailable in his memory) or sociolinguistic competence (how to address a person whose name is not known). Native speakers of a given language tend to use their strategic

competence automatically. The learner has to acquire the ability to bridge possible deficits, a fact that is frequently ignored in current language teaching and testing materials.

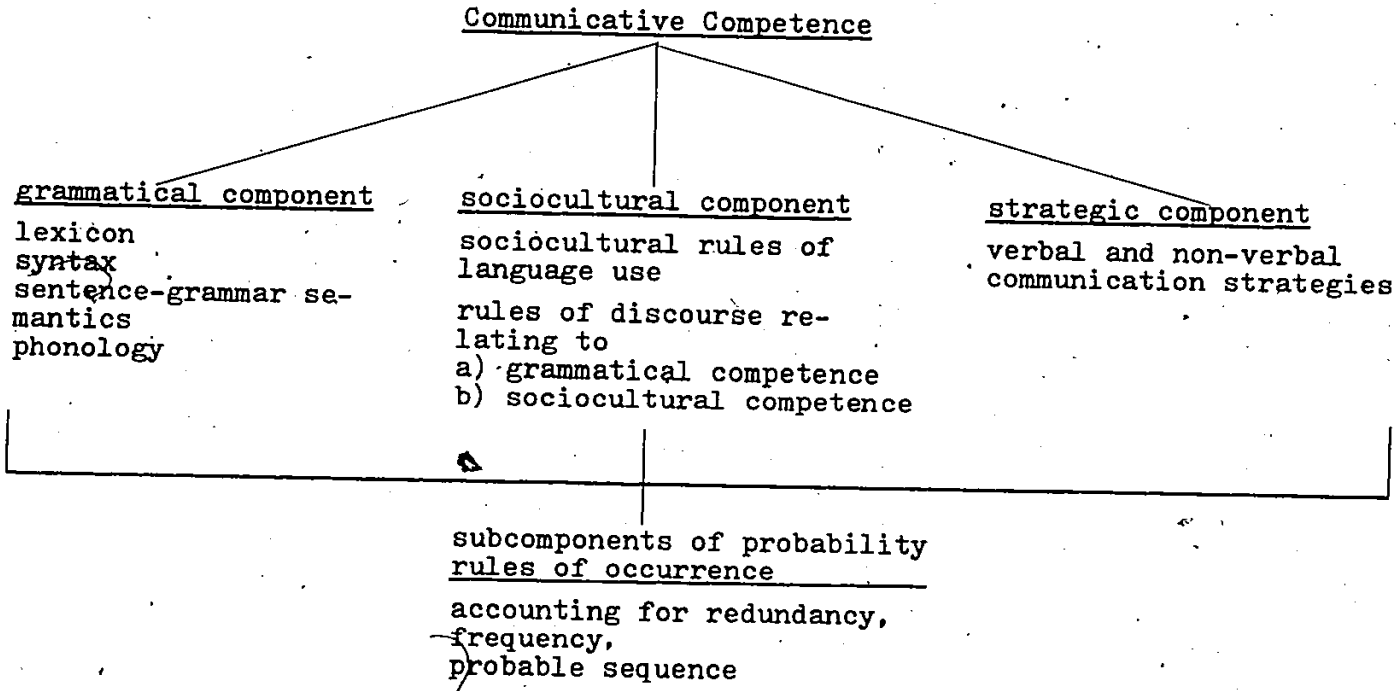
Canale and Swain (1979) also expect subcomponents, the probability rules of occurrence, within each of the above three components. The subcomponents would attempt to account for the redundancy aspect of language as postulated by Spolsky (1968). They account for the fact that the native speaker of a given language not only has the ability to determine the relative frequencies and probable sequences of sounds within a word, words within an utterance (grammatical competence), but also the probable sequences of utterances within a discourse (sociolinguistic competence) and the probable effects of certain strategies in specific situations. Labov (1972) claimed that the frequency of a given rule of grammar is conditioned by a combination of various features of the grammatical and sociolinguistic contexts. He discussed a number of proposals for possible formalizations of probability rules but did not reach a definite conclusion. The form of such rules has yet to be determined.

The framework of communicative competence proposed by Canale and Swain is the most integrative and comprehensive available for the purpose of this summary. (Figure 3.5. summarizes its components.) However, several questions remain (some of these were pointed out above) and will have to be resolved in the light of further research.

3.3. A grammar of expectancy

Oller (1970, 1973a, 1976, 1978, 1979a) avoids the competence/performance dichotomy proposed by Chomsky. He also avoids reference to the terms "sociolinguistic factors" and "communicative competence" as proposed by Hymes. Oller proposes a system in

Figure 3.5. Graphic representation of Canale and Swain's (1979) Theory of Communicative Competence



A component of performance is also proposed in the above framework. It is intended to represent the actualization of the above competencies.

which grammar is understood as a theory of language processing. As such it refers to native speakers of a language. This type of grammar is characterized as a "pragmatically-generated expectancy device" (Oller 1978:39). Oller has claimed repeatedly that adequate tests need to activate the language learners' internalized expectancy grammar. A brief discussion of Oller's conception of pragmatics and the expectancy grammar he postulates seems useful in order to relate his work to that of researchers working within the framework of communicative competence.

Pragmatics derives from the study of formal systems in which syntax is concerned with the structure of expressions. Semantics studies the meaning of expressions within the system while pragmatics deals with the relationship between expressions in the formal system and anything else outside of it (Ingram 1978). Oller (1970) has explained that

> "Pragmatic facts of language are those having to do with the relations between linguistic units, speakers and extralinguistic facts." (Oller 1970:99)

and he has further elaborated

> "It [pragmatics] is concerned with the relationship between linguistic contexts and extralinguistic contexts. It embraces the traditional subject matter of psycholinguistics and also that of sociolinguistics." (Oller 1979a:19)

According to Oller, these "pragmatic facts" have to be built into any account of linguistic contexts because he does not believe that language can profitably be studied as a self-contained system. In this sense Oller's view of pragmatics is quite similar to Hymes' idea of communicative competence (and contrary to Chomsky's conception of language as a monologic system). Hymes' framework, however, is essentially a componential one, i.e. communicative competence consists of several sub-components,

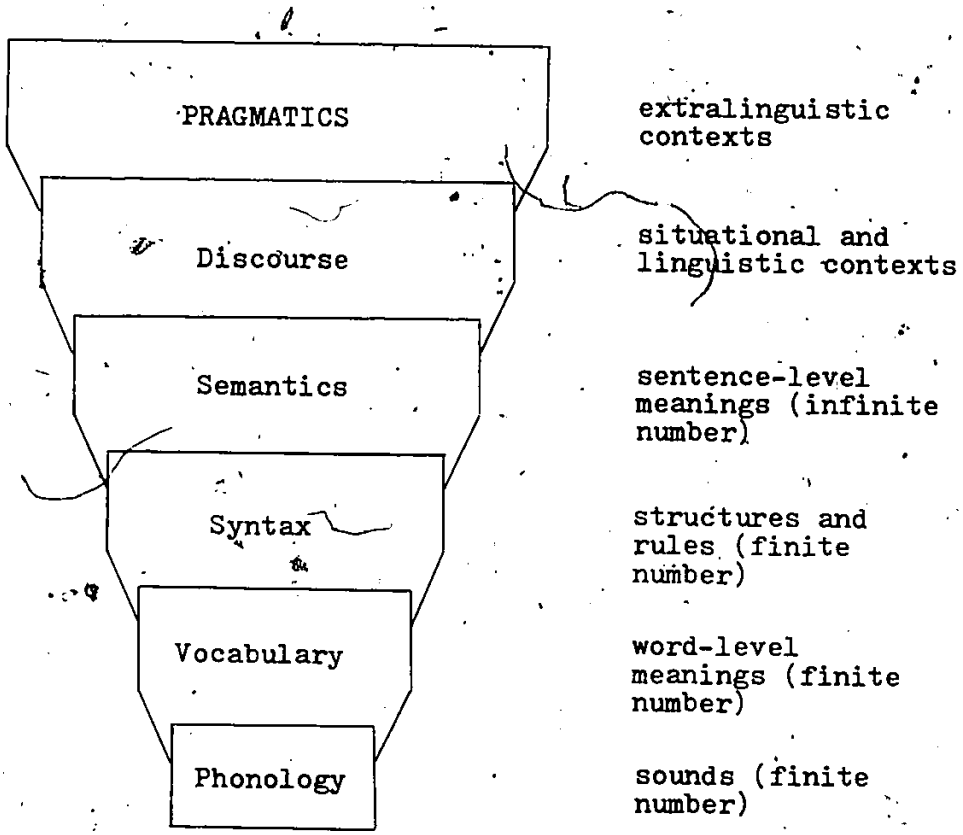
because non-verbal communication is seen as intimately connected with verbal communication. Oller refers to the correspondences between linguistic and extralinguistic contexts as "pragmatic mappings". Linguistic information is mapped onto contexts and, accordingly, the higher the level of contextualization of language, the more efficient language perception and processing are (Oller 1973a).

Oller's theory of pragmatics might be represented in the form of an inverted pyramid (Figure 3.6.). Higher-level linguistic processing builds on and includes lower-level processing, thus each higher level presupposes the lower levels. Pragmatics seems to be the top of the pyramid in Oller's theory (Canale and Swain, 1979, do not expect their communicative competence component to be the most inclusive component).

Oller's theory of an expectancy grammar started with his introduction of the notion "expectancy" as a way to understand the nature of psychologically real processes that underly language use. Listeners expect to hear utterances that make sense. They check the signals they receive against their knowledge of grammar and their experience of the world and expect them to correspond with each other. Chafe (1974) demonstrated that the speaker will also anticipate what the listener will infer from what the speaker says, assessing not only sentence meaning but sentence force (attitudes and subjective evaluations that are emotionally potent, particularly in the realm of human relationships and self-concept (Oller 1976a)).

Oller postulates a grammar capable of generating such expectancies based on contextually dependent information. The proposed framework of an expectancy grammar represents an adaptation of Wood's (1970, 1972), and Woods and Makhoul's (1973) recursive transition network model, which was designed

Figure 3.6. Oller's pragmatic theory, represented as an inverted pyramid



The levels of linguistic processing that have to be activated in communicative tests according to Oller (1973a)

in an attempt at formalizing the kinds of integrative tasks a language user performs in communicative acts. The transition network grammar was subsequently modified by Oller to allow for the incorporation of pragmatic information.

An expectancy grammar would interpret the sentence "Strange mountains loom in the fog" (Oller 1976a) forming hypotheses and rejecting or accepting these according to the specifications incorporated in such a grammar. The following partial example helps to illustrate the principle involved:

- The lexical item "strange" is recognized and identified as an adjective; it is hypothesized that the lexical item forms part of a noun phrase and that a noun, which is expected to be the subject of a declarative sentence, is to follow.
- The lexical item "mountains" is recognized and identified as a plural noun; the hypothesis is formed that a verb phrase is to follow. According to Oller (1976a):

> "The grammar associates the noun phrase with either strange mountains that are in fact visible, or it infers a situation in which strange mountains could be said to appear. This leads to subsequent expectations about the kinds of predicates or modifiers that might follow." (Oller 1976a:289)

The grammar continues the interpretation of the sentence, forming new hypotheses. It rejects them until they can be accepted on the basis of the information incorporated in the grammar. Intonation patterns and pauses indicate the end of a sentence as well as stylistic and attitudinal factors in oral communication. The grammar is expected to have information about, for example, what objects can be referred to as "mountains" and circumstances that may be considered "strange" in metaphorical or actual contexts.

The brief discussion of the principle involved in an ex-

pectancy grammar serves to explain Oller's request for language tests presented in context. Expectancies cannot be instilled in the learner, and his expectancy grammar cannot be invoked, through the manipulation of discrete-point items.

Oller's theory that verbal and non-verbal contexts of communication tend to limit the elements that may follow any given sequence of elements is supported by a number of theorists. Jespersen (1904) stressed that in the course of normal communication a certain connection holds between individual utterances. More recently, the same phenomenon was observed by Watzlawick, Beavin and Jackson (1967) who concluded that

> "...in a communicational sequence, every exchange of messages narrows down the number of possible next moves." (Watzlawick, Beavin and Jackson 1967:131)

In addition, Richards (1970, 1971) has shown that verbal and non-verbal contexts tend to differentiate linguistic elements that might be confused or misinterpreted if used out of context.

Oller's grammar of expectancy is in fact a description of what the native speaker knows, i.e. a model of competence. But the term "competence" used in this sense describes more than Chomsky's model of competence as it accounts for linguistic as well as extra-linguistic factors. An expectancy grammar as postulated by Oller attempts to describe language in context, both with regard to linguistic and situational factors. For the purpose of this study, the componential model as proposed by Canale and Swain (1979) and others will be given preference. Although an expectancy grammar seems to postulate an interesting explanation of what speakers-listeners do, it seems that more research is needed to make it operative.

3.4. A working theory of communicative competence

The preceding discussion of various theories of communicative competence does by no means exhaust the topic. It does, however, allow conclusions to be drawn as to what should or should not be included in a comprehensive theory of communicative competence for the purpose of this paper.

The assumption that the main purpose of language is that of communicating seems to be widely accepted. Kelly (1969) traces the idea that

> "...language is an instrument of human society..." (Kelly 1969:331)

back to the 16th century scholar Vivès. A similar position was later taken up by Rousseau and, more recently, by Searle 1969, Campbell and Wales 1970, Habermas 1970 and others.[4] Communication is seen as a process of tentative interpretation of linguistic forms and social behaviour in different contexts in the course of joint negotiations (Candlin and Breen 1980). Such interpretation between interlocutors does not occur at the word or sentence levels of language but at the discourse level. Communication involves both oral and written channels. Based on Morrow (1977) and Canale and Swain (1979), communication is defined as

- based on sociocultural, interpersonal interaction
- involving a) creativity in Chomsky's sense of the term
  b) unpredictability, although participants are able to predict to some extent what their interlocutor is going to say, they cannot be sure until they have heard it, i.e. the uncertainty is reduced when the answer to a proposition has been given (Goodman 1968, Palmer 1978, Smith 1971).
- taking place in a discourse and sociocultural context, in the sense that the relationship between an utterance and its meaning depends on social parameters and may vary depending on

discourse and sociocultural context (Candlin 1978). The expression "He's lost his marbles" uttered by a 4-year old explaining why his friend is crying, takes on an entirely different meaning when uttered by political advisers hearing their candidate promise potential voters that he would eliminate all taxes if elected!

A theory of communicative competence is expected to interact in a meaningful and observable way with other systems of human action and knowledge (Canale and Swain 1979). The exact nature of such interaction is not clear at this point in time and will need to be determined in the course of future research.

A working theory of communicative competence is assumed to be composed of at least two components:

- grammatical competence, representing the second/foreign language learner's knowledge of the code, discussed in section 3.2.
  Although native speakers of a given language tend to focus more on language use than on grammar in the course of conversation (Widdowson 1978), linguistic competence is nonetheless an important part of communicative competence, especially for the second/foreign language learner. Hymes (1972) pointed out that there are rules of grammar that would be useless without rules of use, in many cases the opposite also applies, i.e. there are rules of use that would be useless without rules of grammar.
- sociolinguistic competence, representing the knowledge of the rules of use in specific social contexts, made up of sociocultural rules and rules of discourse as discussed in 3.2.
  The inclusion of sociolinguistic competence in a theory of communicative competence is as important as the inclusion of linguistic competence. The existence of rule-governed, universal and creative aspects of sociolinguis-

tic competence is illustrated by examples such as Hymes' speaker who produces grammatical sentences at random.

According to Canale and Swain (1979) these components are not in rank order, they are considered equally important.

Psycholinguistic factors such as memory constraints and perceptual strategies and strategic competence are not included in this theory of communicative competence. Although some researchers (Jakobovits 1970, Hymes 1972, Canale and Swain 1979) advocate the inclusion of one or both of these competencies, they seem better accounted for in performance (for the time being at least). Psycholinguistic factors are usually considered to be general psychological constraints on such aspects as production and comprehension of utterances. There seems to be no conclusive evidence that they could be accounted for in some more appropriate way. It would seem possible to account for psycholinguistic factors in terms of the differences between second/foreign language learners' target communicative competence and native speakers' communicative competence models. The advantages of such an account would have to be shown on the basis of actual data.

Canale and Swain's strategic competence component seems best accounted for within the performance component at this point in time. There seems little doubt that such strategies are part of native speakers' communicative competence. The importance of these strategies is, however, insignificant for native speakers if compared with their importance for second/foreign language learners. In other words, the absence of the strategies discussed by Canale and Swain (1979) is far more apparent than their presence. Based on research available, the universality or competence-specific characteristics of strategic competence in a model of communicative competence of the second language learner has not been clearly established. Until such evidence is available, strategic competence as well as psycholinguistic factors seem best dealt with as performance variables.

A theory of communicative competence such as outlined above serves as the basis of a model of communicative competence. Such a model provides the foundation for the testing aparatus needed to determine how far the learner has, at a given point in time, progressed toward his target competence.

4. THE ROLE OF COMMUNICATIVE PERFORMANCE

The purpose of this chapter is to characterize the role of performance in language ability, particularly second or foreign language ability. The theorists discussed in chapters two and three were primarily concerned with competence, paying little or no attention to the performance component of Chomsky's dichotomy. After a brief discussion of some of the research carried out in the area of performance, its importance in linguistic situations will be shown. Furthermore, it will be shown that communicative competence cannot be tested without the intermediary of communicative performance. The remainder of the chapter will point out the main characteristics of a communicative performance test and some of the problems incurred in communicative performance testing.

4.1. Performance and research

Performance has generally received considerably less attention than competence in research dealing with Chomsky's dichotomy. Chomsky (1965) maintains that the study of performance is secondary to the study of competence. In his theory, performance seems like a residual category, worthy of investigation only in as much as it will help clarify competence. Chomsky (1965) does not state exactly what belongs to a theory of performance, giving the impression that anything not accounted for in a theory of competence will be part of performance. It is uncertain that all variables affecting language performance are known. But clearly, a study of performance has to account for a large number of variables. According to Canale and Swain (1979) most researchers accept, at least implicitly, that there is a difference between communicative competence and communicative performance (Halliday is an exception as pointed out in chapter three). Nonetheless, the idea that the performance aspects can

be considered a legitimate area of research is a relatively new one.

Fromkin (1968 and 1971) was one of the first to consider performance questions as valid linguistic inquiry. Her study of spoonerisms and slips of the tongue led her to argue that performance is not random and unpredictable. She arrived at the conclusion that there is a difference between competence and performance, but that there exists a close interrelation between performance rules and rules of competence. Fillenbaum (1971) considered the relationship between competence grammars and performance grammars as being very close. Like Fromkin, he adopted a "correlative" point of view. The precise relationship was, however, still uncertain and even the variables which must be accounted for remained unclear.

Formal linguistics (for example Lakoff 1969, Salomaa 1969, Kay 1970, Lakoff 1971) has tried to deal with restrictions on the use of performance rules rather than rules per se. Results have, however, been disappointing in that only a fraction of the known phenomena could be accounted for.

Some linguists trying to explain performance phenomena (Fillmore 1968, Fraser 1971, Lakoff 1975) began to turn their attention to pragmatics and speech act theories (Searle 1969, Fraser 1974, Sadock 1974) in addition to syntax, phonology and semantics. However, the only clear result of pragmatic investigations up to this point seems to be that speech acts are considerably more complex than linguistic theories might have led to believe. Although some insights have been gained, e.g. in the areas of constraints on verbal behaviour and discourse phenomena, research in pragmatics is generally accepted to be in an embryonic state at this point in time.

4.2. Communicative performance in research

As outlined in chapter three, "communicative competence" as used here, refers to the interaction and relationship between linguistic competence, sociolinguistic competence and rules of discourse. Communicative performance refers to the actualization of these competencies and their interaction in the actual comprehension and production of utterances under general performance constraints. It includes psycholinguistic factors such as memory constraints and perceptual strategies. It further includes communicative strategies, factors of volition and motivation as well as organic and functional pathology, insofar as these affect language behaviour.

Recent research tends to deal primarily with a model for communicative competence. Canale and Swain's paper (1979) is one of the few recent studies to deal, although only briefly, with the theoretical aspect of communicative performance.[5] They have pointed out that there is a fundamental difference between communicative performance and communicative competence that should be maintained in theories of communicative competence.

4.3. Testing communicative performance

Before commencing the discussion on testing communicative performance, a terminological clarification seems indicated. In the preceding sections the term "communicative competence" has been established as referring to underlying knowledge. The term "communicative performance" is used to refer to the performance of the above knowledge as well as specific performance factors. Some researchers use it as a technical term, others treat it as a concept. In order to have a neutral term, that is not necessarily tied in with the dichotomy, to describe the performance of overall language proficiency, the term "communicative ability" is used.

It is generally accepted that communicative competence can only be tested through performance. Although there can be no performance without competence, there can be no indication of competence without performance. Accordingly, performance is directly observable, but any conclusions about competence have to be inferred from performance. Tests sample the learners' behaviour in a given situation, and are designed to evaluate specific knowledge. But not all tests involve performance to the same extent. Neufeld (1978) used the distinction "competence-oriented" and "performance-oriented" to differentiate between two types of tests.

- Competence-oriented tests, according to Neufeld, eliminate performance variables as much as possible, e.g. instead of production tasks such as speaking or writing, the examinee is given multiple choice tests or similar. Neufeld hypothesized that as a multiple choice test does not require the learner to actually produce language, an evaluation of speaking or writing skills is avoided and a clearer reflection of underlying competence can be achieved.
- Performance-oriented tests require the learner to produce language utterances through the actual performance of skills (e.g. writing, speaking), thus a distortion of underlying competence is thought more likely.

Canale (personal communication) also uses the distinction. He considers it particularly important for the purpose of testing. He compares competence-oriented tests and performance-oriented tests to the two components of a driving test, where the pencil-and-paper knowledge test can be said to be competence-oriented. The actual road driving test would, according to Canale, represent a performance-oriented test. The following test item serves to illustrate the difference between competence- and performance-oriented tests.

Instruction to the student: Fill in the blank with the missing word from one of the words given below.

These chairs_____ $ 97.94

a) buy b) cost c) pay d) charge

The above test item tests primarily the students' receptive ability. Performance is limited to recognizing written phrases and to circling correct forms. A similar test item might be constructed but without indicating a number of alternatives. The students would then have to supply the appropriate item from their own language experience. Such a test item tests the learners' ability to produce and write down appropriate items and it is, therefore, more performance-oriented than the example given above. The distinction made between competence-oriented and performance-oriented test items could also be represented in terms of a continuum, ranging from linguistic competence to communicative performance. At the linguistic competence extreme test tasks involve manipulations of the code, at the communicative performance extreme they closely resemble actual communicative situations.

However, in order to measure the students' ability to perform in a communicative situation, their performance in relation to a specified goal must be evaluated. Such a goal should be expressed in functional terms (Canale and Swain 1979, Carroll 1978b, Morrow 1980, Munby 1978, Van Ek 1976, Wilkins 1978) and answer the question: can this learner "do" a specific task. Accordingly, a test based upon functional objectives and claiming to evaluate what learners can do seems by necessity to be a performance-oriented test. It follows that tests based on functional objectives involve the learners' knowledge of the world to a varying degree. Several researchers do not think it advisable to involve the learners' knowledge of the world in language tests. Corder (1973) warned that

> "...[i]f a test is to be truly valid we must avoid, as far as possible, testing such cognitive factors as general intelligence, knowledge of the world and the belief system of the learner, or his general motor-perceptual skills." (Corder 1973:359-360)

Davies (1977) expressed similar concerns. It could, however, be argued that the learners' ability to communicate is to be tested and communication takes place, when a speaker and a listener share or exchange sentiments, observations, opinions or ideas, all of which presume knowledge of the world to a greater or lesser extent. This is the justification for maintaining that tests based on functional objectives can reasonably expect the learners to demonstrate knowledge of the world in the area in which they plan to function. The objective of a functional test is not an evaluation of the learners' potential ability under idealized conditions but their actual communicative performance. Chomsky (1965) maintained that distractions, shifts of attention and interest, memory limitations and errors are grammatically irrelevant (a view to the contrary was put forth by Lakoff 1975). However, they occur in any communicative situation and the language learners have to be able to formulate strategies for dealing with them.

Another view supporting the premise that the learner needs to be allowed to function under performance constraints is Spolsky's observation that

> "...[there is] no evidence for performance in competence, although competence will be established in order to explain performance." (Spolsky 1968:67)

In other words, the fact that a student performs well in a given competence-oriented test, does not indicate that he will also succeed on a performance-oriented test; however, the opposite is true.

4.4. Characteristics of a test of communicative performance

Morrow (1977) lists a number of features present in communicative situations. It would seem that a test claiming to evaluate communicative ability should reflect these characteristics. They are described below and will be used for the evaluation of currently used proficiency tests in Chapter Five. These characteristics

are listed here because they represent a number of aspects of communicative situations learners might reasonably be expected to master. (The descriptions are interpretations of Morrow, 1977)

4.4.1. Interaction-basis

Most instances of language use involve a sender of a message and a recipient. The oral, face-to-face situation is the most obvious, but not sole example of such language use. Other examples include: telephone conversation, letter-writing and less obvious ones, such as writing a report. The recipient of a message has expectations which affect what is communicated to him and how this is done. Interaction involves the learners' receptive and productive abilities and the emphasis may be on any one of the four basic skills (listening, speaking, reading, writing). The sender of a message has to decode what was said to him and then integrate relevant parts into his responses.

The characteristic of interaction-basis represents a vital aspect of communication. It seems difficult if not impossible to envisage communication without interaction. Thus communicative tests need to reflect this interaction-basis. The most obvious way to achieve this would seem to be through some kind of oral interview test.

4.4.2. Unpredictability

Two main factors contribute to unpredictability in communication:

- the creativity of language in Chomsky's sense of the term
- the processing of unpredictable data as fast as it becomes available in communicative exchanges.

The recipient of a message has certain expectations (Oller 1973a, 1973b, 1976a among others), but before he processes what is being said to him, he does not know whether to accept or re-

ject his hypotheses. Communication is explained as a process of gradual reduction of uncertainty (Goodman 1968, Smith 1971, Palmer 1978).

The element of unpredictability is very important, i.e. it serves the purpose of maintaining interest in communication. No recipient of a message can be expected to maintain interest if he already knows all that is being said to him. The most obvious way to achieve unpredictability in a testing situation is, again, through an oral interview.

4.4.3. Contextualization

Communication takes place in a context which determines the appropriateness of specific language forms. Interlocutors have to be able to deal with appropriateness in terms of linguistic context and situational context.

Situational context

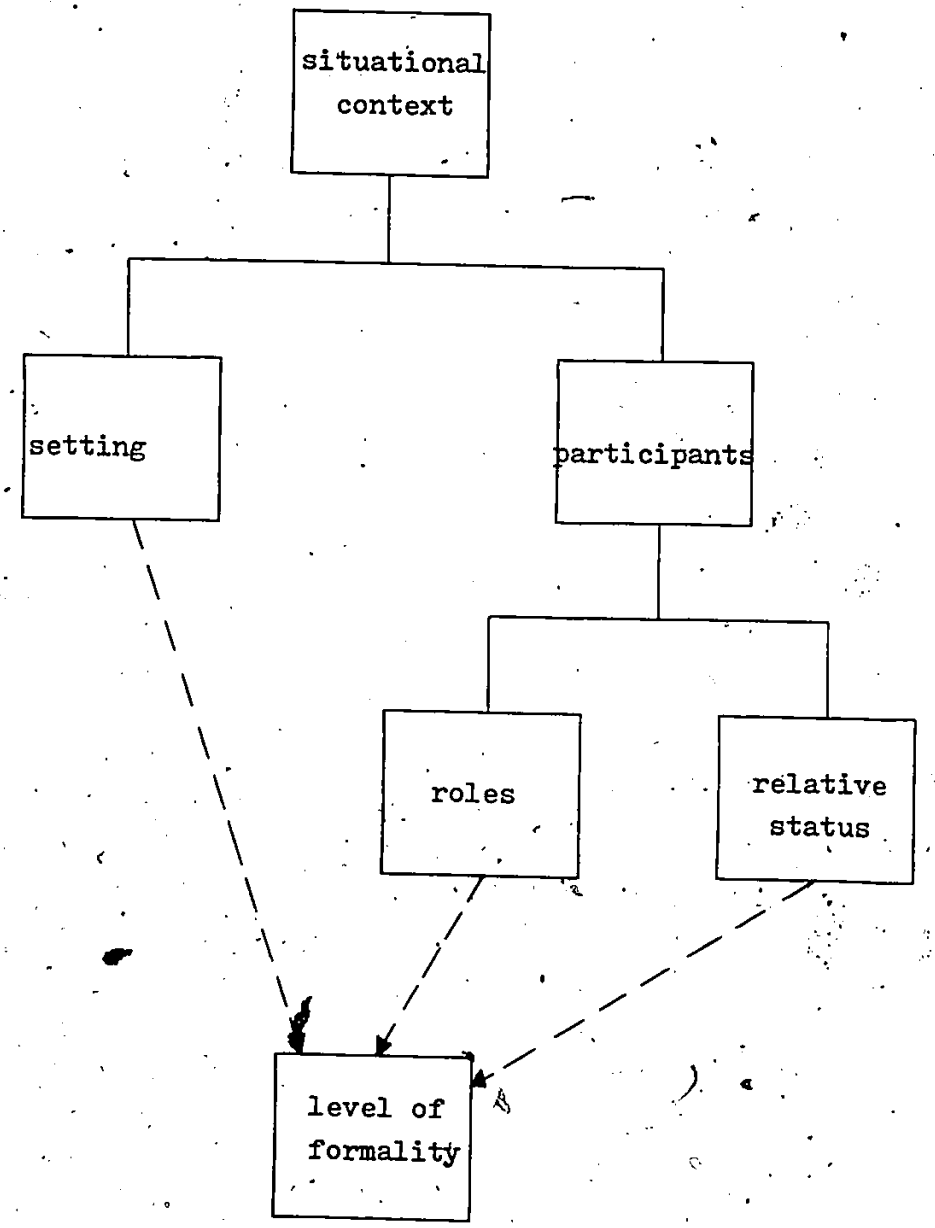
As Figure 4.1. illustrates, communication takes place in a situational context that consists of a setting (e.g. restaurant) in which the participants have roles (e.g. waiter-customer) and relative status (e.g. employee - member of the public), which determine the level of formality (e.g. formal, friendly) appropriate for the interaction.

Linguistic context

Interlocutors express mood and attitude through supra-segmental and kinesic features, choice of lexis etc. The interlocutors' presuppositions about each other and the status of events depend also on context and are expressed through some or all of the above means. As Morrow (1977) pointed out, language learners need to understand the difference between the various possible intonations of a sentence like

"Paul drove the car to work today"

Figure 4.1. The situational context



The determiners of the level of formality

Stress can be placed as follows, depending on intended meaning:

Paul (not somebody else)
drove (not pushed)
the (a very specific one)
car (not truck)
to (not from)
work (not the beach)
today (not last week)

Every utterance is thus shaped linguistically and situationally by the contextual features of communicative situations (Firth 1952). A communicative test should reflect these features and measure the candidates' ability to deal with them, i.e. the learners' ability to understand them as well as their ability to use them to serve their own communicative needs.

4.4.4. Purpose

In normal communication, utterances are made for a reason (Wilkins' 1976 categories of communicative function). Furthermore, an utterance has a topic which might be implicit or explicit. Learners have to be able to recognize the topic of a specific unit of discourse, or the purpose of a certain remark addressed to them. They also have to be able to encode utterances appropriately so that they serve their own purposes.

Reason and purpose are necessary characteristic of a communicative situation. They are likely to help maintain interest and should be reflected in tasks required for communicative testing.

4.4.5. Performance-basis

Communication is interaction-based and as such it involves performance. A young boy is unlikely to become a champion soccer player by merely studying re-plays. He has to go out and convince others through his performance. He also needs the real-life prac-

tice to learn skills in addition to knowing abstract rules and probabilities.

Similarly, second/foreign language learners need to be given the opportunity to practice and to demonstrate their knowledge in situations that reflect the characteristics of communication. Accordingly, tests that are intended to evaluate communicative ability have to be based on performance.

4.4.6. Authenticity

Morrow (1977) and Widdowson (1978) claim that native speakers of a language do not normally simplify their utterances to allow for a learner's incomplete command of the target language. On the rare occasions where native speakers do simplify their language to account for their interlocutor's linguistic abilities, the simplifications do not seem to bear any resemblance to those situations found in teaching and testing materials.

Authenticity of the type of language used in tests is very important because the second/foreign language learners do not need to know some simplified version of their target language that they will never encounter in a communicative situation. They need to know what is actually said or written. Accordingly, every effort should be made to present authentic language in testing materials.

4.4.7. Behaviour-basis

Of prime importance in a communicative act is whether or not a message gets across. A test aiming to assess a learner's communicative ability needs to evaluate what the learner can achieve through the use of the target language. Although this aim puts primacy of content over form, it also indicates that form cannot be neglected entirely. If a learner understands the utterance

"Do you want to close the window"
(intended as a request)

as a question, the behavioural outcome from the speaker's point of view might not be as anticipated and, therefore, be unsuccessful.

In real-life communication the outcome of a communicative situation tends to indicate whether or not the interaction was successful. However, as Morrow (1977) pointed out, it might be difficult to construct tests that reflect the criterion of behaviour-basis, as pragmatic factors play an important role, e.g. a request might be expressed appropriately but ignored by the interlocutor. The speaker needs to mobilize his communicative strategies in order to be able to handle the situation. A communicative test should reflect the fact that different people react in unpredictable ways to any given situation, and examinees need to demonstrate that they can use language to meet their communicative needs.

The seven criteria proposed above for an evaluation of communicative tests are all interrelated and the areas covered by some overlap with areas covered by others. They indicate clearly that a test of communicative ability should evaluate the learners' performance in the context of authentic situations. A pencil-and-paper test based on discrete-point items might at best evaluate some aspect of communication, but it is unlikely to measure communicative ability. Therefore a number of researchers (e.g. Oller 1973a) have proposed tests that require examinees to use more than one of the traditionally recognized components of grammar in the same test, i.e. tests that reflect more accurately real-life communicative situations.

4.5. Integrative tests as tests of communicative ability

Integrative tests are tests which require processing at the discourse level of language organization. An integrative test

is generally accepted as the logical antithesis of discrete-point tests (Oller 1979a). As discussed in chapter 1, discrete-point philosophy advocates testing of one of the traditionally recognized language skills at a time. It further advocates testing of only one of the traditionally recognized components of a hypothetical grammar at a time. Tests are usually low-level, i.e. at the level of syntax or below, with regard to language organization. This practice is consistent with the views of language and testing predominant under the influence of the psychometric-structural trend in testing. Discrete-point tests rarely require comprehension of context beyond the sentence level. Accordingly, they do not reflect normal language use.

Oller (1973b) and Davies (1978) have pointed out that the difference between the two kinds of tests is one of degree rather than type. Both researchers speak in terms of a continuum ranging from discrete-point items at one end to full-scale language use at the other.[6]

When Carroll (1961) proposed the distinction between discrete-point and integrative tests, he suggested that integrative tests measure an aspect of "knowing a language" that is not measured by discrete-point tests. He added that native speakers' ability to integrate a large number of linguistic phenomena at great speed, i.e. real time, is likely connected with that aspect. Oller (1979a) has further pointed out that the language users' ability to make inferences about how the linguistic context of an utterance relates to its appropriate extra-linguistic context is an important part of native speaker ability. Tests that take account of extra-linguistic context are pragmatic tests (Oller 1979a) if they meet Oller's two "naturalness criteria":

- they have to reflect sequential constraints specified by the grammar
- they have to relate linguistic content to states of affairs outside of language.

Furthermore, Oller (1976b) has claimed that the redundancy (or grammatical organization) of verbal material accounts for a large portion of variance found in verbal experiments. From this he has concluded that second/foreign language tests should "involve the examinee's capacity to utilize such organisational constraints." (Oller 1976b:160) Thus they should be not only integrative, but also pragmatic, i.e. they should reflect clear relationships to extralinguistic contexts.

The brief discussion of integrative and pragmatic tests shows that tests of communicative ability must measure more than grammatical accuracy in idealized, context-free language. They have to reflect his ability to understand and produce appropriate and acceptable utterances, i.e. language in use.

4.6. Testing communicative ability: problems

A number of problems and controversies connected with testing communicative ability (or communicative performance) remain and will be discussed briefly below.

4.6.1. Criterion vs. norm-referenced testing

Communicative ability is judged primarily on its outcome. It answers the question: Was the language user successful in performing the task he attempted to perform? Criterion-referenced testing, where the learner's progress toward a specific goal is evaluated, seems ideally suited for the purpose. But criterion-referenced testing presupposes the existence of clearly stated objectives. If there are no clear objectives, criterion-referenced evaluation is not suitable. Criterion-referenced tests have been criticized for a number of reasons recently and some writers have proposed domain-referenced tests instead. Domain-referenced tests are intended to test an entire spectrum of behaviours required to master some specific aspect of knowledge (Educational Technology, 1979).

In norm-referenced tests the level of performance of each individual is compared with the general standard of performance shown by the group he belongs to. The emphasis is on how an individual stands with reference to his peers. Such information might be necessary for administrative purposes, e.g. to determine the instructional level suitable for individual students. An evaluation of communicative ability, however, would seem more useful in terms of a functional, externally predetermined objective.

4.6.2. Evaluation

Evaluation criteria and procedures present another problem in the assessment of communicative performance or, in more general terms, of communicative ability. Evaluation would need to discriminate on a continuum ranging from "yes, the candidate is able to perform task xyz" to "no, the candidate is unable to perform task xyz". Most of the rating scales devised for the evaluation of communicative ability list different levels. There are, however, some inconsistencies between the stated objectives and the evaluation criteria used. The interview test administered by the Foreign Service Institute, for example, ranges from a low level of "Elementary Proficiency" to "Native or Bilingual Proficiency", with several subdivisions, allowing for eleven different levels (Wilds 1975). The corresponding evaluation scale is formulated in functional terms, e.g. "Can he [the candidate] ask and answer questions on topics very familiar to him" etc. The examiners' rating checklist, however, consists of the following parameters:

(Focus)	(rated between)	(weighting)
1. ACCENT	foreign - native	0
2. GRAMMAR	inaccurate - accurate	3
3. VOCABULARY	inadequate - adequate	2
4. FLUENCY	uneven - even	1
5. COMPREHENSION	incomplete - complete	2

(Wilds 1975:38)

The inconsistency is in the apparent evaluation of operational

objectives on the basis of structural aspects of communication. Only the fifth parameter represents a global judgment and thus an integrative aspect of the test. It is unclear, why the examinee's grasp of the grammar of the target language is singled out as the strongest factor involved in successful communication.

A framework of evaluation that is more consistent with the communicative aims in testing is proposed by Garcia-Zamor and Krowitz (1974). Their preliminary framework is aimed at an evaluation of advanced learners' communicative abilities. They based their evaluation on systematic individual job analyses-objectives. Level 1 is labeled as that of

"comparing" : Selects, sorts, or arranges data, judging whether their readily observable functional, structural, or compositional characteristics are similar to or different from prescribed standards." (Garcia-Zamor and Krowitz 1974: 16)

The scale continues hierarchically, i.e. each higher function is understood to include those below it, to level 6

"synthesizing": Takes off in new directions on the basis of personal intuitions, feelings, and ideas (with or without regard for tradition, experience, and existing parameters) to conceive new approaches to or statements of problems or the development of system, operational, or aesthetic "solutions" or "resolutions" of them, typically outside of existing theoretical, stylistic, or organizational context." (Garcia-Zamor and Krowitz 1974:17)

These definitions are complemented by descriptions of what the employee is expected to do. A candidate required to function at level 4 "analyzing"

"...might be assigned the task of evaluating suggestions for improving Tokyo's urban transport system, weighing cost and other factors in order to give a written evaluation of the relative merits of the suggestion." (Garcia-Zamor and Krowitz 1974:17)

The evaluation criteria used include factors like "continuity of discourse", "irritation factors" (e.g. slowness of speech production), "ability to recognize and respond to communication breakdown". No purely linguistic faotors are involved in their evaluation criteria. Thus the above framework for evaluating advanced learners' communicative ability indicates how assessment parameters can be selected to make functional statements about learners' communicative ability. It is aimed at a relatively small and specialized group of language learners both in terms of language ability and profession, but could be adapted to suit the needs of other groups.

A somewhat broader (and simpler) rating scale was developed recently by the British Council's English Language Testing Service. The sample rating scale has been tested extensively already and is documented in various publications (e.g. Carroll 1980). The "General assessment scale" is derived from the sub-scales of an oral interview and an academic writing scale. It ranges from

"Band 1/0 : Non-user. May not even recognize with certainty which language is being used."

to an intermediate stage

"Band 5 : Modest user. Although he manages in general to communicate, often uses inaccurate or inappropriate language."

to a

"Band 9 : Expert user. Communicates with authority, accuracy and style. Completely at home in

idiomatic and specialist English." (Carroll 1980:134)

4.6.3. Reliability

It was pointed out above that an assessment of communicative ability has to be made on the basis of integrative or global tests. This type of evaluation is difficult if not impossible in quantitative terms. Morrow (1980) has pointed out that

> "Performance is by its very nature an integrated phenomenon and any attempt to isolate and test discrete elements of it destroys the essential holism." (Morrow 1980:151)

Thus a qualitative judgement has to be made that might be converted into a quantitative score to facilitate subsequent analyses. A qualitative (or subjective) basis for evaluation has, however, some important repercussions with regard to reliability. Functional dialogue or similar language tests tend to achieve initial low inter-rater reliability scores. Raters need to be trained carefully and usually large investments of time and monetary resources are necessary to achieve high levels of rater reliability. Research in the area of inter- and intra-rater reliability has, however, shown that high levels of reliability can be achieved if raters are trained carefully, e.g. Clark (1975, FSI Oral Interview), Clark (1978b, experimental oral interview tests for TOEFL), National Institute for Educational Measurement of the Netherlands (1979), Farhady (1980).

4.6.4. Validity

The concept of validity is closely connected with that of reliability. Validity indicates the extent to which a test measures what it is supposed to measure. At least two factors affect validity

- the accuracy with which the area to be tested is characterized

- the suitability of the assessment instrument.

Oller and Perkins (1978) pointed out that empirical tests must be applied to the tests themselves in order to determine whether or not they are good tests according to some range of purposes. Validity tests are usually applied to determine this aspect.

Five different types of validity are generally mentioned in the literature on language testing:

- face validity
- content validity
- construct validity
- concurrent validity
- predictive validity

A short discussion of these five types of validity (in the order given above) will serve to point out their relevance in tests of communicative ability.

Harris (1969) considered a test's face validity (i.e. the way a test looks to those taking and administering it) very important. A candidate's performance is likely to be affected either positively or negatively, depending on how he judges a test he is about to take. Performance-oriented tests (as defined in 4.3.) seem to have greater face validity than competence-oriented tests (Carroll 1961, Canale and Swain 1979). Students appear to prefer tests involving performance tasks that are closely related to actual language use. Schulz (1977) found that 79.75% of student responses in her study indicated that candidates prefer communicative testing procedures over those emphasizing controlled, structural manipulation. High face validity of a test should therefore result in more motivated examinees and, probably, higher levels of performance. Lado, however, disputes the value of face validity.

> "I don't think there is any merit in face validity."
> (quoted in Jones and Spolsky 1975:25)

But despite Lado's rejection of face validity, it does seem to be an important feature of a test. As pointed out above, its presence is likely to motivate the examinee positively. Its absence is likely to make him feel that the test is not really relevant or important and as a result the examinee is not likely motivated to perform to the best of his ability on the particular test in question.

A good test of communicative ability establishes its own content validity in the sense that it measures those skills it sets out to measure. A test derived from a careful specification of the learners' needs, which are also used to specifiy the syllabus (as proposed by Munby 1978, Credil 1975, Van Ek 1975), is expected to show very high content validity.

Construct validity tries to determine if a test reflects the theories of language and learning that motivated the test. It can, for example, be achieved by following Munby's (1978) guidelines for test construction:

- elaborate a theory of communicative competence
- specify a model of communicative competence based on the above theory, and
- construct assessment instruments of communicative ability evaluating progress towards model set out above.

Examples of tests based on Munby (1978) are those of the British Council's English Language Teaching Institute, documented by Carroll (1978a, 1978b, 1980); tests based on Van Ek (1975) include the CITO Test (documented by Roos-Wijgh 1978) and the UCLA Test (documented by Farhady 1980).

Concurrent and predictive validity are likely to be more difficult to establish for tests of communicative ability, at least initially. But ultimately communicative tests should cor-

relate with each other. Concurrent validity refers to the correlation between the results obtained using two different tests, i.e. one test is used as an external criterion. Predictive validity refers to the accuracy with which a candidate's future performance can be predicted on the basis of a particular test. If communicative tests are constructed according to the criteria set out above, they are unlikely to measure the same abilities as discrete-item tests. It is as yet not clear how well direct and indirect tests correlate. Further research will have to show how different types of tests are related. Until such information becomes available, it is not clear if and with what tests a particular communicative test should be expected to correlate.

4.6.5. Generalizability of test results

The relationship that exists between a candidate's ability to perform one task and his ability to perform another task is, at this point, not clear (Canale and Swain 1979). Morrow (1980) has referred to the problem of "extrapolation" in this context. Global tasks such as isolated by Munby (1978) require a set of "enabling skills" to be mobilized in order to be completed. A global task such as

"Search for specific information"

requires enabling skills such as

"e.g. - Distinguish main point from supporting details
- Understand text relations through grammatical cohesion devices
- Understand relations within sentences
- Understand conceptual meaning
- Deduce meaning of unfamiliar lexis"
(Morrow 1980:152)

These enabling skills are identified in operational terms, i.e. performance. However, they can be applied to complete a number of different global tasks, thus involving an aspect of what is

generally understood to be competence. Morrow (1980) has further pointed out that an analysis of the global tasks in terms of which the candidate is to be evaluated, are expected to yield a relatively consistent set of enabling skills over a variety of tasks. Data obtained in an evaluation of the learner's ability to use these skills is not limited to a particular instance of performance. They are relevant for a considerable range of global tasks. The whole (i.e. the global task) is accordingly analyzed into individual elements (i.e. enabling skills) in a way reminiscent of discrete-point tests in the psychometric-structural trend, although the way it is presented in order to extract student performance remains integrative. A language learner might be able to handle individual enabling skills successfully yet be unable to synthesize and mobilize them to perform a specific global task. Similarly, he might be able to perform global tasks A, B and C, yet be unable to handle task D which requires enabling skills performed in A, B and C to be mobilized. It is not clear at this point if generalizability at the level of enabling skills is possible or, indeed, necessary. But it seems unlikely that generalizability at the level of tasks can be achieved if the learner is unable to synthesize and mobilize individual enabling skills.

4.6.6. Direct vs. indirect tests

The opposition direct/indirect test refers to how closely a particular test resembles the type of tasks learners are expected to perform in actual communicative situations. No matter how carefully a communicative test is prepared, it still is a simulation. But a test of communicative ability should resemble the target performance as closely as possible, i.e. it should be as direct as possible. It is the real-life communicative situation that ultimately shows how successful language users are in performing their knowledge and their skills. Indirect tests tend to be easier to administer and they might correlate significantly with more direct forms of testing. It is, however, difficult to see how e.g. a cloze test administered to a group of nurses could

indicate how well they could communicate with patients speaking their target language. The integrative-sociolinguistic trend in language testing calls for the evaluation of language as it occurs, not ways of understanding or constructing it (Harrison 1977).

4.6.7. Administrative Problems

A test may have all the qualities of a good test of communicative ability but require either unreasonable time or material investments. Such a test would be practically impossible to apply. If a test is to be useful, limitations imposed by administrative factors have to be considered, although they should not restrict the search for more appropriate tests. Frequently, a test is developed within the restrictions provided by an administrative framework. It would seem more promising, if the problem was approached from the opposite end of the spectrum. If the "ideal" test is devised first, compromises to allow for practical considerations might have to be made subsequently. This approach ensures that the overall view of what is needed to test communicative ability is not obscured by administrative restrictions.

4.7. Summary of desirable characteristics of a communicative test

Communication was defined as requiring performance. A test designed to measure learners' knowledge of a given target language (i.e. communicative competence, overall proficiency etc.) should evaluate learners' ability to communicate in terms of actual performance. Accordingly, a test of communicative ability should be

- interactive
- unpredictable
- contextualized with regard to situational and linguistic contexts.
- purposive
- performance-based
- authentic

- behaviour-based

At the present stage of research in the area of communicative testing there seems to be no data requiring a specific rank order for the above characteristics. They appear to be interrelated, some presupposing the existence of others. The above characteristics are primarily related to the communicative aspect of communicative testing. In addition, a number of characteristics more directly related to the statistical aspect of communicative testing are important. Ideally, a communicative test is also

- integrative
- criterion referenced
- reliable and valid
- generalizable (in the sense that it allows some useful predictions about examinees' success with other performance tasks)
- as direct as possible
- administratively acceptable

5. TESTING COMMUNICATIVE ABILITY

The purpose of this chapter is to evaluate two testing techniques and two well-known standardized testing instruments in terms of Morrow's (1977) characteristic features of communicative situations. It will be determined if and how these tests evaluate communicative ability. All four of these tests are claimed to evaluate overall proficiency in English as a second/foreign language. The tests to be discussed are:

1) dictation
2) cloze
3) Standardized English proficiency tests:
   a) Test of English as a Foreign Language (TOEFL)
   b) Two subtests of The Michigan Proficiency Test Battery:
      - The Michigan Test of English Language Proficiency (MT)
      - Composition

      (A third subtest, The Michigan Test of Aural Comprehension, containing ninety recorded problems with written answer choices was omitted because of its relatively infrequent administration.)

The first two represent the integrative-sociolinguistic trend in language testing. The standardized tests represent the psychometric-structuralist trend but they are, nevertheless, still widely used today to measure overall language proficiency. A fourth, experimental, test (COMTEST) will also be evaluated. This test was designed specifically to measure communicative ability of second/foreign language learners and should, therefore, rate highly when evaluated according to the criteria established in 4.4.1. - 4.4.7.

Each test will be described briefly in terms of
- background (to show, in general terms, the origins and the main proponents of the test under discussion)

- administration procedure (various techniques and evaluation criteria used in the administration of cloze and dictation are likely to produce different results)
- justification (i.e. why proponents of the test in question consider it a valid measure of overall proficiency)
- evaluation, based on Morrow's characteristic features of communicative situations (4.4.1. - 4.4.7.)
- discussion.

5.1. Knowing a language - a terminological clarification

Depending on the school of thought individual researchers adhere to, knowing a language is either synonymous with communicative competence, communicative ability, language proficiency, overall proficiency or some similar term. Details of various theories usually differ considerably but it seems that for most theorists knowing a language would include language users' ability to communicate. This assumption is supported by the generally accepted premise that the purpose of language is primarily for communication (as discussed in 3.4.). Robinson (1972) considered human interaction the prime function of language. Clark (1978a) has confirmed this idea in his discussion of language proficiency testing

> "The purpose of proficiency testing is to determine the student's ability to use the test language effectively for 'real-life purposes', that is to say for vocational pursuits, for travel or residence abroad..." (Clark 1978a: 23)

A proficiency type test would, accordingly, have to evaluate learners' ability to perform in life-like communicative situations.

Brière (1971) asked: "Are we really measuring proficiency with our foreign language tests?" The question here is rephrased: Are

we really measuring communicative (or simply linguistic) ability with our English as a foreign/second language tests?

5.2. Dictation

5.2.1. Background

Dictation has been a very popular form of language testing for many years. It was a frequently used means of evaluation during the pre-scientific trend in language testing. Under the influence of the psychometric-structuralist trend, dictation was considered uneconomical and uninformative (Lado 1961, Rivers 1968, Harris 1969). Lado (1961) further pointed out the following apparent inadequacies of dictation:

- it cannot be used to test listening comprehension or phonological discrimination because the sounds are usually given away by the context
- it cannot be used to test word order because word order is given
- it cannot be used to test vocabulary because the words are already given.

Valette (1967) cautiously wrote that dictation was a very specialized skill and that more research was needed. In the second and revised edition (1977) of her book, however, Valette considered dictation a precise indicator of students' listening skills.

Oller undertook several studies, both alone (1971, 1972a, 1972b) and with colleagues (Oller and Conrad 1971, Irvin, Atai and Oller 1974, Oller and Streiff 1975) to investigate the usefulness of dictation in language testing. He has contributed considerably towards a renewed interest in dictation as a measure of overall language proficiency under the influence of the integrative-sociolinguistic trend in language testing. Spolsky, Sigurd, Sato, Walker and Arterburn (1968), Spolsky (1969) and Johansson (1973) have further experimented with dictation using added noise in order to study redundancy in language and

how it relates to language proficiency. Gaies, Gradman and Spolsky (1977) have since shown that the addition of noise does not significantly affect student ranking. They hypothesized that the lack of context contributed more to low student performance than the added noise. Instead of using discrete sentences, they experimented with short dialogues and mini-lectures, i.e. material that reflected the kind of English second/foreign language learners are likely to encounter. Gaies, Gradman and Spolsky (1977) subsequently reported improved student performance as well as increased face validity of the test in the sense that there was less irritation on the part of the students. Thus contextualization of material emerged as an important factor in language testing, lending support to Oller's (1973a, 1973b and others) call for pragmatic tests.

5.2.2. Administration procedure

Oller (1979a) lists many different types of dictation (e.g. standard dictation, partial dictation, dictation with competing noise, dictation/composition and elicited imitation). He considers standard dictation the most frequently used technique in this group (1972a, 1979a) and describes it as the procedure whereby examinees write verbal sequences of text. These sequences may be spoken by an examiner or played back from a recording. He has pointed out that the rate of speech is the single most important factor to be considered in a dictation test (Oller 1972a, 1973b). Unless it is to be merely a test of spelling, the passage has to be read at conversational speed.[7] The length of sequences between pauses has to challenge the short-term memory of non-native speakers.

The administration procedure used for the evaluation in 5.2.4. below is the one Oller (1972a) described for standard dictation:

- first, the complete passage is read through at conversa-

sational speed
- second, the passage is read again at conversational speed but with pauses at convenient phrase or clause boundaries to allow the students to write down what they hear
- third, the passage is read again at conversational speed with occasional pauses that allow students to make corrections.

5.2.3. Justification

Oller (1971, 1972a, 1972b) has claimed that dictation is a valid testing technique for the assessment of overall language proficiency because it fulfills both of his "naturalness criteria".[8] He suggested that a dynamic process of analysis by synthesis is involved, mobilizing all the different levels of processing shown in Figure 3.6. (Chapter Three). He further claimed that dictation activates and tests the learners' grammar of expectancy. Davies (1978) has pointed out that what Lado (1961) finds objectionable in dictation tests is exactly what Oller seeks in a language test. In Oller's theory a test that does not measure any one particular discrete-item is integrative and thus desirable.

One of Oller's reasons for advocating dictation as a measure of language proficiency is its high correlation with both cloze and standardized proficiency tests (Oller 1971, Irvine, Atai and Oller 1974). However, Alderson (1978) has interpreted these results to indicate that dictation (like cloze) measures lower order skills rather than the higher order skills claimed by Oller.

Oller (1979a) has further claimed that dictation is not subject to practice effects, unlike other testing techniques frequently used (e.g. multiple choice). This claim is supported by data collected by Kirn (1972) who administered dictation tests to ESL

students repeatedly over a period of ten weeks. She concluded that dictation scores did not improve with practice unless there was also an improvement in language proficiency. This conclusion is in contrast with Valette's (1973) warning that dictation scores may improve due to a practice effect in the technique rather than due to improved language ability. Her hypothesis is based on data collected with learners of French as a foreign language (1964).

5.2.4. Evaluation

Figure 5.1. summarizes how a standard dictation test rates when compared to Morrow's (1977) features of communicative situations (as discussed in 4.4.1. - 4.4.7.)

Interaction-basis

A dictation test can be considered interaction-based in the sense that the author of the text chosen has something to communicate to his audience (i.e. readers as well as listeners). However, depending on the text chosen, it could also be indirect interaction, e.g. a recorded dialogue. The interaction-basis of dictation tests is further weakened by the fact that testees do not emit any responses. They listen to and decode messages but they do not need to integrate these messages into responses of their own. Dictation tests, accordingly, measure students' ability to listen to, decode and record spoken text.

Unpredictability

Dictation tasks require examinees to process novel data in real time (i.e. as it occurs). Thus there is a degree of unpredictability for them. Creativity, however, is unlikely to be involved as examinees are expected to write down what they hear and hear what is being said.

Contextualization

The criterion of contextualization depends on the choice of text used. Theoretically, it should be possible to use material

Figure 5.1. Summary of test evaluations

	DICTATION	CLOZE	TOEFL	MICHIGAN COM*	MICHIGAN MT	COMTEST-2
interactive	2	2	0	1	2	1
unpredictable	2	2	0	2	0	1
contextualized	2	2	2	1	2	2
purposive	2	2	2	1	2	1
performance-based	2	2	0	1	0	1
authentic	2	2	2	1	2	2
behaviour-based	0	0	0	1	0	1

Morrow's (1977) characteristic features of communicative situations are reflected in the test:

0 = no
1 = yes
2 = not clearly or not completely (details in text)

*COM can, however, only test a limited range of language functions

that challenges candidates' ability to deal with appropriateness in terms of linguistic as well as situational context. Unless dialogues or similar formats are used, the scope of contextualization is likely reduced to the sentence level only. If contextualization is present it is, however, not possible to evaluate students' response to it because they do not respond by actually producing utterances of their own. It is difficult to see how testees' reactions to pre-supposition or sentence stress can be evaluated when they do not have to respond to such phenomena by actually producing utterances. Sentence stress for example can often drastically affect the intended meaning of an utterance as well as the reaction of interlocutors. It is not clear, how e.g. the difference between

I go to London vs. I go to London

could be reflected in a dictation task.

Another problem with dictation in the context of situational and linguistic contextualization is the fact that a text that is read is not likely to reflect normal sentence stress as it occurs in real-life situations.

Purpose

The criterion of purpose involves the propositional content of an utterance, its value and its modality. Examinees should be able to demonstrate that they understand the reason why an utterance was made. It is difficult to see how dictation can meet this criterion.

Examinees should also understand the topic of a stretch of language. If they are unable to understand what a text is about they are unlikely to be able to render it successfully in written form. Thus dictation is likely to evaluate examinees' understanding of topics. It does, however, not allow examinees to demonstrate their ability to use appropriate language to meet their own communicative needs.

Authenticity

This criterion depends to a large extent on the previous criteria having been met successfully. In addition, Morrow (1977) has criticized the frequent practice of simplifying language for the purpose of second/foreign language testing, adding

> "...there is no useful purpose whatsoever in testing the extent to which candidates can understand texts which are not authentic." (Morrow 1977:44)

While the above involves authentic content of the material presented, there is also authentic presentation of such material to be considered. Standard dictation tasks read at conversational speed require testees to process language input in real time in order to be able to keep pace writing what they think they have heard. Reading dictation tasks at a slower than conversational rate would result in a non-authentic presentation of the target language.

The above discussion implies that only a somewhat limited range of kinds of language could be presented authentically in this format.

Behaviour-basis

This criterion should be judged on the basis of whether or not an interaction was successful. It could be argued that the interaction was successful if examinees were able to process the text read to them and reproduce it in written form. This would show that a candidate is capable of writing down what is dictated to him in the target language. It is, however, unclear how such a dictation test can measure communicative ability. Researchers are as yet unsure of what strategies and processes allow an accurate rendition of a dictation text. Alderson (1978) has claimed that dictation tasks measure low-level linguistic skills, thus requiring linguistic competence only. In his opin-

ion, inference is not tested in a dictation task. This is in direct contrast to Oller's repeated claim (1972a, 1979a and others) that dictation measures higher level skills. However, whether dictation measures lower or higher level skills, it remains unclear how dictation evaluates what the learner can achieve through the use of the target language as he has had no input in the construction of the message, thus behavioural outcomes are unrelated to the nature or purpose of the message.

5.2.5. Discussion

The above evaluation shows dictation to be an unsuitable measure of overall proficiency. Further evaluation of dictation tasks as proficiency measures against some of Oller's own criteria (implicit or explicit) serves to confirm this conclusion. Oller has argued repeatedly that language tests should reflect language as it is used in everyday situations. Criticizing discrete-point tests Oller has pointed out:

> "There is no ordinary discourse situation in which a learner might be asked to listen to and distinguish between isolated minimal pairs of phonological contrasts. There is no normal language use context in which one's attention would be focussed on the syntactic rules involved in placing appropriate suffixes on verb stems...Thus, discrete point tests cannot be pragmatic..." (Oller 1979a:38)

It is difficult to imagine an ordinary discourse situation in which a language user might be asked to write down a text that he can hear. An exception might be a secretary or a court clerk taking minutes, i.e. not really what would be considered ordinary discourse situations, despite Oller's claim that

> "...dictation and closely related procedures probably work well precisely because they are members of the class of language processing tasks that faithfully reflect what people do when they use language for communicative purposes in real life contexts." (Oller 1979a:267)

Although Oller contends that the writing down in a dictation task is simply to show that discourse processing has occurred, the fact remains that it does not reflect real-life communicative situations. The processing of narrative input in itself is not an unusual language activity (e.g. listening to a discussion on the radio etc.), writing down sequences of text spoken or read out aloud seems more like a somewhat contrived and artificial testing situation than actual communication. It is limited in the range of communicative situations it can evaluate.

One of the most important characteristics of a good test of language proficiency according to Oller's theory requires such a test to be pragmatic, i.e. it has to relate linguistic with non-linguistic features. Only a pragmatic language test activates the language users' grammar of expectancy. Oller (1979a) has characterized language proficiency as "consisting of such an expectancy generating system". Proficiency tests that are valid with regard to the theoretical construct of an expectancy grammar as proposed by Oller (see discussion in Chapter Three) have to meet two "pragmatic naturalness criteria". Oller (1979a) has explained these two criteria as follows:

- tests have to challenge learners' ability to process (produce, comprehend, store, recall) texts of language as produced under normal performance constraints (Oller does not refer to performance constraints but rather to linguistic and extralinguistic contextual constraints)
- tests must require the learner to understand the interrelationship between situational and linguistic contexts.

It is uncertain how dictation tasks challenge learners' ability to process texts of language as produced under normal performance constraints. Unless texts consist of recorded actual conversations, they are likely to be presented faultlessly by the examiner or as a taped version. Furthermore, texts are expected to be presented three times at predeter-

mined intervals. Again, it is difficult to envisage a parallel to this in real-life discourse situations.

The second of Oller's criteria is put in doubt by Alderson's (1978) hypothesis that linguistic competence is the prime factor in successful completion of dictation tests. If only linguistic competence is needed, the candidates' understanding of the interrelationship between linguistic and situational contexts are not measured. If more than linguistic competence is required, the extent to which pragmatic clues are used would depend largely on the type of text chosen.

5.3. Written cloze tests

5.3.1. Background

Cloze tests were originally used with native speakers by Taylor (1953) to determine the level of difficulty of different reading materials. The technique was later adapted to measure reading skills in both native and second/foreign languages (Anderson 1971, Davies 1975, Porter 1976[9], Potter 1968, Ruddell 1965, Weaver and Kingston 1963 and others). Oller (1972a, 1973c 1976a, 1976c) has conducted extensive research in the area of second language cloze testing and considers the procedure a valid indicator of examinees' overall ability. In 1973 he claimed:

> "One of the most promising types of integrative skills tests which has been proposed for measuring either achievement or proficiency in foreign language or second language situations is the cloze test." (Oller 1973b:192)

And in 1979, justifying the above claim:

> "Because of the fact that cloze items are usually scattered over an entire text on some fixed or variable ratio method, cloze tests are generally tests of discourse level processing." (Oller 1979a:346)

Oller's point of view is consistent with Hinofotis (1976) who stresses the efficiency and the validity of the cloze procedure in measuring language ability.

Traditionally, a cloze test is intended as an integrative measure. This seems to have been Taylor's (1953) intention

> "...cloze procedure deals with contextually interrelated series of blanks, not isolated ones...the cloze procedure does not deal directly with specific meaning. Instead, it repeatedly samples the likeness between the language patterns used by the writer to express what he meant and those possibly different patterns which represent readers' guesses at what they <u>think</u> he meant." (Taylor 1953:417, his emphasis)

However, some researchers have tried to adapt the cloze procedure for discrete-point tests of vocabulary, structure or morphology. Oller and Inal (1971) report on a cloze test designed to test students' understanding of prepositions. Davies (1975) described a study in which a cloze test was prepared with certain grammatical categories of words deleted. This type of procedure results in a fill-in-the-blank test or, as Taylor (1953:417) put it, in a "sentence-completion test". The purpose of a cloze test as a test of communicative ability is to test the examinees' understanding of discourse level constraints in addition to structural constraints within a sentence, i.e. a cloze test is meant to be integrative.

5.3.2. <u>Administration procedure</u>

There are several different cloze techniques. According to Oller (1979a) the most widely used and best researched technique is the fixed-ratio method. The fixed-ratio method is used in the evaluation below (5.3.4.) and described in the following. Construction of a fixed-ratio cloze test simply requires deletion of every nth word (5th, 6th or 7th mostly) from a passage of

prose. Students are asked to restore the passage by replacing the missing words. Approximately two sentences at the beginning and the end are usually left intact to allow readers to get a feel for the style and subject matter (although Klare, Sinaiko and Stolurow 1972, argued that this is not necessary). A passage length that allows about 50 items is considered the most profitable in terms of analyzable data (Oller 1979a).

One of the following methods of scoring is usually adopted:

- Exact replacement; only the exact words used in the original passage are counted as correct. Taylor (1953) argued that the exact replacement technique would indicate the degree of correspondence between the writer's and the examinee's language system.
- Acceptable substitutes; in addition to exact replacements, any substitute that is contextually acceptable is counted as correct. This technique would seem to allow a less rigid evaluation, taking into consideration Chomsky's (1957, 1965) creative aspect of language. Morrow's (1977) observation that a normal communicative situation presupposes unpredictable (but acceptable) input from all interlocutors involved would seem to favour an evaluation allowing acceptable substitutes. Some doubts about the level of validity of the acceptable word method of scoring remain. Oller (1972b) found that acceptable substitute scoring was better than exact word only for the purpose of validating correlations. Haskell (1973, quoted in Davies 1978) however compared exact replacement with acceptable substitutes methods and found the verbatim methods were best. Irvine, Atai and Oller (1974) seemed to agree with this conclusion, i.e. they achieved better validating correlations with the exact scoring method. More research seems to be required in order to clarify the situation.

5.3.3. Justification

Several studies (Darnell 1968, Anderson 1976, Conrad 1970)

indicated that cloze test scores correlated very highly with results obtained from tests of general language proficiency. Results from these and additional studies have led Oller to claim

> "...the cloze technique has proved to be a very useful measure of language proficiency." (Oller 1979a:354)

Oller has repeatedly advocated cloze tests because they fulfill, in his opinion, his naturalness criteria:

- They test the efficiency (and extent) of the internalized expectancy grammar of individual learners. In replacing the missing items each learner indicates how successful he has been in developing his expectancy grammar. (It should be pointed out that this only applies if the appropriate replacement method of scoring is used. If the exact word method of scoring is used, results are likely to simply indicate how the learners' grammar of expectancy differs from that of the writer of the text.)
- They use texts from prose in the target language, thus reflecting the real world rather than the often artificial context of the classroom.

According to Oller (1979a) another point in favour of cloze as a test of general language proficiency is provided by Kirn's (1972) study, in which it was shown that cloze tests, like dictation, resist practice effects.

5.3.4. Evaluation

It is assumed that the texts to be chosen for cloze tests reflect language as it is used and that they have communicative value.

Table 5.1. shows how cloze tests rate when compared with Morrow's (1977) features of communicative situations.

Interaction-basis

Cloze tests can be considered interaction-based in the sense that the writer of the piece of prose chosen for the test has something to communicate to readers. Testees have to read the text and restore it. Cloze tests, accordingly, evaluate the extent to which examinees understand what is being communicated to them. They are not required to produce language utterances to indicate how they judge a communicative situation.

Unpredictability

Creativity is involved because students have to supply appropriate items according to their experiences of language and the world. Cloze tests are rarely administered under time constraints, thus students do not need to process data in real time. However, there are not normally time constraints imposed on reading tasks in the real world either.

Contextualization

Cloze tests provide the opportunity to present appropriate language in an appropriate context. There is some disagreement among researchers as to how far cloze actually evaluates contextual understanding. MacGintie (1961) studied cloze tests constructed from prose materials up to 144 words in length to determine how far the context on either side of blanks affects their restoration. He concluded that only a context of up to five words has any marked influence on missing items being replaced. However, he did not claim that constraints never operate over more than four or five words but hypothesized that knowledge of the subject matter was likely to play a greater role in learners' ability to replace the missing item (which would, in fact, be an extralinguistic factor). Coleman and Miller (1967) reached a similar conclusion to that of MacGintie, i.e. they maintained that contexts beyond five words had very little effect on cloze scores. However, Oller (1975) and Chihara, Oller,

Weaver and Chavez (1977), using different versions of the "cut and scramble" technique, have shown that cloze passages whose sentence order is changed are more difficult to reconstruct than passages presented in sequence. They concluded that cloze tests do indeed test context beyond the sentence level. There does not seem to be a clear answer to the question of the extent to which cloze measures context beyond the sentence level and further research is required to clarify this point.

Purpose

Teachers using cloze tests found that, as for dictation tasks, examinees writing cloze tests are unlikely to complete the task successfully if they do not understand what the text is about. However, it is not clear how understanding of the propositional context of an utterance is measured in a cloze test. Furthermore, examinees' ability to use language to serve their own purpose, i.e. to talk or write about their own topics, is not challenged.

Performance-basis

Theoretically, a cloze test format allows the presentation of non-idealized language, i.e. false starts, repetitions etc. could be incorporated. A superficial survey of cloze test examples given by researchers mentioned in this section indicates, however, that the majority of cloze tests used do not challenge examinees' strategies for language perception under performance constraints.

Authenticity

This criterion depends largely on the preceding criteria having been met successfully. Furthermore, a cloze test constructed according to the above criteria for text selection is unlikely to present simplified language to allow for examinees' imperfect command of the language. It would be

possible to have an authentic text but an unauthentic presentation.

Behaviour-basis

As pointed out in 5.2.4., this criterion should be judged on the basis of whether or not an interaction was successful. Presumably, the communication has been successful if testees are able to replace the blanks with appropriate lexical items. It is not clear, how the ability to fill in missing items of a cloze test corresponds to any real-life discourse situation examinees are likely to encounter. Alderson (1978) considered cloze a test of linguistic skill, thus it seems to be a competence-oriented test in Neufeld's sense (as discussed in 4.3.). The fixed-ratio method of cloze construction requires every nth word to be deleted, the variable-ratio method calls for other deletion criteria, e.g. words judged to be semantically important (Oller 1979a). Neither of these procedures seems to correspond to a naturally occurring phenomena in communicative situations.

5.3.5. Discussion

Based on Morrow's (1977) characteristic features of communicative situations, cloze tests are rather weak measures of communicative ability. Evaluated against the qualities Oller requires of a good test, this conclusion is confirmed. For example, the same quality Oller required with regard to discrete-point tests, i.e.

> "...the most serious disadvantage of discrete-point tests in general is they fail (in most cases) to reflect language use." (Oller 1973b:189)

seems to be lacking in cloze tests. Replacing missing items in a text seems no less artificial than many of the discrete-point items he has criticized. Because of this it can be concluded

that cloze tests suffer from the same deficit many other tests proposed to evaluate communicative ability or overall proficiency do: they are restricted-answer (i.e. examinees produce little if any language) paper-and-pencil tests and as such they are unable to evaluate communicative ability. In this sense they fail to meet another of the qualities Oller seeks in good tests. In 1979 he stated that

> "...the learner must be motivated to communicate with people in the target language." (Oller 1979a:32)

This criticism of cloze tests as mesures of overall proficiency is supported by Garcia-Zamor and Krowitz (1975), who pointed out that cloze tests fail to test performance in communicative situations. They evaluate primarily linguistic competence, but in an integrative manner.

The usefulness of cloze tests as measures of communicative ability (or proficiency) is further put in doubt by MacGintie (1971). He argued that cloze tests do not require a true understanding of the text. He considered them more suitable as measures of knowledge of redundancy. Rankin (1974) argued along similar lines, suggesting that cloze is best used as a measure of readability rather than language proficiency. Alderson (1978) has concluded that Carroll's claim (Carroll and Freedle 1972) that cloze depends largely on local redundancy is supported in the literature. He has suggested that cloze is associated with linguistic competence as tested by discrete-point tests.

Based on the evaluation criteria set forth in 4.4.1. - 4.4.7. it has to be concluded that dictation and cloze tests are not valid measures of overall language proficiency as they fail to adequately measure testees' communicative ability. Neither type of test challenges candidates' ability to integrate messages they have decoded successfully into an appropriate response. It

can further be concluded from the above discussion that if dictation and cloze tests are evaluated against statements made by Oller in the articles cited, they have to be rejected as measures of communicative ability as they do not meet the set objectives.

5.4. Test of English as a Foreign Language (TOEFL)

5.4.1. Background

The TOEFL was designed during the early 1960s, under the influence of discrete-point philosophy predominant during the psychometric-structuralist trend in language testing. It was never based on contrastive analysis, as its designers always intended it to be used for candidates of heterogeneous language backgrounds. The TOEFL is administered by Educational Testing Services (ETS), Princeton, N.J. According to Gruber and Gruber (1977) it is the most widely used standardized test of foreign language proficiency. It is primarily intended to measure proficiency of non-English speakers seeking admission to post-secondary academic institutions in North America.

The TOEFL was first used internationally in 1964 (Oller and Spolsky 1979). Since then it has undergone relatively minor changes only. As a result of a study carried out by Pike (cited in Oller and Spolsky 1979) some revisions undertaken involved the inclusion of more functionally oriented items. However, the TOEFL still embodies the key-points of discrete-point philosophy mentioned in 1.2.3. The TOEFL is a very large scale project and represents a considerable financial investment. Kaplan (1969) considered this the reason for the relatively slow and minor changes that have been made on the test's basic format despite rapid and often drastic changes in linguistic theory.

5.4.2. Administration procedure

The TOEFL is administered periodically in designated centres. The whole test (consisting of several sub-tests) is administered in one session lasting two hours. A testing manual, published by ETS, insures that standard administration procedures are followed.

Appendix A shows what Gruber and Gruber (1977) considered a typical TOEFL format and typical TOEFL items for individual sub-tests.

5.4.3. Justification

The TOEFL was designed according to discrete-point philosophy. It reflects the view of language predominant under the influence of the psychometric-structuralist trend in language testing (as discussed in 1.2.1.), i.e. language is thought to consist of individual components which can be tested separately. Mastery of these components is assumed to equal mastery of the language in question. The most important features of a language test are thought to be high validity and reliability. One way to achieve the latter is through a large sample of behaviours, therefore a short answer format, including short, unrelated sentences, is desirable. Students are thus asked to mark appropriate boxes (i.e. they never actually produce language) to facilitate scoring procedures.

5.4.4. Evaluation

Figure 5.1. summarizes how the TOEFL rates when evaluated on the basis of Morrow's (1977) characteristic features of communicative situations.

Interaction-basis

Most of the language presented in the TOEFL cannot be considered interaction-based. Unlike normal communicative behaviour,

sentences tend to be unrelated and devoid of context. There is no apparent communicative value in the sentences used. Furthermore, candidates have no input, i.e. they do not produce language. The sentences used in Gruber and Gruber's examples seem to be representative of a testing rather than a communicative situation.

Unpredictability

Morrow's (1977) major factors contributing to unpredictability in communication, namely creativity and language processing in real time, do not seem to be involved in the tasks set by the TOEFL format. Although the total amount of time allowed for writing the test is limited, there is ample opportunity to read and re-read questions.

Contextualization

Contextualization of materials used in the TOEFL is primarily at the sentence level. The sentences are not usually presented in an appropriate context. Accordingly, they do not seem to relate to the real world, i.e.

"Would you like to go.....a ride?"

might involve - some type of vehicle, or
- a metaphorical sense

similarly

"They were here, but they've gone back to.....apartment"

is unclear because "they" does not seem to refer to anybody in particular, no previous subject is given. As none of the items presented in Gruber and Gruber (1977) exceed paragraph-length (approximately 150 words) and most of them are considerably shorter, little opportunity is given for linguistic or situational context.

Purpose

Due to the multiple choice format of the test the purpose and the topic of given items often remain unclear. Comprehension of propositional context is not evaluated and examinees' ability to use language to serve their own communicative needs is not measured.

Successful completion of an item on the vocabulary subtest frequently presupposes comprehension of one particular word rather than of a general topic, e.g.

"That doesn't make sense; I think your reasoning is fallacious"

with the possible answers: A_1 : faulty
A_2 : licentious
A_3 : pretentious
A_4 : officious

Examinees do not have the opportunity to try and find out meaning from context. Normally, successful communication does not depend on the comprehension of one single word but on comprehension of utterances in a particular context.

Performance-basis

There is little evidence of language as it occurs in normal discourse in the sample TOEFL items studied (samples from Gruber and Gruber 1977 as shown in Appendix A). A normal communicative situation would allow testees to mobilize their communicative strategies. They could try to understand individual meanings on the basis of global meaning, likely to be discernable from the context. Such strategies are not challenged within the present framework of the TOEFL.

Authenticity

Although individual sentences used in the test could feasibly have been uttered by native speakers, authenticity is lost due to the use of unrelated statements with the result that neither content nor presentation is authentic. Furthermore, the vocab-

ularly used in the test items is controlled in order to keep the difficulty level down (Oller and Spolsky 1979). According to researchers cited in 4.4.6. such procedure does not reflect real-life communication.

Behaviour-basis

The TOEFL primarily assesses candidates' ability to recognize and identify correct forms of their target language. The aim of communicative testing, however, is an assessment of what the learners can achieve through the use of their target language. The TOEFL requires testees to process language produced by an unidentified writer or speaker, without allowing them any input of their own. Testees do not produce any language and accordingly they cannot achieve anything through the use of the type of language presented in TOEFL items. Thus, the test does not evaluate what it claims to evaluate, i.e. overall proficiency, which would have to include the ability to produce utterances.

5.4.5. Discussion

To conclude the evaluation of the TOEFL it should be pointed out that, due to the nature of discrete-point tests, the TOEFL does not reveal anything about the developmental stage of individual examinees in their second/foreign language. However, this was identified as being one of the aims of language testing in an integrative-sociolinguistic framework of testing (1.3.3.). As the TOEFL is standardized against a population of non-native speakers (Oller and Spolsky 1979), the TOEFL score simply indicates how well each testee succeeded in relation to other non-native speakers. Nothing can be said about testees' level in terms of native speakers. The total score for each examinee is used to determine his level of proficiency, i.e. his overall proficiency is inferred from the number of correct test items answered. This inferred language proficiency also implies the

testees' level of communicative ability, although as the above showed, this ability was not tested. Based on the preceding evaluation, it appears safe to conclude that the TOEFL does not measure global proficiency but linguistic proficiency. Thus, it is primarily a test of linguistic competence.

5.5. The Michigan Proficiency Test Battery (two subtests)

5.5.1. Background

The Michigan Proficiency Test Battery is, like the TOEFL, a product of the psychometric-structuralist trend in language testing. It was prepared at the English Language Institute (ELI) of the University of Michigan. According to the Manual (1977) produced by ELI the purpose of the test battery is to estimate whether a non-native speaker's command of English is sufficient to allow him to pursue academic studies at an English college or university. If an examinee is thought to be sufficiently proficient in English for this purpose, the test results are further expected to indicate how much study the examinee is likely able to handle at his present level of proficiency.

5.5.2. Administration

The Michigan Proficiency Test Battery is administered by the Testing and Certification Division of the English Language Institute as well as by authorized subscribers. Administration procedures are regulated through guidelines published by the Testing Service.

The Michigan Proficiency Test Battery consists of three different subtests:

- The Michigan Test of Aural Comprehension, i.e. 90 recorded problems with written answer choices,
- Composition, i.e. an impromptu composition, to be written within 30 minutes, on an assigned topic,

- The Michigan Test of English Language Proficiency, i.e. a 100 item multiple choice test with sections on grammar, vocabulary and reading comprehension.

The Michigan Test of Aural Comprehension will not be evaluated below because of its relative infrequent administration. The Michigan Test of English Language Proficiency will be referred to as MT, the Composition as COM. Appendix B shows examples of an MT as provided in the Manual (1977).

5.5.3. Justification

Proponents of the Michigan Proficiency Test Battery justify its use as a test of overall proficiency on the basis of prevailing theories during the psychometric-structuralist trend in language testing, i.e. similar to the TOEFL. The Michigan Proficiency Test Battery differs from the TOEFL in that it includes a composition, thus requiring examinees to produce language. According to the test Manual (1977) composition writing is included to indicate how well students are able to indicate their knowledge of subject matter in a course examination type situation. It is not intended as a comprehensive test of English writing ability. Similarly, the aural comprehension subtest is included to evaluate students' comprehension of "basic English Structures" (Manual 1977:1) presented orally as in a lecture type situation. Furthermore, some of the reading selections included are up to 350 words long and thus more integrative than the shorter reading selections included in the TOEFL.

The Michigan Proficiency Test Battery is further advocated due to its high reliability. The test Manual (1977) draws attention to the fact that reliability is high for homogeneous as well as heterogeneous groups and at all levels of proficiency. Except for the COM subtest, students do not produce utterances of their own. They have to identify and mark the correct item. This procedure facilitates scoring and increases reliability.

5.5.4. Evaluation

Figure 5.1. summarizes how the two Michigan subtests rate when evaluated on the basis of Morrow's (1977) characteristic features of communicative situations.

Interaction-basis

The COM subtest (Section I) is interaction-based because examinees communicate their thoughts on an assigned topic. In doing so they have to take into consideration the evaluator's reaction to their communication. This involves deciding what to communicate as well as how to do it.

The MT subtest cannot be considered interaction-based as it involves mostly unrelated short sentences that appear to be examples of the linguistic code rather than of communicative situations. Part III (reading comprehension) is interaction-based to some extent. There are, however, few examples long enough to really meet the criterion as defined in 4.4.1.

Unpredictability

The COM subtest meets this criterion at least partially because it involves creativity in the sense explained in 4.4.2. The remaining sections do not seem to fulfill this criterion because the candidates do not have to produce any language. Furthermore, they are not challenged in their ability to process utterances in real time.

Contextualization

The COM subtest allows for both linguistic and situational contextualization. The remaining subtest sections are contextualized primarily at the sentence level. Part I of Section II involves partial conversations. Two hypothetical interlocutors utter one sentence each ~~in~~ each question. Answers have

to be made on the basis of what is correct linguistically rather than what is appropriate situationally. Part III involves texts of varying lengths (approximately 250 - 300 words each), which are likely to measure candidates' understanding of situational as well as linguistic context to some extent. Context of situation and participant variables however appear to have been incorporated in the test items to a very limited extent only and primarily by implication.

Purpose

Candidates' understanding of purpose, i.e. the reason why an utterance was made, and topic are evaluated in the COM subtest. The MT subtest's items are often unclear with regard to their purpose and topic. Candidates are, in addition, not tested in their ability to use appropriate language for their own purposes.

Performance-basis

Section I (COM subtest) allows for language as it occurs in life-like situations, e.g. letter writing, composition for the purpose of evaluation in academic courses etc. , although it should be added that the language functions that can be tested in this format are limited. The MT subtest does not show any evidence of this characteristic of communication. The candidates' communicative strategies are not evaluated.

Authenticity

Except for the COM subtest, authenticity is lost for the same reasons as those cited in 5.4.4. (TOEFL).

Behaviour-basis

The COM subtest evaluates candidates' ability to do something with their target language, namely communicate their ideas on a particular topic. Examinees are given the opportunity to use language to serve their communicative needs, although the composition

format can only evaluate a limited range of language functions (e.g. appropriate style etc.) as was pointed out above. The ability to incorporate an interlocutors' utterance into their development of the assigned topic is not evaluated in the tasks presented. The MT subtest requires examinees to process language produced by an unidentified writer or speaker (like TOEFL). They have no input of their own and, accordingly, they cannot achieve any communicative goals with the test items.

5.5.5. Discussion

It can be concluded that the COM subtest is the only part of the two subtests evaluated to measure candidates' communicative ability, although not all aspects of this ability are tested. The MT subtest is concerned with the linguistic code, i.e. linguistic competence and, as such, fails to evaluate communicative ability.

5.6. COMTEST (Communication Test)

5.6.1. Background

The COMTEST was designed by Palmer (1972), with time and facilities provided by the Testing and Certification Division at the English Language Institute of the University of Michigan. The purpose of the study that produced the COMTEST was to find out if the Michigan Proficiency Test Battery measures oral communication ability in non-native speakers of English. Another purpose of the study was to determine if such an ability could be evaluated objectively and reliably. This discussion is concerned with the COMTEST as a potential measure of communicative ability only.

The basic aim of the COMTEST is to evaluate examinees' ability to obtain and supply information. This contrasts with more tradi-

tional language tests which require examinees to recognize correct (or sometimes appropriate) forms or to answer questions.

One problem facing the test designer was to evaluate candidates' performance, not a combination of their skill and that of the examiner combined. Palmer prepared a total of three tests: COMTEST-2 and COMTEST-3 are revised and improved versions of COMTEST-1 in the sense that the test situation is more strictly controlled, thus allowing a more objective evaluation of the examinees' performance. The difference between COMTEST-2 and COMTEST-3 is in the type of pictures used, i.e. COMTEST-2 uses all stick drawings (Appendix C) whereas COMTEST-3 uses five stick drawings and five pictures cut out from magazines.

5.6.2. Administration Procedure

Each testee was shown a set of four stick drawings at a time of the type illustrated in Appendix C. The complete test consisted of a total of ten sets of drawings. Each candidate was tested individually, one examiner was required for the administration of the test. The examiner and the candidate were each seated on opposite sides of a desk to avoid the possibility of visual clues. Each of them was provided with a copy of the drawings. Each candidate was told that the examiner had one particular drawing of a set of four in mind. In order to avoid variability in response time due to the order of questioning, the examiner did not actually have a particular drawing in mind in COMTEST-2 and COMTEST-3. The picture identified with the candidate's third question was accepted as being "correct". In order to standardize the test administration sessions (in particular the examiner's replies and questions) as much as possible, guidelines for examiner responses were prepared (Appendix C).

All test administrations were taped and later the tapes were

played back in order to tabulate the scores. For COMTEST-1 the following scores were tabulated for each candidate on each problem:

- Elapsed time until the correct answer was reached
- Number of times the examiner couldn't understand the question and had to ask for clarification
- Number of times that the examiner misunderstood a question and answered it, causing the subject to receive incorrect information
- Number of times the subject gave up
- Number of questions until the correct answer was obtained
- Number of guesses

(Palmer 1972:37)

For COMTEST-2 and COMTEST-3 the time required for the subject to "identify" the criterion drawing was the only variable being measured. Palmer (1972) felt that the length of time required to complete COMTEST-2 was "more systematically related to communication ability than was true of COMTEST-1".

5.6.3. Justification

Palmer (1972) considered COMTEST-2 "a useful and practical measure of oral communication ability" because it measures factors not measured by standardized proficiency tests such as the Michigan Proficiency Test Battery. He observed that measuring the time each candidate needed to complete the task required was an adequate evaluation criterion. If a candidate could not make himself understood easily, or if he was unable to understand what was said to him, he needed more time than a more proficient candidate, thus obtaining a lower score. Palmer does not claim that a testing instrument like COMTEST-2 should replace the more traditional forms of proficiency testing. He seems to think of it in terms of another subtest in a test battery.

5.6.4. Evaluation

Figure 5.1. summarizes how COMTEST-2 rates when evaluated on the basis of Morrow's (1977) characteristic features of communicative situations.

Interaction-basis

COMTEST-2 is interaction based and it involves the interlocutors' receptive and productive communication abilities. COMTEST-2 involves the type of communicative activity that requires interlocutors to decode an utterance and to integrate relevant parts into an appropriate response.

Unpredictability

COMTEST-2 is unpredictable because it involves the creative element in language. Furthermore, candidates need to process data in real time to avoid communication breakdowns. The task set is somewhat like a puzzle and likely to motivate candidates to want to find out the "correct" drawing, at least in the early stages of the test. (Palmer 1972 does not report student response to the test, i.e. perceived face validity of the test.)

Contextualization

The task set allows communication to take place in a specific context that determines appropriacy of specific language forms used. Thus, the situational context is involved. COMTEST-2 allows, in theory, for an evaluation of extralinguistic features too. However, Palmer (1972) does not seem to consider these as important factors of communicative ability. He stated

> "This [being seated on opposite sides of the desk] made it natural for the examiner to look at the pictures and ignore the subject's gestures." (Palmer 1972:38)

In an effort to eliminate as many subjective elements as pos-

sible in his tests, Palmer seems to have chosen to ignore one of the most natural features of oral, face-to-face communication (Savignon 1972 actively encouraged her students to make use of facial expressions, gestures etc.)

Purpose

COMTEST-2 seems to meet the criterion set with regard to purpose. Candidates need to encode utterances appropriately in order to succeed at the task given. They have to recognize the topic of specific units of discourse or they will not be able to respond appropriately.

Performance-basis

Candidates are given an opportunity to demonstrate their ability to communicate in an actual performance test.

Authenticity

Theoretically, the test encourages authentic language. There is, however, a danger that the examiner might try to "simplify" his utterances. The test seems to lose somewhat in authenticity because the exchanges about each set of drawings are likely to be quite short. The quick succession of ten different sets of drawings, producing ten different topics, is not likely to be representative of life-like communication. Thus, authenticity of content as well as presentation seem weakened in COMTEST-2.

Behaviour-basis

The test is designed to evaluate learners' ability to get meaning across and the testees are, in fact, evaluated on the basis of how successful they are at getting their meaning across effectively. More proficient examinees are likely to get their intended meaning across more quickly, thus needing less time and scoring better than less proficient learners.

5.6.5. Discussion

COMTEST-2 is able to meet the criteria set forth for the purpose of the evaluation of communicative tests, at least in theory. Extralinguistic factors (e.g. gestures) could be taken into consideration and authenticity could probably be increased through

- appropriate training of the examiners (as shown by Clark 1975, 1978 and others)
- a modification of the test that allows a less frequent change of topic and thus a more realistic communicative situation.

Although COMTEST-2 seems to evaluate examinees' communicative ability, the testing situation and the task itself seem rather artificial. A testing situation with its own rules undoubtedly increases objectivity with regard to evaluation, but it is not clear how a candidate's success in such a task can be generalized to predict possible success or failure in the performance of other communicative tasks. Furthermore, it has been shown by Clark (1978b) that oral interview type tests of the length of COMTEST-2 (approximately five minutes for test administration) are too short to draw any useful conclusions about candidates' communicative ability. Thus, it can be concluded that COMTEST-2 is an interesting experimental test but, at least in its present form, not suitable to evaluate second/foreign language learners' communicative ability.

6. SUMMARY AND CONCLUSIONS

6.1. Summary

1) Three major trends in second/foreign language testing i.e.

- the pre-scientific trend
- the psychometric-structuralist trend
- the integrative-sociolinguistic trend

were discussed. Close connections between testing techniques used, views of language, learning and teaching were shown to exist.

2) Chomsky's competence/performance dichotomy was discussed and the subsequent survey of some of the most influential critiques showed that most theorists working in the field considered the proposed framework for the dichotomy too restrictive.

3) A discussion of some of the most influential theories of communicative competence showed that various researchers have different views of what is meant by communicative competence, i.e. knowing a language.

4) Knowledge of a language was shown to be most evident in the language users' communicative ability. It was concluded that this ability should therefore be measured by tests claimed to evaluate overall proficiency. Based on the characteristic features of a communicative situation outlined by Morrow (1977) a number of criteria to be met by communicative tests were set forth.

5) Based on these criteria an evaluation of two tests representing the integrative-sociolinguistic trend in second/foreign language testing, and two standardized tests, representing the psychometric-structuralist trend, was carried out. An experimental test of communication (COMTEST, Palmer 1975)

was also evaluated. The first four tests are currently used widely and claim to evaluate overall language proficiency, part of which is communicative ability.

6.2. Results

It is generally accepted that communication is the most important function of language. This point of view is reflected throughout the literature and was adopted for the purpose of this thesis. Consistent with this approach is the assumption that second/foreign language tests should evaluate learners' ability to communicate. On the basis of Morrow's (1977) characteristic features of communicative situations a number of evaluation criteria were set forth. These criteria were expected to indicate if the four currently used tests of language proficiency (which is seen to be more inclusive than communicative ability) do in fact measure candidates' communicative ability. On the basis of the evaluation it was shown that none of the four tests measures candidates' communicative ability because

- the tests do not meet the criteria set
- the tests do not reflect current views in linguistics
- the tests do not evaluate what they claim to evaluate.

In fact, they do not measure global but rather only linguistic proficiency, i.e. they are primarily tests of linguistic competence.

These testing instruments only infer candidates' communicative ability, which was established to be part of overall language proficiency and, therefore, cannot be considered tests of overall language proficiency. It is concluded that the above four testing instruments do not measure what they claim to measure. Thus it is shown, in this thesis, that a considerable difference exists between competence and performance (i.e. communicative competence and communicative ability/performance), but that this difference is not given enough consideration in second/foreign language testing at the present time. In past testing practices

language, i.e. knowledge of the linguistic code, had been considered the target of second/foreign language testing and communication was proposed as the medium to achieve this target. Within a communicative framework, however, the target of second/foreign language testing is communication, while language, i.e. the linguistic code, is the means to reach this target. As Schlissinger (1977) pointed out: language testers have to learn to test communication (as it occurs in real-life) rather than knowledge of the code.

6.3. Suggestions for further research

This thesis on the difference between competence and performance in second language testing gives rise to a number of questions. Some of those directly relevant to the results of this thesis are:

a) How do results learners achieve on cloze (or dictation) tests relate to results they achieve on a test of communicative ability? Or, in more general terms: How do indirect test results relate to more direct ones? Can any predictions be made or is there no discernable relationship between the two?

b) What is the relationship between learners' success in one communicative situation and their success (or failure) in another communicative situation? Can accurate predictions be made? If research should show that extrapolation is not possible, the concept of communicative testing would have to be reevaluated.

c) Is communicative ability primarily a function of language or personality?

Some suggestions for further research in specific areas relevant to communicative ability were mentioned during the course of the preceding discussion, for example:

d) Within the framework of communicative competence:
What is the relationship between the individual components of

communicative competence?
Is there a more satisfactory way to account for psycholinguistic factors and performance strategies than has been proposed so far?

e) Within discourse analysis:
What are the rules of discourse? Should they be accounted for within the sociocultural or within the grammatical component? How are they to be formalized in order to be of help to the language learner?

f) Within Oller's pragmatic theory:
Can a pragmatic expectancy generating device do what it is claimed to do and remain practical?

FOOTNOTES

1 (page 9)

Second or subsequent editions of audio-lingual texts tend to have adopted a more eclectic approach, combining meaningful learning, elements of cognition and habit formation.

2 (page 11)

Some linguists were developing ideas along very different lines. Chomsky (1965) published a theory that is fundamentally opposed to the linguistic theory Lado's work is based on.

3 (page 24)

Humboldt (1836) postulated the concept of "Innersprachen". This refers to a type of inner form of language that informs and energizes language behaviour. The outer form is the actual sentence. Chomsky (1965) draws a parallel between Humboldt's inner and outer form and his own distinction between deep and surface structure.

4 (page 62)

Chomsky (1976b) draws attention to "organizing one's thoughts and non-verbal thinking" as examples of non-communicative uses of language. It could, however, be argued that both are examples of forms of communication with self.

5 (page 68)

Paulston (1974) used the term "communicative performance" in her model for language teaching, but considers her "analogic extension" a case of "conceptual confusion" (personal communication).

6 (page 78)

This could be taken to suggest that discrete-point tests

should be used at the very early stages of second/foreign language learning. Tests would gradually become more integrative. Such a procedure would reflect models of first language acquisition as put forth by Moskowitz (1978), Bloom (1973) amongst others.

7 (page 93)

Conversational speed has been defined in research as:
4.4 - 5.9 syllables per second by Goldman-Eisler (1968)
6.7 - 8.2 syllables per second by Miller (1962)
6± 1 syllables per second by Lenneberg (1967)

8 (page 94)

Oller's "naturalness criteria" are defined as follows:
- language tests "must cause the language learner to process temporal sequences of elements in the language that conform to normal contextual constraints (linguistic and extralinguistic contexts)"
- language tests "must require the learner to understand the pragmatic interrelationship of linguistic contexts and extralinguistic contexts" (Oller 1979a:33).

Oller's naturalness criteria are further discussed in 5.2.4.

9 (page 101)

Other researchers who experimented with cloze in a multiple choice format are Jonz (1976), Griffin et al. (1978)

REFERENCES

Agard, F.B. and H.B. Dunkel, 1948, An investigation of second language teaching, Boston, Ginn & Co.

Alderson, J.C., 1978, A study of cloze procedure with native and non-native speakers of English, unpublished Ph.D. thesis, University of Edinburgh

Anderson, J., 1976, Psycholinguistic experiments in foreign language testing, Santa Lucia, Queensland, University of Queensland Press

Anderson, J., 1971, Research on comprehension in reading, in Bracken & Malmquist (eds.) Improving reading ability around the world, Newark, Delaware, International Reading Association

Ausubel, D.P., 1968, Educational psychology: a cognitive view New York, Holt, Rinehart & Winston

Ausubel, D.P., 1967, Learning theory and classroom practice, Bulletin No. 1, Toronto, Ontario Institute for Studies in Education

Bloom, L., 1973, One word at a time, The Hague, Mouton

Brière, E.J., 1971, Are we really measuring proficiency with our foreign language tests, Foreign Language Annals 4

Brooks, N., 1964, Language and language learning: theory and practice (2nd edition), New York, Harcourt, Brace & World

Brooks, N., 1960, Language and learning, New York, Harcourt, Brace & World

Campbell, R. and R. Wales, 1970, The study of language acquisition, in J. Lyons (ed.), New horizons in linguistics, Harmondsworth, Penguin Books

Canale, M. and M. Swain, 1979, Theoretical bases of communicative approaches to second language teaching and testing, revised draft copy, Toronto, Ontario Institute for Studies in Education

Candlin, C.N., 1978, Discoursal patterning and the equalising of integrative opportunity, Paper read at the Conference on English as an International and Intranational Language, The East-West Center, Hawaii, April 1978, mimeo

Candlin, C.N., 1976, Communicative language teaching and the debt to pragmatics, in C. Rameh (ed.) Semantics: theory and application, Georgetown University Round Table on Languages and Linguistics, Georgetown University Press

Candlin, C.N. and M. Breen, 1980, Evaluating, adapting and innovating language teaching materials, in C. Yorio et al. (eds.) On TESOL '79, The learner in focus, TESOL, Washington

Carroll, B.J., 1980, Testing communicative performance: an interim study, Oxford, Pergamon Press

Carroll, B.J., 1978a, Specifications for an English testing service, London, British Council

Carroll, B.J., 1978b, Guidelines for the development of communicative tests, London, Royal Society of Arts

Carroll, J.B., 1974, Learning theory for the classroom teacher, in G. Jarvis (ed.) The challenge of communication, Skokie, Ill., National Textbook Co.

Carroll, J.B., 1972, Defining language comprehension: some speculations, in J.B. Carroll & R. Freedle (eds.) Language comprehension and the acquisition of knowledge, New York, Winston & Sons

Carroll, J.B., 1968, The psychology of language testing, in A. Davies (ed.) Language testing symposium, London, Oxford University Press

Carroll, J.B., 1967, The foreign language attainments of language majors in the senior year: a survey conducted in U.S. colleges and universities, Cambridge, Mass., Harvard University

Carroll, J.B., 1966, The contributions of psychological theory and educational research to the teaching of foreign languages, in A. Valdman (ed.) Trends in language teaching, New York, McGraw-Hill

Carroll, J.B., 1961, Fundamental considerations in testing for English language proficiency of foreign students, in Testing the English proficiency of foreign students, Center for Applied Linguistics, Washington, D.C.

Carroll, J.B., 1953, The study of language: a survey of linguistics and other related disciplines in America, Cambridge, Mass., Harvard University Press and London, Oxford University Press

Carroll, J.B., 1952, An appraisal of language tests from the standpoint of the psychology of language, Yearbook of the National Council on Measurements Used in Education, 1951-2, 9

Carroll, J.B. and R.O. Freedle, 1974, Language comprehension and the acquisition of knowledge, Washington, D.C., Winston and New York, Wiley

Chafe, W., 1974, Language and consciousness, Language 50

Chastain, K., 1976, Developing second language skills: theory to practice, second edition, Chicago, Rand McNally

Chavez, M.A., T. Chihara, J.W. Oller jr. and K. Weaver, 1977,

Are cloze items sensitive to constraints across sentences? II Paper read at the Eleventh Annual TESOL Convention, Miami, Florida

Chihara, T., J.W. Oller jr., K. Weaver and M.A. Chavez, 1977, Are cloze items sensitive to constraints across sentences, Language Learning 27

Chomsky, N., 1976a, Conditions on rules of grammar, Linguistic Analysis, 2

Chomsky, N., 1976b, On the biological basis of language capacities, in R.W. Rieber (ed.) The neuro-psychology of language, New York, Plenum Press

Chomsky, N., 1975, Reflections on language, New York, Pantheon Books

Chomsky, N., 1972, Language and mind, second (enlarged) edition, New York, Harcourt Brace Jovanovich Inc.

Chomsky, N., 1970, Remarks on nominalization, in R. Jacobs and P. Rosenbaum (eds.) Readings in English transformational grammar, Waltham, Mass., Ginn & Co.

Chomsky, N., 1967, The formal nature of language, in E.H. Lenneberg (ed.) Biological foundations of language, New York, Wiley

Chomsky, N., 1966, Topics in the theory of generative grammar, The Hague, Mouton

Chomsky, N., 1965, Aspects of the theory of syntax, Cambridge, Mass., MIT Press

Chomsky, N., 1964, Current issues in linguistic theory, The Hague, Mouton

Chomsky, N., 1957, Syntactic structures, The Hague, Mouton

Clark, J.L.D., 1978a, Psychometric considerations in language testing, in B. Spolsky (ed.) Approaches to language testing, Advances in language testing series: 2, Papers in Applied Linguistics, Center for Applied Linguistics, Arlington, Va.

Clark, J.L.D., 1978b, Interview testing research at Educational Testing Service, in J.L.D. Clark (ed.) Direct testing of speaking proficiency: theory and application, Princeton, N.J., Educational Testing Service

Clark, J.L.D., 1972, Foreign language testing: theory and practice, Philadelphia, Center for Curriculum Development

Coleman, E.B. and G.R. Miller, 1967, The measurement of information gained during prose learning, Reading Research Quarterly, 3

Conrad, C., 1971, The cloze procedure as a measure of English proficiency, Master's thesis, Abstracted in Workpapers in TESL: UCLA 5

Cooper, R.L., 1968, An elaborated language testing model, in J.A. Upshur (ed.), Problems in foreign language testing, Language Learning, Special Issue No. 3

Corder, S.P., 1973, Introducing applied linguistics, Harmondsworth, Penguin

Corder, S.P., 1967, The significance of learner's errors, International Review of Applied Linguistics, Vol. 5, No. 4

Coste, D., J. Courtillon, V. Ferenczi, M. Martins-Baltar, E. Papo, CREDIF and E. Roulet, 1976, Un niveau-seuil, Strasbourg, Council of Europe

Council of Europe, 1979, A European unit/credit scheme for modern language learning by adults, Report of the Symposium organized by the Government of the Federal Republic of Germany at Ludwigshafen-am-Rhein, 7 - 14 September 1977, Strasbourg, Council of Europe

Council of Europe, 1973, Systems development in adult language learning, Strasbourg, Council of Europe

CREDIL, 1975, Rapport synthèse de l'élaboration du test d'Anglais langue seconde, Québec, CIRB

Darnell, D.K., 1968, The development of an English language proficiency test of foreign students using a cloze-entropy procedure, ERIC ED 024 039

Davies, A., 1978, Language testing, part I and II, Language Teaching and Linguistics Abstracts, London, Cambridge University Press

Davies, A., 1977, The construction of language tests, in J.P.B. Allen & A. Davies (eds.) Testing and experimental methods, The Edinburgh Course in Applied Linguistics, Volume 4, London, Oxford University Press

Davies, A., 1975, Two tests of speeded reading, in R.L. Jones & B. Spolsky (eds.), Testing language proficiency, Arlington, Va., Center for Applied Linguistics

Davies, A., 1968, Language testing symposium, London, Oxford University Press

Dialogue Canada II, Part II, Language Bureau, Public Service Commission, Ottawa, 1976

Diller, K.C., 1971, Generative grammar, structural linguistics and language teaching, Rowley, Mass., Newbury House

Di Pietro, R.J., 1971, Language structures in contrast, Rowley, Mass., Newbury House

Educational Technology, 1979, Volume 14, June

Farhady, H., 1980, Preparing functional tests, TESOL Conference, San Francisco (draft copy)

Fillenbaum, S., 1971, Psycholinguistics, in P.H. Mussen & M.R. Rosenzweig (eds.) Annual Review of Psychology, 22

Fillmore, C.J., 1968, The case for case, in E.Bach & R.T. Harms (eds.) Universals in Language, New York, Holt, Rinehart & Winston

Firth, J.R., 1952, Papers in linguistics 1934 - 1951, London, Oxford University Press

Fodor, J. and M. Garrett, 1966, Some reflections on competence and performance, in J. Lyons & R. Wales (eds.) Psycholinguistic papers, Edinburgh, Edinburgh University Press

Fraser, B., 1974, An examination of the performative analysis, Indiana University Linguistics Club 1971, reprinted in: Papers in Linguistics 7

Freedle, R.O., 1977, Discourse production and comprehension, Vol. 1, Norwood, N.J., Ablex Publishing

Fromkin, V.A., 1971, The non-anomalous nature of anomalous utterances, Language 47

Fromkin, V.A., 1968, Speculations on performance models, Journal of Linguistics 4

Gaies, S.J., H. Gradman and B. Spolsky, 1977, Towards the measurement of functional proficiency: contextualization of the noise test, TESOL Quarterly 11

Garcia-Zamor, M. and M.J. Krowitz, 1974, Contextual testing and communicative competence, Paper presented at the Golden Anniversary Meeting of the Linguistic Society of America in New York (draft copy)

Goldman-Eisler, F., 1968, Psycholinguistics: experiments in spontaneous speech, London and New York, Academic Press

Goodman, K.S., 1968, The psycholinguistic nature of the reading process, Detroit, Wayne State University Press

Greene, J., 1972, Psycholinguistics: Chomsky and psychology, Harmondsworth, Penguin Bookds

Griffin, P., T. Dietrich and C. Freeman, 1978, Assessing comprehension in a school setting, in P. Griffin (ed.) Papers in Applied Linguistics, Linguistics and Reading Series 3, Arlington, Va., Center for Applied Linguistics

Gruber, G.R. and E.C. Gruber, 1977, Test of English as a second language, New York, Monarch Press

Habermas, J., 1970, Towards a theory of communicative competence, In H. Dreitzel (ed.) Recent sociology, No. 2, London, Collier-Macmillan

Halliday, M.A.K., 1978,Language as social semiotic, London, Edward Arnold

Halliday, M.A.K., 1973, Explorations in the functions of language, London, Edward Arnold

Halliday, M.A.K., 1970, Language structure and language function, in J. Lyons (ed.) New horizons in linguistics, Harmondsworth, Penguin Books

Hamp, E.P., 1968, What contrastive grammar is not, if it is, in J.E. Alatis (ed.) Georgetown Round Table 1968, Washington, D.C., Georgetown University Press

Harris, D., 1969, Testing English as a second language, New York, McGraw-Hill

Harrison, A., 1977, Test development: theory and practice, Bulletin pédagogique (langues vivantes), Vol. 51, No. 30, Novembre

Haskell, J.F., 1973, Refining the cloze testing and scoring procedures for use with ESL students, Ed.D. thesis, Columbia University

Hinofotis, F.B., 1976, Cloze as an alternative method of ESL placement and proficiency testing, Paper presented at the Annual Meeting of the Linguistic Society of America, Philadelphia.

Humboldt, W. von, 1836, Ueber die Verschiedenheit des menschlichen Sprachbaues, Berlin, republished: Darmstadt, Claasen & Roether, 1949

Hymes, D., 1972, On communicative competence, in J.B. Pride & J. Holmes (eds.) Sociolinguistics, Harmondsworth, Penguin Books

Hymes, D., 1967, Models of the interaction of language and social setting, in J. Macnamara (ed.) Problems of bilingualism, Journal of Social Issues, 23

Ingram, E., 1978, The psycholinguistic basis, in B. Spolsky (ed.) Advances in language testing series:2, Approaches to Language Testing, Arlington, Va., Center for Applied Linguistics

Ingram, E., 1977, Basic concepts in testing, in J.P.B. Allen & A. Davies (eds.) The Edinburgh course in applied linguistics: testing and experimental methods, London, Oxford University Press

Ingram, E., 1968, Attainment and diagnostic testing, in A. Davies (ed.) Language testing symposium, Language and Language Learning Series, 21, London, Oxford University Press

Irvine, P., P. Atai and J.W. Oller jr., 1974, Cloze, dictation and the Test of English as a Foreign Language, Language Learning, 24, 2

Jakobovits, L.A., 1970, Foreign language learning, Rowley, Mass., Newbury House

Jesperson, O., 1904, How to teach a foreign language, London, Longman

Johansson, S., 1973, An evaluation of the noise test: a method for testing overall second language proficiency by perception under masking noise, International Review of Applied Linguistics, 11

Johnson, K., 1977, The adoption of functional syllabuses for general language teaching courses, Canadian Modern Language Review, 33,5

Jones, R.L., 1978, The oral interview of the Foreign Service Institute, in B. Spolsky (ed.) Some major tests, Papers in Applied Linguistics: Advances in Language Testing Series:1, Arlington, Va., Center for Applied Linguistics

Jones, R.L., 1977, Testing: a vital connection, in J.K. Philips (ed.) The language connection:from the classroom to the world, ACTFL Review of Foreign Language Education, Vol. 9, Skokie, Ill, National Textbook Co.

Jones, R.L. and B. Spolsky, 1975, Testing language proficiency, Arlington, Va., Center for Applied Linguistics

Jonz, J., 1976, Improving on the basic egg: the m-c cloze, Language Learning, 26

Jorstad, H.L., 1974, Testing as communication, in G.A. Jarvis (ed.) The challenge of communication, ACTFL Foreign Language Education Series, Vol. 6, Skokie, Ill., National Textbook Co.

Kay, M., 1970, Performance grammars, The RAND corporation, Paper P-4391

Kelly, L.G., 1969, Twenty-five centuries of language teaching, Rowley, Mass., Newbury House

Kempson, R.M., 1977, Semantic theory, Cambridge, Cambridge University Press

Kempson, R.M., 1975, Presupposition and the delimitation of semantics, Cambridge, Cambridge University Press

Kettering, J., 1974, Communication activities, Pittsburgh, English Language Institute

Kirn, H., 1972, The effect of practice on performance on dictations and cloze tests, unpublished master's thesis, University of California, Los Angeles, abstracted in Workpapers in TESL: UCLA 6

Klare, G.R., H.W. Sinaiko and L.M. Stolurow, 1972, The cloze procedure: a convenient readability test for training materials and translations, International Review of Applied Psychology, 21

Labov, W., 1972, Sociolinguistic patterns, Philadelphia, University of Pennsylvania Press

Lado, R., 1964, Language teaching: a scientific approach, New York, McGraw Hill

Lado, R., 1961, Language testing, New York, McGraw Hill

Lado, R., 1950, Measurement in English as a foreign language with special reference to Spanish-speaking adults, unpublished Ph.D. dissertation, Michigan, University of Michigan

Lakoff, G., 1971, On generative semantics, in D.D. Steinberg & L.A. Jakobovits (eds.) Semantics: an interdisciplinary reader in philosophy, linguistics and psychology, London, Cambridge University Press

Lakoff, R., 1975, Linguistic theory and the real world, Language Learning, 25, 2, December

Lakoff, R., 1969, Some reasons why there can't be any some-any rule, Language, 45

Lambert, W.E. and G.R. Tucker, 1972, Bilingual education of children: the St. Lambert experiment, Rowley, Mass., Newbury House

Lenneberg, E.H., 1967, Biological foundations of language, New York, J. Wiley & Sons

MacGintie, W.H., 1971, Comments on Professor Coleman's paper, in E.Z. Rothkopf & P.E. Johnson (eds.) Verbal learning research on the technology of written instruction, Columbia University, Teachers' College Press

MacGintie, W.H., 1961, Contextual constraint in English prose paragraphs, Journal of Psychology, 51

Macnamara, J., 1973, Nurseries, streets and classrooms: some comparisons and deductions, The Modern Language Journal, 57, 5-6, September-October

Macnamara, J., 1966, Bilingualism and primary education: a study of Irish experience, Edinburgh, Edinburgh University Press

Michigan Test of English Language Proficiency, Manual, 1977, Division of Testing and Certification, English Language Institute, University of Michigan, Michigan

Miller, G.A., 1962, Some psychological studies in grammar, American Psychologist, 17

Morrow, K.E., 1980, Communicative language testing: revolution or evolution? In C. Brumfit & K. Johnson (eds.) The communicative approach to language teaching, London, Oxford University Press

Morrow, K.E., 1977, Techniques of evaluation for a notional syllabus, University of Reading, Reading, Centre for Applied Language Studies (Study commissioned by the Royal Society of Arts, draft copy)

Moskowitz, B.A., 1978, The acquisition of language, Scientific American, Vol. 239, No. 5

Moulton, W.G., 1961, Linguistics and language teaching in the United States, 1940 - 1960, in C. Mohrman, A. Sommerfelt & J. Whatmough (eds.) Trends in European and American linguistics 1930 - 1960, Utrecht, Spectrum

Munby, J., 1978, Communicative syllabus design, Cambridge, Cambridge University Press

National Institute for Educational Measurement of the Netherlands (CITO), unpublished reports on functional dialogue language tests, Arnhem, Netherlands

Neufeld, G., 1978, Notes of lectures in error analysis (LIN 4951) University of Ottawa, Ottawa, (unpublished)

Newmark, L., 1966, How not to interfere with language learning, International Journal of American Linguistics, 32, 1, Part II

Newmark, L.K. and D.A. Reibel, 1968, Necessity and sufficiency in language learning, International Review of Applied Linguistics 6, 2, May

Oller, J.W., jr., 1979a, Language tests at school: a pragmatic approach, London, Longman

Oller, J.W., jr., 1978, Pragmatics and language testing, in B. Spolsky (ed.) Advances in language testing series:2, Approaches to language testing, Arlington, Va., Center for Applied Linguistics

Oller, J.W., jr., 1976a, Language testing, in R. Wardhaugh & H.D. Brown (eds.) A survey of apllied linguistics, Ann Arbour, University of Michigan Press

Oller, J.W., jr., 1976b, A program for language testing research, in H.D. Brown (ed.) Papers in second language acquisition, Language Learning, Special Issue No. 4

Oller, J.W., jr., 1976c, Evidence from a general language proficiency factor: an expectancy grammar, Die Neueren Sprachen, 25

Oller, J.W., jr., 1975, Cloze, discourse, and approximations to English, in M.K. Burt & H.C. Dulay (eds.) New directions in second language teaching and bilingual education: on TESOL '75, Washington, D.C., TESOL

Oller, J.W., jr., 1973a, Pragmatic language testing, Language Sciences, December

Oller, J.W., jr., 1973b, Discrete-point tests versus tests of integrative skills in J.W. Oller jr. & J.C. Richards (eds.) Focus on the learner: pragmatic perspectives for the language teacher, Rowley, Mass., Newbury House
(Revised version of a paper presented at the TESOL Convention, New Orleans, 1971)

Oller, J.W., jr., 1973c, Cloze tests of second language proficiency and what they measure, Language Learning, 23

Oller, J.W., jr., 1972a, Dictation as a test of ESL proficiency, in H.B. Allen & R.N. Campbell (eds.) Teaching English as a second language, New York, McGraw Hill

Oller, J.W., jr., 1972b, Scoring methods and difficulty levels for cloze tests of ESL proficiency, Modern Language Journal, 56

Oller, J.W., jr., 1972c, Assessing competence in ESL: reading TESOL Quarterly, 6

Oller, J.W., jr., 1971, Dictation as a device for testing foreign language proficiency, English Language Teaching, 25, 3

Oller, J.W., jr., 1970, Language communication and second language learning, UCLA Workpapers in TESL, 4

Oller, J.W., jr. and C.A. Conrad, 1971, The cloze procedure and ESL proficiency, Language Learning, 21, 2

Oller, J.W., jr. and N. Inal, 1971, A cloze test of English prepositions, TESOL Quarterly, 5, 4

Oller, J.W., jr. and K. Perkins, 1980, Research in language testing, Rowley, Mass., Newbury House

Oller, J.W., jr. and K. Perkins, 1978, Language in education - testing the tests, Rowley, Mass., Newbury House

Oller, J.W., jr. and B. Spolsky, 1979, The Test of English as a Foreign Language (TOEFL), in B. Spolsky (ed.) Some major tests, advances in language testing, series:1, Papers in Applied Linguistics, Arlington, Va., Center for Applied Linguistics

Oller, J.W., jr. and V. Streiff, 1975, Dictation: a test of grammar-based expectancies, English Language Teaching Journal 30, 1

Palmer, A.S., 1972, Testing communication, International Review of Applied Linguistics, Vol. X/1, February

Palmer, A.S., 1978, Measures of achievement, communication, incorporation, and integration for two classes of formal EFL learners, Paper read at the fifth AILA Congress, Montreal, August 1978

Paulston. C.B., 1974, Linguistic and communicative competence, TESOL Quarterly, Vol. 8, No.4, December

Pilliner, A.E.G., 1952, The application of analysis of variance to problems of correlation, British Journal of Statistical Psychology, 5, 31, 38

Politzer, R.L., 1964, The impact of linguistics on language teaching: past, present and future, Modern Language Journal 48

Porter, D., 1976, Modified cloze procedure: a more valid reading comprehension test, English Language Teaching, 30

Potter, T., 1968, A taxonomy of cloze research, part I: readability and reading comprehension, Inglewood, California, Southwestern Regional Laboratory for Educational Research and Development

Quine, W.V., 1960, Word and object, Cambridge, Mass., Massachusetts Institute of Technology Press and New York, Wiley

Richards, J.C., 1970, A noncontrastive approach to error analysis, Paper presented at TESOL Convention, San Francisco, March 1970, appeared in English Language Teaching, 25, 1971

Richards, J.C., 1971, Error analysis and second language strategies, Language Sciences, 17

Richterich, R. and J.L. Chancerel, 1978, Identifying the needs of adults learning a foreign language, Strasbourg, Council of Europe

Rivers, W., 1973, From linguistic competence to communicative competence, TESOL Quarterly, 7, 1

Rivers, W., 1968, Teaching foreign language skills, London and Chicago, University of Chicago Press

Robinson, P., 1973, Oral expression tests, English Language Teaching, 25

Robinson, W.P., 1972, Language and social behaviour, Harmondsworth, Penguin Education

Roos-Wijgh, I.F., 1978, Testing speaking proficiency through functional dialogues, in J.L.D. Clark (ed.) Direct testing of speaking proficiency, Princeton, N.J., Educational Testing Service

Roulet, E., 1976, Présentation et guide d'emploi pour "un niveau seuil", Strasbourg, Council of Europe

Ruddell, R.B., 1965, Reading comprehension and structural redundancy in written material, Proceedings of the International Reading Association, 10

Russell, B., 1940, An inquiry into meaning and truth (The William James lectures, 1940), Allen & Unwin, London, 1966

Sadock, J.M., 1974, Towards a linguistic theory of a speech act, London and New York, Academic Press

Salomaa, A., 1969, Probabilistic and weighted grammars, Information and Control, 15

Savignon, S.J., 1972, Communicative competence: an experiment in foreign language teaching, Philadelphia, Center for Curriculum Development

Schlissinger, J., 1977, Franchir le seuil en langue seconde pour un adulte français en formation. Comment en attester, Bulletin pédagogique (langues vivantes), Vol. 57, No. 33

Schulz, R.A., 1977, Discrete point vs. simulated communicative testing in foreign languages, Modern Language Journal, Vol. LXI, No. 3, March

Searle, J.R., 1969, Speech acts, Cambridge, Cambridge University Press

Selinker, L., 1972, Interlanguage, International Review of Applied Linguistics, Vol. 10, No. 3

Sinclair, J.M. and R.M. Coulthard, 1975, Towards an analysis of discourse: the English used by teachers and pupils, London, Oxford University Press

Skinner, B.F., 1957, Verbal behavior, New York, Appleton-Century-Crofts

Smith, F., 1971, Understanding reading, New York, Holt, Rinehart & Winston

Spolsky, B., 1978, Introduction: linguists and language testers, in B. Spolsky (ed.) Advances in language testing, series:2, Approaches to Language Testing, Arlington, Va., Center for Applied Linguistics

Spolsky, B., 1975, Language testing: art or science, Paper read at the fourth International Congress of Applied Linguistics, 1975 in: Proceedings of the 4th International Congress of Applied Linguistics, Stuttgart, Germany, Sonderdruck, 1976

Spolsky, B., 1969, Reduced redundancy as a language testing tool, Paper read at the Second International Congress of Applied Linguistics, Cambridge, England, ERIC ED 031 702

Spolsky, B., 1968, Language testing - the problem of validation, TESOL Quarterly, 2

Spolsky, B., B. Sigurd, M. Sato, E. Walker and C. Arterburn, 1968, Preliminary studies in the development of techniques for testing overall second language proficiency, in J.A. Upshur & J. Fata (eds.) Problems in foreign language testing, Language Learning, Special Issue Number 3

Stevick, E., 1976, Meaning, memory and method, Rowley, Mass., Newbury House

Strevens, P., 1977, New orientations in the teaching of English, London, Oxford University Press

Swain, M., 1972, Bilingual Schooling: some experiences in Canada and the United States, Ontario Institute for Studies in Education

Taylor, W.L., 1953, Cloze procedure: a new tool for measuring readability, Journalism Quarterly, 30

Trim, J.L.M., 1980, Developing a unit/credit scheme of adult language learning, Oxford, Pergamon Press

Upshur, J.A., 1971, Productive communication testing: progress report, Paper presented at the Second International Congress of Applied Linguistics, Cambridge, England, September 1969, in G. Perren and J.L.M. Trim (eds.) Selected papers of the Second International Congress of Applied Linguistics, Cambridge, Cambridge University Press

Upshur, J.A., 1962, Language proficiency testing and the contrastive analysis dilemma, Language Learning, 12

Valette, R.M., 1977, Modern language testing, second edition, New York, Harcourt, Brace, Jovanovich

Valette, R.M., 1973, Developing and evaluating communication skills in the classroom, TESOL Quarterly, 7

Valette, R.M., 1967, Modern language testing - a handbook, New York, Harcourt, Brace and World

Valette, R.M., 1964, Evaluating the objectives in foreign language teaching, International Review of Applied Linguistics 2

Van Ek, J.A., 1980, Waystage English, Oxford, Pergamon Press

Van Ek, J.A., 1976, The threshold level for modern language learning in schools (with contributions by L.G. Alexander), Strasbourg, Council of Europe

Van Ek, J.A., 1975, The threshold level in a European unit/credit system for modern language learning by adults, Strasbourg, Council of Europe

Watzlawick, P., J. Beavin and K. Jackson, 1967, Pragmatics of human communication, New York, Norton

Weaver, W.W. and A.J. Kingston, 1963, A factor analysis of the cloze procedure and other measures of reading and language ability, Journal of Communication, 13

Widdowson, H.G., 1978, Teaching language as communication, London, Oxford University Press

Widdowson, H.G., 1975, Two types of communication exercise, in C.N. Candlin (ed.) The communicative teaching of English, London, Longman

Wilds, C.P., 1975, The Oral Interview Test, in R.L. Jones & B. Spolsky (eds.) Testing language proficiency, Arlington, Va. Center for Applied Linguistics

Wilkins, D.A., 1978, Approaches to syllabus design: communicative, functional or notional, in K. Johnson & K. Morrow (eds.) Functional materials and the classroom teacher: some back-

ground issues, University of Reading, Reading, Center for Applied Linguistics and Language Studies

Wilkins, D.A., 1976, Notional syllabuses, London, Oxford University Press

Woods, W.A., 1972, An experimental parsing system for transition network grammars, Cambridge, Mass., Bolt, Beranek & Newman

Woods, W.A., 1970, Transition network grammars for natural language analysis, Communications of the ACM, 13

Woods, W.A. and J. Makhoul, 1973, Mechanical inference problems in continuous speech understanding, Cambridge, Mass., Bolt, Beranek and Newman

APPENDIX A

- Example of typical TOEFL format
- Examples of typical TOEFL items (Gruber and Gruber 1977)

The Format of a Recent TOEFL

4 SECTIONS OF TOEFL	NUMBER OF QUESTIONS	NUMBER OF MINUTES
SECTION ONE LISTENING COMPREHENSION PART A Understanding short statements. PART B Understanding short conversations. PART C Understanding brief talks or conversations.	50	40
SECTION TWO ENGLISH STRUCTURE Grammar and sentence structure.	40	25
SECTION THREE VOCABULARY AND READING COMPREHENSION PART A Completing a sentence with the correct word. PART B Reading different types of material and interpreting them.	60	45
SECTION FOUR This section varies from test to test since there are different forms of TOEFL. Section Four on the test you will take will probably be a repeat of the above Section Two (English Structure) or the above Section Three (Vocabulary and Reading Comprehension).	50	30
TOTAL NUMBER OF QUESTIONS	200 QUESTIONS	140 MINUTES (2 HOURS AND 20 MINUTES) — TOTAL TIME

SECTION TWO: ENGLISH STRUCTURE

Time: 25 minutes for all of Section Two.

PART A

Directions: In Part A, each question consists of an incomplete sentence. In your test booklet there will be four words or phrases. Choose that one word or phrase which best completes the sentence according to standard written English. Mark your choice on your answer sheet.

EXAMPLE 1

If you would rather, we can take.....

(A) of the bus
(B) the bus
(C) in the bus
(D) for the bus

ANSWER

Choice B is correct because it provides a direct object for the verb "take." The other choices are incorrect grammatically.

EXAMPLE 2

John has not seen Mary .....

(A) during a week
(B) next week
(C) last week
(D) for a week

ANSWER

Choice D is correct because it completes the sentence with the only expression that makes sense with the verb form used in this sentence.

1. Would you like to go ..... a ride?

(A) with
(B) on
(C) for
(D) in

A B C D
1 ▯ ▯ ▯ ▯

2. They were here, but they've gone back to ..... apartment.

(A) they're
(B) theirs
(C) his
(D) their

A B C D
2 ▯ ▯ ▯ ▯

PART B

Directions: In this part, each question consists of a sentence with four parts underlined. Select that part (A *or* B *or* C *or* D) which is *not acceptable* for standard written English.

EXAMPLE 1

Although (A) George has attended (B) college until (C) recently, he left without (D) getting his degree.

A B C D
▯ ▯ ▯ ▯

ANSWER

Choice B is incorrect because the past tense ("attended") — not the present perfect tense ("has attended") — must be used to express a completed action in the past.

EXAMPLE 2

Yale's (A) football captain could pass (B), block, and tackle better (C) than anyone (D) on the team.

A B C D
▯ ▯ ▯ ▯

ANSWER

Choice D is incorrect because "anyone else" — not "anyone" — should be used in this case since the captain is incorrectly included with the other players on the team.

16. When (A) the movie was (B) over, Joe and me (C) went for a quick (D) walk.

A B C D
16 ▯ ▯ ▯ ▯

17. Mrs. Adams was surprise (A) that her (B) son and his friend had gone (C) to the mountains to ski (D).

A B C D
17 ▯ ▯ ▯ ▯

18. Michael himself (A) could not gone (B) to the concert, which (C) was held (D) at the school, so he gave his sister his tickets.

A B C D
18 ▯ ▯ ▯ ▯

19. Although (A) Elmer did not look like (B) his brother, there (C) personalities were (D) very similar.

A B C D
19 ▯ ▯ ▯ ▯

20. Alice was having (A) trouble controlling the children because (B) there (C) was (D) so many of them.

A B C D
20 ▯ ▯ ▯ ▯

SECTION THREE: VOCABULARY AND READING COMPREHENSION

Time: 45 minutes for all of Section Three.

PART A

Directions: In Part A, each question presents a sentence, with a word or phrase underlined. Below each sentence there are four other words or phrases. Find the word or phrase which would best maintain the original meaning of the sentence, if it were substituted into the original sentence in the place of the underlined word or phrase. Mark your choice on your answer sheet.

EXAMPLE

My sleeve needs to be repaired.

(A) replaced
(B) mended
(C) measured
(D) lengthened

ANSWER

The answer to the example is Choice B because the new sentence — "My sleeve needs to be mended." — means the same as the original sentence — "My sleeve needs to be repaired."

1. I don't mind giving him things, but he never does the same in return.

(A) retaliates
(B) reciprocates
(C) replaces
(D) recognizes

A B C D
1 ▯ ▯ ▯ ▯

2. That doesn't make sense; I think your reasoning is fallacious.

(A) faulty
(B) licentious
(C) pretentious
(D) officious

A B C D
2 ▯ ▯ ▯ ▯

3. Alex knew that he must breathe nothing of this to Nancy.

(A) tell
(B) secrete
(C) talk
(D) believe

A B C D
3 ▯ ▯ ▯ ▯

SECTION FOUR: ENGLISH STRUCTURE

Time: 30 minutes

Directions: In this part, each question consists of a sentence with four parts underlined. Select that part (A *or* B *or* C *or* D) which is *not acceptable* for standard written English.

1. An old miser <u>which</u> (A) picked up yellow pieces of gold had <u>something</u> (B) of the <u>simple ardor</u> (C) of a child who <u>picks out</u> (D) yellow flowers.

   1 A B C D

2. If we here in America <u>cannot live</u> (A) <u>peaceably</u> (B) and happily <u>together,</u> (C) we cannot hope that nations which have different living conditions — different economic standards, different aspirations, different mores, different interests — <u>to live</u> (D) peaceably with us.

   2 A B C D

3. Although Marilyn <u>was not invited</u> (A) <u>to the wedding,</u> (B) she would <u>very much</u> (C) have liked <u>to have gone.</u> (D)

   3 A B C D

4. Every man, woman, and child <u>in this community</u> (A) <u>are</u> (B) now aware of the <u>terrible consequences</u> (C) <u>of the habit</u> (D) of smoking.

   4 A B C D

5. The inexperienced teacher had difficulty in controlling the students <u>whom</u> (A) she <u>was escorting</u> (B) on a visit to the chemical factory, because it <u>stunk</u> (C) <u>so.</u> (D)

   5 A B C D

6. The question <u>arises</u> (A) <u>as to who should go out</u> (B) (C) this morning in this below-zero weather to clean the snow from the garage entrance, you or <u>me.</u> (D)

   6 A B C D

7. Since I <u>loved</u> (A) her very much when she was alive, I prize my <u>mother's-in-law</u> (B) picture and I <u>wouldn't</u> (C) sell it for <u>all the money</u> (D) in the world.

   7 A B C D

APPENDIX B

- Examples of typical Michigan Test of English Language Proficiency items

  (Manual 1977)

Appendix

Sample Questions and Answer Sheet

MICHIGAN TEST OF ENGLISH LANGUAGE PROFICIENCY

PART I. GRAMMAR

EXAMPLE A.
"What is that thing?"
"That _____ a spider."
a) to call
b) for calling
c) be called
d) is called
The correct English sentence is: "That is called a spider." To show that d, is called, is the best answer to this example a cross has been made next to d for Example A on the answer sheet.

Problems

1. "When will Fred leave town?"
"He _____ to leave tomorrow."
a) has
b) must
c) will
d) shall
2. "What did you do last night?"
"I _____ at a concert."
a) sing
b) sung
c) sang
d) singing
3. "Is Jack a good student?"
"No, he is _____ in the school."
a) worse
b) worst
c) bad
d) the worst
4. "How did you know Helen was here?"
"She _____ by some of her friends."
a) seen
b) is seeing
c) has seen
d) was seen
5. "Has Chet started teaching?"
"No, he is _____ in school."
a) yet
b) then
c) as
d) still

PART II. VOCABULARY

There are two types of vocabulary items in the test. In the first type you are given a sentence followed by four words or phrases. You are to find the word or phrase that is closest in meaning to the underlined word (or words) in the sentence and that could be used in the sentence without changing its meaning greatly.

EXAMPLE B.
It's too windy to go for a stroll.
a) swim
b) sail
c) drive
d) walk
The word "walk" means about the same thing as "stroll" in this sentence. The sentence "It's too windy to go for a walk" means about the same thing as "It's too windy to go for a stroll." To show that d, walk, is the correct answer, a cross has been made in the space next to d for Example B on the answer sheet.

In the other type of item you are given a sentence with one word omitted and list of four words. You are to find the word that would best complete the sentence.

EXAMPLE C. Because of the storm and rough waves, it would be foolish to go out sailing today in a small _____.
a) automobile
b) house
c) boat
d) beast
The word "boat" fits best in the sentence so that it reads "Because of the storm and rough waves, it would be foolish to go out sailing today in a small boat." To show that c, boat, is the correct answer, a cross has been made in the space next to c for Example C on the answer sheet.

Problems

6. When the boat left everyone was very joyful.
a) happy
b) sad
c) tired
d) angry
7. We found about a score in the box.
a) one
b) ten
c) twelve
d) twenty
8. Do not leave by this door; take the other _____.
a) departure
b) exit
c) outline
d) relay
9. Write carefully; don't _____.
a) skulk
b) scribble
c) skimp
d) skin
10. The floor wasn't clean so Joe _____ it.
a) swept
b) kindled
c) erased
d) smoothed

PART III. READING COMPREHENSION

EXAMPLE D. While getting ready to go to town one morning last week, my wife handed me a little piece of red cloth and asked if I would have time during the day to buy her two yards of cloth like that. I told her I would be glad to do it. And putting the piece of cloth into my pocket, I took a train for town.

The person telling the story is. . .

a) a married lady.
b) an unmarried lady.
c) a married man.
d) an unmarried man.

You know that the person telling this story is a married man because he says ". . . my wife handed me . . ." Because c, a married man, is the correct answer, a cross has been made in the space next to c for Example D on the answer sheet.

EXAMPLE E. The author was given a red cloth. . .

a) in the morning.
b) at noon.
c) in the afternoon.
d) in the evening.

The passage says, ". . . one morning last week, my wife handed me a little piece of red cloth . . ." To show that a, in the morning, is the correct answer, a cross has been made in the space next to a for Example E on the answer sheet.

The influenza virus is a single molecule built from many millions of individual atoms. You must have heard of the viruses, which are sometimes called "living molecules." While bacteria can be considered as a type of plant, secreting poisonous substances into the body of the organism they attack, viruses are living organisms themselves. We may consider them as regular chemical molecules, since they have a strictly defined atomic structure, but on the other hand we must also consider them as being alive, since they are able to multiply in unlimited quantities.

11. Bacteria are . . .
    a) poisons.
    b) larger than viruses.
    c) very small.
    d) plants.
12. The writer says that viruses are alive because they . . .
    a) have a complex atomic structure.
    b) move.
    c) multiply.
    d) need warmth and light.
13. The atomic structure of viruses . . .
    a) is variable.
    b) is strictly defined.
    c) cannot be analyzed chemically.
    d) is more complex than that of bacteria.

ANSWER SHEET

Part I. Grammar

Ex. A	a( ) b( ) c( ) d(x)	3.	a( ) b( ) c( ) d( )
1.	a( ) b( ) c( ) d( )	4.	a( ) b( ) c( ) d( )
2.	a( ) b( ) c( ) d( )	5.	a( ) b( ) c( ) d( )

Part II. Vocabulary

Ex. B	a( ) b( ) c( ) d(x)	8.	a( ) b( ) c( ) d( )
Ex. C	a( ) b( ) c(x) d( )	9.	a( ) b( ) c( ) d( )
6.	a( ) b( ) c( ) d( )	10.	a( ) b( ) c( ) d( )
7.	a( ) b( ) c( ) d( )		

Part III. Reading Comprehension

Ex. D	a( ) b( ) c(x) d( )	12.	a( ) b( ) c( ) d( )
Ex. E	a(x) b( ) c( ) d( )	13.	a( ) b( ) c( ) d( )
11.	a( ) b( ) c( ) d( )		

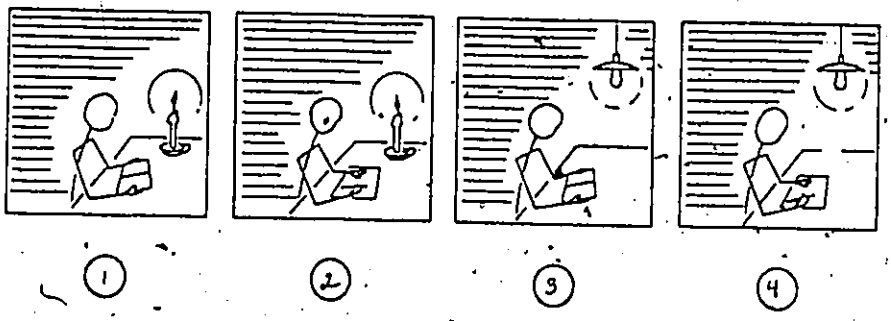
KEY

1(a), 2(c), 3(d), 4(d), 5(d), 6(a), 7(d), 8(b), 9(b), 10(a), 11(d), 12(c), 13(b).

APPENDIX C

- Example of type of stick drawings used in COMTEST-2
- Example of rules used during test administration (Palmer 1972)

Sample Test Item



Guidelines for COMTEST-2

1. Do not answer questions which require the examiner to identify the "correct" picture. Explain that this kind of question is not permitted.

2. If the question is about the position or size of the picture, say: I CAN'T ANSWER THAT KIND OF QUESTION, YOU MUST ASK ME ABOUT THE DETAILS OF THE PICTURE.

3. Do not penalize the student for time involved in explaining the rules (as in 1. and 2. above).

4. If the question is within the rules of the test (i.e. does not violate guidelines 1 and 2. above) answer it directly, and answer only the question which is specifically asked.

a) If the question is irrelevant to distinguishing the pictures, answer it anyway.

b) If the question is incorrectly put, but if it is clear nonetheless, answer it anyway.

c) If the question is so vague or incorrect that it is not clear what the subject is asking, then say: I DON'T UNDERSTAND THE QUESTION.

d) If the question deals with a distinction "X" which you do not see, say: THEY ALL LOOK THE SAME "X" TO ME. If the subject can clarify what he means by the distinction, then answer the question.

5. Do not volunteer extra information unless it is an explanation of the rules of the test. If this sort of explanation is necessary, do not count the time required as part of the timed portion of the test.

6. Policy on misunderstanding: If the subject misunderstands your answer, keep in mind that his subsequent questions are based upon this misunderstanding, and his identification of the "correct" picture will be wrong. At this point, repeat your answer which he misunderstood, or re-phrase it so that the subject understands where he went wrong. Require the subject then to continue asking questions based upon the corrected misunderstanding.

7. If the subject guesses, say: WHY DO YOU THINK THAT PICTURE IS CORRECT? IT COULD BE ANY OF THE PICTURES. If the subject has already eliminated one or two pictures, but guesses between the remaining ones, say: WHY DO YOU THINK THAT PICTURE IS CORRECT, IT COULD BE EITHER PICTURE X OR Y.