THE SYNTACTIC DEVELOPMENT OF GRADE 12 ESL LEARNERS

K. HATTINGH Hons. B.A.

Dissertation submitted for the degree Master of Arts in English at the North-West University, Potchefstroom Campus

Supervisor: **Prof. J.L.van der Walt**

Assistant Supervisor: **Mr. J.R. Pheto**

May 2005

Potchefstroom

Acknowledgements

My sincere thanks to the following persons who contributed greatly towards this dissertation:

- My parents for always being supportive and for their encouragement to further my studies.

- Prof. J.L. van der Walt, my supervisor, for his guidance and advice, and whose time, patience and enthusiasm throughout the duration of this study have been invaluable.

- Mr. J.R. Pheto, my assistant supervisor, for his confidence and motivation.

- Prof. H.S. Steyn for his time, effort and assistance with the statistical analyses.

- Tienie Buys and Klaasje Benadé for their friendly assistance in preparing the document.
- Prof. A.M.D de Lange for teaching us to *read*, giving me the confidence to write and for always being sincerely interested and supportive.

- The National Research Foundation (NRF) for the financial support provided for this project.

- Our Father for the ability, strength and perseverance to complete this study.

Abstract

Keywords: interlanguage, language development; syntactic development; level of development; T-unit, index, error analysis, grammar teaching.

The primary aim of this study is to determine the level of syntactic development in English of South African matriculants. The ESL standard in South Africa has been criticised, but no objective data are available. This study provides relevant data, based on an index, and indicates shortcomings in learners' syntactic competence. Poor school-leaving standards in English are a cause of great concern in South Africa, and complaints about school leavers' standard of English have increased over the last few years. Yet, apart from generally being labelled as "poor", little is known about the actual level of development reached by ESL learners. Comments are often based on subjective impressions.

This study focuses on syntactic development in writing and aims to determine the level of syntactic development of Grade 12 ESL learners in an objective way. Interlanguage, the concept of 'stages of development' and fossilization are discussed. The need for an index that can measure language development objectively is considered. General means of measuring syntactic development are evaluated and an index formula is established by means of statistical analyses. This formula is based on the T-unit and assigns numerical values to levels of development. The index formula is used to determine the level of syntactic development of a group of Grade 12 ESL learners. The compositions that were analysed were obtained from six provinces in South Africa. Index values are calculated for Higher and Standard Grade, for the group as a whole, and each of the six provinces. An Error Analysis is conducted and

error frequencies are reported. Problem areas in syntax are identified. The implications of the findings are considered and recommendations are briefly made for the teaching and learning of grammar.

Uittreksel

Sleutelwoorde: tussentaal, taal ontwikkeling; sintaktiese ontwikkeling; vlakke van ontwikkeling; T-eenheid, indeks, fout analise, grammatika onderrig

Die primêre doelwit van hierdie studie is om die vlak van sintaktiese ontwikkeling in Engels van Suid-Afrikaanse matrikulante te bepaal. Die standaard van Engels as tweede taal in Suid-Afrika word gekritiseer, maar geen objektiewe data oor die werklike standaard is beskikbaar nie. Hierdie studie bied relevante data wat gebaseer is op 'n indeks en wys tekortkominge uit in die sintaktiese bevoegdheid van leerlinge. Swak standaarde van skoolverlaters is 'n bron van groot kommer in Suid-Afrika en veral klagtes oor die standaard van skoolverlaters se Engels het toegeneem oor die laaste paar jaar. Buiten dat dit as 'swak' bestempel word, is min egter bekend oor die werklike vlak van ontwikkeling wat Engels tweedetaal leerders bereik. Uitlatings in hierdie verband word dikwels gemaak op grond van subjektiewe indrukke.

Hierdie studie fokus op sintaktiese ontwikkeling in skryfwerk en beoog om die sintaktiese ontwikkelingsvlak van Graad 12 Engels tweedetaal leerders op 'n objektiewe wyse te bepaal. Tussentaal, die konsep van 'n 'vlak van ontwikkeling' en fossilering word bespreek. Die tekort aan 'n indeks wat taalontwikkeling op 'n objektiewe wyse kan meet, word ook bespreek. Algemene metodes wat gebruik word om sintaktiese ontwikkeling te meet word geëvalueer en 'n indeksformule word voorgestel deur middel van statistiese analises. Die indeks is gebaseer op die T-eenheid en word gebruik om numeriese waardes aan vlak van ontwikkeling toe te ken. Die indeksformule word gebruik om die vlak van sintaktiese ontwikkeling van 'n groep Graad 12 Engels tweedetaal leerders te bepaal. Die opstelle wat geanaliseer is,

is verkry vanaf ses provinsies in Suid-Afrika. Indekswaardes word bereken vir die Hoër en die Standaard Graad, vir die hele groep, en vir elk van die ses provinsies. 'n Foute-analise word gedoen en die frekwensie van foute word gerapporteer. Probleemareas in sintaks word geïdentifiseer. Die implikasies van die bevindinge word oorweeg en voorstelle word kortliks gemaak vir die onderrig en leer van grammatika.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Problem Statement

At the end of every year, the issue of matriculation pass rates becomes a widely discussed topic. Subjects like Mathematics, Science and English usually elicit the most discussion. A general feeling that matriculation standards are dropping seems to prevail amongst the public.

This study focuses on the standard of English achieved by second language learners, as English has been singled out for attention in the matriculation year. For example, at the press conference where the 2002 matriculation results were announced, the minister of education stated that English as a second language was an area of growing concern. Others, such as Professor Dlamini, Vice-Chancellor of the former University of Zululand (*The Sunday Times*, 8 March, 1998) and Gruenewaldt (1999), stressed the problem of learners who struggle with their school subjects because they have not mastered the language of instruction, which is English. This affected matriculation standards. Weideman and Van Rensburg (2003: 155) found that the academic language proficiency of prospective university students was low. After testing a group of first year English Second Language university students in 2000, their results showed that, of the 476 students tested, 78% had proficiency levels in English of Grade 10 standard or lower. They emphasized the importance of language proficiency at university level, because of the high correlation between language proficiency and academic performance. Recent newspaper articles reflect further concern for the standard of English as a second language. Dr Peliwe Lolwana, executive officer of the certification council, Umalusi, called for a national strategy to be implemented in an attempt to improve standards of English. She stated that the challenge for schools to improve the national standard of English was even tougher than improving standards in Mathematics and Science (*Beeld*, 7 September, 2004).

The question that arises is: What is the level of English of the Second Language learner in South Africa? It is assumed that Grade 12 learners of English as a second

language should be able to perform at an advanced level, but what does it mean to perform at an "advanced" level? The term "advanced" is often taken to be a reference to the number of years a student has been learning English as a second language. It should, in actual fact, be a reference to the student's ability to perform or communicate successfully in the target language. In other words, it should reflect the learner's actual level of proficiency. There is, however, no standard definition of what it means to be at an "advanced level" of language development.

What is required is an objective definition of the term 'advanced level of development'. No such characterisation of the Grade 12 ESL learner exists in South Africa. Describing a Grade 12 learner's point of acquisition implies determining what the learner knows and does not know.

Unsworth (2002) claims that ways in which proficiency levels are measured are often inadequate. Methods of measuring language development include error analysis, mean length of utterance, standardised tests and rating scales. All of these methods have been criticised and they have proven to be insufficient for measuring a learner's development in English Second Language. As Larsen-Freeman (1976; 1978b) points out, what is needed is a yardstick which will allow researchers to attribute numerical values to different points along a second language development continuum. These numerical values would be correlates of the developmental process (Larsen-Freeman & Long, 1992: 43).

One way of approaching this problem is to argue that Grade 12 ESL learners should be able to function with larger units in language, for example clauses and sentences. It is only after automatising the rules and restrictions governing smaller units that the learner will be able to handle larger units successfully. Therefore, this study proposes to limit itself to syntactic development.

Gass and Selinker (2001: 50) point out that a more accurate means of measuring syntactic development is the so-called T-Unit. Larsen-Freeman and Strom (1977) also regard the T-Unit as the most suitable measure of development in their search for a Second Language Acquisition (SLA) index (Larsen–Freeman & Long, 1992: 43). The T-unit was developed by Hunt (1965), who was searching for a unit by which

linguistic maturity could be measured. Hunt convincingly demonstrated the use of T-units as the basic unit of measurement of syntactic development (Vavra, 2000).

The problem to be addressed in the present study is the establishment of a measure of syntactic development of the Grade 12 English Second Language learner in South Africa. It is necessary to establish an index formula according to which the actual standard or level of development can be calculated. The index formula should produce a numerical value that can be plotted on a continuum which represents the actual level of syntactic development.

The issues of concern regarding this study are the following:

- How can syntactic development be measured objectively?
- What formula can be used as an objective index of syntactic development?
- What is the level of syntactic development achieved by ESL students at the end of Grade 12 in the English language?
- What are the implications for grammar teaching and learning

1.2 Aims of the study

The aims of this study are to:

- Evaluate means of measuring syntactic development;
- Establish an index formula for measuring syntactic development;
- Determine the syntactic level of development of a group of Grade 12 ESL learners;
- Determine the implications for ESL grammar teaching and learning.

1.3 Method of Research

Relevant literature on English Second Language learning was reviewed. This included topics such as interlanguage, measures of language development, indices of language development; and the T-unit. The design of the study included a cross-sectional description and a quantitative analysis (cf. Larsen-Freeman & Long 1992). The

creative writing scripts of 216 Grade 12 English Second Language learners were collected from six different provinces. The learners came from different native language groups. The compositions were initially divided into five groups with Group 1 containing the "poor" compositions and Group 5 the "excellent" or "advanced" compositions. The compositions were then analysed in terms of their T-units. An Analysis of Variance (ANOVA) and Tukey's Studentized Test were used to determine which dependent variable discriminated best between the five groups. In order to determine which dependent variable/s would best serve as an index formula, a Multiple Regression Analysis and a Discriminant Analysis were conducted. An index formula was then proposed, and the index values for the compositions were calculated. An Error Analysis was made to determine error frequencies.

1.4 Programme of Study

Chapter 2 gives an overview of studies of interlanguage and the learning of a second language. The concept of 'stage of development' is discussed.

In Chapter 3 general measures of language development are discussed. The measures that are considered are Mean Length of Utterance, Error Analysis, Standardised Tests and Rating Scales.

Chapter 4 stresses the need for an index of language development and the T-unit is suggested as a possibility. The terms 'language development' and 'language proficiency' are defined and measures of development are classified.

In Chapter 5, the development and the use of the T-unit as measure of language development is discussed. Hunt's (1965) study is reviewed.

Chapter 6 gives an overview of a number of studies that have employed accuracy developmental measures, such as error-free T-units.

Chapter 7 describes the method of research that was used in the present study. The collection and analyses of the data are described.

In Chapter 8, the results of the analyses are reported and discussed. The characteristics of second language writing that are best suited to base an index on are identified. An index formula for measuring syntactic development in ESL is established index values are reported for the group of compositions analysed.

Chapter 9 concludes the study and makes recommendations for further research.

Chapter 2

Interlanguage

2.1 Introduction

In order to determine the level of development a learner has reached, one first has to understand what is meant by 'levels' or 'stages of development'. Such an explanation cannot be attempted without referring to certain aspects connected with the study of interlanguage. In this chapter, a brief overview is given of some of the most influential studies in early second language acquisition (SLA) research, which led to the area becoming an independent field of study. The views of Corder (1978), Nemser (1971) and Selinker (1970) with regard to interlanguage are referred to in particular. Both cross-sectional and longitudinal studies providing evidence for a natural sequence of development, as well as possible explanations for such a sequence, are discussed. The concept of 'stage' or level is then discussed. This chapter serves as a short historical overview of the development of second language acquisition research.

2.2 Discovering Interlanguage: A Historical Overview

In the early years of research in the field of language acquisition, applied linguists such as Weinreich (1953) and Lado (1957) perceived the learning of a second language as nothing more than a relationship between the first and the second language. The key concept in Weinreich's (1953) book, *Languages in Contact*, is interference. While Lado's (1957) book, *Linguistics Across Cultures*, complements Weinreich's book in some ways, Lado's overall objective is helping language teaching (Cook, 1993: 10). Consequently, not much attention was paid to how a second language is acquired by a learner. After it was realised that the assumption that interference was the major determining factor in learning a second language was an oversimplification of the process of acquiring a second language (L2), the field of Second Language Acquisition research started to develop and eventually grew into an independent discipline.

It became very clear to researchers that a modified version of the explanation of how a mother tongue is acquired is insufficient for explaining the complex process involved in mastering a second language. Only through treating language learners' language as a phenomenon to be studied in its own right could it be possible to understand the process of second language acquisition (Corder, 1978).

The initial development of this field can be divided into three periods, which roughly equate with the 1960s, 1970s and the 1980s.

2.2.1 The 1960s: From Behaviourism to Mentalist Theory

Initially, this period is characterized by an attempt to relate behaviouristic habit-formation psychology and structural linguistics to the learning and teaching of a foreign language. Linguists compiled lists of the grammatical differences between a particular first and second language. The applied linguists then studied these lists in an attempt to predict those areas in which the learner is most likely to experience learning difficulties. The main objective was to produce a scientific foundation on which to base the development of teaching techniques and materials.

In time it became clear that the process underlying acquisition was much more complex than the mere formation of habits, as behaviourists and structuralists claimed. Also, the negative interference effects of the native language were no longer regarded as a sufficient explanation for the occurrence of errors (Sharwood-Smith, 1994). The grammatical lists pointing out the differences between two languages were not a credible source for predicting areas in the target language in which the learner was most likely to produce errors.

The first seminal article in the field of Interlanguage studies, *The significance of learner's errors,* was published by Pit Corder (1967). In it, he argued that second language research should view the development of the learner's acquisition of a second language as a development of underlying linguistic competence. As in first language research, the learner should be regarded as in charge of making the rules. These rules are produced through a process of hypothesising[1]. Systematic errors

[1] A more thorough explanation of the notion of 'hypothesising' is provided later in this chapter.

should be regarded as evidence for the learner's developing competence. When a learner makes errors, it reflects attempts to make sense of or organise input provided by the language he is exposed to.

Corder (1967) further argued for the existence of a 'built-in' syllabus, or an internal programmed sequence for learning different aspects of the target language. This syllabus may or may not agree with the order in which the teacher addresses grammatical rules in the classroom. This implies that no matter which rule is taught in class, the learner will only acquire it if his internal system is ready to do so. The system is seen as 'ready' when the learner is at "the appropriate point in the predetermined built-in learning programme" (Sharwood-Smith, 1994: 24).

Sharwood Smith(1994: 26) presents the essence of Corder's proposal in a diagram as in Figure 2.1.

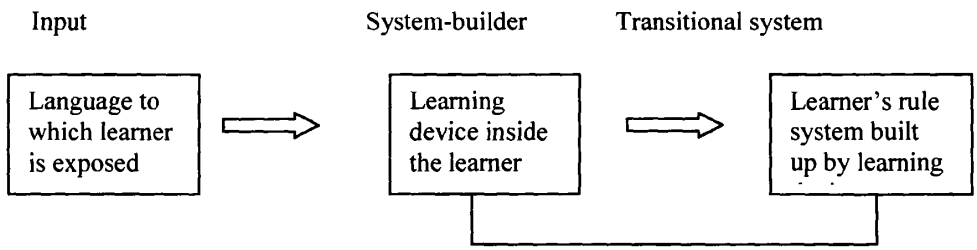

Figure 2.1 The creation of transitional systems: a partial model.

Sharwood Smith(1994: 25) explains the figure as follows:

> [It] shows how learning is seen as an operation whereby the learning device (or LAD...) builds a system using the language to which the learner is exposed (input) and creating out of that input in a particular transitional system. This transitional system is then modified when new input encountered by the learner is seen to conflict with the rules of that system: the learning device's 'hypotheses' are disconfirmed. The learning device thus functions as a little linguist in the head.

The process is an unconscious one. The learner does not consciously make the 'linguist' in his head process the input.

Chomsky (1965), who was a critic of the behaviouristic approach, introduced the term 'Language Acquisition Device' (LAD). The LAD refers to the mechanism inside the mind that is believed to be responsible for the acquisition of one's mother language. It is said to create a grammar from the input of the young child's experience, and is specifically designed to create linguistic knowledge. Chomsky's (1965) view is represented in Figure 2.2.

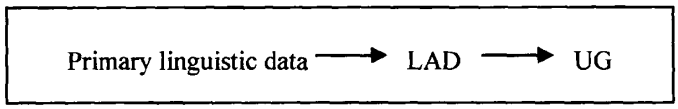

Figure 2.2 Chomsky's model of First Language Acquisition.

According to Chomsky, the child draws his knowledge of the mother tongue from a Universal Grammar, which specifies the essential form that any natural language could take. The Universal Grammar exists as a set of innate linguistic principles. It constitutes the child's initial state and controls the form which the sentences of any given language could take. The 'initial state' refers to the knowledge the child has before receiving input (White, 2003: 2).

White (2003) investigates the question of whether the UG is the same for L2 acquisition as it is for L1. She considers a number of proposals about the initial state. All of these assume that the UG is constant, i.e. unchanged because of the L1 acquisition, that the UG is different from the L1 grammar, and that the UG constrains the L2 learner's interlanguage grammar. Two possibilities about what initial state comprises of are considered: UG is the initial state, or the grammar of the L1 is the initial state for the acquisition of an L2.

A set of discovery procedures is part of the Universal Grammar. It is used to relate universal principles to the data provided by exposure to a natural language. The acquisition device contains the Universal Grammar. In order for the LAD to work, comprehensible input is necessary. The input acts as a trigger which activates the device, enabling it to shape the acquisition process (Ellis, 1996).

Chomsky stresses the active contribution of the child, and minimises the importance of imitation and reinforcement stressed by behaviourists. The child is said to build up

knowledge of his/her mother tongue through a process of hypothesis testing (Ellis, 1996).

The notion of hypothesis testing entails that a learner forms ideas or hypotheses about the structure of a language. They test these hypotheses by producing target language output, following the rules they have hypothesised. They then consider the feedback they receive on the utterance. Positive feedback supports and strengthens the hypothesis, while negative feedback leads them to discard the hypothesis and form a new one in accordance with the new information. The primary linguistic data are critical for the child to be able to determine the exact form that the grammar must take. White (2003: 2) explains that a language-specific lexicon is built up as the child takes account of the input. The parameters or options of the Universal Grammar are set appropriately for the language in question. As the child responds to different properties of the input over time, the Grammar is restructured until the child reaches a 'steady state grammar' for the mother tongue.

Gass and Selinker (2001: 279) explain that it does not mean that learners test hypotheses every time they make an utterance. Rather, learners can be made aware of the hypotheses that they have formed through negotiation and feedback when they produce the language.

The term LAD places the focus on the process of language development "without involving the idea that the learner is consciously and intentionally controlling every aspect of learning" (Sharwood-Smith, 1994: 19).

2.2.2 The 1970s: Nemser, Selinker and Interlanguage

The second historical phase marks the beginning of a split between the applied and theoretical aims of researchers (Sharwood-Smith, 1994). Learners were now regarded as independent creators of language systems. Nemser's and Selinker's studies followed that of Corder. They each published similar proposals concerning language learning. Nemser (1971) proposed that the learner's second language development should be regarded as continually evolving systems which take the learner further from the source system (L1) and closer to the target system (L2). These learner-based

systems process input from the target language in small processable doses. Nemser called these systems 'approximative systems '.

These approximative systems form a 'language' which is somewhere between the learner's L1 and the target language. As Cook (1993: 17) points out, an L2 learner at a particular point in time or stage of development actually uses neither the L1 nor the L2, but a completely different language system. In describing this system in terms of the L1 and the L2, one misses the distinctive features of second language learning. Some properties of this system are present in neither the L1 nor the L2.

The approximative system gradually approaches the target language. However, it hardly ever reaches it one hundred percent. In most cases, fossilization[2] takes place at some point of the L2 learner's development before the approximative system has combined completely with the target language.

Eventually, the term 'interlanguage' proposed by Selinker (1972) became the universally accepted term for referring to learner language. For a time, the entire field of L2 research was effectively named 'interlanguage studies' (Cook, 1993: 18). However, the term refers to two different, yet related concepts. According to Ellis (1996: 47), the term 'interlanguage' refers to the structured system which the learner creates at any given stage in his development (i.e. an interlanguage). It also refers to the series of interlocking systems which form what Corder (1967) called the learner's 'built-in syllabus' (i.e. the interlanguage continuum).

Nemser (1971) clearly states the assumptions underlying the theory of interlanguage:

- at any given time the approximative system is different from both the first and the second language;
- the approximative systems form an evolving series; and
- learners at the same stage of proficiency have roughly matching approximative systems in any given contact situation.

---

[2] Fossilization is discussed in more detail later in this chapter.

This account of interlanguage closely follows the principles of the mentalist theories of language acquisition. An emphasis on hypothesis testing and internal processes, as well as the insistence on the notion of a continuum of learning which involves through consecutively restructuring an internal system, is borrowed directly from the theory of L1 acquisition.

The construct *interlanguage* emphasises the particular significance and the structural independence of various developmental stages:

> The process of acquiring a second language is seen as a successive chain of 'interlanguages' whose grammatical structures are partly determined by features of the learner's mother tongue ... Selinker merely states that the learner's native language constitutes a powerful source of influence and is to a certain extent responsible for the particular structure of the interlanguage ... but does not contribute to a clarification of the basic problem of how and to what extent interference affects second language learning (Felix, 1980: 96).

Selinker's (1972) seminal article provided not only the theoretical framework for the interpretation of second language acquisition as a mentalistic process, but also the framework for the empirical investigation of learner language. Selinker searched for the origin of interlanguage in the process through which the mind acquires a second language (Cook, 1993: 18). Although both the terms 'interlanguage' and 'approximative systems' stress the changes that occur in a learner's language system over time, the terms should not be understood to mean the same thing. Selinker (1992: 225) states that L2 learners do not (always) approximate closer and closer to the target language until their target language production is indistinguishable from native production. Interlanguage does not necessarily reach the point of the target language, due to fossilization.

Cook (1993: 18) summarises the five central processes which Selinker (1972) suggests are part of the "latent psychological structure"[3] on which Interlanguage is dependent:

---

[3] The role of the latent psychological structure in second language acquisition is discussed later in this chapter.

- Language transfer: Some features used appropriately in the L1 are transferred into the L2. The learner may not yet have acquired the knowledge of the appropriate rules of the L2 for the situation. In order to overcome this problem, the learner applies the L1 rule while using the L2.
- Overgeneralization of L2 rules: The L2 rule is used in ways that are not allowed and without reference to the L1 system. The learner applies the rule to more cases than are valid.
- Transfer of training: The teacher, intentionally or unintentionally, creates rules which are not part of the L2. These are then acquired by the learner. Also, an overemphasis of certain structures which are thought to be difficult might lead to non-native like frequent use.
- Strategies of L2 learning: The language may, for example, be simplified. Learner strategies may also be conscious, for example rote repetition of an important structure.
- Communication strategies: These are used when a learner is experiencing specific difficulty in expressing his/her meaning and they find ways around the problem.

These central interlanguage processes characterise the psychological basis for L2 behaviour. They form the way in which a second language learner tries to make learning easier and to internalise the target language system. When L2 learners produce errors while using the target language, it cannot really be regarded as mistakes, since they are the correct form of the learner's interlanguage at that point of development. Cook (1993: 20) states that "Sentences produced by L2 learners are signs of their underlying interlanguage, not of the deficient control of the L2".

Corder's, Nemser's and Selinker's proposals shared certain essential assumptions:

- The existence of a complex, creative learning device;
- Internal coherence in the learner's language system;
- The independent character of the learner's system.

All three supported the idea of a complex mental process whereby linguistic input is organised into interlanguage systems. The learner is regarded as a creative selector and organiser who sifts input from his environment. A linguistic system is built up in a subconscious and very complex way from the input. The process of building up a linguistic system taking place inside one learner correlates with the same process taking place inside other learners, although their environmental experiences are different. A learner's interlanguage may be seen as one dialect of a specific target language.

The acquisition process is a creative one. Two kinds of creativity can be found in the process, namely 'developmental' and 'structural' creativity. Developmental creativity refers to the forming of original rules that may challenge the evidence of language encountered by the learner. A learner's language acquisition device is developmentally creative in that it uses available evidence to create new forms and rules for which evidence is not yet available. Structural creativity refers to the learner's ability to produce new and entirely original systematic utterances, whether they conform to the target language or not.

In summary, although following a similar trail of thought, each of the three arguments discussed has a specific point of focus. Whereas Corder focuses on the dynamic, developmental nature of the learner system, Nemser highlights the notion of a journey with a destination. Selinker emphasises the autonomy of the system as a 'language' in its own right and its more static aspects. It is best to regard interlanguage terms of a gradual progression, i.e. growth slowly progresses towards native-like target language proficiency along a continuum (Sharwood Smith, 1994: 32).

2.2.3 The 1980s: Krashen's Input Hypothesis and Beyond

Researchers in the field of second language learning began exploring other ways of accounting for learner behaviour, which elaborated or deviated from standard accounts. The 1970s marked the beginning of further development, with an interest in variable behaviour, more sophisticated aspects of linguistic competence, the role of consciousness and cross-linguistic influences. All of these suggest that the process of

learning a second language is more complex than was first assumed (Sharwood Smith, 1994: 119).

Krashen's (1979) 'Input Hypothesis' was a very controversial theory that had a great influence on the field of SLA in the 1980s. Krashen's distinction between acquisition and learning refers to two separate ways in which an adult develops skills and knowledge in a second language. Initially, Krashen's account of SLA (put forward during the late 1970s) was known as the 'Monitor Model' after its claim about the role of monitoring in learning a language. This model was expanded in the early 1980s into a broader-based model, which is described in Krashen (1982). Krashen and Terrell (1995: 59) provide a short summary of the five hypotheses:

- The Acquisition/learning Hypothesis: Adults develop competence in a second language in two different ways, namely 'acquisition' and 'learning'. 'Acquisition' refers to using language for communication in real situations. The language is developed or 'picked up' by using it in natural, communicative situations. 'Learning' refers to knowing about a language, e.g. being able to state grammatical rules etc. Children acquire their first language, while learning refers to having a conscious knowledge about grammar.
- The Monitor Hypothesis: Learning only functions as a monitor or editor that checks and repairs language used in real communication. The learner uses what he/she has learnt about a language to check his/her actual use of language in a conversation.
- The Natural Order Hypothesis: The rules of a language are acquired in a set and predictable order. Any attempt to learn or teach a rule in a different order to that of the natural sequence will fail (cf. 2.3 below).
- The Affective Filter Hypothesis: Internal factors, for example the student's level of motivation, self-confidence and anxiety (as well as external factors, e.g. noisy/uncomfortable environment) are regarded as an adjustable filter. It may create a mental block that prevents input form reaching the language acquisition device. If the filter is down, meaning the learner is confident, relaxed etc., the learner is more receptive towards input.

- The Input Hypothesis: Humans can only acquire a language by understanding messages or by receiving comprehensible input. A language is best acquired by understanding input that is slightly beyond the learner's current level of competence (i+1). If the amount of comprehensible input is sufficient, i+1 will usually be provided automatically. Comprehensible input refers to input understood by the learner based on the context of the input.
- Comprehensible input, though necessary, is not sufficient for acquisition. There are some affective prerequisites, such as that the learner has to be 'open' to the input in order to fully utilize it for acquisition. A positive orientation to the culture of the target language-speakers, self-confidence and acquiring in a low anxiety situation are examples of contributing factors. Krashen and Terrell (1995: 20) note that "spoken fluency in second language is not taught directly. Rather ... [it] ... emerges by itself, after sufficient amount of competence has been acquired through input".
- Structural linguistics in the 1960s failed to predict learner errors reliably. This failure led to a general desire to create second-language based tools for analysing non-native linguistic behaviour. Sticking to analysing simple structures (e.g. morpheme studies) conveniently meant that it was not necessary to go too deeply into the technicalities of linguistics. It did not seem possible to apply linguistics in any direct way to predict learner performance.

Since the 1980s, researchers have been turning back to theoretical and descriptive linguistics for ways of asking precise questions about L2 development. When looking at more complex structures in language in a more advanced stage of development, it becomes clear that a proper linguistic framework is essential. This implies that theoretical concepts and descriptive techniques from linguistics should be applied to the performance of second language speakers (Sharwood Smith, 1994).

One example of how linguistics has stimulated thinking in different ways since the abandonment of the Behaviourism of the 1950s and 1960s, is the notion of 'markedness'. Although various definitions of linguistic markedness exists, Ellis (1994: 713) defines the term as "referring to the idea that some structures are 'special' or 'less natural' or 'less basic' than others". For example, the use of 'break' in 'she

broke my heart' can be considered marked in relation to how 'break' is used in 'she broke a cup'. Some forms or structures seem to be more basic than others – 'house' seems to be the basic form from which the plural 'houses' is derived. The label 'marked' is therefore used to refer to forms that require the extension of a rule. It has attracted particular attention in Second Language Acquisition research, because it can help to account for patterns of attested L2 acquisition.

In terms of acquisition, markedness has been identified as a possible element in the learner's strategies in coping with system building. Sharwood Smith (1994: 126) points out that learners may prefer to begin with more basic, unmarked structures as a way of simplifying their learning task. According to Zobl (1984), learners are more likely to transfer unmarked native language features than marked ones, especially if the corresponding feature in their native language is marked.

Gass and Selinker (2001: 60) describe an unmarked form (it may be phonological or syntactic) as one that is more usual in languages. A marked form is used less often. Applied to phonology, there are some sounds that are common in many languages. These are the unmarked sounds. Those that are less common are the marked sounds.

In principle, linguistic markedness has nothing to do with language acquisition, but is part of an analysis of the commonalities observed in the languages. Sharwood Smith (1994: 129) explains the basic idea behind linking markedness and language learning:

> The basic idea ... is to try to find an explanation regarding the typical early assumptions or hypotheses made by given learners which can be tied up meaningfully with those features of the L2 that are relevant for a given theory of markedness. This is to see if markedness affects the order of acquisition. Do learners, in an attempt to simplify the learning task, 'demark' (Zobl, 1984) their early L1 grammar ... or do they initially assume/hypothesise that L2 will have the same degree of markedness or unmarkedness in the system as the L1 does?

Gass and Selinker (2001: 160) explain how markedness impacts on second language learning. They predict that it will be easier for learners with a more marked native language (more marked contrast) to learn a target language that is less marked.

By the 1980s, the notion of interlanguage and the existence of a natural order of acquisition were firmly established. Researchers then focussed more on the explanation of these phenomena. The present study, however, is concerned with determining the specific stage or level that a group of ESL learners have reached. Therefore, the earlier studies of interlanguage are more relevant. The remainder of this chapter refers to studies conducted during the 1970s, which provided evidence for the existence of a natural order of acquisition. The definition of what is a 'stage' or 'level' is also considered.

2.3 Sequences in Second Language Acquisition

2.3.1 Evidence for the existence of a sequence of acquisition in L2

Before continuing this discussion, it is necessary to clarify what is meant by certain closely related, yet different terms. One should distinguish between the order and the sequence of development in second language acquisition. It is possible to consider acquisition in terms of the order in which different features are acquired. It can also be considered in terms of the sequence of the stages evident in the acquisition of a single feature. 'Order' refers to the 'rank' according to which related structures, grouped together in hierarchies (that resemble individual stages), are acquired. The 'rank' in which these hierarchies are acquired is referred to as the 'sequence' of acquisition. Thus, 'order' refers to the acquisition of all elements or structures in a specific group or hierarchy, while 'sequence' refers to the acquisition of the full set of structures of one hierarchy before progressing onto the next level or stage which is represented by the next hierarchy. Each structure in that hierarchy then again needs to be acquired before the learner can progress to the next stage. The process is illustrated in Figure 2.3.

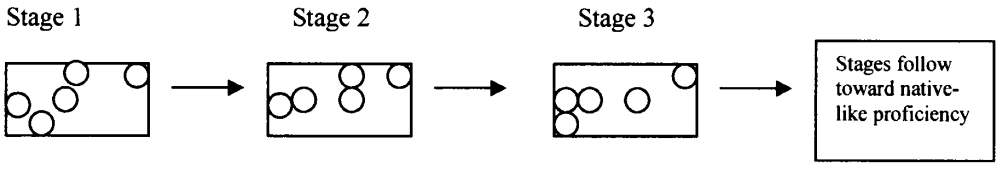

Figure 2.3 Illustration of difference between 'order' and 'sequence'.

Figure 2.3 illustrates the difference between 'order' and 'sequence'. The blocks represent the hierarchy or stage acquired in sequence, and the circles inside each

block represent each individual structure that is acquired at roughly the same time as the others in that hierarchy. These structures are acquired in a natural order.

Ellis (1994: 82), for example, discusses some characteristics of learner language in the early stages of second language acquisition in natural settings. He identifies a silent period, a period during which formulaic speech is used and one of structural and semantic simplification. After acquiring these structures, the learner will move to more complex structures in the more advanced stages of the acquisition process. After the silent period, the learner goes through a stage in which unanalysable wholes are used on particular occasions (Lyons, 1968: 177). This speech is called formulaic speech. A stage in which learners produce cryptic and fragmented utterances may follow. This type of speech is referred to as 'simplified speech' and is commonly used in the unplanned speech of both adult and child L2 learners.

The remainder of this chapter briefly considers some of the studies which have provided evidence for the existence of a universal natural order of development. Since this chapter only serves as background the studies are not discussed in detail.

2.3.1.1 Cross-sectional studies

In the early 1970s, a number of morpheme studies were conducted in order to investigate the order of acquisition of a range of grammatical morphemes in the speech of second language learners. The hypothesis that there was a set order in SLA which was the result of universal processing strategies, similar to those observed in first language acquisition, served as motivation for these studies (Ellis, 1996: 55).

Dulay and Burt (1973: 256) established that a common order of acquisition for certain structures in L2 acquisition exists. This conclusion was reached after they had performed a cross-sectional morpheme study using 151 Spanish-speaking children between the ages of six and eight who were English Second Language learners.

The results of this study were important to the development of SLA as an independent study, since demonstrating the existence of an L2 order of acquisition proved that it was important to develop SLA research separately from the study of the L1 and the

L2, as well as from first language acquisition. In other words, second language learners had interlanguages of their own and these were valid objects of study.

Studies in this regard include Dulay and Burt (1973; 1974), Bailey et al. (1974), Larsen-Freeman (1976) and Krashen et al. (1978). Generally, the picture that emerged from these studies was that the acquisition order for various grammatical functors is more or less the same. The order is irrespective of the subjects' language backgrounds, their age, and whether the medium is speech or writing (Ellis, 1996: 56). Krashen and Terrell (1995: 29) point out that "Several studies, starting with the 1974 study by Bailey, Madden and Krashen, show for adults what Dulay and Burt showed for children: subjects who speak different first languages show remarkably similar difficulty orders".

However, even in those studies showing statistical significance, the order was not entirely invariable. Dulay and Burt (1975) therefore proposed to rather group the items together than list the morphemes in order of accuracy. Each group, then, reflects a clear developmental stage. Also, the morphemes within each group are estimated to be acquired at roughly the same time. Grouping the morphemes into hierarchies helped clear up possible confusion about whether morphemes are acquired linearly (one after another), which may not be the case (Ellis, 1996: 94)

At first glance, the morpheme studies seem to provide sufficient evidence of a natural order (acquisition of structures in hierarchies) and sequence (acquisition of complete hierarchies) of development in SLA. L2 learners seemed to progress along a similar way towards the end of the interlanguage continuum, irrespective of learner differences. Thus, the main hypothesis for interlanguage theory was supported (Ellis, 1996).

Morpheme studies, however, were criticised for being English-specific, i.e. they were not replicated with other second languages (Cook, 1993: 31). Researchers such as Hakuta (1974) and Rosansky (1976) objected to the validity of the results produced by the morpheme studies and suggested that 'accuracy order' and 'acquisition order' are not the same thing. The evidence of some longitudinal studies runs counter to that of the morpheme studies. Furthermore, whereas the L1 order is "an order of

acquisition based on the chronological points when the forms attain a certain level of accuracy in the children's speech, the L2 order is an order of difficulty of production based on the scores of learners on a single testing occasion" (Cook, 1993: 33).

Pienemann and Johnson (1987: 50) believe that morpheme studies are not valid studies of language development or acquisition. They describe the main flaw of morpheme studies as follows:

> ...they provide no motivated hypothesis for choosing to define acquisition in the way they do, and, as a result of their methods of deciding whether an item is 'acquired' or not, reduce the acquisition process to one which can only be described in terms of accuracy.

2.3.1.2 Longitudinal studies

One major advantage of longitudinal studies over cross-sectional studies and Error Analysis is that data from different points in time are presented, enabling a reliable profile of the SLA to be constructed by individual learners. Longitudinal studies based on L2 English as well as other languages such as German give support to the general finding that "the acquisition of L2 syntax involves a series of transitional stages which are more or less universal" (Ellis, 1996: 59).

The longitudinal studies focussed on the acquisition of a particular grammatical subsystem, such as negatives, interrogatives and relative clauses. These grammatical subsystems are all examples of transitional constructions in SLA. Dulay et al. (1982: 121) define 'transitional constructions' as "the language forms learners use while they are still learning the grammar of a language".

Wode (1981) took a more complex approach to the acquisition of negation and established a sequence involving five stages after conducting a longitudinal study. These are: anaphoric negation; non-anaphoric external negation; internal "be" negation; internal full verb negation, and "don't" imperative; and suppletive non-imperative "do". Further work with the acquisitional sequences of negation was done by Hyltenstam (1977), Cancino, Rosansky and Schumann (1978) and Eubank (1987). In most cases, progress through the stages was found to be gradual with each new rule being used more frequently over time and in a greater number of contexts.

These studies all provide evidence for the existence of a common sequence of L2 acquisition. They lead to the conclusion that a common grammar at particular stages of language development is available to the learner, more or less regardless of the L1 or L2 (Cook, 1993). Ellis (1994: 99) gives the following account of the sequence of development in the acquisition of transitional constructions such as negation:

> Negation is an example of a transitional structure...That is, it involves a series of forms or structures which learners use en route to mastering the TL form. These interim forms are indicative of the developmental stages that learners pass through on the way to TL competence...

Stages of development are not clearly defined, but overlap considerably. Also, development does not occur in sudden jumps. Progression through these stages is a gradual process of reordering rules formed earlier in favour of ones formed later. Some learners may take years to complete this process, while others will never reach the final destination. Learners also show differences amongst one another (Ellis, 1996: 60).

2.3.2 Explanations of the Sequence of Development

Although much research has been done in order to prove the existence of a common sequence of L2 acquisition, Cook (1993: 43, 50) points out that researchers need to focus their attention on explaining this phenomenon. Felix (1984: 133) is also of the opinion that it may be more beneficial to look at both what he calls the 'logical' problem and the 'developmental' problem of language acquisition simultaneously. The first entails the basic problem of how it is possible to acquire a language, while the second is concerned with how the regularities observed in the actual process of acquisition can be explained. The focus of the developmental problem is on why children acquire a language in the way they do.

More specifically, why does a learner acquire a second, or even a first, language in stages? Cook (1993: 44) points out that the order of acquisition is not the reason why learners make errors, but "a generalisation about errors which still lacks reason". It is

not only the researcher's duty to find data, such as the discovery of an order of acquisition, but also to explain the data, in this case explaining the cause of this order.

Brown (1973: 105-106) simply suggests that "... children work out rules for the speech they hear, passing from levels of lesser to greater complexity, simply because the human species is programmed at a certain point in its life to operate in this fashion in linguistic input". This theory contributed to the initial studies of L2 learning sequence by raising questions such as whether a common order of acquisition exists for certain L2 structures and if so, whether it would be any different from that found in L1 acquisition (Burt & Dulay, 1980).

According to Felix (1984: 134), the developmental problem has mainly been studied under two aspects, namely the influence of external factors and certain psychological variables. Certainly one of the most remarkable discoveries of developmental psycholinguistics has been that "all instances of language acquisition seem to follow a certain basic structural pattern, commonly referred to as 'developmental sequence' or 'order of acquisition'" (Felix, 1984).

Some significant aspects with respect to the problem of why a language is acquired in the way that it is are summarised by Felix (1984: 134):

- The acquisition process is highly systematic in the sense that certain properties of this process hold across languages and across learners;
- Language acquisition proceeds in developmental stages; i.e. a given structure is used consistently and regularly for a certain period of time and is then replaced by some other structure;
- The sequence of stages is ordered; i.e. the structure(s) characterising any stage x emerge(s) only after the structure(s) characterising stage x-1 has/have been productively used;
- Children make regular use of constructions which are ungrammatical from the point of view of the adult, i.e. they produce structures which do not reflect rules of the input language in any direct way.

With these invariant developmental patterns in mind, Felix (1984) attempts to explain why language acquisition proceeds in developmental stages. He first investigates the question of why a child uses 'ungrammatical' sentences while they do not hear such structures being spoken by their parents in the language they are learning.

Secondly, Felix (1984) investigates the reason for the progression at regular intervals from one stage to the next, i.e. why do children give up the use of one structure in favour of another? Traditionally, the answer may be that the child is driven forward by his/her need for effective communication and social acceptance. The question arises, however, of why the child uses the 'ungrammatical' structure for an extended period of time? Does the learner ignore his/her communicative needs for some time and then decide that the structure he/she is using is not socially and communicatively sufficient or correct? Or does the child's linguistic awareness develop in stages? Why does a child's desire to communicate effectively come in stages? This line of questioning brings Felix (1984) to the conclusion that a child's need for effective communication, although influential, cannot adequately and fully explain the matter of stage-transition that occurs in language acquisition.

With regard to the theory of hypothesis testing (cf. 2.2.1), Felix (1984: 137) also offers criticism:

> The view of language acquisition as nothing but a continuous process of simply testing hypotheses against adult output presupposes a satisfactory answer to a problem, for which, as far as I know, nobody has so far offered an adequate solution; namely how does a child find out, in a principled way, if the rules of his own grammar are in conflict with those of the adult language ... As far as I can see, there is only one type of evidence that would tell the child unambiguously that a given structure is ungrammatical; namely, if somebody came along and told him so, i.e. if the child had access to negative evidence.

White's (1982) suggestion that children's grammars are 'possible grammars' just as much as adult grammars are is also criticised by Felix (1984). White's argument entails that a grammar is always a grammar of something and in this case, it is the grammar of a certain set of data. Whenever new input is received, the grammar has to be replaced with a new grammar that is optimal for that specific set of data. This entails that for each stage, the child constructs an optimal grammar for that specific

stage, not an optimal grammar as such. Therefore, White argues that the child will perceive linguistic input in a different way at different stages. Thus, the fact that a child moves from one stage to the next is a matter of changing perceptions.

Felix (1984), however, is of the opinion that White places the solution to the developmental problem, particularly the problem of stage transition, outside the domain of the theory of grammar and into the domain of the theory of perception.

Cook (1993: 44) refers to another possible explanation illustrated by Ellis (1985) for the order of acquisition in an L2. Using the example of the developmental structure "no" + Verb Phrase (such as in "No finish bath"), Ellis shows how the same fact might be explained by any of the following: a transfer strategy; a production strategy; an acquisition strategy; or an interactional strategy.

The idea of explaining the negative sequence through acquisition strategy, in other words, a strategy reflecting a universal part of the human mind, is advocated by Felix (1987), who argues that "the basic developmental pattern ... is seen as something that specific properties of human mental organisation impose upon the acquisition process". This may seem obvious at one level – if not caused by the environment the sequence can only be caused by the mind – but on another it is not precise, since it fails to specify which part of the mind is involved and how.

Larsen–Freeman (1976) has suggested frequency as an explanation for the sequence of grammatical morphemes. However, explanations based on the frequency of morphemes in the input need more precise information about the statistical properties of speech which is addressed to the L2 learner in various situations. Van Patten (1984) suggests that those morphemes that are essential for communication are acquired first, since they are the most important for communication.

Clearly, the question of how and why the development of a child's language acquisition process progresses through stages has not yet been answered adequately and definitively. Many theories are available on this subject, but what remains most relevant to this study is the fact that acquisition develops in stages. A person learning a language, be it a first or a second language, does so by progressing through a

number of stages towards native-like competence. It is determining which level of acquisition a learner has progressed to which is of importance in this study.

2.4 Stage of development

This study has already referred to numerous studies (e.g. Dulay & Burt, 1973; Hyltenstam, 1977; Cancino et al., 1978; Wode, 1981; etc.) that have produced evidence that stages or levels of acquisition exist and that these stages are acquired in a natural order. The learner gradually proceeds through each stage and continues to the next until reaching a certain point of acquisition. This point is not necessarily that of native-speaker competence. Certain structures are acquired at approximately the same time, but their acquisition proceeds in a natural order (one is acquired only after the other). These structures of language that are acquired at approximately the same time have been grouped into hierarchies (cf. 2.3.1).

All learners progress to a certain stage in the sequence of development. However, while very few achieve actual target-like proficiency, every learner reaches a certain stage in the developmental continuum, which may be different from the stage reached by another learner of the same age, language background, etc. This stage or level is often referred to in terms of levels such as intermediate, upper-intermediate and advanced. Harmer (2004: 44) says that the problem with these labels is that they mean different things to different people. There are no standard definitions of these levels. It is especially difficult to define what an advanced level means. Van der Walt (1981: 188) points out that the advanced learner should have reached a degree of competence which permits him to function most of the time at sentence level (the intermediate student still has to exercise most of his choices at word or even morpheme level because his knowledge of the grammar cannot be recalled automatically). In theory, a Grade 12 learner should be at the advanced level, after 12 years of English tuition. However, this is often not the case, as is evidenced by Nkani (1998):

> The alarming matric failure rate is identified as key education problem in the Department of Education. Insufficient command of English is identified as a key cause of this state of affairs by researchers such as Coetzee-Van Rooyen & Verhoef (2000), Finn (1986), Greenbaum & Mbali (2002), Mawasha (1982) and Nwaila (1993). Particularly disconcerting are the results of Coetzee-Van Rooyen and Verhoef (2000), which provide evidence that the majority of

> students entering tertiary education have English language below the level of matric second language and therefore do not have sufficient command of English to afford reasonable opportunity to succeed in tertiary education.

As pointed out in Chapter 1, the purpose of the present study is to determine the syntactic developmental level in an objective way. In this study, an attempt is made to determine the level of syntactic development of Grade 12 ESL learners. This would be beneficial in many ways. For example, the accuracy of the mark allocated by the teacher can be tested by comparing the mark to the stage of development. Such an index will also give an indication of those areas of syntax which learners at the Grade 12 level have not yet acquired.

The next question, however, is "What is a stage?" In first language acquisition research, Ingram (1989) finds six meanings of "stage":

- Continuous stage – a measure such as age is used to compare learners;
- Plateau stage – a characteristic behaviour remains constant;
- Co-occurrence stage – two or more arbitrary behaviours overlap;
- Succession stage – one group of behaviours succeeds another;
- Principle stage – several different behaviours are related by one principle;
- Implicational stage – one aspect of behaviour necessarily implies another.

In SLA research, the stages are essentially *post hoc*, meaning that the stages are worked out from the data, rather than the data being searched for forms that support a prior hypothesis. For example, the research of the L2 grammatical morpheme first started out with 'continuous' stages, after which the acquisitional hierarchy was used to group behaviours together into 'co-occurrence' stages employing several grammatical features. Cook (1993: 46) deems it necessary to consider why the sequence of stages is important, especially after taking into account the feature that makes L2 acquisition unique – that there is no common final stage at which all L2 learners have the same knowledge.

As far as transitional structures are concerned, a 'stage' or 'level' consists of a period during which learners use a particular form or structure in a systematic way, without necessarily excluding other forms or structures (Ellis, 1994: 111).

2.4.1 Fossilization

Selinker (1972) refers to the phenomenon of a learner becoming stagnant at a certain stage or level of acquisition, before reaching target-like proficiency, as fossilization. No matter how much more comprehensible input the learner is exposed to, no more input is processed and interpreted. If the input was interpreted, the learner would be able to progress to the next stage. Fossilized structures may be correct forms of the target language, or be realised as errors, depending on the level of development at which fossilization occurs. Selinker (1992) describes fossilization as the situations in which non-target-like forms persist and will permanently exist in a learner's speech, despite attempts to eradicate them.

Selinker and Lamendella (1978) argue that both internal and external factors cause fossilization. Fossilization may, for example, occur due to the learner's satisfaction with his/her current level of development. The learner may feel that he/she is proficient enough in the target language to communicate effectively in order to fulfil or complete his/her purpose or task. Alternatively, fossilization may occur due to changes in the neural structure of the brain as a result of age, restricting the operation of the hypothesis-testing mechanism.

Selinker (1972; 1992) addressed what became the central issue in mentalist accounts of SLA, namely how it is possible for an adult to succeed in learning a second language if the LAD is no longer functional after puberty. Selinker argues that those adults who achieve native-like proficiency, do so by reactivating the acquisition device. Lenneberg (1967) referred to it as the 'latent language structure'.

Those that do not achieve native-like acquisition of the target language rely on or fall back on what Selinker calls a "latent psychological structure". Ellis (1996: 49) explains that this is a more general cognitive mechanism which does not involve resource to the universal grammar. It is presumed to be responsible for other types of

learning apart from language. Dulay and Burt (1977), referred to this second mechanism as the 'cognitive organiser'. This term later became the popular description for this type of learning.

Kuo (2003: 2) identifies four main characteristics of fossilization: repetitiveness, permanency, persistency and fluency. Repetitiveness refers to the regular reappearance of erroneous target language forms in interlanguage speakers. Permanency refers to "the everlasting existence of erroneous forms", whereas persistency refers to the "persistent failure to achieve complete mastery". Brown (1993) notes that fossilization often occurs at the point where learners have reached fluency in the target language, but minor errors occur repeatedly and unconsciously. The learners often achieve a level of competence which ensures communicative success, but the grammar may not be similar to that of the target language.

One can only reach the conclusion that a certain aspect is indeed fossilized when further interlanguage learning continues not to occur, in spite of the learner's ability, opportunity and motivation to learn the target language. Kuo (2003) distinguishes between 'fossilization' and 'stabilisation'. "It is only when the 'stabilisation' of learners' TL production is assumed to be permanent that the issue of fossilization arises" (Kuo, 2003: 3). Destabilisation is possible, whereas defossilization is not. Destabilisation is positive in the growth of the interlanguage system, since it brings about plateau-like learning development (Selinker & Lamendella, 1979).

With regard to how fossilization occurs, Kuo (2003) refers to Vigil and Oller's (1976) distinction between affective feedback and cognitive feedback. Vigil and Oller (1976) regard fossilization as resulting from an excess of extrinsic feedback the L2 learner receives for erroneous forms. Affective feedback, passed on by paralinguistic devices such as facial expressions, is said to be able to override the effect of the cognitive feedback and determines the degree to which learners internalise the correct forms. In other words, the requirement of positive affective feedback will encourage the learner to make additional efforts to change the "inaccurate assegais" that is indicated through negative cognitive feedback (Kuo, 2003: 4).

Selinker and Lamedella (1979) argue that it is not extrinsic feedback alone, but also the motivational and attitudinal factors of the individual which controls the onset of fossilization. In support, Krashen and Terrell (1995) note that comprehensive input is not sufficient for acquisition to take place. Some affective factors have an influence on the success of the target language acquisition process. Krashen's Affective Filter hypothesis (cf. 2.2.3 above) states how these factors relate to the second language acquisition process. Most of the studies concerned with affective variables can be placed into one of three categories (Krashen, 1982: 31):

- Motivation: Usually learners who have high levels of motivation perform better in acquiring a second language than those who are less motivated.
- Self-confidence: Self-confident learners and those with a good self-image tend to perform better in SLA.
- Anxiety: Low anxiety levels appear to be conducive to second language acquisition, whether it is measured as personal or classroom anxiety.

The Affective Filter Hypothesis expresses the relationship between these affective variables and the SLA process. It suggests that acquirers vary with respect to the strength or level of their Affective Filters. Those learners with optimal attitudes tend to seek more input and have lower Affective Filters, which results in the input being absorbed and reaching the language acquisition device. A high Affective Filter will 'block out' the input, although the student may understand the message, keeping it from reaching the language acquisition device (Krashen, 1982).

Krashen (1982) argues that the filter hypothesis explains why fossilization may still occur even though the learner receives enough comprehensible input. The hypothesis claims that the "effect of affect is 'outside' the language acquisition device proper" (Krashen, 1982: 32). It maintains that input is the primary contributing variable in SLA and that affective variables act to delay or facilitate the delivery of input to the language acquisition device (cf. 2.2.3).

Although it may not be possible to 'defossilize' erroneous target language forms (according to Selinker and Lamedella (1979), this is by definition not possible), Kuo

(2003) identifies five aspects that help teachers 'destabilise' the errors before fossilization sets in.

Firstly, teachers need to develop the ability to distinguish between 'mistakes' (occurring in random fashion) and 'errors' (occurring systematically and repetitively). The teacher can help to reduce or remove ambiguity that surrounds the single encounter of an error by observing carefully, developing intuitional judgement from experience and theoretical foundations. Teachers need to encourage students to consciously try to change errors, and to influence the attitudes of students positively by being an expert facilitator, rather than merely providing remedies (Kuo, 2003: 12).

Secondly, Kuo (2003: 12) argues that the responsibility for improvement should be placed on the students, which is effective both in terms of learners' acceptance of the instruction and the improvement of their motivation. The goals of learning should also be varied according to each learner's individual differences.

An optimal amount of feedback is essential in order to be effective in assisting the learner in reaching native-like proficiency. The primary task of the teacher whose students have developed fossilized L2 forms is to separate the optimal 'tension' between positive and negative feedback:

Teachers need to strike the balance between offering enough positive feedback to encourage continued communication, but not so much that crucial errors become internalized and results in fossilization. Also, in order to destabilise incorrect forms, the negative feedback must be sufficient to call learners' attention to crucial errors, but not so much that the learner becomes discouraged from attempting to speak at all (Kuo, 2003: 13).

Teachers should motivate learners to accept the identity or culture of a new language. Learners also need to be willing to participate in the social network of the culture whose language they are learning. It is the teacher's duty to help lower the psychological inhibitions of students and to foster 'ego flexibility'. The key to initiate changes lies in having an adjustable ego and in helping the learners to overcome the inhibition related to fears of appearing foolish (Kuo, 2003: 13). In other words, the

learner needs to be made comfortable with the target culture and not to resist accepting the challenge of taking on the culture.

Researchers (e.g. Ellis, 1996; Kuo, 2003) argue that focusing on form rather than on meaning will help learners discover irregularities and patterns of the L2. Kuo (2003: 14) holds that "with the focus on conversational and discourse-level grammar, the teacher bases the language instruction on the content and specific linguistic skills upon the conversational framework". Structurally organised communicative instruction such as this promotes fluency as well as accuracy of the target language forms.

2.5 Conclusion

The aim of this chapter was to provide a short background of the influential theories of Corder, Nemser and Selinker that settled SLA research as an independent field of study. The discovery of interlanguages and the question of how a learner progresses through different stages during the acquisition process and how these stages are ordered have lead to a vast field of research which include many interrelated topics. As Ellis (1996) points out, early interlanguage theory provided momentum for empirical research into both the sequence of development and the nature of errors in SLA. "Interlanguage was the theoretical construct which underlay the attempts of SLA researchers to identify the stages of development through which L2 learners pass on their way to the L2 (or near-L2) proficiency" (Ellis, 1996: 42). The term 'interlanguage' was used to refer to the systematic knowledge of language which is independent of the learner's first language and the target language system. This research led to the suggestion that L2 learners followed a 'universal route' of development which was to a large degree uninfluenced by factors such as age and the learner's background.

Furthermore, a short overview of practical studies by Dulay and Burt (1973) as well as Brown (1973), supporting the main hypothesis of the theory of interlanguage, was given.

The concepts of sequence and order of development were discussed. The conclusion was reached that 'sequence' refers to the acquisition of structures in or during a specific stage or level of the acquisition process. On the other hand, 'order' refers to the interlanguage continuum, or the progression through one stage or level onto the next towards native-like proficiency of the target language. The concept of fossilization and the role it plays in the learner's progression through the developmental stages was also discussed.

The following chapter considers a number of conventional methods of measuring language development.

Chapter 3

Measures of Language Development

3.1 Introduction

The previous chapter provided a broad historical overview of the growth of SLA into an independent discipline. The main focus of the chapter was on the discovery of interlanguages and its effect on the field of second language research. It was pointed out that there is a natural order of grammatical development, and that a learner proceeds through stages or levels of development.

This chapter focuses on four common means of measuring language development, namely Mean Length of Utterance, Error Analysis, Rating Scales and Standardised Tests. Each of these measures is described and evaluated. It is concluded that none of them is an adequate measure of second language development.

3.2 Mean Length of Utterance

Mean Length of utterance or MLU has been used to calculate the linguistic development of children in first language acquisition. MLU is calculated by collecting enough consecutive utterances to make up 100 completely understandable utterances (following Brown, 1973). These are analysed by counting the morphemes in each utterance. The researcher adds up the number of morphemes for all 100 utterances. Then, the total is divided by 100 to get the mean or average. The lowest possible MLU count is 1.

Brown's (1973) study has been very influential in this field. According to Chengappa et al. (2002), Brown's contribution lies in his development of five stages of sentence construction. These stages seem to parallel overall language development. Brown (1973) first noted that children use the same grammatical structures at comparable MLUs, up to the MLU of about Stage IV.

In 1962, Brown, Bellugi and Fraser began a longitudinal study of the development of English as a mother tongue in the early stages of childhood. The subjects were three

preschool children with English as their first language. They were named Adam, Eve and Sarah. The children were not the same age, since earlier studies had shown that children acquire language at widely varying rates. The subjects could rather be compared according to the mean length of their utterances and the upper bound or longest utterances they produced. "Two children matched for MLU are much more likely to have speech that is, on internal grounds, at the same level of constructional complexity than are two children of the same chronological age" (Brown, 1973: 55).

Brown (1973: 53) describes MLU as "an excellent simple index of grammatical development because almost every new kind of knowledge increases length". For example, the number of semantic roles expressed in a sentence, the addition of essential morphemes and embedding and coordinating all increase the length of the surface structure of the sentence, especially when measured in morphemes.

The five stages of sentence construction that Brown defines are:

- Stage I: Semantic roles and syntactic relations. The child puts the noun-verb sequence together. (MLU= 1.0-2.0 morphemes or 1,75 morphemes.)
- Stage II: Grammatical morphemes and modulation meaning. The child starts to change the endings of words to portray grammar. (MLU=2.0-2.5 with average of 2.25 morphemes.)
- Stage III: Modalities of simple sentences. The child begins to use questions and imperatives. (MLU=2.5-3.25 with average of 2.75 morphemes.)
- Stage IV: Embedding. The child begins to use complex sentences. (MLU=3.25-3.75 with average of 3.5 morphemes.)
- Stage V: Co-ordinating. The child may start to use connectors and more functions. (MLU=3.75-4.25 with average of 4 morphemes.)

These stages should not be considered to be discrete. Rather, linguistic development should be regarded as continuous and that the stages allow comparison and characterisation at different levels of language proficiency (Chengappa et al., 2002).

Brown (1973: 54) and his colleagues drew up a list of rules for calculating the mean length of utterance and longest utterance for a transcription to make one child's data comparable to another. These rules were not set to be followed precisely (Brown, 1973: 54). However, the rules assisted them in making one set of data, one project and, to a limited degree, one language comparable to one another.

Brown (1973: 54) lists the following rules for calculating MLU:

- Start counting on the second page of the transcription (in case the child has to 'warm up' first), or start with the first recitation-free stretch. Count the first 100 utterances that conform to the following rules.
- Only use utterances that are fully transcribed, not those containing blanks. However, doubtful transcriptions that may be entered in parentheses (portions of utterances) are used.
- Include all exact utterance repetitions. When stuttering occurs, count the word once in the most complete form produced. If a word is repeated for emphasis, count each occurrence.
- Don't count fillers such as *mm* or *oh*, or transcriptions such as *no*, *yeah* and *hi*.
- Count all compound words, proper names and "ritualised reduplications" as one word, for example, *birthday*, *choo-choo*, and *night-night*.
- Count irregular past forms of a verb as one morpheme.
- Count diminutives, such as *doggie*, as one morpheme.
- All auxiliaries and all catenatives, such as *gonna* and *wanna,* are counted as single morphemes. Inflections, such as possessive [s], plural [s], third person singular [s] and regular past [d], are counted as separate morphemes.

The range count follows the above rules, but is always calculated for the total transcription rather than for 100 utterances.

As children progress through the stages of development, they become able to express much more complex ideas. As the child becomes more proficient and reaches Stage V, he/she produces constructions of such a great variety that the type of interaction becomes more influential than the child's knowledge of the language structures. What

the child says and the length of the utterances produceD start to depend more on the type of conversation situation than on the knowledge the child has. At this point, the index starts to lose its value as an indicator of grammatical knowledge.

Chengappa et al. (2002) also employed MLU to measure development in their study. MLU was used to compare the speech of normal children between the ages of 4 and 11 with the speech of mentally retarded children of the same ages. The aim was to determine the relationship between IQ and sentence complexity in mentally retarded children. Overall language development and the rate at which it occurs are slower than the rate of intellectual development in mentally retarded children. Chengappa et al. (2002) point out that when mentally retarded children are assessed, there is always inconsistency between the child's competence and performance.

Therefore, it is necessary to follow a systematic and structural approach to studying the language development from the beginning of the single word utterance to the adult speech model. MLU is therefore not a sufficient measure for measuring L2 development. Another measure is needed that can measure the development from the beginning to the more advanced stages.

Critique

Mean length of utterance has been criticised as a means of calculating the level of a child's linguistic development. It is not always easy to decide what counts as an utterance in the speech of a child and what does not. The rules drawn up by Brown (1973), as mentioned above, are not all-inclusive and can only be used as a guideline for decisions. One researcher may decide to count an utterance, while another may decide not to. Also, it is difficult to decide what can be counted as a morpheme. For example, past tense forms of irregular verbs are counted as one morpheme, while those of regular verbs are counted as two. MLU should therefore be used cautiously (Baker & Mc Enery, not dated).

MLU can only be applied to first languages, not second or foreign languages. It is a means used to measure the linguistic development in the native language of subjects in their early childhood. Brown (1973) explains that, after a child reaches the fifth stage of development, MLU starts losing its credibility and value as an index of

growing grammatical knowledge. Generally, L2 learners are cognitively at a much more complex level when they start learning the second language than a child who is learning to speak his/her native language. Larsen-Freeman (1978: 127) points out that L2 learners can produce more complex constructions and sentences from a very early stage in the learning process, unlike a child learning to speak his/her first language. Gass and Selinker (2001: 49) also point out that the syntactic sophistication of L2 learners' utterances differ right from the beginning stage. Furthermore, each learner creates his/her personal interlanguage, which may differ from those of others.

3.3 Error Analysis

Error has always been an important issue in the analysis of second language development (Wolfe-Quintero et al. 1998: 34). Error analysis, which focuses on the errors that learners make, has been very popular. It differs from contrastive analysis in that the errors that a learner makes in the target language are compared to the target language itself. Similar to the weak form of contrastive analysis, error analysis starts from learner production data.

Contrastive analysis attributes all errors to language transfer from the L2 learner's native language. Error analysis provides a broader range of possible explanations for researchers to account for learner errors. Two main error types can be identified within an error analysis framework, namely interlingual and intralingual errors. Whereas interlingual errors can be attributed to the native language, intralingual errors can be ascribed to the language being learnt (Gass & Selinker, 2001: 80).

Error analysis allows researchers to focus on the learner's own system and the way it develops (Cook, 1993).

Selinker (1972) was concerned with establishing the study of interlanguage (cf. 2.2.2). He considered any attempt to communicate meaningfully in the target language as appropriate evidence for second language learning. The method of comparing the speech of different L2 learners and target language speakers form part of the tradition associated with Corder (1967; 1971; 1981). He suggested that learners (both L1 and L2) form an internal grammar by applying learner strategies to language input.

Sentences that are constructed incorrectly compared to the target language should not be interpreted as errors, but as evidence of the process of acquisition. In terms of the learner's interlanguage, the sentence is correct. It is only when compared to the target language that the utterance seems wrong. "Sentences produced by L2 learners are signs of their underlying interlanguage, not of their deficient control of the L2" (Cook, 1993: 20).

This view of errors influenced the methodology for conducting an error analysis. It approaches L2 learning through a detailed analysis of the learner's language. Error Analysis consists of the following three basic stages (as proposed by Corder 1971):

- Recognising idiosyncrasies: the researcher compares the learner's utterances to see whether it conforms to the L2 grammar in terms of its use or meaning.
- Accounting for the learner's idiosyncratic dialect: the researcher determines how the learner's interlanguage can be described. The researcher still needs to know the L1 of the learner to be able to account for the errors made.
- Explanation: the researcher attempts to explain the reason for deviation from the L2 grammar (the main explanation often being transfer from the L1).

The goal of error analysis is often that of pedagogical remediation. Remediation based on the frequency of and type of error is carried out (Gass & Selinker, 2001: 79).

Critique

Two main methodological problems are advanced by Cook (1993: 21) with regard to the three stages proposed by Corder (1971):

Corder (1971) reserves the term 'error' for referring to systematic or competence errors, while he refers to errors of performance as mistakes. In short, the first problem is concerned with the distinction between performance and competence and what classifies as an error (Cook, 1993: 21).

The second problem concerns the nature of an error. How can the researcher determine when the learner has made an error? What does it consist of? The

researcher needs to determine what the learner actually meant by an utterance. 'I want to know the English' may, for instance, either mean 'I want to know the English *people*' or 'I want to know the English *language* '(Cook, 1993: 21).

Other points of criticism have also been raised against Error Analysis. Schachter and Celce-Murcia (1983) identify a number of weaknesses of the procedure:

- With the focus on errors, only partial accounts of interlanguages are given. Little is said about what the learner actually does right;
- Errors are often classified subjectively due to analysts' lack of familiarity with the range of native languages the sample represents. The learner's first language is often underestimated as a source of errors;
- Analyses are often unquantified. Those that are quantified are often done incorrectly or vaguely so;
- Explanations of errors are often impressionistic and vague. In cases where two or more explanations for errors would be valid, only one was sometimes presented. Identifying the cause of an error should be attempted only after a description of the error is given;
- Studies suffer from bias samples in which certain L1s may be over-represented, as well as certain types of subjects and certain types of performance data preventing valid generalisations;
- The phenomenon of avoidance is ignored because the focus is on what the learner does wrong.

Larsen-Freeman and Strom (1977) are not optimistic that the developmental reduction in errors in a particular grammatical structure or structures will provide the answer to the quest for an index of development. "If we used errors in article usage alone as our index of development, we would have no way of discriminating between the 'poor' and 'good' levels of writing" (Larsen-Freeman & Strom, 1977: 127). Neuman (1977) also concludes, after attempting to characterise the problems ESL learners experience at intermediate level by means of error analysis, that errors cannot be used to distinguish between beginning, intermediate and advanced levels.

Error analysis has proved to be important in recognising that language learners are more than "passive hiccups of the NL [native language] forms and functions" (Gass & Selinker, 2001: 86). However, its methodology "falls short in the analysis of second language data in that it only sees a partial picture of what a learner produces of the second language. One cannot hope to appreciate the complexity of the learning situation by studying one biased part of it" (Gass & Selinker, 2001: 86).

3.4 Standardised Tests

Standardised tests are often used for measuring proficiency levels. Learners may, for example, be classified as being advanced after scoring above a certain level in a standardised test such as the TOEFL (Test of English as a Foreign Language) (cf. Gass & Selinker, 2001: 37).

Brooks (1964) lists the steps that must be taken in order for a test to be considered a standardised test. Developers need to reach a preliminary agreement on specifications, such as the types of items to be used, the quantity and the arrangement of the items. Competent teachers should write the items and representatives of the scientific measure should supervise the process. Items should be well integrated and organised. For instance, items of the same type should be grouped together and presented under a single set of instructions. In order to pre-test the test, a group of students typical of the entire target population should write the test. The pre-test should be officially scored and a statistical analysis of the pre-test items should be done. According to the results of the pre-testing, the final editing of items should be done (Brooks, 1964: 207 - 212).

Alderson et al. (1987) review a number of well-known and frequently used English standardised tests. They point out that there is no such thing as the perfect test - none of these tests is flawless or perfectly reliable and valid. The following are examples of standardised tests:

Royal Society of Arts

The Royal Society of Arts Examinations in the Communicative Use of English as a Foreign Language (RSA CUEL) are designed, as the name implies, to measure the ability of ESL/EFL speakers to perform tasks in English at different levels in four

independent modes: reading, writing, listening and speaking. Although the test is primarily designed for use in Great Britain, it is also used outside of Great Britain at basic, intermediate and advanced level. As far as the oral and writing sections are concerned, open-ended responses are matched to perform specifications for each level and criterion targets determine the pass mark. Materials for the writing test often take the form of extracts from advertisements, surveys and informational brochures. The differences in levels are determined by the complexity of the tasks involved (Hargreaves, 1987: 32).

Although centres outside of Britain have the option of designing versions of the test more suited to their needs, the examples of source texts in the RSA CUEL indicate a distinct bias towards content relevant to the use of English in Britain (what the information pamphlet refers to as 'in-Britain English'). Neither integrative or discrete tasks are solely dependent on the skill being assessed; for example, a certain level of reading competence is assumed in the listening, writing and speaking tests, even though its communicative importance is played down.

Traditional norm-referenced tests are more straightforward than the criterion-referenced CUEFL on setting pass/fail scores. As the test is intended to measure a student's ability to perform certain tasks in English, it is difficult to establish whether a candidate needs to reach a specific criterion in each of the tasks in the test, or whether overachievement in one question can compensate for deficiencies in others. Also, failure to reach the criterion at one level says nothing about performance at other levels. The specifications for different levels may be accumulative, but assessment of one level is not related to others. The implication is that it is possible to perform tasks in English at basic, intermediate and advanced level. Students can almost 'choose' different combinations of levels to meet personal requirements (Hargreaves, 1987: 33).

Test of Ability to Subordinate

The Test of Ability to Subordinate (TAS) is designed to test strengths and weaknesses in writing sentences in English at intermediate or advanced level. Responses are compared to suggested responses and then scored as correct or incorrect. The format consists of nine subgroups with five items each. These subgroups represent different

subordinate structures. The writing maturity of the learners is assessed by means of sentence combining tasks requiring subordinate structures. The test is constructed in such a way that it contains structures expected to (and do) appear in the writing of proficient, native first year college students (Mullen, 1987: 73).

It is important to take the technique, which is based on three assumptions, by which these structures were determined, into consideration. The author of the test assumes that written texts will cover all types of a given structure indicating mature writing. It is also assumed that structures that function differently can be put in the same category. Thirdly, it is assumed that structures that function similarly can be considered members of different structural patterns. Although some subordinate structures are less complex than others, no means is provided for assessing the level of syntactic maturity according to the level of structural complexity mastered. Additionally, no distinction is made between the two major functions of subordinated units. In other words, if a student shows evidence of mastering one function and not another, there is no system of evaluation to show this.

Mullen (1987) argues that the testing sample should be larger and more representative of ESL learners in different programmes. When calculating the correlation coefficient, the inter-rater reliability of the writing sample and the inter-item reliability of the TAS need to be taken into account (Mullen, 1987: 75). The procedure for establishing concurrent validity needs to be refined. It is not clear whether students will be given the test again, after necessary instruction, if they are found lacking in control over these subordinate structures. Students may perform well on sentence combining tasks, but not be able produce them in spontaneous writing. The composition may also contain many examples of the correct use of the subordinate structures, but show deficiencies beyond sentence level which will result in low ratings.

There is no analysis of items to determine if those examinees with high scores get the items correct and whether those with low scores miss them (Mullen, 1987: 75).

Test of English as a Foreign Language

The test is recommended for ESL students at 11^{th} grade level or above (i.e. more advanced learners). It is designed to measure English proficiency of non-native

speakers. Tests are either scored manually or by means of computer. Learners have approximately 180 minutes to complete the test (this includes time for administration). Since 1976, the test has consisted of three sections, yielding scores for each section and a total score. A new, equated form is presented each month.

Each learner has to pay a fee for the International Programme, which includes the examinee handbook, two copies of the Examinee's Score Confirmation Record and official score reports that are sent to up to three designated institutions.

The programme is jointly sponsored by the College Board, the Graduate Record Examinations Board and the Educational Testing Service (Stevenson, 1987).

The test is most often used for people who want to study in North America. The test is only applicable to learners of English as a Second Language, and not to learners of second languages in general. The costs associated with the writing of the test may also be high for foreign learners going to North America, depending on international currency rates.

Critique

Standardised tests in general have been criticised for the components the tests consist of. For example, Scott and Stokes (1995) examined the subtests in which the syntactic knowledge and performance of learners are typically targeted by three standardized tests. Their analysis revealed that all subtests do not tap into the same variety of syntactic forms and operations. Furthermore, the types of language behaviour generated in subtests differ from the types of behaviour that occur in normal daily communication. Learners will, for example, not be asked to make sentences from randomly ordered words or to identify two sentences meaning the same thing in real life (Scott & Stokes, 1995: 314).

Scott and Stokes (1995: 314) argue that, contrary to what may be assumed, all response channels (listening, writing, speaking, reading) test syntax in a different way. Each channel employs very different formats and processing requirements. Furthermore, channels very often target different sets of syntactic structures and operations. Thus, it is incorrect to assume that Written, Spoken, Reading and

Listening Grammar tasks are the written, spoken, listening or reading equivalent of each other.

Scott and Stokes (1995: 315) also question the extent to which tests predict syntactic performance in a more natural environment.

In a study by Kobayashi (2002), the structures of texts and the way in which tests are formatted were found to have an effect on learner performance. The study draws on previous studies on readability and text characteristics suggesting that organisation and coherence of reading texts and test formats are significant factors which influence reading comprehension. A third research variable was included in the study, along with text organisation and response format, namely learners' language proficiency. Learners of different proficiency levels are affected in different ways by the organisation and coherence of texts.

Kobayashi (2002: 198) formulated two null hypotheses:

- There is no interaction between reading-comprehension, learner performance in tests and the way texts and the response format of tests are organised.
- There is no interaction between reading-comprehension test performance, learners' language proficiency level, and text organisation and/or response format.

Kobayashi (2002) rejects both hypotheses. With regard to the first hypothesis, it is suggested that it is essential to take the type of reading passage into account when learners' reading comprehension is tested by means of summary writing (Kobayashi, 2002: 198). The implication is that reading comprehension performance will reflect learners' proficiency more accurately if texts used in comprehension tests are well structured.

With regard to the second hypothesis, it is suggested that the way in which reading passages that are used as input for summary writing are structured does not matter to less proficient learners, but it matters greatly to more proficient learners. One

implication of the findings is that the effects of the format in which learners produce results (whether cloze format, open-ended questions or summary writing) are more evident when the proficiency levels of learners are higher:

> In other words, learners with higher English language ability are more susceptible to different test formats and may not be able to demonstrate their reading comprehension skills when an unsuitable kind of format is used to measure them (Kobayashi, 2002: 209).

The study by Kobayashi (2002) supports the view that tests cannot be used to accurately determine the level of language proficiency. If the way a text is structured, for example, in a comprehension test has such a great effect on the learner's performance, the results on a test cannot possibly be regarded as an accurate reflection of the learner's level of proficiency.

Further critique on the use of standardised tests are, for example, that research projects that used standardised tests often report results in an unsatisfactory way. In general, it has been found that some researchers only reported the minimum and maximum scores achieved by the whole group examined. Other studies divided the subjects into levels of development and reported scores as averages for the whole group without any indication of how the levels were defined, or without mentioning ranges or standard deviations in the group. Sometimes, research results are only given for a part of the group. It is not even always specified for which part of the group the results are valid. Results from studies that use standardised tests are thus often of little value to others working in the same field (Thomas, 1994). Therefore, standardised tests are not very good indicators of language development.

3.5 Rating Scales

Language proficiency is often conceptualised in rating scales (Wolfe-Quintero et al., 1998: 6). They entail matching verbal descriptions to the performance of an examinee. Categories are predetermined and each category is accompanied by a description of what is expected of a performance at that level. Rating scales can be defined as offering definitions of learner proficiency at successive levels of ability (North & Schneider, 1998: 219).

Rating scales are increasingly being included in tests of second language performance. In fact, objectively scored items, such as discrete-point tests, are more and more being replaced by rated tasks, such as in the IELTS (International English Language Testing System). This is not, however, a new idea - teachers have been rating written compositions for years. Now, tests designed to measure more integrated use of language also make use of ratings.

Many examples of rating scales are available in modern textbooks on language testing. These examples usually include a series of short descriptions of the different levels of language ability. The purpose of a scale is to describe what a learner's expected abilities are at a certain level. This makes it easier for the teacher to decide what level or score should be attributed to each learner in a test. Textbooks may supply examples of scales, but the procedure to be followed for the development of a scale is generally not included. Guidelines for the construction of scales are not readily available (Upshur & Turner, 1995).

The growing popularity of rating scales can be attributed to their numerous functions, such as:

- providing a standard or stereotype against which learners can roughly measure their performance and compare their self image;
- providing a common standard for judgements, increasing the reliability for subjectively judged ratings;
- provide guidelines for the construction of tests;
- reporting results of teacher assessment, self assessment, test scores and rated tests in terms of the same instrument;
- providing coherent internal links within an institution between pre-course testing, syllabus design, organising materials, progress assessment and certification;
- establishing a framework of reference which describes achievement in an educational system in terms that are meaningful to all the parties involved;
- providing a common yardstick for comparing systems or populations (North & Schneider, 1998: 219).

Critique

Regardless of their growing popularity, rating scales do not provide the perfect solution when it comes to measuring second language development. Matthews (1990) notes that the criteria for developing rating scales are often arbitrary and inconsistent. Furthermore, the validity and the reliability of rating scales are questionable. Fulcher (1987) argues that scales are based on theory without the adequate empirical support. He is concerned about the possibility that scales reflect what researchers think happens, instead of reflecting what actually does happen. Fulcher suggests the need for data-based criteria for the evaluation of performance. Upshur and Turner (1995: 4) are of the opinion that the lack of direction and clear procedure for developing rating scales is, to a large extent, is reason for the criticism against rating scales.

Upshur and Turner (1999: 84) further argue that the method of testing influences the strategies for assessing language performance. Second Language Acquisition research focuses on examining the effect that testing methods has on what the students produce and is unconcerned with scores. Language-testing research, however, examines the effects the method of testing has on the scores, without considering what the learners actually produce. If the effect that the method of testing has on the score is ignored, the score is only a function of the examinee and the test task. When one considers the effect that the method of testing may have on the results, a score is no longer only a function of the learner's characteristics or test ability and the test requirements, but also of other non-characteristic abilities and requirements. "Tests are never perfectly valid. Every method of testing requires a variety of abilities [i.e. characteristic and non-characteristic of the learner] for satisfactory performance" (Upshur & Turner, 1999: 84).

The method of testing has potential effects on the contents of rating scales as well as on the final rating score achieved by the learner. Studies demonstrating that the methods used to measure language ability influence performance on language tests include Bachman (1990), Bachman and Palmer (1981; 1982; 1996), Brutsch (1979) and Clifford (1978; 1981). However, any method may include a number of characteristics that may influence the scores. Bachman (1990: 163) discusses variables that may influence the performance of a learner, such as the testing

environment, task requirements, scoring procedures and rating scales, as well as test stimuli.

Turner (2000: 3), however, points out that Bachman does not include any variables involved in the development stage of a test, such as a rating scale. For example, the team of scale developers and the specific sample of learner productions naturally affect the development process of scales that are derived empirically. Turner (2000) asks to what extent the variables make a difference in the final scale and, in turn, on the final score ratings.

Upshur and Turner (1999) note that no explanation is available for this relationship. There is no theory or method to explain how particular aspects of test method affect what the learners produce and how the differences in these utterances are reflected in test scores. "We propose the presence of different rating strategies as a result of the way task type determines the importance of various discourse characteristics. We find no theory to explain this relationship" (Upshur & Turner, 1999: 106).

Furthermore, there is no developed explanation of how characteristics of the rater and the learner interact with each other and with the characteristics of the utterances to produce ratings. There is also no explanation of how tasks and well-functioning rating scales relate to one another. In other words, rating scales ignore many influential relationships that exist when it comes to rating a learner's performance. In short, Upshur and Turner (1999: 106) argue for "developing and evaluating a more sophisticated model of performance testing" which seems to be "an important challenge for LT [language teaching] and for SLA".

Another problem regarding ratings scales concerns their validity and reliability. This is especially true for scales that are used in the testing of performance in a second language. Some of the problems encountered are listed by Upshur and Turner (1995: 5-6):

Problems concerning the reliability of rating scales are:

- Grading standards change as students improve. Teachers may unconsciously lift their standards as the learners move to a higher level of ability. A teacher of a class at a higher level may attribute the same average score to her/his class as a teacher who uses the same rating scale, but teaches a class with learners at a lower level of ability.
- Raters disagree on what the description of levels provided by a scale mean. A learner may be rated highly by one rater and much lower by another rater.
- Only a portion of scales published in textbooks is applicable to the classroom situation. Generally, published scales are too broad.
- Scales developed in one class setting may not be appropriate for use in another class setting. This description of the highest or lowest scale category may be too high or too low for the new setting.
- A rater's standards may change during one rating session. Scale descriptions aren't precise enough to keep the rater from being influenced by other learners' performances in the same rating session. A learner's performance may rather be measured against another learner's performance graded minutes before.

Problems concerning the validity of rating scales are:

- A teacher's own objective is often different to the description provided by the scale. For example, not all the features considered necessary to be incorporated by the rating scale may be included in the teacher's objectives of the course.
- Scale points are often described by a list of features that do not actually co-occur in the performance being rated.
- Features are included in the descriptions provided by the scales that actually do not occur in learner performance. It may be predicted, for example, that learners make certain mistakes at a particular level that actually do not occur in their performance.
- Some descriptions reflect questionable assumptions about the order of second language development.

Lumley (2002) examines the process by which raters of ESL texts make scoring decisions (by means of an analytic rating scale). Lumley explains that although raters seem to follow a similar procedure in three stages, the relationship between the content of the scale and the quality of the text remains obscure. Raters have to merge their own impression of the text, the specific features of the text and the descriptions of the rating scale, and then produce a score for each learner. Problems arise because the rules and the scale do not cover every situation or possibility. This results in tension between the rater's personal impression and the rules. A somewhat undetermined process resolves this tension.

Criticism specifically related to the inefficiency of rating scales as measurement of writing proficiency includes that of Kondo-Brown (2002). He explores the possibility of using a modified version of Jacobs et al.'s (1981) rating scale for norm-referenced decisions about the writing ability of Japanese L2 learners. There is no established scale for assessing Japanese L2 learners' writing ability. The study also investigates the bias of trained raters towards certain types of candidates and certain criteria in assessing Japanese second language learner writing. The raters that were used in the investigation were self-consistent and produced highly correlated scores. However, big differences were encountered in the over-all sternness with which the learners were rated. Some candidates and criteria were scored more harshly or leniently and each rater followed an individual bias pattern, which was different from the others.

The reliability of writing performance assessment can be improved by training and better specifying scoring criteria and tasks. However, the validity of writing assessment has not received enough attention. A number of researchers (e.g. Huot, 1993; Vaughen, 1991; Hamp-Lyons, 1990; Cumming, 1997; Kroll, 1998) have called for further investigation of the rating process in the context of ESL assessment. They are especially concerned about the superficiality of rating scales in comparison with the complexity of written texts and readings made of them (Lumley, 2002). Subjectivity also remains a concern:

> [J]udgements must be made by human specialists...They cannot agree on the absolute quality or relativity of essays, nor can they agree on the specific qualities in essays that make them

> good, worse, or worst. If their judgements correlate at .70, they are agreeing only 49% of the time...we do not share a construct of writing quality (Hamp-Lyons, 1990: 80).

Subjective raters show concern about the validity and the reliability of rating scales as means of assessing written performance. Rating scales cannot possibly include every situation, and each rater may interpret the scale differently. Every rater still makes a personal decision at some point.

North and Schneider (1998: 219) hold that the main weakness of this type of scoring is that "there is no guarantee that the description of proficiency offered in a scale is accurate, valid or balanced". They echo a definition by Clark (1985: 348) that suggests the main weakness of the rating scales:

> [Rating scales offer]...descriptions of expected outcomes, or impressionistic etchings of what proficiency might look like as one moves through hypothetical points or levels on a developmental continuum.

Raters can be trained to think in a similar way and learners are able to actually interpret a scale very successfully for the purpose of self-assessment. However, this does not mean that what the scales say are valid reflections of proficiency.

It is often not clear why and how certain statements at specific levels are arrived at (i.e. why and how it was decided that statement D should resemble level 3 and statement K resemble level 4). Furthermore, many proficiency scales claim to offer criterion-referenced assessment, while this is in fact not the case. Firstly, this is because the meaning of descriptors at one level depends on a reading of the descriptors at all other levels. Secondly, descriptors are often formulated 'overtly relativisticely' or norm-referenced. For example, phrases such as 'better than level 2' and expressions such 'good', 'poor' or 'weak' are used. Many descriptors do not present an independent criterion, which can stand sensibly and descriptively on its own and generate a simple yes/no response from the rater (North & Schneider, 1998: 220).

The production of valid, reliable and objectively rateable proficiency scales seem to be close to impossible. According to North and Schneider (1998: 220), most scales

appear to have been "produced pragmatically by appeal to intuition, the local pedagogic culture and those scales to which the author had access". During the process of development, the following are often ignored:

- Using a model of communicative competence and/or language use;
- Ensuring that the different categories and the formulation of descriptors are relevant and clearly understandable to the user;
- Making use of a model of measurement;
- Avoiding the danger of taking the content of one rank-ordered scale and using it inappropriately in another (North & Schneider, 1998: 220).

In short, different scales that are all intended to rate written proficiency should produce comparable scores or at least comparable rankings. Looking at the extent to which rankings differ, however, scales either rate different constructs or error in the measurement is dominant. According to Hamp-Lyons (1990), both these factors that spoil valid and reliable writing assessment are occurring:

> Our constructs of good writing are inadequate and incomplete, and we have severe measurement error in our tests. Our mistake to date has been in focussing our attention on the scoring procedure as the source of measurement error or accuracy; ... the reader, the human instrument, is the source of much error (Hamp-Lyons, 1990: 81).

The process of developing language proficiency rating scales, the validity and reliability of the scales, the objectivity with which rating scales can be used and the accuracy of the results are all questioned. Basically, every aspect of rating scales seems to be unreliable.

3.6 Conclusion

In this chapter, four means of measuring language proficiency were discussed, namely Mean Length of Utterance, Error Analysis, Standardised Tests and Rating Scales. Each of these measures has been used to determine the development of second language acquisition in learners. All of them, however, are rough measures and inadequate as measures of development.

A more exact means for measuring development, particularly syntactic development, is necessary. This measure should take the form of index according to which levels of development can be determined. The next chapter discusses the need for such an index.

Chapter 4

The Need for an Index of Development

4.1 Introduction

It was pointed out in the previous chapter that a more exact means for measuring syntactic development is required. With various complaints and comments about dropping standards in South African school education, it has become necessary to determine the level or stage of acquisition that learners of English as a second language have reached at the matriculation level. What is required is an index with which to measure linguistic development. The focus of this chapter is on the need for such an index.

This study primarily aims to measure language development and not language proficiency. The difference between these two concepts is also discussed in this chapter.

4.2 Developmental Level and Proficiency

According to Wolfe-Quintero et al. (1998: 2), "language development refers to characteristics of a learner's output that reveal some point or stage along a developmental continuum". Learner language develops through stages or levels and can be categorised into hierarchies. This process, during which the learner progresses from one stage to the next towards native-like use of the target language along the interlanguage continuum, is what is called 'language development'.

The present study focuses on determining the level of language development that Grade 12 ESL learners have reached. This will be done by means of an index. This index will not measure language proficiency, but language development.

'Language proficiency' refers to a much broader concept than 'developmental level'. But it is not an easy term to define, and little consensus has been reached on this.

During the 1950s and 1960s, proficiency was understood to refer to "the degree of correctness in certain formal aspects of L2" (Pienemann & Johnson, 1987:67). The hypothesis that language proficiency consisted of a single unitary characteristic, and a quantitative, statistical research methodology, dominated language-testing research.

During the 1970s the proponents of 'communicative competence' adopted a broader view of language ability. However, attention shifted towards pragmatic factors in theoretical linguistics, and with the introduction of the notion of 'communicative competence', the argument arose that correctness on the learner's part does not necessarily indicate the learner's ability to use a second language. The work of applied linguists such as Widdowson (1978; 1979; 1983), Savignon (1972; 1983) and Canale and Swain (1980) had a strong influence on language testing. They proposed a more wide-ranging model of language. They saw language ability as 'multi-componential' and dynamic. Language was regarded as the creation of discourse or the situation negotiation of meaning. This view required language testers to consider the speech and sociolinguistic aspects of language and the context in which it takes place. It forced them away from their narrow conception of language as an isolated feature (Bachman, 1999).

Bachman's (1990) model of communicative competence and performance, however, is the most comprehensive model to date. The model regards language proficiency as "an expanded conception ... whose distinguishing characteristic is its recognition of the importance of context beyond the sentence to the appropriate use of language" (Bachman, 1990: 82).

Canale (1994) suggests a definition of language proficiency (usage) that is based on a socio-theoretical foundation. These two definitions, along with the model proposed by Bachman (1990), share at least two critical features. Each definition accommodates the four linguistic modalities: speaking, listening, reading and writing. Each definition also places language within a specific context, for example the educational setting (Kaur, 2000).

Wolfe-Quintero et al. (1998: 2) describe proficiency as being related to dividing language users into cross-linguistic groups according to the normal distribution of

their language abilities (i.e. put them into groups on account of how well they use the target language).

Pienemann and Johnson (1987) separate the terms 'proficiency' and 'development' to the point of "severance". They argue that it is not possible to speak meaningfully of "development of proficiency". 'Proficiency' cannot be "successfully defined in an objective, non-relational (i.e. non-subjective) way and ... the most that can be expected of such a term is that it can be reduced to the way one group of people (raters) behaves in regard to a second group (ratees)" (Pienemann & Johnson, 1987:46). The closest to a definition of proficiency provided by Pienemann and Johnson is "that which is measured by proficiency scales such as the Australian Second Language Proficiency Ratings (ASLPR), the American Council for the Teaching of Foreign languages (ACTFL), or the Foreign Service Institute (FSI) scale".

They say: It is possible, however, to speak meaningfully of 'language development'. In order for this to happen, a theoretically logical and empirically provable model of language development needs to be established. Such a model will have to be based on observed learner behaviour and deal in objectively quantifiable units or values that have a universal status (Pienemann & Johnson, 1987: 46).

Unsworth (2002: 2) describes L2 proficiency as a learner's "accurate and sophisticated knowledge of target-language grammar". The focus is on grammar, rather than on skills covered by terms such as 'communicative competence'. The focus is on morpho-syntax and lexicon, while phonology is excluded.

4.3 The Need for an Index of Syntactic Development

The need for an index or way of measuring a language learner's (syntactic) development or progress in acquiring a second language is not new. Brooks (1964: 199) notes that

> [s]ince the beginning of this [the previous] century there has been an increasing need for a thorough and a systematic program of tests designed to measure success in many subject areas ... Why, it may well be asked, is such an elaborate program of measurement necessary? Can

> we not choose what is easiest to measure and let that be an index of over-all accomplishment? Such a plan may indeed be workable if there has been appropriate and sustained emphasis upon each of the four skills from the beginning of a learner's classroom experience. But it is quite possible to develop one of the skills to a considerable degree while paying little or no attention to the others...

Brooks (1964) calls for a programme that will enable us to answer questions such as: How can tests be used to provide a steady index of the learner's progress as he develops from one stage to the next? How can tests help us to identify areas of strengths and weaknesses in what the student is gaining and what the programme offers? And how can the test programme review whole areas of language learning, cultural and literary studies, and assure the researcher that the student's achievements are satisfactory and is therefore qualified, if he wants to, to go on to more advanced work?

Labels such as 'beginner', 'advanced' and 'intermediate' are not very precise descriptions of the level of development an L2 learner has reached. Larsen-Freeman (1978: 440) describes these labels as very subjective and argues that they are too vague to express the actual level of competence of a particular group of students. An index of second language development would provide a more precise description.

It is important to be precise when labelling a learner's level of syntactic development for various reasons. For example, different factors influence the process of second language acquisition depending on the learners' level of proficiency. Larsen-Freeman and Long (1992: 42) point out some occasions when researchers need to be able to measure a learner's or learners' progress in their second language development. For instance, the sort of information provided by an index will be necessary in:

- characterising a specific subject population (i.e. at what stage of acquisition of the L2 are they?);
- in order to determine the effect of a particular treatment on learners' acquisition (i.e. did the treatment cause the learners to move from stage 1 to stage 2?); or

- when different rates of development between learners are being measured (i.e. how long did learner A and B respectively take to move from stage 1 to stage 2 and what could account for the difference?).

Learner proficiency is sometimes subjectively gauged and subjects are merely classified as being beginning, intermediate or advanced. Particularly, vagaries such as this and recognising the need for a measure that is sensitive for differences over time led to Larsen-Freeman's (1976a) call of an SLA index of development (cf. Larsen-Freeman & Long, 1992: 43).

Larsen-Freeman and Strom (1977: 124) argue that studies have often been hampered due to the researcher's inability to determine whether the subjects are at the same level of development. For example, Hukata (1975) conducted a longitudinal study in which he attempted to examine the issue of language transfer as a factor in the process of learning a second language. He was specifically interested in the phenomenon of learners avoiding using certain structures. Hakuta decided to compare the L1 subject's use of relative clause with that of the L2 learner subject. However, no reliable means was available to determine whether the two subjects were at similar stages of development in English. Therefore, no reliable comparison could be made between the two subjects (Larsen-Freeman and Strom, 1977: 124).

Thomas (1994) argues that it is necessary to regard the performance of a learner or a group of learners in a certain context or on a specific task in the light of their knowledge of the L2, because different factors (such as first language transfer and overgeneralization) may play different roles at different stages of language acquisition:

> We do not know enough about whether, and if so, how knowledge of a second language is acquired, organised, integrated, stored, or mobilised differently at different levels of proficiency. However, so long as individual research results are not keyed to learners at identifiable stages in L2 development, it will remain difficult to investigate these issues (Thomas, 1994: 309).

In other words, it is necessary to assess and report the developmental levels in order to generalise research results. It is also important to understand how knowledge and language use may change during the course of acquisition.

Assessing proficiency is also important in research concerned with developmental sequences. Furthermore, it is vital in order to have control over the levels of development of subjects that are part of a study of L2 acquisition. In other words, researchers should make sure that subjects are at similar stages of proficiency at the beginning of the study in order for results to be significant. In addition, it is sometimes necessary for researchers to determine whether one aspect of the L2 has been acquired in relation to another aspect of the L2 that serves as the research target (Thomas, 1994).

Larsen-Freeman (1983) aims to establish an ESL index that is not based on the structure of the English language itself, so that it will be universally applicable. "Such an index would consist of a description of the stages through which second language learners pass on the way to acquiring any second language" (Larsen-Freeman, 1978: 127).

Unsworth (2002) identifies a number of characteristics that are considered desirable for an index of development:

- It should not be tied to any specific L1 or L2. Speakers of different first and second languages should be able to use the index;
- The index should be readily available, i.e. no specific instrument is needed in order to carry out the analysis;
- The measure should not be age-specific;
- Different educational backgrounds should not cause interference;
- The measure should be applicable *post hoc* to previously collected data;
- The analysis can be performed at the same time as the experimental task, as to ensure that the student's knowledge of that specific time is reflected;
- The measure should be relatively quick to be carried out and to analyse;

- The index should reflect global L2 'competence' (i.e. influence of metalinguistic knowledge should be minimised).

In short, there is a need for establishing a second language acquisition index of development. "Such an index would be a developmental yardstick by which researchers could expediently and reliably gauge a learner's proficiency in a second language (Larsen-Freeman, 1978: 439).

4.4 Classification of Measures of Development

Wolfe-Quintero et al. (1998) divide measures used in second language development studies into three major categories. Each category corresponds with a different aspect of development. The three categories are: fluency, accuracy and complexity. "This classification captures three types of intuitions that have been assumed in studies of second language development in writing" (Wolfe-Quintero et al., 1998: 4):

- As L2 learners become more proficient, they write more *fluently*, or write more in the same number of time.
- As they become more proficient, L2 learners write more *accurately*, or produce fewer errors, than before.
- As they become more proficient, L2 learners produce more *complex* sentences - both grammatically and lexically - than before.

These three characteristics of language development are assumed to progress together. More proficient learners are assumed to write more accurately, fluently and more complex than less proficient learners. Individual variation may, however, occur at any point.

Wolfe-Quintero et al. (1998: 4) attempt to find fluency, accuracy and complexity measures "that work together in order to define a global picture of language development in second language writing". In the traditional sense, Lennon (1990: 390) defines these three concepts as follows: 'Fluency' refers to speaking with "native-like" rapidity; 'accuracy' refers to being "error-free"; while 'complexity' refers to "using a wide range of structures and vocabulary".

These components of development appear to correspond with two separate concepts of language processing, namely language representation and language access. 'Language representation' is associated with declarative knowledge, linguistic representation, usage or competence. 'Language access' is associated with procedural knowledge, control of representation and use or performance.

> In this view, complexity and accuracy reflect the second language learner's current level of language knowledge; while complexity reveals the scope of expanding or restructured second language knowledge, accuracy shows the conformity of second language knowledge to target language norms. In comparison, fluency is a function of the control in accessing that knowledge, with control improving as the learner automatizes the process of gaining access (Wolfe-Quintero et al., 1998: 4).

The purpose of the present study is to establish a measure according to which it is possible to determine the point of development achieved by a group of ESL learners. The focus is only on measuring accuracy. This study focuses on measuring syntactic accuracy, in order to determine the stage of development reflected in the writing of Grade 12 ESL learners.

Accuracy can be defined as "freedom of error" (Foster & Skehan, 1996). According to Pica (1983), the degree to which language is error-free can be measured by analysing target-like use of a language, while taking both the context and uses of the structure into account. From this, Wolfe-Quintero et al. (1998: 33) formulate the following definition of 'accuracy': "accuracy is the ability to be free from errors while using language to communicate in either writing or speech".

Pienemann and Johnson (1987) point out that in many studies of second language acquisition, correctness has been used as a measure of development. They argue that accuracy is "quite an erroneous criterion for development" (Pienemann and Johnson, 1987: 69), since it has been proven that learners at a low stage of acquisition quite commonly produce very accurate structures, while learners at higher stages may produce many inaccurate structures. Some learners prefer to simplify certain structures to communicate more effectively, while others prefer to concentrate on being accurate rather than communicate effectively. In other words, a learner's level

of development cannot be determined accurately by merely counting the number of errors a learner makes. Additional measures need to be considered.

A distinction can be made between 'developmental features' and 'variational features' in language acquisition. The first type of feature is acquired chronologically, whereas the second type is produced, or not produced, at different rates by different learners. Accuracy increases in both dimensions. This implies that accuracy is not a linear phenomenon. Pienemann and Johnson (1987) argue that an increase in accuracy does not represent a simple movement along either the developmental or variational dimension (Pienemann & Johnson, 1987: 70).

It is necessary to be clear about what exactly is being measured and what means of measuring are being used. For example, is 'proficiency' being measured, or is the researcher in fact measuring development or accuracy? Measures of language development cannot be used to measure language proficiency. Also, because 'proficiency' cannot be regarded as only one object of study, one should be specific about which subset of proficiency is being studied. This study focuses on measuring language development by means of analysing written samples of speech (i.e. written proficiency).

This study acknowledges Pienemann and Johnson's objection against associating accuracy with development. This study does not *equate* accuracy with development as a whole, but rather regards it as one aspect of development. Lightbown (2002: 532) points out that accuracy is one way to assess language acquisition. This study focuses on the development of syntax and grammatical accuracy. Accuracy measures determine to what extent a learner has mastered the target grammatical system. The type of all-encompassing study required if all aspects of language development were to be addressed or taken into consideration would be of much larger scale than the present study.

4.5 A Suggestion for an Index of Development

Hunt (1965) suggests the length or number of T-units[4] that a learner produces as a possible measure that can be used to establish different stages of development. Scott and Tucker (1974) found that the number of error-free T-units that learners produce increased for more advanced learners. Larsen-Freeman and Strom (1977) and Larsen-Freeman (1978) regard measures of length (such as the average number of words written per T-unit) and the number of error-free T-units produced by a learner as valid measures to base an index on. The T-unit is discussed in detail in the next chapter.

4.6 Conclusion

The purpose of this chapter is twofold. It distinguishes between the concepts of 'development' and 'proficiency'. Language development refers to the process a learner goes through while acquiring a second language. It refers to the progress a learner makes from one stage to the next along the interlanguage continuum towards native-like use of the target language. The language a learner produces displays characteristics that indicate the learner's stage of development. In the present study, proficiency is used as a more general term. It refers to a number of subsets or areas in which a learner can develop second language skills, such as speaking, listening, reading and writing. This study focuses on measuring the language development of learners in written proficiency.

This chapter also emphasises the need to establish an index according to which learners' language development can be classified numerically.

In the context of this study, 'accuracy' is not equated with language development, but is used to refer to one category (the others being frequency and complexity) into which measures of development can be classified. Accuracy is regarded as an important measure of the degree to which a learner has acquired the grammatical system of the target language.

---

[4] A more detailed explanation and definition of a T-unit follows in 5.6.

A researcher must make sure of what is being measured in a study and how to measure it (i.e. the instrumentation used must not in fact measure something completely different.) Although the terms accuracy and correctness basically mean the same thing, accuracy is used rather to refer to one category (the others being frequency and complexity) into which measures of development can are classified (according to Wolfe-Quintero et al., 1998).

The next chapter discusses the T-unit as a measure on which to base an index of language development. The original study that introduced this measure is discussed.

Chapter 5

The T-unit as Measure of Language Development

5.1 Introduction

A number of conventional measures of language development have been considered, but found insufficient for the purposes of this study. The previous chapter also stressed the need for an index to measure development and to determine the stage of acquisition. The T-unit was suggested as an adequate measure on which to base an index (cf. 4.5) Larsen-Freeman (1978) proposes the length or number of T-units produced by learners as possible measures to establish different stages of development (Larsen-Freeman, 1978: 134). Scott and Tucker (1974) noted a significant increase in the production of error-free T-units as students become more developed, both orally and in writing.

This chapter is concerned with the origin and development of the T-unit measure. Hunt's (1965) seminal study is discussed, and an overview of the development of the T-unit is given.

5.2 A Criterion for an Index of Development

Larsen-Freeman (1978: 134) supplies the following basic criterion for a valid, reliable and unbiased index of development. She explains that it is necessary to look for something more global than learner performance on a specific structure or structures: "... it would be difficult to choose structures which were representative of overall competence in the English language and not be biased against any particular language group." Larsen-Freeman suggests that the index should not be based on the English language itself to ensure that the index is one of second language acquisition and not only of ESL. She suggests that a measure of syntax will be necessary to measure levels of development.

5.3 The Development of the T-unit

The T-unit was developed by Kellogg Hunt and first presented in a study entitled *Grammatical Structures Written at Three Grade Levels* (1965). Hunt received an award for a research project which contributed significantly to the teaching of English. The award is presented annually by the Executive Committee of the National Council of Teachers of English. The Committee unanimously decided that Hunt's study provided new tools for language research and suggested new objectives in language teaching, which were of potentially major significance (Carlsen, 1965: ii).

During the early 1960s, numerous studies speculated about the differences between child and adult language performance. Other studies tried to confirm these speculations without contributing much to explaining the changing process from child to adult performance. The grammatical systems that were used to analyse language structures were part of the problem. These served only as very rough yardsticks. Furthermore, the classification of structures in the language of children was often based on the subjective decisions of researchers. In some cases, structures were only classified after the researcher had changed the actual language produced by the child.

Hunt's study was one of the few of that time that offered solutions:

> Hunt has worked out a scheme for the analysis of written work that seems to minimise the subjective interpretations and arbitrary classifications that earlier investigators have ultimately had to use. His construction of the T-unit, for example, overcomes the problem of what is or is not a sentence and therefore establishes a base unit that can be systematically investigated (Carlsen, 1965: i).

According to Carlsen (1965: ii), Hunt's study offered a possible operational definition of language objectives for the schools. School programmes should aid learners in developing mature writing patterns. "Hunt's study does offer the possibility of establishing objective criteria for measuring language change that may make the testing of curricular theories possible" (Carlsen, 1965: ii).

5.4 An Introduction to Hunt's study

5.4.1 The Procedure

Hunt (1965: 1) describes the purpose of his study as twofold:

- To provide a coherent, systematic and broad method of procedure, which can be refined to accommodate details for the quantitative study of grammatical (syntactic) structures.
- To search for developmental trends in the frequency of grammatical structures written by students in the fourth, eighth and twelfth grades.

Following procedures for this type of study, Hunt (1965) determined the sentence length, clause length, subordination ratio and kinds of subordination clauses. He also studied a different unit and analysed structures within the clause (e.g. coordinated structures, nominals, the verb auxiliary, the main verbs and complements, etc.).

In order to conduct an intensive analysis, the total corpus of writing was kept small, while a large sample of each individual's writing (1000 words) was analysed. Teachers were instructed merely to collect the writing samples of the selected students without making any changes to the student's own writing. The study was conducted over a long time span in order to improve the chances of finding significant differences between grade levels.

The fourth, eighth and twelfth grades were chosen as starting, check and finishing points respectively. All three grades wrote the Short Form of the California Mental Maturity Test, in order to keep the specifications for writing the same for all three grades. The chosen subjects all scored between 90 and 110 on the test. As far as possible, an equal number of girls and boys (nine each) were chosen from each grade. The names of the children were substituted for a number code on each paper for identification purposes. Each grade wrote on topics that were characteristic for that grade to write on and were uninfluenced by the topics given to the other two grades. The argument was that it is possible that grade twelfths characteristically write on topics that fourth graders cannot be expected to write about. Each grade wrote on a

variety of topics. Any group of words that the investigator could not understand was labelled a 'garble'.

5.4.2 Sentence Length as Inadequate Measure of Language Development

Sentence length was regarded as a natural place to begin a quantitative study of writing development. Hunt (1965: 7) notes that it is first necessary to know what is regarded as a sentence before any meaningful discussion on sentence length can be attempted. It may be assumed that children write shorter sentences than adults. However, some children may not use full stops to indicate the end of a sentence, while others may string a sentence together using numerous *and*s where an adult would have used a full stop:

> If a sentence is defined as whatever is written between terminal marks, and if sentence length is assumed to be an index of language maturity, then the child who under-punctuates the most or uses *and* the most will, regrettably, be credited with the greatest language maturity (Hunt, 1965: 8).

Hunt (1965: 8) argues that by inserting punctuation marks where necessary, the researcher will weaken the validity of the results. Thus, the concept of 'sentence' presented a problem to the investigator. The subjective option of editing the samples was rejected. The objective unit (meaning the unedited sentence as produced by the learner), however, could not be expected to be a very good index, because younger writers tend to use too many *and*s and not enough punctuation marks. It was therefore necessary to find a better unit.

Hunt (1965: 8) calculated the average length of sentences as originally produced by the students (ignoring under punctuation and/or too many *ands)* by dividing the total number of sentences – those passages starting with a capital letter and ending with a full stop – by the total number of words for each student. Hunt calls this the 'average punctuated sentence length'. Results showed that the average was larger for each successively older group: fourth graders average 13.4, eighth graders average 15.9, while the average for grade twelve students was 16.9. Although the average sentence

length did increase with successively older grades, Hunt argues that sentence length is not a satisfying index of individual maturity.

The reason for this, according to Hunt (1965: 10), is the lack of punctuation and the use of periods by younger writers. For example, the student with the highest average sentence length was found to be a fourth grader, who wrote one sentence consisting of 77 words without a period. The same student wrote another sentence of 68 words without full stops. Two other fourth grade students each used more than 70 *and*s between main clauses, while the average twelfth grader used only 5. Therefore, while sentence length may be a more satisfying index of language maturity for superior students, it is not so for young average students. (Hunt acknowledges that the overuse of *and* is not a grammatical offence, but rather a stylistic one.)

5.5 Hunt's Search for a More Valid Index

After concluding that sentence length was not an objective index of language maturity, Hunt (1965) continued the search for a more valid index. This involved determining both the mean length of clauses, as well as the ratio for subordinate clauses for all three the grades.

5.5.1 Mean Length of Clauses

Hunt (1965: 15) regards a clause as "a structure with a subject and a finite verb". Coordinated subjects and any coordinated part of the verb only lengthen the clause. Hunt divided the total number of words by the total number of clauses to determine the mean clause length for each student. The average clause length for fourth graders was found to be 6.6 words. Eighth graders averaged 8.1 and twelfth graders 8.6 words, with differences significant at the 0.01 level (Hunt, 1965: 15). In other words, there was an increase in the mean length of clause with each successively older grade. Hunt's results, however, showed considerable overlapping of grades, where 15 out of 18 eighth graders overlapped twelfth graders, which made mean clause length a less accurate index of language maturity.

5.5.2 Subordination Ratio

The ratio between subordinate clauses and all clauses (both subordinate and main) was also calculated. The frequency of subordinate clauses was found to increase from grade to grade. The ratio reported by Hunt (1965: 18) is as follows: for fourth graders, 22.2%, eighth graders 28.8% and for twelfth graders 40.1% with differences significant at the 0.01 level. Again, there is considerable overlapping between the grades (Hunt, 1965: 18).

5.6 A More Promising Index

Hunt regarded the finding that both mean length of clause and the number of subordinate clauses increase from one grade to another significant:

> If it is true that 1) the average main clause written by successively older students has more subordinate clauses attached to it, and if, in addition, 2) those clauses are longer, then the total length of such a unit would of course increase as a result of the two subsidiary kinds of lengthening. This unit whose total length is being discussed contains, to repeat, one main clause with all the subordinate clauses attached to it...The length of such a unit might turn out to be a good index of maturity (Hunt, 1965: 20).

Hunt suggests that a whole piece of writing should be divided into such units. He predicts no difficulty in deciding whether an expression should go with the preceding or the following main clause. An *and* or other conjunction between two main clauses will always be counted with the following main clause. The lack of full stops in proper places will not interfere with the "slicing up" procedure, unless the passage is already an "unintelligible garble" (Hunt, 1965: 20).

Hunt (1965: 20) presents the following sample produced by a fourth grader. It is punctuated as one sentence consisting of 68 words.

> I like the movie we saw about Moby Dick the white whale the captain said if you can kill the white whale Moby Dick I will give this gold to the one that can do it and it is worth sixteen dollars they tried and tried but while they were trying they killed a whale and used the oil for the lamps they almost caught the white whale.

Sentences such as this one will obviously be problematic when one attempts to determine language maturity by means of sentence length. Therefore, Hunt proposes that the researcher slice sentences such as this one up into smaller meaningful parts. The "slicing up" is presented as follows, where every unit now begins with a capital letter, is ended by a full stop and a new clause is indicated by a forward slash.

> I like the movie / we saw about Moby Dick, the white whale./
> The captain said / if you can kill the white whale, Moby Dick. / I will give this gold to the one/ that can do it.
> And it is worth sixteen dollars.
> They tried and tried.
> But / while they were trying / they killed a whale and used the oil for the lamps.
> The almost caught the white whale (Hunt, 1965: 21).

The original sentence is now sliced up into units that can be considered as the shortest possible grammatically allowable sentences into which the passage could be divided, without creating fragments. Several of the units contain only the main clause, just like a simple sentence. Multi-clause units, like complex sentences, also occur. However, no units like compound sentences or compound-complex sentences are included, since these would have to be sliced up into two or more parts so that each will only contain one main clause.

In naming these "units", Hunt (1965: 21) rejects the option of "minimal sentences". Too many problems are associated with the word "sentence", such as the different meanings attributed to it. A more practical option would be "'minimal terminable units'. These would be minimal as to length and it will be possible to start each unit with a capital letter and end it with a full stop. The minimal terminable sentence is referred to as the "T-unit" in short. Hunt is positive about the T-unit's ability to act as an index of language maturity:

> As a potential index of maturity, the unit has the advantage of preserving all the subordination achieved by a student, and all of his co-ordination between words and phrases and subordinate clauses (1965: 21).

If any co-ordination that the student established between main clauses or T-units (as in numbers 3 and 5 above) is broken, it is regarded as a gain rather than a loss in the search for an index of maturity. Co-ordination with *and* between T-units (referring to when T-units are strung together by numerous *and*'s) was, in fact, found to be an index of immaturity significant at the 0.01 level.

Hunt (1965) calculated the mean T-unit length for each student and grade. Fourth graders scored a mean of 8.6, eighth graders 11.5 and twelfth graders 14.4. Calculated in percentages, the increase is from 60% for fourth graders to 80% for eighth graders to 100% for grade twelves. (Note that Hunt calculated percentages throughout his study with the twelfth grade taken as 100% every time.) The differences are significant for grade at the level 0.01. Although overlapping still occurs between students of different grades, with this index it is much less than with any of the other indexes employed in Hunt's study.

5.6.1 The T-unit as Index of Development

Hunt (1965) tested all four indexes (mean sentence length, subordinate ratio, clause length and T-unit length) statistically in order to determine which indexes correlated the best and would best indicate a student's grade level. The score of each student on each index was subjected to a chi-square analysis. If the chi-square was found to be significant at the 0.5 level, a contingency coefficient was calculated. The T-unit length proved the best index (0.694), with mean clause length second best (0.616), subordination ratio third (0.523) and sentence length fourth (0.489) (Hunt, 1965: 23).

T-unit length

Young children generally seem to write shorter T-units than older children or adults. It may be more correct to say that the continual usage of short T-units is more characteristic of younger children's language use. Hunt (1965) found that fourth graders produced twice as many short T-units than eighth graders do, while eighth graders produce more than twice as many T-units as produced by twelfth graders. (Any number of words less than 9 words is taken as a short T-unit.) "So typical of grade level is the number of these short T-units, that it provides a better index of grade level than the widely used sentence length" (Hunt, 1965: 29). When figures of

the number of T-units written by each learner were subjected to a contingency coefficient analysis, the score was 0.70 with a chi-square of 57.87. In order to illustrate the dominance of short T-units in the writing of younger children, one can say that 43% of all words expressed by fourth graders are expressed in short T-units. Eighth graders only express 21% of words in such units, while grade twelfths express only 10% of their written words in short T-units. Writing many short T-units is a definite characteristic of younger writers.

Hunt (1965) found a correlation of 0.902 between the number of short T-units and the average length of T-units. The contingency coefficient was just as high. In other words, for practical purposes, the number of short T-units appears to be as good a measure as the average length of T-units. If the number of words per passage is known, counting the number of short T-units may be a valuable time-saving procedure.

As far as T-unit length is concerned, 9 to 20 words can be considered as a 'mid-length' unit, while more than 20 words can be considered as 'long'. Hunt's (1965) results show that all three grades write approximately the same number of mid-length units in a composition. However, fourth graders write more short units, while twelfth graders write more long units. Long units are as characteristic of twelfth graders as short units are of fourth graders. Hunt (1965: 33) reports the ratio of short units to long units as a dramatic 42 to1 for fourth graders, 5.3 to 1 for eighth graders and 1.4 for twelfth graders.

Ratio of Clauses per T-unit

Hunt (1965) suggests that the subordination ratio be replaced by a new measure, namely the ratio of clauses per T-unit. This measure gives information comparable to the old measure, but in a form that is more widely usable. The ratio of clauses per T-unit is defined as "the number of all clauses (both subordinate and main) divided by the number of T-units – or, since the number of main clauses is identical with the number of T-units, divided by the number of main clauses. The results are expressed in decimal" (Hunt, 1965: 33). The formula can be presented as follows:

- Subordination ratio = (subordinate clauses) / (subordinate + main clauses)

- Ratio of clauses to T-units = (subordinate + main clause) / (main clause)

Like the old subordination ratio, this new ratio is an index of maturity with a high level of significance (for each of the three grades) at the 0.01 level. The ratio of clauses per T-unit has three uses: 1) it provides an arithmetic bridge between clause length and T-unit length; 2) it conveniently expresses the frequency with which a subordinate clause is added to a main clause; 3) it serves as a significant index of one kind of grammatical development (Hunt, 1965: 36).

The ratio of clauses per T-unit enables the researcher to see whether language development is rather characterised by the tendency of older writers to produce longer clauses, or by their tendency to produce a larger proportion of subordinate clauses. Hunt (1965: 37) found that clauses are lengthened by the same percentage with which subordinate clauses are added to T-units.

Ratio of T-units per Sentence

This ratio would provide information about the characteristics of even mature writers in relation to whether they produce long sentences by joining many T-units, or by expanding T-units. The ratio is said to serve as an arithmetic bridge from T-unit length to sentence length. The ratio is not a significant index of maturity or immaturity, although it is useful in describing characteristics of a writer. It is indicative of how often T-units are co-ordinated and never has a value smaller than 1. To calculate the average frequency with which T-units are added to the first one in a sentence, one subtracts the value of 1 from the ratio (Hunt, 1965: 39).

Generally, Hunt's argument that syntactic complexity increases with age is supported, although results for adults were not always clear-cut. Steward (1978), for example, found that surface structure clause length in words can be as important as the length of T-units in the writing of adults. Steward does, however, still argue that T-unit length is the best indicator of syntactic maturity for writers moving from high school to university (cf. Mills, 1990: 107). The students who wrote the essays analysed in the present study fall precisely in this category.

The T-unit has also been found a significant index of syntactic development for kindergarten and elementary school children. O'Donnell (1976) describes the T-unit as "a simple, objective, and apparently valid measure of development of syntactic control". Therefore, T-unit length seems to be an adequate index of development to employ.

5.7 Intercorrelation of Five Synopsis Factors

Hunt (1965) investigated and related five clause-to-sentence factors, namely, average length of clauses, ratio of clauses per T-unit, average length of T-units, ratio of T-units per sentence and average length of sentences. Of these factors, T-unit length shows the greatest percentage increase and is statistically the best indicator of grade level. The intercorrelation of these factors is relevant to the researcher concerned with the achievement of individuals at a particular grade level. A number of conclusions can be drawn from the results obtained from the intercorrelation.

Hunt (1965: 44) suggests that one of the two factors – either the length of the T-units or the ratio of T-units per sentence - contributing to T-unit length may be a better index of individual achievement than T-unit length itself. For example, fourth graders lengthen their clauses to such a limited degree that any individual differences among them in this regard do not show up in T-unit length. When all the individuals are taken together, however, T-unit length is highly correlated with both contributing factors. The two factors contributing to T-unit length are unrelated to each other at each grade. However, they are significantly related when all individuals are taken together.

The ratio of T-units to sentences doesn't correlate significantly with any other factor except sentence length. However, when all individuals are taken together, this ratio correlates (though negatively) with T-unit length and its components, thus supporting evidence that over the years, a high score on the factor ratio of clauses to T-units indicates immaturity. Further evidence was found that sentence length, at least in early grades, is not an index of maturity, though it does have a significant contingency coefficient over a long span of time (Hunt, 1965: 47).

5.7.1 Summary of Clause-to-Sentence Findings

Hunt's study continues with a more in-depth analysis of characteristics of mature fiction and investigates how far the performances of twelfth graders are from the standard of superior writers. Although interesting, the remainder of Hunt's study is not relevant to the present study. Therefore, no further mention will be made of the second part of Hunt's (1965) study. This section provides a summary of the main ideas and findings from the first part of the Hunt (1965) study that are relevant to the present study.

The "minimal terminal unit" or T-unit was introduced and defined. This unit is described as "a grammatically discrete unit intervening in size between the clause and what is punctuated as a sentence" (Hunt, 1965: 49). The unit is formally defined as "one main clause, plus any subordinate clauses or non-clausal structure that is attached to or embedded in it" (Hunt, 1970: 4).

Three mean lengths were calculated for each grade, namely the mean length of clauses, of T-units and of punctuated sentences. A clause is defined as a structure containing a subject and a finite verb phrase. The mean length of clause for each student can be calculated by dividing the total number of words by the total number of clauses. In this case, a punctuated sentence is defined as words written between a capital letter and a full-stop or other terminal punctuation.

Besides the three mean lengths, two ratios were calculated for each student and each grade. The two ratios are mean number of clauses per T-unit and mean number of T-units per punctuated sentence.

The three lengths and the two ratios for each student were subjected to a statistical analysis of variance, and tested by means of the contingency coefficient technique, in order to determine which was the better indicator of a student's grade level. According to the results on the contingency coefficient test, the T-unit length is the best indicator of grade level, followed by clause length, then ratio of clauses to T-units followed by punctuated sentence length as the least adequate index.

5.8 Conclusion

The above synopsis indicates a number of things. For one, the main clause length increases more in earlier stages of language development than it does in later stages. However, the proportion of subordinate clauses increases less in earlier stages than in later stages. These two gains produce the gain in average T-unit length, which is therefore smoothed out to be equal for the two periods. Much of the gain in T-unit length is counteracted before it is carried forward to increase mean sentence length. This counteraction is a result of the decline in average number of T-units per sentence (Hunt, 1965: 50).

Both the ratio of clauses to T-units and the ratio of T-units to sentences were examined in greater detail. Hunt (1965) examined the ratio of clauses to T-units by determining how many clauses each grade used in a one-clause T-unit, a two-clause T-unit, a three-clause T-unit, etc. The ratio of T-units to sentences was examined by noting how many T-units each grade used in one-T-unit sentences, two-T-unit sentences, etc.

With regard to the average length of T-units, it was found that all three grades produced the same number of mid-length T-units, while fourth graders produced more short T-units and twelfth graders more long T-units. The number of short T-units was found to be a significant index of maturity with a chi-square of 52.87 and contingency coefficient of 0.70. Expressed in percentages, each successively older grade used half the percentage short T-units of the previous grade. This extended analysis of T-unit length adds significant information to the increase in mean length from grade to grade:

> ...the increase is due to a major reduction in the number of very short T-units, which are usually single, independent clauses, and a corresponding increasing long units, with only a relatively minor change in middle length units (Hunt, 1965: 52).

In other words, the combination of fewer short T-units and more longer T-units results in an overall increase in the average length of compositions produced from grade to grade.

Hunt's (1965) synopsis of clause-to-sentence length factors provided the first stages of a systematic, coherent and broad analysis, which is still capable of sufficient refinement to incorporate details. The T-unit should be thought of as consisting of a single short independent clause, capable of being expanded at several points, and is, in fact, expanded and lengthened by every writer. Older writers simply expand the unit more often and to a greater extent.

The single short independent clause must contain certain elements that are essential, such as: a subject nominal, a finite verb or verb phrase and certain objects, complements etc., depending on the verb. That is all that is necessary for a clause to be classified a grammatical sentence in English. Other optional elements may be added, such as modifiers. Older students write longer T-units. This must surely mean that it is characteristic for older students to add more of these optional elements.

In conclusion, Hunt determined that the T-unit was the best measure of syntactic development available. The syntactic level of development a learner has reached can be determined by calculating the average number of words written per T-unit produced in a composition (Mills, 1990: 107).

The following chapter reviews a number of studies in which accuracy measure have been used to measure language development.

Chapter 6

A Review of Accuracy Measure Studies

6.1 Introduction

In Chapter 4, Larsen-Freeman's (1978) call for a more precise, numerically based, global and independent index of development was mentioned. Such an index could be used to measure a learner's or subject population's language development. In the previous chapter, an overview was given of the development of the T-unit as measure of development.

This chapter provides a brief review of a number of studies in which language development is measured by means of accuracy measures. Most of the studies used measures based on the T-unit, with some based on clauses and sentences. These accuracy measures include frequency measures (e.g. Homburg, 1984), ratio measures (e.g. Larsen-Freeman & Strom, 1977; Larsen-Freeman, 1978; Bardovi-Harlig & Bofman, 1989; Mills, 1990) and indices (e.g. Gipps & Ewen, 1974; Evola, Mamer & Lentz, 1980; Arnaud, 1992; Engber, 1995 and Kaczmarek, 1980).

The employment of accuracy measures based on the T-unit in the present study is discussed, as well as the issue of defining "an error" and "an error-free T-unit".

6.2 Developmental Measures

Developmental measures can be distinguished according to the way they are calculated. Wolfe-Quintero et al. (1998) distinguish three calculations. The first is a frequency count (e.g. Homburg, 1984 analysis 1), which is a simple type of calculation. It merely involves counting the number of units the composition is made up of, such as sentences, clauses or T-units. They are problematic, because they are influenced by the time provided for the students to finish the compositions.

The second calculation, ratio measures (e.g. Larsen-Freeman & Strom, 1977; Larsen-Freeman, 1978; Bardovi-Harlig & Bofman, 1989; Mills, 1990 and Homburg 1984),

entails the presence of one unit being expressed as a percentage of another type of unit. In other words, one type of unit is divided by the total number of comparable units. Ratio measures are useful for measuring the length of a unit (e.g. words per clause), the rate of accuracy within a given unit (e.g. error-free units per sentence or T-unit), or the complexity of a given unit (e.g. clauses per T-unit). The ratio can be multiplied by 100 in order to give a percentage. Wolfe-Quintero et al. (1998: 10) consider this to be a more accurate measure than frequency counts.

A third type of calculation can be done by means of an index based on a formula that produces a numerical score (e.g. Evola, Mamer & Lentz, 1980; Arnaud, 1992; Engber, 1995; Kaczmarek, 1980). This type of index is usually developed when the features that are being examined belong to a hierarchy of complexity or sophistication (Wolfe-Quintero et al., 1998: 10).

6.3 Frequency Measures

Wolfe-Quintero et al. (1998) compared a number of studies in which different accuracy frequency measures (e.g. error-free T-units, error-free clauses, errors and first-, second- and third-degree errors) were employed. They conclude that the error-free T-unit is the best measure of language development (Wolfe-Quintero, 1998: 38). The number of error-free T-units appears to discriminate between proficiency levels.

The validity of accuracy frequency measures can be doubtful, because there is no one fixed denominator as found in ratio measures. Error-free T-units, as well as other frequency accuracy measures, work best when used to analyse compositions produced within a set time limit (Wolfe-Quintero et al., 1998: 38-39).

6.3.1 Homburg (1984)

Homburg (1984) investigated whether a reliable and valid objective grading scheme exists and how this grading scheme might relate to certain objective measures applicable to ESL compositions. The Testing and Certification Division of the English Language Institute at The University of Michigan developed the holistic marking scheme used in Homburg's study. The graders or raters holistically grade compositions produced by foreign applicants to American universities by assigning

ordinal scores to the compositions on a scale from one to ten. These graders have been found to score relatively high on both inter-rater and intra-rater reliability.

The features and criteria that these graders use to evaluate compositions holistically were investigated. Homburg aimed to identify the criteria which graders consider important. The objective measures used in Homburg (1984), in an attempt to validate this particular holistic grading procedure used at the University of Michigan, represent what the graders regard as the most important features of ESL compositions (Homburg, 1984: 88).

The relationship between the objective measures and the grading procedure followed at the university should indicate the nature of the subjective grading method that the graders follow (i.e. how the graders make decisions about grading). Homburg (1984) attempts to clarify the nature of ESL writing proficiency assessment by applying these objective measures to compositions.

Some researchers (e.g. Arthur 1980; Flahive & Snow 1980; Kaczmarek 1980; Mullen 1980; Nas 1975) support objective measures of writing proficiency such as: measures of composition length (the writer is only allowed a limited time to write), measures of syntactic complexity (e.g. mean length of T-units and subordinate indexes) and measures of errors.

The compositions that were analysed were composed within 30 minutes and consisted of at least 150 words. Homburg (1984) only investigated those compositions awarded scores of 5, 6 and 7 by the graders (holistically graded according to a scale of 1 to 10).

Homburg (1984: 91) employed the following objective measures: measures of length, measures of subordination and relativisation, measures of sentence connectors and measures of number and types of errors. The T-unit was used additionally as measure of length and as measure of standardising other measures.

Measures of Length

Homburg (1984) employed the following measures of length: number of words per composition, T-units per composition, sentences per composition, number of T-units per sentence, words per sentence and the number of words per T-unit. Results indicated apparent differences between the three scores for the number of words per composition, per sentence and per T-unit. Generally, these measures clearly distinguished compositions rated 5 from those rated 6 or 7 (on the holistic rating scale from 1 to 10).

Subordination and Relativisation Measures

Two measures dealing with the ability to join independent and dependent clauses were employed. They were: 1) the number of dependent clauses per composition and 2) the ratio of the number of dependent clauses per independent clause. Results show a general increase across the three score levels (5, 6, 7) for these measures. These results indicate that subordination and relativisation increase as holistic scores increase.

Connector Measures

The connector measures employed in the study are not relevant to the present study and are not discussed in further detail.

Error Measures

Homburg regards it necessary to employ errors as an index of ESL writing proficiency. The problem, however, is in determining what constitutes an error. Homburg (1984: 94) adopts the view that readability and comprehensibility are both important in determining characteristics and definitions of errors.

Errors are classified, according to Nas's (1975) classification of errors, into three degrees, based on their effect on comprehensibility. First-degree-errors do not cause much confusion, while third-degree-errors may make a composition very hard to understand. The number of error-free T-units was used along with the grading of errors into the three degrees. Results show that the number of errors per T-unit decreases as the score becomes higher, i.e. compositions rated 7 contain fewer errors per T-unit than those rated 5 or 6.

Three statistical analyses were run on the data. These were an Analysis of Variance (ANOVA), Orthogonal Comparisons (independent t-tests) and a Discriminant Analysis. Although an ANOVA would indicate any significant trends across the three scoring levels for a measure, Orthogonal Comparisons provide more information about the effectiveness of the various measures. However, the ANOVA and the Orthogonal Comparison do not provide information about which measure best discriminates between the three scoring levels, as Discriminant Analysis does.

Analysis of Variance

Results from the ANOVA across all three scoring levels were found to be at or below the 0.05 significance level for the following measures: number of words per composition, words per sentence, words per T-unit, number of dependent clauses per composition, number of dependent clauses per independent clause, the number of second-degree errors per composition, number of error-free T-units per composition and the number of second-degree errors per T-unit.

Orthogonal Comparison

The orthogonal comparison results indicate how these measures separate the three composition scores. The number of words per T-unit seems to separate scores 5 from 6 and 7, but do not separate score 6 compositions from score 7 compositions. The measure of error-free T-units appears to discriminate among all three composition score levels.

Discriminant Analysis

According to the results of the Discriminant Analysis, five measures differentiated among all three composition score levels: First, second-degree errors per T-unit; second, dependent clauses per composition; third, words per sentence; fourth, co-ordinating conjunctions; and fifth, error-free T-units per composition. Homburg (1984) concludes that these results seem to support the notion that these measures capture what trained ESL compositions graders look for when holistically evaluating ESL compositions written levels 5, 6 and 7 of proficiency (Homburg, 1984: 101). In other words, a composition should score the same, regardless of whether it was rated by the graders using the holistic procedure or others using the objective measures

mentioned. The same features are considered important by graders using either procedure.

Specifically regarding the T-unit (as relevant to the present study), Homburg (1984: 90) acknowledges Ney's (1966) criticism on the T-unit as measure of writing development and communicative competence, but agrees with Gaies (1980) that the T-unit has proven to be a good measure of syntactic complexity. Homburg (1984) concludes that T-unit measures can be used as an aid to determine some ESL writing features. However, it is not the only measure found to discriminate clearly between scoring levels. Therefore, Homburg (1984: 103) proposes that a combination of features should determine the final assessment.

6.4 Ratio Measures

A wide range of accuracy ratio measures have been employed in various studies. These include: error-free T-unit ratio, error-free T-units per sentence, error-free T-units per word, error-free sentence ratio, error-free clauses per sentence, error-free clauses per T-unit, errors per T-unit, errors per clause, etc.

In general, Wolfe-Quintero et al. (1998) found that results for the use of error-free T-units ratio were positive.

6.4.1 Larsen-Freeman and Strom (1977)

Larsen-Freeman and Strom (1977) examined written compositions of 48 non-native English speakers. The students were at undergraduate and graduate level at UCLA. The only criterion was that no more than two compositions written by speakers of the same native language would be accepted. The compositions were written under examination conditions with a minimum length assigned, but without time restrictions.

Two researchers independently and impressionistically classified each composition into one of five categories, namely poor, fair, average, good or excellent. (Inter-researcher reliability scored significantly at $r = 0.97$.) After this classification was completed, the researchers attempted to identify certain characteristics which made

each group unique. Features such as punctuation, capitalisation, spelling, organisation, clarity and syntactic sophistication etc. were considered. Each composition was "scrutinised" for the author's ability to write grammatically, focussing on morphology, syntax, prepositions, tenses, aspect, articles, subject/verb agreement, case and negation.

Certain problems were identified when an index was based solely on these features and on the errors responsible for making a composition more or less native-like. The first problem entailed the way in which errors and positive features would be weighted. Secondly, if an index was based on the decreasing number of errors committed with regards to any linguistic structures, one would have to assume that language acquisition develops linearly. This, however, is not a valid assumption (Larsen-Freeman & Strom, 1977; Henning, 1976; Hakuta, 1975; Neuman, 1977). Furthermore, by basing an index on the absence of errors in certain structures, certain language groups may be discriminated against. Some language groups may experience more L1 interference with a specific structure than other language groups. In order to avoid these difficulties, the researchers turned to a more objective feature, namely measures of length.

The number of words and sentences in each composition was counted, as well as the total number of T-units and the average word length per T-unit.

> The length of T-units were found to be important because Hunt found that as children learning English as their first language mature, they tend in their writing to reduce main clauses to structures such as words and phrases which are incorporated within other main clauses. This tendency, Hunt reports, increases almost in a straight line (Larsen-Freeman & Strom, 1977: 128).

The number of error-free T-units was also determined. A T-unit was counted as error-free only if it were perfect in all respects, including spelling and punctuation.

The analysis of errors and positive features did not reveal any significant or out of the ordinary trends. The results of the quantitative analysis produced more interesting information. First and most obvious, even without time limitations, the older students produced more words per composition than younger ones. Results produced by

Neuman (1977) support this finding. An analysis of variance revealed that the difference between the five groups was significant at the 0.001 level. Although a good deal of variance was noted, an unmistakable trend towards a greater number of error-free T-units with increasing proficiency levels was identified.

Larsen-Freeman and Strom (1977) conclude that they are not optimistic about the possibility of error analysis being the answer to their search for an index of development, but suggest making use of measures of length and error-free T-units:

> ...the measures of length and error-free T-units seem to be much more viable contenders on which to base an index of development since they are easily quantified, linear progression does seem possible and they would seem to be impervious to differences in language backgrounds (Larsen-Freeman & Strom, 1977: 132).

6.4.2 Larsen-Freeman (1978)

Larsen-Freeman (1978) reports on the progress made in the attempt to establish an index of development which would serve as a yardstick researchers could use to "expediently and reliably gauge a learner's proficiency in a second language" (Larsen-Freeman, 1978: 439).

She reports the success of numerous first language acquisition researchers (O'Donnell, Griffin & Norris 1967; Mellon 1969; O'Hare 1973) in using T-unit length to measure written maturity, with specific mention made of Loban's (1976) study. This study may best reflect the reason for researchers' enthusiasm in using T-unit length as an index of development in first language acquisition. Loban initially selected 338 children at kindergarten level and studied their progress up to grade twelve or the age of eighteen. The students were divided into three groups: one consisting of students with high language ability, one with students with low language ability and another with random students to represent the total group. Though some minor backsliding was recorded, the length of the T-unit increased for all three groups into which the students were categorised.

Larsen-Freeman (1978) reviews the success of a number of researchers of second language acquisition development who also employed the T-unit length as a measure

of development. These include Thornhill (1969), Crymes (1971), Pike (1973), and Gaies (1976; 1977) for ESL; Cooper (1976) and Akin (1976) for German; and Monroe (1975) for French. Furthermore, Cooper (1976), Monroe (1975) and Gaies (1976; 1977) found T-unit length to discriminate between second language learners at different levels of proficiency. Thornhill (1969) also found that the mean length of T-units was a useful measure of development towards second language maturity. Others, such as Pike, did not find significant results. (The studies by Crymes and Akin did not relate directly to the focus of Larsen-Freeman's (1978) study.)

Farhady (1978) examined written compositions of 28 Persian students. A significant correlation was found between the length of their T-units and five independent measures of their writing ability – namely organisation, logical development, grammar, writing mechanics and style – that were based on the subjective evaluation of three judges of the students' compositions.

Larsen-Freeman (1978) examined the compositions of 212 ESL speakers written as part of the UCLA placement examination. The minimum number of words required was 200 and the time was restricted to 30 minutes. Both a quantitative and a qualitative analysis were performed. The compositions were divided into five groups according to the score achieved on the entire placement examination. Group 1 contained those who scored lowest, while group 5 contained those who scored highest.

First, the length of each composition was determined, the argument being that subjects with a higher proficiency tend to write longer compositions. An analysis of variance test resulted in an F ratio of 14.792, significant at the 0.001 level of probability.

The average number of words per T-unit was also determined by dividing the total number of T-units per composition into the total number of words written in that composition. Again, the results showed statistically significant differences between the five groups, with $F = 6.10$.

The percentage of error-free T-units was calculated and a definite linear trend towards a higher percentage of error-free T-units with an increase in learner proficiency was identified. An ANOVA showed the results to be statistically significant with an F = 31.29.

Since increasing length alone may not be a sufficient measure, the total number of words written per error-free T-unit was also determined. (After T-unit boundaries have been agreed upon, this measure is quick and easy to use, and no testing instrument is needed.) Results of the ANOVA indicated highly significant differences between the five groups with F = 18.432.

Scheffe's method of post-hoc comparisons was used to determine whether the difference between each of the group means was significant. The average number of words per Error-Free T-unit (N) for groups following one another was not found to be significantly different, but the number for non-adjacent groups was, as shown in Table 6.1.

Group	N	Mean	Standard Deviation
1	37	4.61	4.12
2	39	7.52	4.51
3	45	9.26	6.30
4	56	10.77	3.89
5	35	13.20	3.00

Table 6.1 Average number of words per error-free T-unit (Larsen-Freeman, 1978)

Larsen-Freeman (1978: 446) concludes that she has "been careful to note that at least some of these measures work quite well for discriminating proficiency differences among groups". The size of the standard deviations shows that the groups are not entirely homogeneous with regards to proficiency. Groups overlap, meaning that one subject may fit into more than one group. She therefore feels that it is still necessary to refine the procedure to find an index that works well for both groups and individuals.

6.4.3 Bardovi-Harlig and Bofman (1989)

This study examines the relationship between syntactic development (or complexity) and overall accuracy of the writing of adult foreign language learners. Two types of studies have typically been identified in research on grammatical development in second language acquisition.

The first type of study is concerned with examining formal features. This entails isolating particular grammatical morphemes or syntactic structures for detailed examination. The second type describes developmental progress independent of any grammatical aspect of the interlanguage. The Bardovi-Harlig and Bofman (1989) study, as well as the present study, falls into the second category (mainly), as does the work of Larsen-Freeman (1978, 1983) and Larsen-Freeman and Strom (1977). The measure of average length of error-free T-units is the one most widely used. "By taking into account both length and accuracy of the language sample, this index measures movement toward target-like production" (Bardovi-Harlig & Bofman, 1989: 18).

Bardovi-Harlig and Bofman's (1989) study differs somewhat from the others mentioned in that it combines aspects of studies from both types. It seeks to describe three aspects, namely morphology, syntax and errors, or "overall accuracy with regard to the distribution of errors" (Bardovi-Harlig & Bofman, 1989: 18). Furthermore, the study only focuses on learners at one level of proficiency – the advanced adult learner. "Advanced adult learner" is defined for the Bardovi-Harlig and Bofman study as one having a TOEFL score of approximately 550.

A sample of 30 essays was collected from applicants to Indiana University who met the above-mentioned criterion. The essays were written between 1983 and 1987 as part of a placement examination. Students were allowed 45 minutes in which to finish a composition on one of three general topics. The learners who wrote the compositions were from five different native language backgrounds: Arabic, Chinese, Korean, Malay and Spanish. Six learners represented each language group.

Written rather than oral samples were collected, since written samples of advanced learners can be evaluated in more detail. Also, written language produced to communicate and oral language have been found agreeable to the same analyses and can be compared (Bardovi-Harlig & Bofman, 1989: 19).

Two parameters were used to score essays, namely syntactic complexity and surface errors. Syntactic complexity was calculated as the number of clauses per T-unit. Clauses were identified by verb phrases. Errors were scored as morphological, syntactic or lexical, while errors in spelling were not counted.

Bardovi-Harlig and Bofman (1989: 22) reported errors and complexity scores in separate tables, rather than combining them into one measure, such as the average number of words per error-free T-unit. This would facilitate the examination of syntactic complexity and the number, type and distribution of errors. Error-free T-units focuses on measuring error-free production, whereas the Bardovi-Harlig and Bofman study investigates the errors themselves as indicators of areas of incomplete development in the interlanguages of advanced learners.

Syntactic Complexity

Results show that of all the T-units produced by the learners as a group, 57% contained two or more clauses, with a mean complexity of 2.73 clauses per T-unit. Across the five language groups, syntactically complex T-units were produced with the same frequency.

Errors

The analysis of errors showed that the greatest number of errors was produced in grammatical morphemes, with fewer errors being produced in lexical choice and the smallest number produced in syntax. Larsen-Freeman (1983) also reports that morphological errors are the most common in oral and written tasks across a variety of competence levels. In other words, advanced learners seem to perform relatively well in syntax, but perform weaker in their production of grammatical morphemes.

According to Dulay et al., 1982: 191), errors can be divided into 'local errors' and 'global errors'. Local errors essentially include errors in grammatical morphemes that

affect a single constituent, whereas global errors are identified as syntactic errors, errors in word order, sentence combining and omission of major constituents. According to Burt and Kiparsky (1972), learners need to control global grammar in order to communicate successfully, but they can do so without controlling local grammar. The "strong syntax – weak morphology stage" identified by Bardovi-Harlig and Bofman (1989), is consistent with the communicative approach to language learning. In other words, learners may be more likely to concentrate on being understood in a foreign language before attending to more technical issues that may be less important communicatively.

Bardovi-Harlig and Bofman (1989) concluded that learners from all five language groups showed similarity in terms of syntactic complexity and error distribution. The areas of syntax and morphology in the grammatical systems of adult language learners seem to develop unevenly. Bardovi-Harlig and Bofman (1989: 29) note that it is not only advanced adult foreign language learners who perform strong in syntax and weak in morphology – a variety of learners share this characteristic.

6.4.4 Mills (1990)

Mills (1990) collected a sample of 177 essays from the Exit Examination of the Freshman English programme at the University of Cincinnati. The essays were analysed from a number of perspectives in order to test various models of composition research. Most of the study was devoted to measuring syntactic complexity in various examination essays, because "the dominant paradigm in composition research is, following Hunt (1965, 1970), based on measurement of syntactic complexity" (Mills, 1990: 109). The essays were produced under tightly controlled circumstances. All students had the same essential training, the same preparation and the same specifications for the writing task.

Mean sentence, T-unit and clause lengths were calculated. Essay lengths varied from 12 sentences to 72, while the mean sentence length was calculated at 18.82, with a Standard Deviation of 3.316. Although the correlation between grade and mean length of sentence of 0.227 was significant, Mills (1990: 110) regards the correlation

coefficient as hardly spectacular and reminds the reader that 'statistically significant' is not always the same as 'explanatorily significant'.

The MLTU (mean length of T-unit) was found to range from 11.327 to 26.667, with the average at 16,049 (Mills, 1990: 110). This average is higher than the 14.4 average found for twelfth graders by Hunt (1965). According to Mills (1990: 110), the mean length of T-unit turns out to be the "best single predictor of grade on these essays", although the correlation coefficient is, again, not spectacular at 0.278 ($R^2 = 0.077$).

The average mean length of clause was 9.923 words per clause. The correlation between mean length of clause and grade was 0.23999, which is better than that for mean length of sentence, but lower than mean length of T-units.

After completing a Multiple Regression analysis, Mills (1990) found the best predictor of grade between the three measures to be mean length of T-unit, with mean length of clause a distant second. Mean length of sentence did not even reach an F-level or tolerance level for it to be included in the regression.

As far as these three measures are concerned, Mills (1990: 110) agrees with "most researchers" who "have argued that mean length of T-unit ... is the best indicator of syntactic complexity". Some argue that the T-unit is the best objective measure of writing quality. Mills further entered the variable of 'position' – that is, the position writers took on the issues they were instructed to write about. However, "... the pragmatics of language use, rules for getting things done in language, outweigh all but one syntactic variable as predictors of writing effectively" (Mills, 1990: 115). Even with the extra variable entered into the equation, mean length of T-unit turned out to be the best predictor of syntactic complexity.

6.5 Accuracy Indices

Indices are based on a formula. For each index, an individual formula is used. Formulas can be based on different units, such as T-units, clauses, words and sentences. The studies discussed in this section all employed different indices, based on different units.

6.5.1 Gipps and Ewen (1974)

Gipps and Ewen (1974) analysed the writing of child second language learners. They applied a scale of intelligibility (IntellIndex = sum of points for the intelligibility of each T-unit) to T-units, with unintelligible T-units rated 0, partly intelligible T-units rated 1 and completely intelligible T-units rated 2. Completely accurate T-units were rated 3. Results showed a high interrater reliability on the classification of T-units. Furthermore, the intelligibility between three groups of children – based on how long the learners had been using the L2 – was significantly different.

6.5.2 Evola, Mamer and Lentz (1980)

Evola et al. (1980) used written compositions produced by learners at various proficiency levels. They created an error formula (EF1) to analyse the correctness of particular grammatical morphemes, such as conjunctions and pronouns. The formula calculates the number of correct usage, minus the number of errors, divided by the number of words produced (EF-E/W). No significant relation was found between holistic ratings and the error formula. They attribute the results to the fact that the formula is a finely tuned measure, as opposed to holistic scales.

6.5.3 Arnaud (1992)

Arnaud (1992) used a lexical quality formula. Other lexical ratio measures showed insignificant correlations with discrete-item tests. The lengths of the compositions were limited. Arnaud only used the shortest essays from the samples of two timed writing sessions (208 and 241 words). A formula was used to determine the lexical quality. The formula combines lexical variation and lexical accuracy. It is calculated by adding the total number of lexical word types and the total number of rare word types (words that do not occur in the syllabus within the first fours years of secondary education in France), then subtracting two times the number of erroneous lexical words on same length samples (LWT+RWT-2LWE). Arnaud (1992) found a significant correlation between the lexical quality formula and scores on a discrete item vocabulary test. It is, however, not clear how this measure will relate to more global measures of proficiency level.

6.5.4 Engber (1995)

Engber developed an index that combines lexical variation and lexical accuracy. The sample used for the analysis was made up of the compositions of Farsi and Arabic writers distributed across five programme levels. The formula captures the relationship between lexical errors and lexical word types. The lexical accuracy index calculates the number of lexical word types, minus the number of lexical errors, divided by the number of lexical words (LWT-LE/LW). This formula is based on the idea that writers with a higher level of proficiency will produce more lexical variation with fewer lexical errors, which results in a higher score. Results showed a significant relation to holistic ratings.

6.5.5 Kaczmarek (1980)

Kaczmarek (1980) was interested in whether raters are influenced by the number of errors in an ESL composition when rating a composition. She collected a sample of essays from university level students enrolled in an ESL programme. Kaczmarek created an error index, which she compared to two holistic rating scales (ability and comprehensibility) that were applied to the compositions. Calculations included determining the number of error-free words and the number of errors (including unnecessary and missing words and non-idiomatic sequences). The Error Index is defined as the number of error- free words, minus the number of errors, divided by the number of error-free words (EFW-{E/EF}). Results showed a correlation between the error index and the holistic ratings. This indicates that raters do pay attention to errors in the writing of second language learners.

6.6 The T-unit as Accuracy Measure of Language Development

This study employs the T-unit as a measure of language development, because it is not dependent on any grammatical structure of the language. Therefore, it does not discriminate against any native language group. A number of studies (cf. chapter 5 and above) have found that measures based on the T-unit, such as error-free T-units, distinguishes effectively between levels of development. Furthermore, measures based on units such as sentences are not as objective as those based on the T-unit, since consensus has not been reached about the definition of "a sentence".

It is generally assumed that as learners become more proficient, they become more likely to produce fewer errors. As Gaies (1976), Larsen-Freeman (1978) and Wolfe-Quintero et al. (1998) point out, structural errors in second language writing indicate lack of syntactic control. This lack of knowledge can be noticeable in simplification (regarding complexity) or errors (regarding accuracy).

It is sometimes assumed that as a learner becomes more proficient, he/she will necessarily produce fewer errors. However, it may be that a more proficient learner trying to write a more complex sentence will produce more errors than a learner at a lower proficiency level (Foster & Skehan, 1996). Furthermore, fluency does not necessarily increase with accuracy and complexity. If a learner is trying to write a more complex sentence, using language that may not be fully automatised, more errors may occur (MacKay, 1982).

A developmental analysis of the accuracy of second language writing involves counting the errors produced by the learner in some way. Two approaches to counting errors have been developed. One approach involves determining whether a structural unit, e.g. sentences, clauses or T-units, is error-free. For this approach, measures such as the number of error-free T-units per T-unit (EFT/T), or the number of error-free clauses per clause (EFC/C) are typically used. In this case, what is regarded as 'an error' needs to be defined clearly (Wolfe-Quintero et al., 1998: 35).

For Larsen-Freeman (1978; 1983) and Larsen-Freeman and Strom (1977), a T-unit had to be perfect before counted as error-free. They included all errors, from morphosyntactic and syntax errors to spelling and punctuation errors, in their definition of an error. In the Homburg (1984) study, all of these were counted, except punctuation errors. Ishikawa (1995) counted morphosyntactic, lexical and discourse errors, whereas Scott and Tucker (1974) only counted morphosyntactic errors. This difference in the definition of an 'error-free' unit influences the results of the studies and makes it difficult to make meaningful comparisons.

Wolfe-Quintero et al. (1998: 36), however, argue that error-free measures alone may not be sufficient enough to determine how accurate a composition is or is not. Gaies

(1980) argues for the development of a hierarchy of error types so that the number of errors and the error types that occur are considered.

Researchers (Bardovi-Harlig & Bofman 1989; Homburg 1984) developed a second approach to a developmental analysis of the accuracy of second language writing. This approach entails determining the number of errors that occur in relation to production units (such as words, clauses and T-units). The approach is only concerned with identifying quantities of errors and not strings of error-free words. All errors can be considered. For example, the number of errors per word (E/W) can be calculated, or various types of errors can be considered, for example, the number of morphological errors per clause (MorE/C) can be calculated.

It is possible to classify errors as morphological, lexical, semantic, etc., or on some other basis, as Homburg (1984) does. He divided errors into first-, second- and third degree-errors. These classifications supply information about possible types of errors that may occur (Wolfe-Quintero et al., 1998: 36). Classifying errors may make more precise identification of errors and developmental patterns possible. In the present study, errors are classified into a number of categories (cf. 7.4 & 8.10).

Bardovi–Harlig and Bofman (1989) used the T-unit as measure of development, since it takes both the length and accuracy of a language sample into account. It measures development towards target-like production. Therefore, measures based on the T-unit, such as the number of error-free T-units per composition, can be regarded as effective for measuring development.

6.7 Conclusion

This chapter provided an overview of a number of studies that have employed different accuracy measures of development. Developmental measures can be classified into three categories on accounts of the way that they are calculated, namely frequency, ratio measures and indices. The studies that were discussed included studies that used accuracy measures based on sentences, clauses and T-units. Results from the studies show that T-unit measures, especially error-free T-units, discriminate well between levels of development.

Accuracy measures based on the T-unit are discussed as objective and the most favourable for the purpose of the present study.

The problem defining a T-unit as 'error-free' was discussed. Some studies (e.g. Larsen-Freeman and Strom, 1977; Larsen-Freeman, 1978) only counted a unit as error-free if it was one hundred percent correct. The other studies allowed for some errors, such as spelling and punctuation. It was determined that it is necessary to establish categories for classifying error types. Also, the number of errors per T-unit should be taken into consideration.

The next chapter discusses the method of research employed in the present study in order to establish an index of syntactic development and determine the level of syntactic development of Grade 12 ESL learners. The index will be based on the T-unit.

Chapter 7

Method of Research

7.1 Introduction

This chapter discusses the method of research that was followed to collect and analyse the necessary data for this study. The first section describes how the sample was collected and processed. The second section describes the statistical procedures conducted. The method of research is discussed under the following headings:

- Design
- Compositions
- Processing of compositions
- Statistical Procedures

7.2 Design

The design involves a cross-sectional description and a quantitative analysis (cf. Larsen-Freeman & Long 1992). Dreyer and Van der Walt (1994:121) describe quantitative analyses as "a powerful research tool which allows researchers to go beyond the identification and linear description of language learning phenomena".

7.3 Compositions

216 Creative writing compositions written in 2003 by Grade 12 ESL learners were collected from 6 provinces. The essays had been written under examination conditions towards the end of the year. The essays were collected from different provinces in order to make the sample a more representative cross-section of the Grade 12 ESL learner in South Africa.

The only criterion for choosing the compositions was that they had to have been written by Grade 12 English Second Language (ESL) speakers. The length of the composition required was 250 – 300 words (similar to the length required in the

matriculation examination) and the learners were allowed one hour to complete the compositions.

The learners came from different language backgrounds. The first languages of the learners were isiZulu, Setswana, isiXhosa and Afrikaans. The learners therefore did not share the same native language, but it must be pointed out that native language does not play any significant role in the analysis of T-units (cf. Larsen-Freeman, 1978). It is therefore a non-discriminatory measure as far as first languages are concerned. Any index of development can thus be applied to the group of learners as a whole.

7.4 Processing of Compositions

Two persons analysed the data, namely the researcher and an experienced English teacher. Each person analysed the Higher and Standard Grade papers from three provinces. Two people were used for the analysis to speed up the process and to increase the validity and reliability of the results. The two people also exchanged compositions that had been analysed in order to ensure the validity of the analyses.

The two persons recorded the following data on each composition: a composition code, category, length, average number of T-units, a number for each T-unit in the essay, the number of error-free T-units, the average number of words per error-free T-unit, and the type of errors encountered in a T-unit. This information was recorded on data sheets.

The first step was to code all the compositions. The code was used for purposes of identification and categorisation. From it, the researcher could then identify the province from which the essay came (although the provinces are not identified in the results), whether it was Higher Grade or Standard Grade, and the number of the composition in the sample.

The second step was to divide all the essays into one of five groups. All the compositions had been assessed by experienced, and therefore taken as reliable, matriculation examiners. The markers allocated marks according to the English

Additional Language Higher and Standard Grade Grids used in the matriculation examination (cf. Appendix A). The grid has two axes: The horizontal one specifies language use, while the vertical one specifies the content levels. As the markers were all experienced teachers and assessors, the mark allocated was considered to be a reliable one. The compositions were marked out of 40.

The compositions were grouped according to the mark allocated by the markers. Five groups or levels of proficiency were identified: Poor, fair, average, good and excellent (cf. Larsen-Freeman, 1978: 440).

The compositions with a mark between 0% and 19% were placed into Group 1 and labelled 'poor'; subjects with a mark between 20% and 39% were placed in Group 2 and labelled 'fair'; those between 40% and 59 were put in Group 3 ('average'); 60% to 79% were put into Group 4 and labelled 'good', and those that scored between 80% and 100% were put into Group 5 with the label 'excellent'.

Figure 7.1 illustrates the spread of the groups in the sample among the five groups.

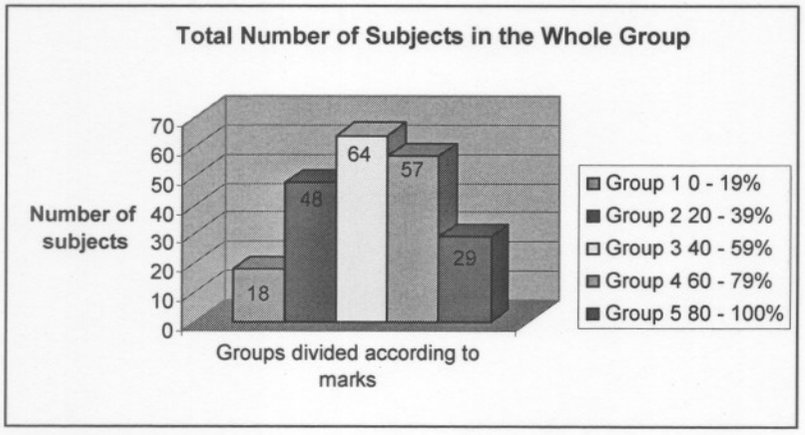

Figure 7.1 Total number of subjects per group.

Figures 7.2 and 7.3 illustrate the number of compositions in Higher and Standard Grade respectively.

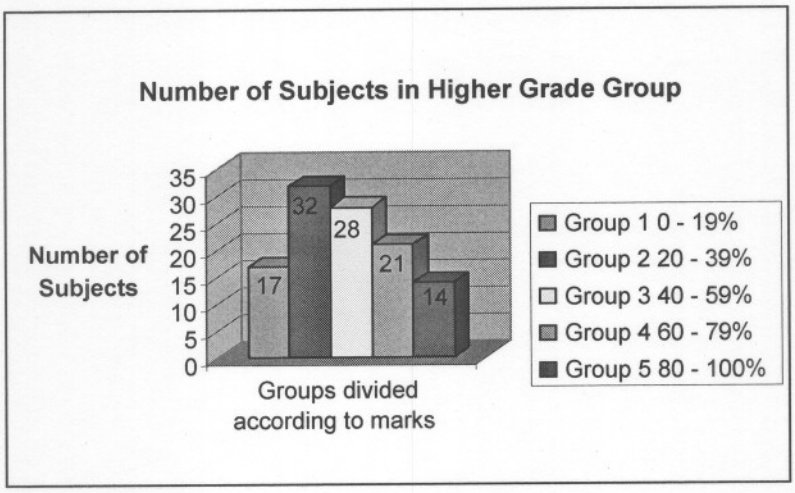

Figure 7.2 Higher Grade groups.

In the Higher Grade, the most learners were rated 'fair', scoring between 20% and 39%.

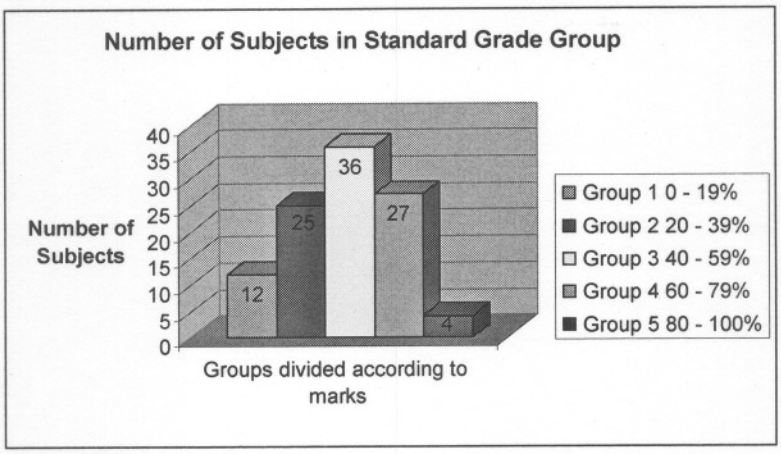

Figure 7.3 Standard Grade groups.

Most learners in the Standard Grade scored in the average group, with marks between 40% and 59%.

The third step in the analysis procedure was to determine the length of each essay. This was done electronically by means of the Microsoft Word 'word count' function.

The fourth step was to divide all the essays into T-units (cf. 5.5). Before starting the procedure, the two analysts discussed the criteria for dividing sentences into smaller units. The main criteria for deciding how a sentence could be 'sliced up' were:

Each unit should be able to function as a complete grammatically correct sentence on its own if punctuated like a sentence (started with a capital letter and ended with a full stop). If the sentence becomes fragmented, the sentence is not 'sliced up'. Conjunctions such as 'and', 'but' and 'because' between two main clauses are always counted with the following clause.

For example, the following sentence taken from the sample is divided into two T-units. The end of each T-unit is indicated with a forward slash:

- I told him everything of course / because I could see that I did not have to hide anything from them/.

Each T-unit can function as a grammatically complete sentence on its own. It is possible to say "I told him everything, of course." It is also correct to say "Because I could see that I did not have to hide anything from them". This could be the answer to a question such as "Why did you tell him everything?" Therefore, the first criterion is met. (If uncertain, it sometimes helps to 'leave out' the conjunction. If the remainder of the sentence makes sense as a sentence the T-unit is correct.) Both the T-units make complete sense even when separated. There is no need to add any words to make either one of the sentences complete. The second criterion is therefore also met. The sentence is cut before the conjunction 'because'. The conjunction is not left out, but counted as the first word of the next T-unit. The units still make sense and no fragments have occurred.

At first, problems were sometimes encountered when dividing more complex sentences into T-units. For example:

- After that he paid me / and told me that I was a very beautiful girl.

If this sentence were to be divided as illustrated, the second unit would not make sense on its own, either with or without the conjunction 'and'. For the second unit to make sense one would have to add a word or words to the unit. If the sentence is

divided into T-units, the second unit becomes a fragment. The conjunction does not make the second unit more sensible. Therefore, the whole sentence was counted as one T-unit.

In his explanation of what a T- unit is, Hunt (1965: 21) states that no units like compound or complex-compound sentences should be left. Yet, Hunt (1965: 21) also states that several multi-clause units such as compound sentences do occur. Some units may be more complicated and contain a number of clauses that are dependent on the main clause and cannot be separated. Hunt (1965: 21) provides the following illustration:

- But while they were trying they killed a whale and used the oil for lamps.

If one divides this sentence before the 'and', the second unit will not be a full sentence on its own. Therefore, the whole sentence is regarded as one T-unit. The criterion stating that no fragmentation should occur is considered 'stronger' or 'more important' than the 'slicing up' of compound and compound-complex sentences.

Some essays had to be discarded, as they made no sense, and it was impossible to divide them into T-units. The following extract is an example of what Hunt would probably consider as an "unintelligible garble" (Hunt, 1965: 20). The learners who wrote them appear to be illiterate. The following paragraph is an extract from such a composition:

> And there she was any worst nightmare had come. True this is the beginning of that topic, in a tight-hearted. the ideas men and women have about the need to communication in a relationship of marriage. you are a very successful and well-known business person or sports man, woman in South Africa.
>
> The high school were you matriculated has invited you to be guest speaker at their annual Price giving ceremony they have asked you to bring a message of hope to the learners showing them that through hard work they can became successful no matter what matter what their background.

> Everything is fine now you get up and slowly walk to the window affert you have put down the receive of the telephone you stare unseeingly in front of you and your customers and there she was my worst night mare had come true,

Step five of the analysis entailed determining the number of T-units in each composition. The two analysts counted the T-units in every essay by hand. (The end of each T-unit was indicated with a forward slash.) Each T-unit was numbered, starting at 1 in each composition, and continuing up to however many T-units the composition contained. These numbers were recorded in a separate column on the worksheet. Each column had a number of rows. The number of rows for all the columns depended on the number of T-units in each composition. Each T-unit number was recorded in a new row. All the information about, for example, T-unit number 9 was recorded in the same row and entered under the respective column (cf. Appendix B).

Step six was to determine the average number of words per T-unit. This was calculated by dividing the length (number of words per composition) by the number of T-units. After step six, the two analysers exchanged compositions in order to cross-check each other, in an attempt to ensure that the analysis was reliable. Very little variation between the analysers was found.

Steps seven, eight and nine were conducted simultaneously. Step seven entailed determining how many T-units per composition were error-free, i.e. did not contain any errors. The focus was on grammatical accuracy only. Spelling and punctuation errors were ignored. If these had been considered, the analysers would have found very few correct T-units. Spelling is a definite problem with many words being spelt in short-hand, sms style. For example, 'you' is often written as 'u'. Punctuation also had to be ignored, since a number of learners do not use capital letters and/or full stops to indicate sentences. In some cases, learners used a full stop in the middle of a sentence, instead of a comma.

Each T-unit was considered in context when errors were determined. For example, pronouns had to be correct in the context of the paragraph. Step eight was to count the number of words in each error-free T-unit. This was done manually.

Step nine entailed doing an Error Analysis. The analysers recorded the type and number of errors made in each T-unit. The error analysis procedure was conducted twice, with the second a refinement of the procedure. First, the error analysis was performed following to the 'pre-selected approach' (Norrish, 1983: 81). The error categories were based on the researcher's conception of the types of errors the subjects would produce. The researcher chose the categories and as the analysis progressed, categories were added or changed as necessary. A table with three columns was used for the Error Analysis. The first one was for the category of error, the second for an example of the error and the third to indicate the frequency with which the error occurs. The analysers exchanged the data of two of the provinces they had annotated and analysed, and then conducted a refined analysis. Each analyst received a list of error categories (cf. Appendix C) that the researcher had set up, revised and refined. The errors that the learners made determined the categories for the second analysis (cf. Norrish, 1983).

The revised categories and the different errors in each category were given a number as a code. These codes were then entered onto the data sheet as the errors occurred. Each error was entered into as many categories as necessary. In other words, one error did not mean that only one code would necessarily be entered. Because it was not always easy to know what the learner had meant to write, all possible meanings had to be considered. This process is somewhat similar to the one Norrish (1983: 86) describes as 'the quick check' method.

The researcher computerised all the data collected on the sheets and stored them in a separate file for each composition in Microsoft Excel. Then, two files were compiled for each province – one containing all the Higher Grade compositions and one containing all the Standard Grade ones. In the end all the data were stored in 12 files. These 12 files were analysed by the Department of Statistical Services of North-West University, Potchefstroom campus.

7.5 Statistical Programmes

Statistical programmes were used in the second part of the analysis procedure. The mark was used as the independent variable in all the analyses. The dependent variables were: the average length of the compositions, the average number of T-units per composition, the average number of words written per T-unit, the percentage error-free T-units per composition and the percentage of words produced per T-unit.

First, an Analysis of Variance test (ANOVA) (cf. Seliger & Shohamy, 1989: 234 - 235; Hatch & Farhady, 1982: 129) was done to compare the mean characteristics of the essays. This was followed by a multiple comparison using the method proposed by Tukey (cf. Miller, 1981: 37 – 48) in order to see which of the five groups differed significantly for each characteristic. The GLM procedure of the SAS programme was used for the ANOVA and the Tukey Test. Both these tests were conducted at a 5% level of significance. The tests were repeated for all the data, for Higher and Standard Grades, for the whole group, and for the five groups.

A multiple regression analysis (cf. Hatch & Farhady, 1982: 233 – 235; Seliger & Shohamy, 1989: 222-226) was conducted. This was done in order to see to what extent the data can predict the mark. The multiple regression analysis was done for all the data, for higher grade and standard grade and for each province, but not for the five groups (poor – excellent).

A Stepwise Discriminant analysis (cf. Seliger & Shohamy, 1989: 226) was done in order to find the characteristics that would make up the best index formula for the level of development. Two dependent variables were identified as the best characteristics to be used for discriminating between levels of development. In other words, two variables were identified according to which compositions can be divided into the five groups more accurately.

7.6 Conclusion

This chapter provided a description of the method of research followed in this study. A full account is provided of the steps taken in the processing and analyses of the

compositions. The next chapter reports on the results of the analyses and interprets the findings.

Chapter 8

Results of the Study

8.1 Introduction

The previous chapter described the method of research. The procedure followed for processing the compositions and the statistical analyses were discussed. This chapter reports and interprets the results. Results are reported for the analyses of the characteristics across the five groups. The results for ANOVAs and Tukey's Standardized Test are also reported.

A Stepwise Regression Analysis and a Discriminant Analysis were done and the results are reported and interpreted. Following the results of the Stepwise Regression and the Discriminant Analysis, two indicators of development are suggested for use as an index. This index is used to determine the standard of syntactic development of the Grade 12 ESL learners whose compositions were analysed. This standard of English can be used as an index with which to compare successive years' matric results.

The results of the Error Analysis and its implications are briefly discussed.

8.2 Characteristics of the Compositions

The compositions were analysed according to six characteristics (cf. 7.4, paragraph 2). Table 8.1 provides a general summary of these characteristics. Results are for the compositions for all six the provinces, groups and grades. The total number of compositions (N), the mean, standard deviation, sum, maximum and the minimum are reported for each characteristic.

Variable	N	Mean	Standard Dev	Sum	Maximum	Minimum
Length	216	276.61	87.81	59748	815.0	102.0
Mark	215	19.43	8.9	4178	39	0
Number of T-units	216	28.27	9.54	6106	72.0	8.0
Words per T-unit	216	10.18	2.63	2198	22.14	4.11
% EFT	216	40.91	23.22	8839	91.89	0
% words /EFT	216	35.94	22.78	7763	107.64	0

Table 8.1 Simple statistics: characteristics of compositions.

Table 8.1 shows that the longest composition was 815 words. This was an exception. It was written by a Standard Grade learner. The second longest essay consisted of 572 words. It was also a Standard Grade composition. The third longest essay, which consisted of 510 words, was a Higher Grade one. Other long essays consisted of 490, 465 and 406 words. Candidates who exceeded the required length mostly wrote between 350 and 410 words.

The marks that the teachers allocated varied from 0 to 39 out of 40. No student achieved full marks. The average mark was 19.43 out of 40 or 48.6%.

The composition of 815 words influences the maximum number of T-units in the sample. The maximum of 72 T-units in Table 8.1 is therefore also exceptional. The essay consisting of 572 words had 48 T-units and the essay of 510 words had 39 T-units. The fewest number of T-units in an essay were 8.

The following section reports the results for characteristics over the five groups. The number of compositions, standard deviation and coefficient of variance is reported for each.

8.3 Analysis of the Characteristics for the Five Groups

The data were first analysed for each of the five groups (cf. 7.2). The results for each group are compared to determine whether the groups differ significantly.

8.3.1 Average Number of Words per Composition

First, the average number of words per composition was calculated. It is generally believed that more developed learners write longer compositions (e.g. Hunt, 1965: 7-8; Larsen-Freeman & Strom, 1977: 129). Learners who are more proficient are generally more self-confident, show a better command of syntax and have a wider range of vocabulary, which lead them to write longer compositions. Table 8.2 presents the results for the average length for each group.

Group	N	Mean length	Standard Deviation	Coefficient of Variance
1	29	222.14	88.33	39.76
2	57	279.86	83.01	29.66
3	64	287.25	97.2	33.84
4	48	266.27	56.88	21.36
5	18	343.83	86.05	25.03

Table 8.2 The average number of words per composition as calculated for groups.

The results support the argument that learners who write longer compositions also on average score higher marks. The results show a clear increase in the average number of words per composition for consecutive groups, with the exception of Group 4. The weakest group (Group 1) on average wrote 222.14 words per essay. This average is shorter than was required. The averages of Groups 2 to 4 are within the required length; between 250 and 300 words per composition. Students in Group 5 wrote longer compositions than required, with an average of 343.3 words per composition.

An analysis of variance test showed that $F(4; 211) = 6.44$ with $p < 0.0001$ (cf. Seliger & Shohamy, 1989: 234 -235; Hatch & Farhady, 1982: 129). Table 8.3 reports the results for the Tukey test (cf. Miller, 1981: 37 – 48) comparing the mean lengths of compositions over the groups. Groups with the same superscript letters do not differ significantly.

Characteristics	Group 5 Excellent	Group 4 Good	Group 3 Average	Group 2 Fair	Group 1 Poor
Mean length	343.83 [a]	266.27 [bc]	287.25 [b]	279.86 [b]	222.14 [c]

Table 8.3 Results for Tukey's Studentized Test comparing mean lengths over groups.

The mean length of compositions distinguishes Group 5 from all the other groups. Group 4 is not distinguished from Groups 3, 2 and 1. Group 3 is not significantly different from Group 2, but Groups 2 and 3 are significantly different from Groups 1, 4 and 5.

8.3.2 Number of T-units per Composition

The average number of T-units per composition across the groups generally shows a similar trend as length of composition. Table 8.4 presents these results.

Group	N	Mean	Standard Deviation	Coefficient of Variance
1	29	23.76	9.06	38.14
2	57	29.91	10.18	34.04
3	64	29.19	9.80	33.58
4	48	26.85	7.28	27.09
5	18	30.83	10.56	34.24

Table 8.4 Average number of T-units per composition as calculated for groups.

There is an increasing trend for non-adjacent groups. There were no significant statistical differences between the groups for the average number of T-units. The average number of T-units per composition is reported for the sake of completeness. No ANOVA or Tukey results are reported here.

8.3.3 The Average Number of Words per T-unit

The average T-unit length is an important measure in both L1 and L2 acquisition. A number of studies (e.g. Hunt, 1965; Larsen-Freeman and Strom, 1977; Larsen-Freeman, 1978; Homburg, 1984; Bardovi-Harlig & Bofman, 1989) have shown that it is a good indicator of syntactic development.

The total (N), the mean, standard deviation and coefficient of variance were calculated for the number of T-units. The results are presented in Table 8.5.

Group	N	Mean	Standard Deviation	Coefficient of Variance
1	29	9.64	2.62	27.20
2	57	9.86	2.53	25.69
3	64	10.18	2.54	24.92
4	48	10.36	2.24	21.58
5	18	11.85	3.70	31.18

Table 8.5 Average number of words per T-unit as calculated for consecutive groups.

Although the results for the groups are clustered closely together, the average number of words per T-unit increases for each consecutive group. All five groups produced what Hunt (1965) refers to as mid-length T-units (cf. 5.5.1).

An analysis of variance test was conducted to determine the statistical significance of this measure. The analysis resulted in $F(4; 211) = 2.92$ with $p<0.0224$. Tukey's Studentized Test was conducted in order to see whether any of the groups' results differed significantly. The results are reported in Table 8.6.

	Group 5 Excellent	Group 4 Good	Group 3 Average	Group 2 Fair	Group 1 Poor
Words/T-unit	12.0^{a}	10.28^{b}	10.18^{b}	9.78^{b}	9.63^{b}

Table 8.6 Results for Tukey's Studentized Test for mean lengths of T-units as calculated for groups.

Groups with the same superscript letters do not differ significantly. The mean length of T-units per composition distinguishes Group 5 significantly from all the groups. Groups 4, 3, 2 and 1 do not differ significantly. The average length of T-units per composition seems to distinguish well between excellent compositions and those in other groups, but not so much between the groups ranging from good to poor.

8.3.4 Percentage Error-free T-units per Composition

Subjects who write longer T-units do not necessarily write correct T-units (cf. 7.4). As Larsen-Freeman (1978: 445) points out, although increasing length may be a satisfactory indicator of increasing proficiency, length alone might not be a sufficient measure for L2 acquisition. Structural errors should also be taken into account, since they indicate a lack of syntactic control (cf. Gaies, 1976 & Larsen-Freeman, 1978).

Longer T-units tend to be more complex and learners can more readily produce errors. It is easier to make errors when writing longer, more complex structures than when writing short and simple ones. Therefore, the percentage of the T-units that are error-free per essay was calculated for each group. The results are shown in Table 8.7.

Group	N	%	Standard Deviation	Coefficient of Variance
1	29	21.06	18.85	89.49
2	57	34.30	21.51	62.72
3	64	40.10	18.04	45.00
4	48	51.06	22.24	43.56
5	18	69.75	12.04	17.26

Table 8.7 The percentage error-free T-units per composition as calculated for each group.

There is a clear linear trend toward a higher percentage of error-free T-units. This indicates that learners in Group 5 not only write longer T-units, but do so with fewer errors. An analysis of variance test showed the results to be statistically significant with $F(4; 211) = 21, 76$ and $p < 0.0001$.

Tukey's Studentized test was used to determine if the groups differed significantly. Results are shown in Table 8.8.

	Group 5 Excellent	Group 4 Good	Group 3 Average	Group 2 Fair	Group 1 Poor
%EFT	69.75^{a}	51.06^{b}	40.09^{bc}	34.30^{c}	21.065^{d}

Table 8.8 Results for Tukey's Studentized Test for the percentage of error-free T-units per composition as calculated for groups.

The multiple comparisons distinguished Group 5 from all the other groups. Groups 4 and 3, 3 and 2 and 2 and 1 do not differ significantly. However, Groups 5, 3 and 1 are distinguished clearly. According to these results, the percentage of EFTs per composition discriminates well between non-adjacent groups, but not between adjacent groups. In other words, the percentage of EFTs a learner produces can be used to determine whether the composition is at the level of Group 1, 3 or 5.

8.3.5 Total Number of Words per Error-free T-unit

Results for the total number of words per EFT support the findings reflected in the previous tables. It is clear from Table 8.9 that learners from each consecutive group write more words per EFT than the previous group.

Group	N	Total	Standard Deviation	Coefficient of Variance
1	29	39.38	45.89	116.53
2	57	80.79	59.85	74.08
3	64	95.22	53.09	55.75
4	48	126.83	69.44	54.75
5	18	219.44	78.61	35.82

Table 8.9 The total number of words per EFT as calculated for consecutive groups.

The results for the total number of words are reported for the sake of completeness. No ANOVA or Tukey results are reported here.

8.3.6 Percentage Words per Error-Free T-units

Table 8.10 reports the percentages of words per composition in error-free T-units and shows results similar to Table 8.7.

Group	N	%	Standard Deviation	Coefficient of Variance
1	29	17.02	17.62	103.57
2	57	29.63	22.25	75.09
3	64	34.33	16.89	49.20
4	48	46.71	21.67	46.39
5	18	63.41	13.60	21.44

Table 8.10 The percentage of words used in EFTs as calculated for groups.

The increase from one group to the next is very clear. This implies that as learners develop, they produce more words in error-free T-units. An ANOVA resulted in $F(4; 211) = 21.19$, which is statistically significant, with $p < 0.0001$.

Tukey's Studentized Test showed which groups are significantly different for the percentage words per EFT. The results are reported in Table 8.11.

	Group 5 Excellent	Group 4 Good	Group 3 Average	Group 2 Fair	Group 1 Poor
% words/EFT	62.8^{a}	47.0^{b}	35.7^{bc}	25.7^{cd}	15.6^{d}

Table 8.11 Results for Tukey's Studentized Test to discriminate between groups for the percentage of words used in EFTs.

The table shows that Group 5 is once again clearly distinguished from all the other groups. The percentage of words per error-free T-units does not distinguish clearly between Groups 4 and 3, 3 and 2, and 2 and 1. However, it discriminates between Groups 5, 3 and 1. The percentage of words per EFTs discriminates clearly between non-adjacent groups, but not between adjacent groups. The percentage words per EFTs can be used to divide compositions into Groups 1, 3 or 5.

8.3.7. Summary of Results

From the results reported in the tables above, a number of conclusions can be drawn. Learners of English as a second language in the matric year display certain characteristics in writing. Subjects at a higher level of syntactic development tend to write longer essays than those at lower levels. Learners of non-adjacent groups (1, 3 and 5) write more T-units per composition. Consecutive groups write longer T-units. As subjects progress through levels of development, they tend to produce more correct (error-free) T-units. Both the average number of words per T-unit and the T-unit accuracy ratio show that the more advanced a learner becomes, the longer the correct T-units become. They produce on average more words per error-free T-unit. In other words, they use more words correctly in T-units than incorrectly.

These results support the original findings of Hunt (1965), as well as Larsen-Freeman and Strom (1977) and Larsen-Freeman (1978). Larsen-Freeman and Strom (1977) also found that the number of words per composition increased for more advanced learners. The average number of words per T-unit did not show significant differences between groups. (i.e. the groups were not statistically different). They suggest measures of length and error-free T-units as the best variables on which to base an index of development, since they are easy to quantify, linear progression seems possible and they seem impervious to different language backgrounds (1977: 132).

Larsen-Freeman (1978) found an F = 14.79 for the length of the compositions, which was significant at the level 0.001. For the average number of words per T-unit, F = 6.10. The groups did not differ significantly.

Results for the percentage error-free T-units showed an upward linear trend for consecutive groups. The error-free T-unit ratio was calculated. This is calculated by dividing the number of error-free T-units by the total number of T-units (EFT/T). With an average percentage of error-free T-units of 40.92%, the accuracy ratio for the whole study sample was 0.409.

The independent variable (mark) and the dependent variables (cf. 7.5) were correlated. The correlations are reported in Table 8.12.

	Number T-units	Length	**Mark**	Number words per T-unit	% words per EFT	% EFT
Number T-units	1.0000	0.7494 < 0.0001	0.1004 0.1424	0.4447 < 0.0001	-0.0333 0.6267	-0.0516 0.0451
Length	0.7493 < 0.0001	1.0000	**0.2603** 0.0001	0.2064 0.0023	0.9433 0.1672	0.1329 0.0510
Mark	0.1004 0.1424	0.2603 0.0001	1.0000	0.1929 0.0045	0.5153 < 0.0001	0.5229 < 0.0001
Number words per T-unit	-0.4447 < 0.0001	0.2064 0.0023	0.1929 0.0045	1.0000	0.1743 0.0103	0.2491 0.0002
% words per EFT	-0.0333 0.6267	0.0943 0.1672	**0.5153** < 0.0001	0.1742 0.0103	1.0000	0.9706 < 0.0001
% EFT	-0.0515 0.4507	0.1329 0.0510	**0.5229** < 0.0001	0.2490 0.0002	0.9706 < 0.0001	1.0000

Table 8.12 Correlations between the mark and characteristics.

The percentage EFTs shows a moderate to high correlation (cf. Wolfe-Quintero et al., 1998: 38) with the mark at 0.5229. Other studies that also found a moderate correlation include Arnaud (1992), Larsen-Freeman (1978) and Larsen-Freeman and Strom (1977). Some studies, such as Hirano (1991), Kawata (1992) and Tomita

(1990), found a high correlation (cf. Wolfe-Quintero et al., 1998). However, these studies did not all share the same definition of what constitutes an error.

The correlation between the length of the composition and the mark is weak, with a correlation at 0.2603 (cf. Wolf-Quintero et al., 1998: 38).

The percentage words per EFTs also correlate moderately with the mark at 0.5153. The results showed that the number of words per error-free T-units distinguishes between the five groups with F = 18.43. Adjacent groups were not statistically different, but non-adjacent groups were.

8.4 Selecting the Best Predictors of the Level of Syntactic Development

The results in the previous section show that some characteristics discriminate better among the five groups than others. In order to establish an index, one needs to determine which characteristic or combination of characteristics would best predict the level of syntactic development. The mark was used as the independent variable or criterion (cf. 7.4 and 7.5). A Stepwise Regression Analysis (cf. Hatch & Farhady, 1982: 233 – 235; Seliger & Shohamy, 1989: 222-226) and a Discriminant Analysis (cf. Seliger & Shohamy, 1989: 226) were conducted to determine which characteristic, or combination of characteristics, was the best predictors of the mark. This section discusses the results of these analyses.

8.4.1 Stepwise Regression

The Stepwise Regression was conducted to identify the most important predictors of the criterion (the mark). The Stepwise Regression included the STEPWISE and REG procedures of the SAS system.

First, the Regression identified the two best predictors of the criterion. The result indicated that the length of the composition and the percentage error-free T-units per composition may be the best predictors.

A coefficient of determination (R^2) of 0.363 was obtained for both the predictors, while R^2 for the percentage EFT alone was only 0.320. The percentage EFTs proved

to be the most significant of the two variables with $F(2; 211) = 99.63$ with $p < 0.0001$, with the value for composition length being $F(4,211) = 59.96$.

To verify the above selection of characteristics, all possible subsets of regression were performed. The process produced essentially the same results.

Length and the percentage of error-free T-units therefore seem to be the best predictors of the level of syntactic development. In other words, these two characteristics would constitute the best index to determine the level of syntactic development that a learner has achieved in writing.

A correlation procedure was conducted to see how these two independent variables (length and %EFT) and combinations of them correlate with the mark. As mentioned above (cf. 8.3.7), the percentage EFTs correlated moderately with the mark. The correlation was 0.5229. The length correlated rather weakly at 0.263. The combination of the two predictors correlated with the mark at 0.589. According Wolfe-Quintero et al. (1998: 38), this is a moderate to high correlation. This combination, (Length) × (%EFT), correlates better with the mark than the two variables taken separately. This suggests that (Length) × (%EFT) is a good indicator of the level of development.

The Stepwise Regression results suggest that the product of the length and the percentage EFTs would be the best indicator of syntactic development. However, if the percentage of EFTs is used, the values of the index become very big. The ratio of EFTs is calculated by dividing the number of EFTs by the total number of T-units. By using the accuracy ratio (i.e. the EFT/T ratio instead of the percentage EFTs), the index values will be more practical to work with. Therefore, the error-free T-unit ratio (EFT/T) will be used instead of the percentage EFTs. In terms of the results of the Stepwise Regression, the following formula can be used to determine the level of syntactic development a learner has reached:

- (Length) × (EFT/T)

The Stepwise Regression showed that these two characteristics are good predictors of the mark. Also, this formula correlates highly with the mark at 0.589. One could therefore base an index of development on these two predictors, using the suggested formula.

8.4.2 Discriminant Analysis

The results of the Stepwise Discriminant Analysis also identified the percentage of error-free T-units (i.e. accuracy ratio of T-units) as the best indicator of the level of development. The classification used in the Discriminant Analysis was related to the original division into the five groups (poor–excellent) to determine the validity with which the compositions were classified into groups.

The Discriminant Analysis entailed determining the accuracy with which compositions were classified into groups. The DISCRIM procedure was used to determine which compositions were scored incorrectly to the extent that they should have been classified into a different – better or poorer – group. Only two groups, namely 'poor' and 'good' were used. Groups 1 and 2 of the original division into 5 groups were combined to form the 'poor' group, while Groups 4 and 5 were combined to form the 'good' group. The results of the new classification into the two groups were then compared with the original division to compare the accuracy with which the compositions were grouped. The new classification, using cross validation (cf. Huberty, 1994), was based on composition length and percentage EFTs, and the results were then compared with the original division to determine the accuracy with which the compositions were initially grouped.

The linear Discriminant function for the variables and the groups produced the following formulas:

- "Good" group: y = -3.475 + (0.124)(% EFT)
- "Poor" group: y = -0.948 + (0.065)(% EFT)

These two formulas can be used to classify compositions into the two Groups, 'poor' and 'good'.

The results of the DISCRIM procedure appear in Table 8.13.

Classification by means of Discriminant Function

	From Group	Good	Poor	Total
Original Classification	**Good (original)**	46 69.7%	20 30.3%	66 100%
	Poor (original)	21 24.4%	65 75.58%	86 100%
	Total (original)	67 44.07%	86 68.8%	152 100%

Table 8.13 Number of observations and percent classified into groups.

Table 8.13 shows that the 'Good' group contained a total of 66 compositions and the 'Poor' group a total of 86. Using (Length) x (EFT/T), 46 (69.7%) of the compositions that were originally classified as 'good' were again classified as 'good'. However, the DISCRIM procedure classified 20 (30.3%) of the compositions originally regarded as good, into the 'poor' group. Of the 86 compositions originally regarded as poor, 65 (75.58%) were reclassified into the poor group and 21 (24.42%) were classified into the 'good' group. Differences in classification may be attributed to the fact that the original classification also took the content and creativity of the compositions into account. Nevertheless, the reclassification suggests a good validity for the original grouping. The majority of the essays were divided into the same groups during the reclassification as they were divided into during the original classification.

The reclassification suggests a good validity for the original grouping. The majority of the essays were divided into the same groups during the reclassification as they were divided into during the original classification. This also implies that the EFT/T is a good predictor of the mark.

The results of the Discriminant Analysis support the result of the Stepwise Regression. The product of length and EFT ratio (%EFT) is identified as the best predictor of the criterion (the mark). Therefore, the formula that was suggested by the Stepwise Regression, (Length) × (EFT/T), is supported.

8.4.3 The Index

The results for the characteristics, the ANOVAs and Tukey tests, the Stepwise Regression and the Discriminant Analysis have all been considered. The results show that the best indicator of the level of development is a combination of the length of the essay and the error-free T-units ratio (EFT/T). The length of compositions and the percentage of error-free T-units both increase for the five consecutive groups. In other words, subjects who scored higher marks both wrote longer essays and had a higher percentage of error-free T-units.

The two characteristics combined, (Length) × (EFT/T) (cf. 8.3.7), produced a moderate to high correlation (cf. Wolfe-Quintero et al., 1998: 38) with the mark of 0.589 with $R^2 = 0.365$. The multiple correlation from the Stepwise Regression produced a correlation of 0.603, with $R^2 = 0.364$. The two values for the squares (R^2) are very similar. These findings increase the validity of the suggested formula, (Length) × (EFT/T), as an index formula to predict the mark.

The results of the Discriminant Analysis further support the suggestion that the accuracy ratio of EFTs is a good predictor of the mark and as a measure of development. The reclassification of the compositions into the groups 'Poor' and 'Good', according to the accuracy ratio of EFTs, resulted in a division similar to the original classification.

Thus, the suggested index formula, (Length) × (EFT/T), is supported for two reasons:

- The length and the EFT accuracy ratio are good predictors of the mark (Stepwise Regression and Discriminant Analysis);
- The two characteristics combined correlate with the mark at 0.589, with $R^2 = 0.0365$, which does not differ much from the multiple correlation of 0.603, $R^2 = 0.0364$.

8.5 An Analysis of the Higher Grade and Standard Grade Groups

This section separates the Higher Grade and the Standard Grade groups of the sample. Results are reported for the two variables, average length of composition and the percentage of error-free T-units, as calculated individually for the Higher Grade and Standard Grade groups. The index formula is used to calculate an index value (i.e. determine the level of syntactic development) for Higher Grade and Standard Grade learners.

8.5.1 Higher Grade

8.5.1.1 Average Number of Words per Composition in Higher Grade

Table 8.14 reports the results for the average number of words per composition for Higher Grade papers in the five groups.

Group	N	Mean	Standard Deviation
1	17	214.7	90.4
2	32	292.0	62.0
3	28	305.2	72.0
4	21	298.6	58.9
5	14	329.4	67.1

Table 8.14 Number of words per composition for Higher Grade groups.

There is a general increase in the average number of words per composition written by students at the higher grade level. The mean of Group 3 is slightly higher than the mean of Group 4. The standard deviation, however, is quite large. It may be that many of the essays in Group 3 consisted of fewer than 305 words. It is possible that the average might have been pushed up by only a couple of long essays. The overall pattern shows an increase of the number of words per composition for consecutive groups.

An analysis of variance was conducted to compare the mean length over the Higher Grade groups. The analysis resulted in F (4; 107) = 6.52 with $p < 0.0001$. The results for Tukey's multiple comparison test is reported in Table 8.15.

	Group 5 Excellent	Group 4 Good	Group 3 Average	Group 2 Fair	Group 1 Poor
Mean length	329.36[a]	298.57[a]	305.21[a]	292.0[a]	214.71[b]

Table 8.15 Results for Tukey's Studentized Test for mean lengths in Higher Grade groups.

According to these results, the average length of the essays clearly discriminates between the poorest group and the other groups. It does not distinguish clearly between Groups 2 to 5.

8.5.1.2 Percentage Error-free T-units in Higher Grade

As pointed out (cf. 8.3.4), longer T-units seem to be a characteristic of learners at a higher level of development. Also, writing longer T-units may result in producing more errors. Table 8.16 shows the results for the percentage EFTs for Higher Grade.

Group	N	% EFT	Standard Deviation
1	17	19.70	19.32
2	32	40.18	22.41
3	28	41.17	18.38
4	21	53.32	19.59
5	14	68.30	11.99

Table 8.16 Percentage of T-units that is error-free for Higher Grade groups.

There is clearly a linear increase in the percentage of error-free T-units for consecutive groups. The learners at a higher level of development produce more correct T-units than those at a lower level.

An ANOVA for the Higher Grade produced an F (4; 107) = 13.81, significant at the 0.0001 level of probability.

Tukey's multiple comparison test was conducted to determine which groups differ significantly. Table 8.17 shows the results.

Characteristic	Group 5	Group 4	Group 3	Group 2	Group 1
% error-free T-units	68.30 [a]	53.315 [ab]	41.17 [b]	40.18 [b]	19.71 [c]

Table 8.17 Results of Tukey's Studentized Test for Higher Grade and the percentage of error-free T-units.

The percentage error-free T-units does not discriminate between Groups 2, 3 and 4. It does discriminate between Groups 5, 3 and 1. It is therefore a better discriminant than the average number of T-units per composition.

8.5.1.3 An Index Value for HG

The following formula for an index has been suggested:

- Index = (Length of the composition) × (EFT/T).

This formula can now be applied to the results for the Higher Grade in order to produce a numerical value that represents the level of syntactic development achieved.

The average lengths and the average percentage of EFTs for the whole higher Grade group were calculated. This was done by first calculating the average length and the average percentage EFTs for each individual composition in the sample. Then the average for the whole Higher Grade group was calculated, using the averages for each individual composition. The SAS system was used for these calculations.

Table 8.18 reports the Index value for the Higher Grade group.

N	Index Value	Standard Deviation	Coefficient of Variation	Maximum	Minimum
112	130.26	82.36	63.23	398.05	0.0

Table 8.18 Descriptive statistics of index for Higher Grade.

The index value for the Higher Grade group is 130.26 for the sample of compositions in this study.

The index value can be used in future to determine whether the standard for the Higher Grade has improved or not.

8.5.2 Standard Grade

This section applies the index to the Standard Grade to determine the level of syntactic development the learners have reached.

8.5.2.1 Average Number of Words per Composition in Standard Grade

The average number of words per composition for the Standard Grade group is reported in Table 8.19.

Group	N	Mean length	Standard Deviation
1	12	232.67	88.21
2	25	264.32	103.23
3	36	273.28	112.02
4	27	241.15	40.99
5	4	394.50	134.25

Table 8.19 The number of words per composition over Standard Grade groups.

The same pattern that is visible in the Higher Grade papers can be seen in the Standard Grade compositions. Group 3 again has a higher average (this time much higher) than Group 4. But once again the standard deviation for Group 3 is much larger than that for Group 4.

The averages of the Standard Grade groups are generally a little lower than those for the Higher Grade groups. The exception is Group 5, where the Standard Grade group averaged 394.5 words and the Higher Grade group only 329.4 words per composition. This might be attributed to the Higher Grade group's ability to control the language better, since Higher Grade learners are expected to be at a more advanced level of

syntactic development than Standard Grade learners. Also, the composition of 815 words may have influenced the average length in the Standard Grade.

An analysis of variance test resulted in F (4; 99) = 2.73 with p < 0 0332. The Tukey multiple comparison test results are reported in Table 8.20.

Characteristic	Group 5	Group 4	Group 3	Group 2	Group 1
Mean length	394.5[a]	273.28 [b]	264.32 [b]	241.15 [b]	232.67 [b]

Table 8.20 Results for Tukey's Studentized Test comparing mean length over Higher Grade groups.

According to the Tukey test results, the mean length of compositions for Standard Grade distinguishes Group 5 from all the other groups. It does not discriminate between Groups 1 to 4.

8.5.2.2 Percentage Error-free T-units in Standard Grade

The percentage error-free T-units for Standard Grade groups was then calculated. The results are reported in Table 8.21.

Group	N	Mean	Standard Deviation
1	12	22.99	18.83
2	25	26.78	18.06
3	36	39.26	17.99
4	27	49.31	24.33
5	4	74.82	12.41

Table 8.21 Percentage T-units that are error-free for Standard Grade groups.

Results for the Standard Grade show a similar trend as the results for the Higher Grade. For each consecutive group, the percentage of the error-free T-units in Standard Grade essays increase.

An ANOVA resulted in F (4; 107) = 13.81 with p < 0.0001. The results for Tukey's Studentized Test for the percentage error-free T-units are reported in Table 8.22.

Characteristic	Group 5	Group 4	Group 3	Group 2	Group 1
% error-free T-units	68.30 [a]	53.32 [ab]	41.17 [b]	40.18 [b]	19.71 [c]

Table 8.22 Results for Tukey's Studentized Test for percentage error-free T-units for Standard Grade.

Results for the percentage of error-free T-units show significant differences between Groups 5 and 4, Groups 5 and 1, and Groups 4 and 1. Groups 2, 3 and 4 are not distinguished clearly. Group 1 is significantly different from all the other groups.

8.5.2.3 An Index Value for SG

The same procedure was followed for the Standard Grade as for the Higher Grade. Table 8.23 reports the index value as calculated for the Standard Grade group.

N	Index Value	Standard Deviation	Coefficient of Variation	Maximum	Minimum
104	100.41	75.25	74.94	466.94	0.0

Table 8.23 Descriptive statistics of index for Standard Grade.

The index value of syntactic development for Standard Grade level of development is 100.41 for the sample of compositions in this study.

8.6 An analysis of the Whole Group

In this section the index formula is used to calculate the index value for the complete sample of compositions analysed. The index value indicates the average level of development of the whole group of Grade 12 ESL learners.

Table 8.24 reports the index value calculated for the whole sample.

N	Mean	Standard Deviation	Coefficient of Variation	Maximum	Minimum
216	115.89	80.24	69.24	466.94	0.0

Table 8.24 Descriptive statistics of index for complete sample.

The index value for the whole sample of compositions for this study is 115.89. This value is the first objective measurement of the standard of ESL syntactic development

of Grade 12s in South Africa. It can be used as a benchmark in future to compare results for matric groups in order to determine whether the standard of syntactic development for the group has improved or not.

8.7 A Comparison of the Provinces

This section compares the index values calculated for the six provinces. Results are presented in Table 8.25.

Province	N	Index Value	Standard Deviation
A	39	110.94	87.15
B	33	100.44	68.93
C	38	118.28	76.03
D	40	139.57	96.22
E	26	92.94	73.72
F	40	122.41	69.04

Table 8.25 Index values for six provinces.

The table shows that the standard of syntactic development differs from province to province. Province E produced the lowest index value of 92.94. This value is lower than that of the Standard Grade group (cf. 8.5.2.3). Province D produced the highest value of 139.57, which is higher than that of the Higher Grade group (cf. 8.5.1.3).

8.8 Summary of Results

This section reported the index values for Higher and Standard Grade, the group as a whole and for each of the six provinces. The index formula, (Length) × (EFT/T), was used to determine the level of syntactic development achieved by each Grade, the group as a whole and the provinces. The results can be compared to establish the difference in the levels of syntactic development achieved by each Grade. The index values show that the syntactic development of the Higher Grade learners are more advanced, as is to be expected, than the Standard Grade learners in Grade 12 for English Second Language. The index produced a value of 130.26 for the Higher Grade and 100.41 for the Standard Grade. The index value for the whole group was

115.89. There is clearly a difference in standard from province to province with index values ranging from 92.94 to 139.57.

8.9 Refined Ways of Determining an Index Formula

Although it may be expected that, as the percentage EFTs in a composition increases, the mark will increase, the same cannot necessarily be expected for the length. These possible non-linear trends were investigated by means of various non-linear regression models and the results are reported in this section.

Non-linear Multiple Regression techniques are sometimes used in cases where the relationships between variables cannot be assumed to be linear. Experiments to fit various models to the data relating mark to the variables composition length and %EFT resulted in the following prediction formula that gave a reasonable fit in terms of R^2

Estimated Mark = -2.78505 + 0.07514 (length) – 0.00008044 (length_square) + 0.19301 (%EFT)

R^2 for this model is 0.3931, which is slightly higher than that obtained for the simpler model giving the index in paragraph 8.4.1.

The value of R^2 indicates the validity of the mark as allocated by means of the formula. A high R^2 value indicates good validity of the mark. The value produced for R^2 is still fairly low. This can be attributed to the fact that the mark is not a completely objective dependent variable. The marks were allocated according to a standard marking grid (cf. 7.4 & Appendix A). When marks are allocated to a composition, the content is also taken into account, as is shown on the vertical axis of the marking grid. The horizontal axis of the grid represents the language use in the compositions. The R^2 value would be higher if the mark was based only on grammar.

The formula provided by the Multiple Regression (Mark = 5.4143 + 0.02195 x Length + 0.19303 x % EFT) can be used as an alternative means of determining a more

refined index for the grammatical section (represented on the horizontal axis of the marking grid).

The mark calculated for the language use in a composition can then be placed on the horizontal axis of the marking grid. The marking grid can then be used accordingly to consider the content and creativity of compositions.

Another way of employing the index formula (Length) × (EFT/T) is by presenting the variables in a table. Some markers may prefer the use of a table according to which they can estimate what mark to allocate, whereas others might prefer using the index formula. Such a table would represent the estimated mark and given values of the other two variables, namely the composition length and the % EFT.

A table has been drawn up by means of which the marker can predict the mark according to the percentage of EFTs and the length of a composition. (The table can easily be adjusted for compositions with a different total. For example, the estimated marks can be multiplied by 2.5 to count out of 100.) Tables 8.26 and 8.27 report the predicted marks for compositions based on the percentage of error-free T-units per composition and the length of the composition.

The tables show the following performance pattern: a higher mark is predicted for a longer composition with a higher percentage EFTs. However, the increase is not linear with regard to the length of the composition. In other words, a higher mark is not necessarily predicted for very long compositions. A mark of 31/40 is predicted for compositions that consist of 440 words with 70% error-free T-units. However, a composition with 75% EFTs consisting of 800 words will only receive a mark of 25/40. The learner who produces longer compositions will score higher marks, as long as the percentage error-free T-units increases, but only up to a point. When compositions become too long, the length influences the mark negatively. Learners who produce short compositions with a high percentage of EFTs will also receive a lower mark. Table 8.26 shows the increasing trend of the marks as the % EFTs and the length increase up to 400 words. Table 8.27 indicates the turning point at 500 words and the decreasing trend as compositions become longer.

Figure 8.1 illustrates this relationship between the estimated mark (z axis), length (y axis) and the percentage error-free T-units (x axis). The surface plot has a curved shape, which illustrates a general increase in the dependent variable (estimated mark) and the two other variables (%EFTs and Length). The figure shows that, for a given essay length the mark increases steadily, but for a given percentage EFT the mark first increases but then reaches a maximum and starts decreasing for longer essays. This latter tendency is more marked for lower values of % EFT.

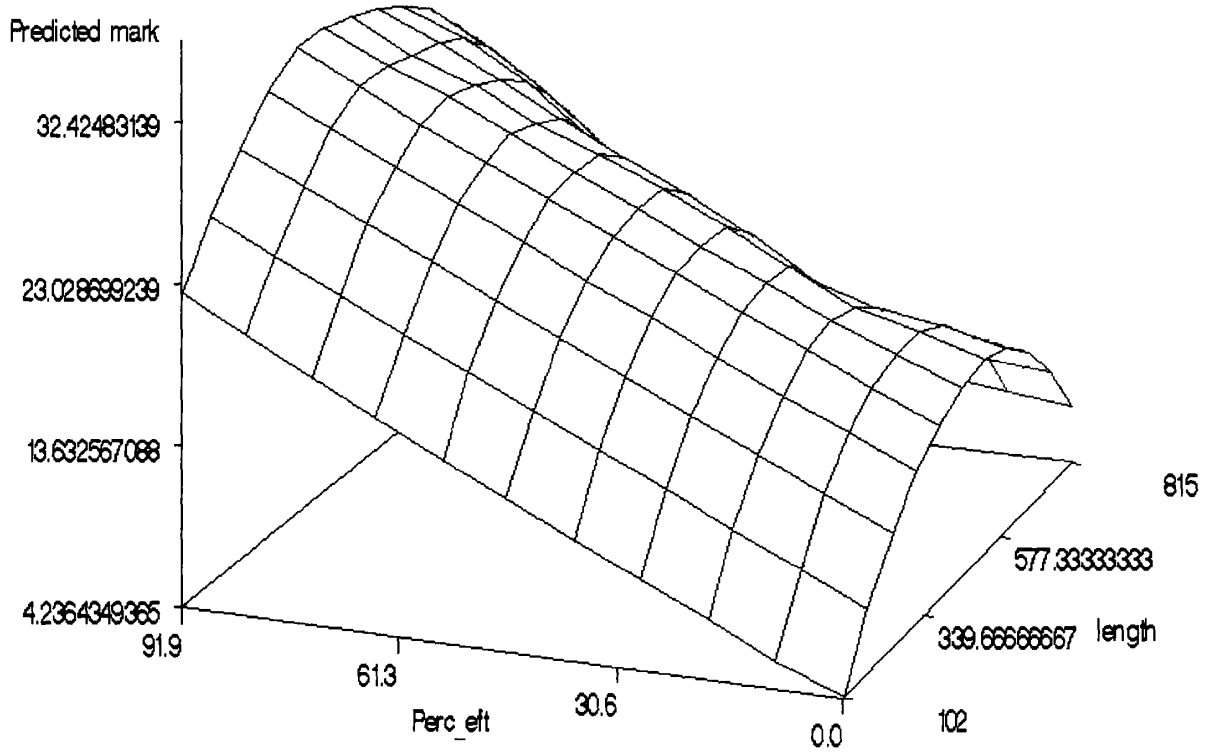

Figure 8.1 Surface plot: the relationship between mark, % EFT and average composition length.

%	LENGTH IN WORDS																			
EFT	**0**	**20**	**40**	**60**	**80**	**100**	**120**	**140**	**160**	**180**	**200**	**220**	**240**	**260**	**280**	**300**	**320**	**340**	**360**	**380**	**400**
100	13	15	17	19	20	22	23	25	26	27	29	30	31	32	33	34	34	35	36	36	37
95	13	14	16	18	19	21	23	24	25	26	28	29	30	31	32	33	33	34	35	35	36
90	12	14	16	17	19	20	22	23	24	26	27	28	29	30	31	32	32	33	34	34	35
85	11	13	15	16	18	20	21	22	24	25	26	27	28	29	30	31	31	32	33	33	33
80	11	12	14	16	17	19	20	22	23	24	25	26	27	28	29	30	30	31	31	32	32
75	10	12	13	15	17	18	19	21	22	23	24	25	26	27	28	29	29	30	30	31	31
70	9	11	13	14	16	17	19	20	21	22	23	24	25	26	27	28	28	29	29	30	30
65	9	11	12	14	15	17	18	19	20	21	22	23	24	25	26	27	27	28	28	29	29
60	8	10	11	13	14	16	17	18	19	21	22	23	23	24	25	26	26	27	27	28	28
55	8	9	11	12	14	15	16	18	19	20	21	22	22	23	24	25	25	26	26	26	27
50	7	9	10	12	13	14	16	17	18	19	20	21	22	22	23	24	24	25	25	25	26
45	6	8	9	11	12	14	15	16	17	18	19	20	21	21	22	23	23	24	24	24	25
40	6	7	9	10	12	13	14	15	16	17	18	19	20	20	21	22	22	23	23	23	23
35	5	7	8	9	11	12	13	14	15	16	17	18	19	19	20	21	21	22	22	22	22
30	4	6	7	9	10	11	12	14	15	15	16	17	18	19	19	20	20	20	21	21	21
25	4	5	7	8	9	11	12	13	14	15	15	16	17	18	18	19	19	19	20	20	20
20	3	5	6	7	9	10	11	12	13	14	15	15	16	17	17	18	18	18	19	19	19
15	3	4	5	7	8	9	10	11	12	13	14	14	15	16	16	17	17	17	18	18	18
10	2	3	5	6	7	8	9	10	11	12	13	14	14	15	15	16	16	16	17	17	17
5	1	3	4	5	6	8	9	10	10	11	12	13	13	14	14	15	15	15	15	16	16
0	1	2	3	5	6	7	8	9	10	10	11	12	12	13	13	14	14	14	14	14	14

Table 8.26 Prediction table: number of words (from 0 - 400) x % error-free T-units.

%	LENGTH IN WORDS																			
EFT	**420**	**440**	**460**	**480**	**500**	**520**	**540**	**560**	**580**	**600**	**620**	**640**	**660**	**680**	**700**	**720**	**740**	**760**	**780**	**800**	**820**
100	37	38	38	38	38	38	38	38	38	38	38	38	37	37	36	36	35	34	33	32	31
95	36	36	37	37	37	37	37	37	37	37	36	36	36	35	35	34	33	33	32	31	30
90	35	35	36	36	36	36	36	36	36	35	35	35	34	34	33	33	32	31	30	29	28
85	34	34	34	35	35	35	35	34	34	34	34	33	33	32	32	31	30	29	29	28	27
80	33	33	33	33	33	33	33	33	33	33	32	32	31	31	30	29	29	28	27	26	25
75	32	32	32	32	32	32	32	32	32	31	31	30	30	29	29	28	27	26	25	24	23
70	30	31	31	31	31	31	31	31	30	30	30	29	29	28	27	26	26	25	24	23	22
65	29	29	30	30	30	30	29	29	29	29	28	28	27	26	26	25	24	23	22	21	20
60	28	28	28	28	28	28	28	28	28	27	27	26	26	25	24	23	23	22	21	20	18
55	27	27	27	27	27	27	27	27	26	26	25	25	24	24	23	22	21	20	19	18	17
50	26	26	26	26	26	26	26	25	25	24	24	23	23	22	21	20	19	19	17	16	15
45	25	25	25	25	25	25	24	24	24	23	23	22	21	21	20	19	18	17	16	15	13
40	24	24	24	24	23	23	23	23	22	22	21	21	20	19	18	17	16	15	14	13	12
35	22	22	22	22	22	22	22	21	21	20	20	19	18	18	17	16	15	14	13	11	10
30	21	21	21	21	21	21	20	20	20	19	18	18	17	16	15	14	13	12	11	10	9
25	20	20	20	20	20	19	19	19	18	18	17	16	16	15	14	13	12	11	10	8	7
20	19	19	19	19	19	18	18	17	17	16	16	15	14	13	12	11	10	9	8	7	5
15	18	18	18	18	17	17	17	16	16	15	14	14	13	12	11	10	9	8	6	5	4
10	17	17	17	16	16	16	15	15	14	14	13	12	11	10	9	8	7	6	5	3	2
5	16	15	15	15	15	14	14	13	13	12	11	11	10	9	8	7	6	4	3	2	0
0	14	14	14	14	14	13	13	12	12	11	10	9	8	7	6	5	4	3	2	0	-1

Table 8.27 Prediction table: number of words (420 - 820) x % error-free T-units.

.

8.10 Error Analysis

This section reports the results of the Error Analysis and briefly discusses its implications.

As mentioned in Chapter 7, the types of errors were coded. The frequencies for all types of errors recorded for the group as a whole (i.e. Higher and Standard Grades were not distinguished) are reported in Table 8.28. The table only reports the error frequencies. Obligatory occasions were not taken into account. Only those errors that occurred 40 times or more are reported. Because of the many different types of errors, the percentages for the individual types are low. However, when considering that the percentage refers to the percentage of times that a specific type of error occurred, the results become more significant. For example, 7.4% of the errors that occurred was a tense error (as opposed to any of the other types of errors that were recorded), namely the present tense being used instead of the past tense. (The complete list of Error Categories is available in Appendix C.) The table indicates the code of the error and the error type, as well as the number of times the error occurred throughout the complete sample, and the percentage.

Type of Error	Frequency	Percentage
22b Tense: Present Tense used instead of Past	460	7.4
5c Article: omitted	330	5.31
4 Verbs: disagreement of subject and number/verb person	277	4.45
17a Prepositions: wrong preposition used	272	4.38
24a Verbs: wrong verb used	196	3.15
24b Verbs: main verb omitted	194	3.12
2 Use of adverb or adjective instead of the other	193	3.1
15b Nouns: Plural "s" omitted	188	3.02

12a Idiomatic speech: sentence/phrase does not make any sense	168	2.7
24f Verbs: omit 3rd person singular –s	161	2.59
1 Adjectives: wrong adjective used	151	2.43
12b Idiomatic speech: expression/idiom used incorrectly, one does not speak like that	128	2.06
6b Auxiliary omitted	123	1.98
19c Personal & possessive pronouns: 'it' omitted.	115	1.85
13d Infinitive: incorrect form used instead of the infinitive	109	1.75
24c Verbs: incomplete	109	1.75
19b Personal & possessive pronouns: omitted	106	1.7
26 Word order incorrect	106	1.7
19a Personal & possessive pronouns: incorrect	107	1.73
19d Personal & possessive pronouns: subject pronoun used as redundant element	104	1.67
15e Nouns: incorrect noun used	97	1.56
25 Vocabulary: wrong word used (e.g. homophones, homonyms etc.)	96	1.54
17b Prepositions: used but unnecessary	90	1.45
17c Prepositions: omitted	88	1.41
22c Tenses: Past used instead of Present	88	1.41

5b Articles: used, but unnecessary	86	1.39
13a Infinitive: used incorrectly or unnecessary	79	1.27
11b Gerund: omitted – simple verb used instead of -ing	76	1.22
16 Noun and verb confused: one used instead of the other	76	1.22
5a Article: wrong one used	74	1.19
5d Article: confused with another part of speech	70	1.13
15a Nouns: plural 's' attached wrongly/unnecessarily	65	1.05
20a Reflexive pronoun: wrong one used	64	1.03
3a Adverbs: wrong one used	58	0.93
21a Relative pronoun: wrong one used	56	0.9
3d Adverb: confused with another part of speech	56	0.9
21b Relative pronoun: omitted	54	0.87
9a Conjunctions: omitted	52	0.84
18 Possessive pronoun and possessive adjective confused: one used instead of the other	51	0.82
19e Personal & possessive pronouns: confused with another part of speech	51	0.82
21d Relative pronoun: confused with another part of speech	51	0.82
3b Adverbs: omitted	50	0.8

11a Gerund: used incorrectly or unnecessarily	48	0.77
24j Verbs: confused with another part of speech	43	0.69
15f Noun: omitted	42	0.68
24d Verb: unnecessary or redundant	42	0.68
24e Noun: wrong (/unnecessary) attachment of 3rd person singular –s	41	0.66
20b Reflexive pronoun: omitted	40	0.64

Table 8.28 Error frequencies for Grade 12 ESL learners.

Table 8.28 shows that the error produced most frequently by Grade 12 ESL learners is a tense error. Instead of using the Past tense, the Present tense is used. In many of these cases, the learner omitted the –ed at the end of the verb which is needed to form the Past tense. The next most frequently occurring error is articles that are omitted – a typical feature of Black South African English (cf. Van der Walt & Van Rooy, 2002). The learners also made many prepositional mistakes – such as using the wrong preposition, e.g. "...vote to president Mbeki"– and concord errors. These are also typical of Black South African English.

This table suggests the areas of grammar that should be focused on in the classroom in order to improve the learners' level of syntactic development. This raises the question of how grammar should be taught and learned. This has been the subject of considerable debate in recent years. The debate is reflected in Wasanasomsithi's (1998) review of a number of studies. Some are in favour of a focus-on-form approach (e.g. Higgs & Clifford, 1982; Pica, 1983; Lightbown & Spada, 1990;), while others support a focus-on-meaning approach (e.g. Hymes, 1971; Savignon, 1972; Hammond, 1988). Wasanasomsithi (1998) supports an approach which integrates a focus on grammar instruction within a communicative approach. This approach is generally supported in the literature (e.g. Harley et al., 1993; Ellis 1995; Fotos & Ellis, 1991; Fotos 1994). Lightbown (2002: 533) also favours the general principle of

the integration of form and meaning instruction. She points out, however, that she does not reject separate grammar lessons completely, as they may be necessary to prepare learners to notice certain language features when they are encountered in communicative interaction and to know what they are being corrected on if the teacher provides feedback 'on the fly'.

Harley et al. (1993) explain that learners may benefit from formal instruction if those activities that focus on form and those that focus on communication are more closely related. They argue that more opportunities to practise a target language alone are not enough to promote development (1993: 77). Learners need to be motivated to use correct, appropriate and coherent language. Ellis (1995) agrees that acquisition in the classroom occurs when learners take part in interaction that provides comprehensible input and output. However, he says that some grammatical structures, especially those in higher levels of development, require direct instruction.

According to Ellis (1994: 659), formal instruction increases accuracy and speeds up progress through developmental stages. Form-focused instruction can therefore be used to develop the syntactic competence of Grade 12 learners. It is likely to facilitate natural language development.

The question is, however, whether the errors identified in this study have become fossilized (cf. 2.4.1). Although learners may never achieve native-like proficiency, formal instruction in the areas suggested in the table above can help to improve the syntactic standards of ESL learners. An improvement in syntactic control can contribute to the improvement of general standards of ESL in South Africa.

8.11 Conclusion

This chapter reported and discussed the results of the statistical analyses. The significant findings include: Consecutive groups tend to write longer essays; there is an increasing trend in the number of T-units per composition for consecutive groups; learners of non-adjacent groups write more T-units per composition; consecutive groups also write longer T-units; and subjects at higher levels of development tend to produce more correct T-units. Both the average number of words per T-unit and the

percentage of words per T-unit show that as a learner becomes more advanced he/she uses more words per error-free T-unit.

According to the results of a Stepwise Regression and a Discriminant Analysis, the length of the composition and the percentage error-free T-units are the best predictors of the mark. The subset-regression of the Stepwise Regression analysis suggested that a combination of the length of compositions and the percentage EFTs would be a more accurate index than the two separate characteristics. The index formula was established as:

$$\text{Index} = (\text{Length}) \times (\text{Error-free T-unit Ratio})$$

A distinction was made between the Higher Grade and Standard Grade groups. Results showed that the average length and the percentage EFTs per composition increase for consecutive groups in Higher and Standard Grade groups, as well as for the group as a whole.

Index values were calculated for the Higher Grade, Standard Grade, the group as a whole and the individual provinces. The value for the Higher Grade was calculated at 130.26. The index value for the Standard Grade was calculated at 100.41. The whole group had a value of 115.89. The individual provinces produced a range of values from 92.94 to 139.57.

The frequencies of errors were reported. The errors that occurred most frequently were tense-related errors, use of the wrong article, concord errors and incorrect preposition use. ESL teachers should focus on these grammatical structures in an attempt to improve learners' levels of syntactic development.

Chapter 9

Conclusion

9.1 Introduction

This chapter briefly reviews the main findings of the study. The implications of the findings are considered and suggestions for further research are made.

9.2 Main findings

The main aim of this study was to establish an index formula that can be used to measure the syntactic development of ESL learners objectively and to employ this index formula to determine the syntactic standard of a group of Grade 12 ESL learners in South Africa.

Interlanguage is a characteristic of second language learners. A learner's interlanguage refers to the developing grammatical system which differs from both that of the learner's L1 and the target language. The acquisition of a second language progresses in stages along the interlanguage continuum. These stages are difficult to define objectively, but it is necessary to attempt to do so in order to determine the stage of a learner's development at a particular level. The T-unit is a useful measure in such an analysis.

Language development can be defined as the process during which a learner progresses from one stage of the interlanguage continuum to the next, towards native-like use of the target language. It refers to those characteristics of the language that a learner produces at some point or stage along a developmental continuum. Language development therefore differs from language proficiency.

The following index formula was established as an objective measure of ESL syntactic development:

$$\text{Index} = (\text{Length}) \times (\text{EFT/T}).$$

The length of a composition is multiplied by the ratio of Error-free T-units in a composition. This index formula does not discriminate against any native language group. Vague and subjective descriptions of the standard of ESL can now be replaced with a concrete numerical value – indicating the actual level of syntactic development – which is easily comparable with results produced by other (also international) students.

The syntactic level of development for the whole sample of compositions of Grade 12 ESL learners was calculated at 115.89. The index value for the Higher Grade was calculated at 130.26 and for the Standard Grade at 100.41. The individual provinces produced a range of values from 92.94 to 139.57.

The Error Analysis identified a number of grammatical structures that are problematic for Grade 12 ESL learners. Teaching these structures in class should be a priority for the ESL teacher as it may assist in improving the level of syntactic development of the Grade 12 learners.

9.3 Implications of the Results

Certain grammatical structures have been identified that learners at the Grade 12 level have not yet fully acquired. Most of the errors produced by learners were of the following types: tense-related errors (use of the Present Tense instead of the Past), use of the wrong article, concord errors and incorrect use of prepositions. 21% of all errors produced were in one of these four structures.

These grammatical errors, along with those suggested in Table 8.28, should be attended to in the ESL classroom, following a form-focused approach. Formal instruction in these structures may help prevent fossilization of the incorrect structural forms. Improving the syntactic level of the learners may help learners to perform better in other areas of the target language, which can contribute to the improvement of general standards of English as a second language in South Africa.

The index values calculated in the present study can serve as a benchmark for future calculations, making it easy to determine in an objective way whether standards are improving or not.

9.4 Limitations of the Study

One limitation of the present study is that it is only a cross-sectional study. The ideal would be to conduct a longitudinal study, in which the level of development is calculated for a number of consecutive years. This will make it possible to determine whether standards are improving.

The sample used in the present study consisted of sample compositions from six provinces in South Africa. The results cannot be generalised to all Grade 12 ESL learners.

Ellis (1996: 71) describes the concept of 'stage' as 'fuzzy'. It remains an approximation. Furthermore, the development of accuracy in learners occurs in a U-shape, following peaks and valleys. The Error-free T-unit ratio may develop in a zigzag pattern (Wolfe-Quintero et al., 1998: 39). It is not possible to say whether the compositions that were analysed were produced during a period of regression, or at a developmental peak.

9.5 Suggestions for Further Research

As suggested in 9.4, a longitudinal study should be conducted. The index values can then be compared with those of subsequent years.

This research should be expanded to include all the provinces in South Africa. The sample should include more compositions from each province in order for the results to be more representative of the all matriculation candidates. The samples should be random ones, representing different levels of performance.

The validity of the index formula proposed in this study should be further investigated by making use of comparable data.

The possibility of non-linear trends in language development should be considered.

9.6 Conclusion

The accuracy with which a learner produces a second language is often regarded as the only concrete evidence of development. Measuring accuracy as such on its own, however, may not be a sufficient means to determine the level of language development. Measuring the accuracy as well as the length of the language produced by L2 learners gives a more accurate account of the level of development. Bardovi-Harlig and Bofman (1989) consider the T-unit as an accurate measure of movement towards native-like production of the target language, since it takes both the length and accuracy of a language sample into account.

It is necessary for researchers and teachers to work together towards improving the standard of English as a second language in South Africa. The first step towards improving the standard is to determine the actual standard of English as a Second Language for the Grade 12 learners. This study has shown how the level of syntactic development can be determined and how problem areas can be identified. If the index values are used as benchmark for future research, and the necessary attention is paid to the development of learners' syntactic development, standards of English usage can be improved in the country.

Appendix A: Marking Grid for Higher Grade and Standard Grade

		Excellent	Very Good	Good	Above Average	Average	Below Average	Poor	Very Weak	Unacceptable
ENGLISH ADDITIONAL LANGUAGE HIGHER GRADE GRID: **40 MARKS**	LANGUAGE	- excellent command of English idiom. - perfect agreement. - uses tenses effortlessly. - complex sentences. - language evokes vivid images. - language manipulated skilfully. - virtually no errors.	- as for *excellent* but there are a few more errors. - errors are not glaring.	- generally correct idiom. - agreement generally correct. - tenses well handled. - good sentence structure. - pleasing vocabulary. - good use of pronouns, prepositions, parts of speech, punctuation and spelling.	- adequate use of idiom. - some errors in agreement. - can generally handle tenses. - adequate vocabulary. - few errors in pronouns, prepositions, parts of speech, punctuation and spelling.	- slightly incorrect use of idiom. - mother-tongue influence slightly impedes flow. - several errors in agreement. - tenses inconsistent. - mostly simple sentence structure. - ordinary vocabulary. - some errors in pronouns, prepositions, parts of speech, punctuation and spelling.	- noticeably incorrect use of idiom. - mother-tongue influence impedes flow. - many errors in agreement. - tenses switched continually. - simple sentence structure and some incomplete sentences. - some inappropriate vocabulary. - several errors in pronouns, prepositions, parts of speech, punctuation and spelling.	- poor use of idiom. - extensive mother-tongue influence. - major problems in agreement. - poor verb tenses. - sentences confused and often incomplete. - inadequate vocabulary. - frequent errors in pronouns, prepositions, parts of speech, punctuation and spelling.	- unidiomatic. - governed by use of mother-tongue. - very little agreement. - very weak verb tenses. - muddled and incomplete sentences. - inappropriate vocabulary. - countless errors in pronouns, prepositions, parts of speech, punctuation and spelling.	- hardly any use of linguistic options. - very difficult to understand. - verging on unintelligible.
CONTENT		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
EXCELLENT - exceptional development of topic. - original, creative, imaginative. - *(non-narrative)*: argument logically developed and ideas thoroughly explored. - *(narrative)*: strong story line building to a striking climax. - excellent introduction and conclusion. - vivid detail and/or examples. - realistic, sincere, mature.	A+	**36 - 40**	**34 - 38**	**32 - 35**						
VERY GOOD - topic well developed. - thought-provoking, interesting, convincing. -*(non-narrative)*: several points explored and logically developed. - (*narrative*) : story flows well, refreshing, unusual. - very good introduction and ending. - very good detail and/or examples.	*A*	**34 - 38**	**32 - 35**	**30 - 33**	**28 - 31**					
GOOD - content sound. - (*non-narrative*) : well-planned, a few points developed. - (*narrative*): story flows well, shows potential. - good introduction and conclusion. - attempt at interesting detail and examples.	B	**32 - 35**	**30 - 33**	**28 - 31**	**26 - 29**	**24 - 27**				
ABOVE AVERAGE - adequate interpretation of topic. - (*non-narrative*):evidence of planning, some points developed. - (*narrative*): story reads easily. - relevant introduction and conclusion. - details adequate.	C		**28 - 31**	**26 - 29**	**24 - 27**	**22 - 25**	**20 - 23**			

AVERAGE - appropriate if ordinary content. - (*non-narrative*): some relevant points but not developed, may be repetitive. - (*narrative*): ordinary, pedestrian. - introduction and conclusion may be omitted or inadequate.	D			**24 - 27**	**22 - 25**	**20 - 23**	**18 - 21**	**16 - 19**		
BELOW AVERAGE - attempt at addressing topic but not always clear, dull. -(*non-narrative*): very few ideas, often repetitive, sometimes off topic, general line of thought can be followed. - (*narrative*): story line can be followed but not always clear. - introduction/conclusion omitted or inadequate. -	E				**20 - 23**	**18 - 21**	**16 - 19**	**14 - 17**	**12 - 15**	
POOR - slight understanding of topic, but ideas are muddled, tedious. - argument and/or story line can barely be followed. - little evidence of planning.	F					**16 - 19**	**14 - 17**	**12 - 15**	**10 - 13**	**8 -11**
VERY WEAK - a vague attempt to answer the question. - contents irrelevant or disconnected. - no evidence of planning.	G						**12 - 15**	**10 - 13**	**8 - 11**	**4 - 9**
INADEQUATE - difficult to connect contents to topic. - pre-learnt. - off topic. - too little written to assess.	H							**8 - 11**	**4 - 9**	**0 - 7**

		Excellent	Very Good	Good	Above Average	Average	Below Average	Poor	Very Weak	Unacceptable
ENGLISH ADDITIONAL LANGUAGE STANDARD GRADE GRID: **40 MARKS**	LANGUAGE	- excellent command of English idiom. - perfect agreement. - uses tenses effortlessly. - complex sentences. - language evokes vivid images. - language manipulated skilfully. - virtually no errors.	- as for *excellent* but there are a few more errors. - errors are not glaring.	- generally correct idiom. - agreement generally correct. - tenses well handled. - good sentence structure. - pleasing vocabulary. - good use of pronouns, prepositions, parts of speech, punctuation and spelling.	- adequate use of idiom. - some errors in agreement. - can generally handle tenses. - adequate vocabulary. - few errors in pronouns, prepositions, parts of speech, punctuation and spelling.	- slightly incorrect use of idiom. - mother-tongue influence slightly impedes flow. - several errors in agreement. - tenses inconsistent. - mostly simple sentence structure. - ordinary vocabulary. - some errors in pronouns, prepositions, parts of speech, punctuation and spelling.	- noticeably incorrect use of idiom. - mother-tongue influence impedes flow. - many errors in agreement. - tenses switched continually. - simple sentence structure and some incomplete sentences. - some inappropriate vocabulary. - several errors in pronouns, prepositions, parts of speech, punctuation and spelling.	- poor use of idiom. - extensive mother-tongue influence. - major problems in agreement. - poor verb tenses. - sentences confused and often incomplete. - inadequate vocabulary. - frequent errors in pronouns, prepositions, parts of speech, punctuation and spelling.	- unidiomatic. - governed by use of mother-tongue. - very little agreement. - very weak verb tenses. - muddled and incomplete sentences. - inappropriate vocabulary. - countless errors in pronouns, prepositions, parts of speech, punctuation and spelling.	- hardly any use of linguistic options. - very difficult to understand. - verging on unintelligible.
CONTENT		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
EXCELLENT - exceptional development of topic. - original, creative, imaginative. - *(non-narrative)*: argument logically developed and ideas thoroughly explored. - *(narrative)*: strong story line building to a striking climax. - excellent introduction and conclusion. - vivid detail and/or examples. - realistic, sincere, mature.	A+	**38 - 40**	**34 – 38**	**32 - 36**						
VERY GOOD - topic well developed. - thought-provoking, interesting, convincing. -*(non-narrative)*: several points explored and logically developed. - (*narrative*) : story flows well, refreshing, unusual. - very good introduction and ending. - very good detail and/or examples.	*A*	**34 - 38**	**32 – 36**	**30 - 34**	**28 - 32**					
GOOD - content sound. - (*non-narrative*) : well-planned, a few points developed. - (*narrative*): story flows well, shows potential. - good introduction and conclusion. - attempt at interesting detail and examples.	B	**32 - 36**	**30 – 34**	**28 - 32**	**26 - 30**	**24 - 28**				
ABOVE AVERAGE - adequate interpretation of topic. - (*non-narrative*):evidence of planning, some points developed. - (*narrative*): story reads easily. - relevant introduction and conclusion. - details adequate.	C		**28 – 32**	**26 - 30**	**24 - 28**	**22 - 26**	**20 - 24**			

AVERAGE - appropriate if ordinary content. - (*non-narrative*): some relevant points but not developed, may be repetitive. - (*narrative*): ordinary, pedestrian. - introduction and conclusion may be omitted or inadequate.	D			**24 - 28**	**22 - 26**	**20 - 24**	**18 - 22**	**16 - 20**		
BELOW AVERAGE - attempt at addressing topic but not always clear, dull. -(*non-narrative*): very few ideas, often repetitive, sometimes off topic, general line of thought can be followed. - (*narrative*): story line can be followed but not always clear. - introduction/conclusion omitted or inadequate. -	E				**20 - 24**	**18 - 22**	**16 - 20**	**14 - 18**	**12 - 16**	
POOR - slight understanding of topic, but ideas are muddled, tedious. - argument and/or story line can barely be followed. - little evidence of planning.	F					**16 - 20**	**14 - 18**	**12 - 16**	**10 - 14**	**8 -12**
VERY WEAK - a vague attempt to answer the question. - contents irrelevant or disconnected. - no evidence of planning.	G						**12 - 16**	**10 - 14**	**8 - 12**	**4 - 10**
INADEQUATE - difficult to connect contents to topic. - pre-learnt. - off topic. - too little written to assess.	H							**8 - 12**	**4 - 10**	**0 - 8**

Appendix B: Example of Composition Data Worksheet

Essay code	Group	Length	Avg words/T-unit	Words/EFT-unit	T-unit no	EFT-unit	Error Types
gha 080	3 21	302	8.62	0	1		15d 15g
gha 080	3 21	302	8.62	0	2		17a
gha 080	3 21	302	8.62	4	3	eft	
gha 080	3 21	302	8.62	19	4	eft	
gha 080	3 21	302	8.62	3	5	eft	
gha 080	3 21	302	8.62	0	6		19c
gha 080	3 21	302	8.62	0	7		25
gha 080	3 21	302	8.62	6	8	eft	
gha 080	3 21	302	8.62	0	9		3a
gha 080	3 21	302	8.62	2	10	eft	
gha 080	3 21	302	8.62	6	11		
gha 080	3 21	302	8.62	0	12		12b 15e
gha 080	3 21	302	8.62	0	13		22b
gha 080	3 21	302	8.62	0	14		19c
gha 080	3 21	302	8.62	5	15		
gha 080	3 21	302	8.62	0	16		5c 17a
gha 080	3 21	302	8.62	4	17		
gha 080	3 21	302	8.62	4	18		22b 21a 6a 22b
gha 080	3 21	302	8.62	0	19		5d 19e 22f
gha 080	3 21	302	8.62	0	20		24a 25
gha 080	3 21	302	8.62	0	21		6b 24c 22o 17a 9a
gha 080	3 21	302	8.62	7	22	eft	
gha 080	3 21	302	8.62	7	23		
gha 080	3 21	302	8.62	11	24		
gha 080	3 21	302	8.62	0	25		26 25
gha 080	3 21	302	8.62	0	26		15c
gha 080	3 21	302	8.62	0	27		5d 9c 5c
gha 080	3 21	302	8.62	0	28		5d 19e
gha 080	3 21	302	8.62	10	29		
gha 080	3 21	302	8.62	5	30		
gha 080	3 21	302	8.62	0	31		25 17b 12b
gha 080	3 21	302	8.62	4	32		
gha 080	3 21	302	8.62	5	33		
gha 080	3 21	302	8.62	10	33		
gha 080	3 21	302	8.62	0	35		24b 6b 17c

Appendix C: Error Categories

1 Adjectives	
1a wrong	I like a woman who bought a house...*except* a man.
1b omitted	The crime was very ( ) because of the vote.
1c unnecessary	...people who are *very* suffering. ..are *very* dead.
1d confused	The crime was very *danger* (dangerous).
2 Adjectives/adverbs	
2 use an adverb or adjective instead of the other	You done your job smooth (smoothly). My step mom treated us nice (nicely).
3 Adverbs	
3a wrong one used	---
3b omitted	A few years (ago) the Lord blessed this place. ...cattles were looked (after) by him.
3c unnecessary	...because now people they do horrible things.
3d confused with another part of speech/word	...to give them same (some) money.
4 Agreement	
4 disagreement of subject and number/verb person	The death penalty are solution. ...there are a large number ..the president will satisfy their people needs. ..of the whole things. ...but they does not play the role.
5 Articles	
5a wrong article used (a/an/the)	It is the must to bring back the death penalty. ..is a taxi a express transport.
5b article used, but unnecessary	The shoot of people is like the drink(ing) was punished by *a* white people.
5c article omitted	..buy something like ( ) car. (The) Death penalty....
5d confused with another part of speech (e.g. adverb/adjective/pronoun etc)	The (It) was important to vote.
6 Auxiliary **to have/ to be/shall/ will**	
6a wrong	All work was not belong to me.
6b omitted	black people (will) never rule..
6c unnecessary	I was vote to president Thabo Mbeki.

6d incomplete	...those who were died... ...crime rate has to (be) very strong. ..people will not (be) killed.
must/ have to/ need to etc	
6e wrong	I would go to work in the morning and they will...
6f omitted	After you get money you (must) buy a new taxi.
6g unnecessary	...after fighting most people you must die.
6h incomplete	----
7. Comparison	
7a wrong	The *worse* thing in all of this is... The people died as cattle
7b incomplete	...and I felt (as) if my life was over.
7c unnecessary	...but than when children feel...
8. Conditionals	
8a wrong	If I am far away from home I felt free.
8b omitted	---
8c unnecessary	...but than when children feel...
8d confused	I (If) you want your child...
9. Conjunctions	
9a omitted	They want us to buy for them clothes, (and a) car..
9b unnecessary	...and etc.
9c confused	Sins/since If a wife and husband have some rules...than (then) no one can come between them.
10. Count/non-count	
10 use non-count as a count noun	Those killers think that it's *a* freedom to kill. You don't have these money.
11 Gerund	
11a wrong/ unnecessary	It is advantaging to be a single mother.
11b omitted (simple verb used instead of –ing)	The disadvantage of to be a single parent.. If there is no vote(ing)... I was suggested (suggesting)...
12. Idiomatic speech	
12a sentence/phrase just does not make sense at all	No one same running and can not tometel. don't that if you husband or is not accept kid. Comprein to work baut this time it is very hungry very march
12b you don't speak like that (e.g. idioms/expressions	

used wrongly)	Even me did not go for marriage (get married) ..we were born in the number of ten people. Death penalty must bring back the spare of rod not the roar. All work was(did) not belong to me.
13. Infinitive 13a wrong/ unnecessary 13b incomplete 13c "to" omitted 13d incorrect form instead of infinitive	 I believe to vote (in voting) The father has also to a big role...(a big role to play) -- She did not even sent...
14. Interrogatives 14a wrong 14b omitted 14c unnecessary 14d confused (other part of speech used in its place)	---
15. Nouns 15a plural 's' attached wrongly/unnecessary 15b plural 's' omitted 15c possessive 's omitted 15d possessive 's attached wrongly/ but unnecessary 15e wrong noun used 15f noun omitted 15g unnecessary	 Inside the taxis (referring to one) is big watt speakers. ...of the whole things/ cattles We as children have question. / cause of the death ...to have the son('s) respect. ...satisfy their people('s) needs. ...since the early seventy's... ...a lot of vegetable's. ...she must buy some grocer./ ...because of hardness (hardship) I feel. ...they are transported to and from (the hospital). ...do some ironing work.
16. Confuse Noun and Verb 16 nouns or verb used instead as the other.	The *prevent* (prevention) was (done) very carefully. If there is something outside like *theft* (a thief)... ...my *advise* would be...

17 Prepositions 17a wrong 17b used but unnecessary 17c omitted 17d confused with another part of speech	The people died as cattle ...vote to president Mbeki. ...building the school of (for) the country. ..to be complete(ly) *by* build. She did not look (to) another person. ...it is important that they limit *to* (the) crime.
18 Pronoun/ adjective 18 use either possessive adjective or possessive pronoun instead of the other.	And *you* opinion is yours. ..with *you* own money. If you are ill you take *you* money...
19 Personal & Possessive Pronouns 19a wrong 19b omitted 19c "it" omitted 19d Subject pronoun used as redundant element 19e confused with another part of speech	The fathers must also help where *he* can. Now (I) know that leaving school... Now that (it) is up to the children.. ..(It) is important that they limit... ... what is (it) to vote ..now the people *they* do horrible things. The worse thing in all of *them* (this) are the children.. There is only one thing that makes *my* worried. *Their* must be a very good relationship...
20. Reflexive Pronouns 20a wrong 20b omitted 20c unnecessary 20d confused	the president will satisfy their people needs. ..looks after his car and also his drinks with their friends. ...when a wife asks his husband... ...now the people the (they) do horrible things.
21. Relative Pronouns 21a wrong 21b omitted 21c unnecessary 21d confused	...and there is nothing (that) will be stolen. ...about the lazy people (who) do not succeed... There is only one thing who (that) lets me worry. ...the worse thing in all of them are the children.
22. Tense Sequence 22a Incorrect Use of Continuous (progressive)	I am having standard six...

	By the time I was having a husband things were easy.
22b Present instead of Past	
22c Past instead of Present	I forget to vote...24 April 1996.
	If her child stays with other children he/she did not know that
22d Pres Continuous instead of Present	he/she did not have a father.
22e Present instead of Present Continuous	They say we are voting, but have no jobs, no money.
22f Past instead of Past Continuous	
22g Past Continuous instead of Past	
22h Present instead of Past Cont	
22i Past Continuous instead of Present	
22j Past instead of Present Continuous	
22k Present Continuous instead of Past	
22l Present Continuous instead of Past Continuous	
22m Past Continuous instead of Present Continuous	
22n Past Perfect instead of Past	
22o Past instead of Past Perfect	
22p Other	
	I lived there since..
	..the president was promised (promised) to build..
23 Tense Sequence: mix two tenses together	
23a Past & Present	
	..to be completely build.
23b Present & Past Perfect	Have you ever notice..
	They were send...
23c Past Continuous & Present	black people was punish..
	The people were vote...
23d Past Continuous & Past	The present was satisfy their people needs.
23e Present Continuous & Past	I was suggested that..
	People who are suffered (suffering)
	...trying to gave..
	He promised that everything is going to be right.
23f Other	----
24 Verbs	
24a wrong verb used	The children are (have) many problems.
	They must let (make) their mother aware...

24b main verb omitted 24c incomplete 24d verb unnecessary/redundant 24e wrong (/unnecessary) attachment of 3rd person singular –s 24f omit 3rd person singular –s 24g apostrophe 's wrong/ unnecessary 24h apostrophe 's omitted 24I omission of "to" 24j confuse	..they say we are voting but (have) no jobs, no money. People but you not at all. ...why they (are) having problems. Just like when you are want need to buy something.. ...and prevent should satisfy needs. At least is you are single. ...the children throws a tantrum. These things reflects... ..because she know(s) that she is a single person. That's what he promised Go ( ) play. You (are) gonna (going to) spread it ...a person may *he* killed his/her own family. My friend grown up
25. Vocabulary 25 wrong word used (e.g. homophones, homonyms etc.)	..by (with) your own *many* (money). I *want* (went) out. ..because *they* (there) are a large number... If *the* (there) is no vote... South Africans died as *cattlers* (cattle). He promised that everything is going to be *write.*
26. Word Order 26 word order is wrong	They want us to *buy for them clothes*, car. ...who were died in the past the ones... (ones who died in the past were)

Bibliography

AKIN, J. 1976. Enhancing the Syntactic Fluency of Beginning Foreign Language Learners through Sentence- Combining Practice. Georgia: University of Georgia (Dissertation - Ed.D).

ALDERSON, J.C., KRAHNKE, K.J. and STANSFIELD, C.W. 1987. *eds*. Reviews of English Language Proficiency Tests. Washington, DC: Teachers of English to Speakers of Other Languages.

ARNAUD, P.J.L. 1992. Objective Lexical and Grammatical Characteristics of L2 Written Compositions and the Validity of Separate-component tests. (*In* Arnaud, P.J.L & Bejoint, H. *eds*. Vocabulary and Applied Linguistics. London: Macmillan.)

ARTHUR, B. 1980. Short-term changes in EFL composition skills. (*In* Yorio, C., Perkins, K. and Schachter, J. *eds*. On TESOL '79. Washington D.C.: TESOL.)

BACHMAN, L.F and PALMER, A.S. 1981. The Construct Validation of the FSI Oral Interview. *Language Learning*, 31, 67-86.

BACHMAN, L.F. and PALMER, A.S. 1982. The Construct Validation of some Components of Communicative Proficiency. *TESOL Quarterly*, 16: 449 – 465.

BACHMAN, L.F. and PALMER, A.S. 1996. Language Testing in Practise. Oxford: Oxford University Press.

BACHMAN, L.F. 1990. Fundamental Considerations in Language Testing. Oxford: Oxford University Press.

BACHMAN, L. F. 1999. Language Testing at the Turn of the Century: Making sure that what we count counts. Los Angeles: University of California. *American Association for Applied Linguistics (AAAL) Letter,* 20 (2). [Web:]: http://aaal.lang.uiuc.edu/letter/20.2/testing.html [Date of access: 15 March, 2004].

BACHMAN, L.F. and PALMER, A.S. 1996. Language Testing in Practise. Oxford: Oxford University Press.

BAILEY, N., MADDEN, C. and KRASHEN, S. 1974. Is there a 'Natural Sequence' in Adult Second Language Learning? *Language Learning,* 24: 235 – 243.

BAKER, P. and MCENERY, T. not dated. Breaking Through – When does the shift from the twp word stage occur? Language Acquisition- Session 8. [Web:] http://www.ling.lancs.ac.uk/chimp/langac/LECTURE8/8break.htm [Date of Access: 2 June, 2004]

BARDOVI-HARLIG, K and BOFMAN, T. 1989. The Attainment of Syntactic and Morphological Accuracy by Advanced Language Learners. *Studies in Second Language Acquisition,* 11: 17 – 34.

BROOKS, N. 1964. Language and Language Learning: Theory and Practice. New York, Burlingame: Harcourt, Brace and World.

BROWN, R. 1973. A First Language: The Early Stages. London: Allen and Unwin.

BROWN, H. D. 1993. Principles of Language Learning and Teaching. 3rd ed. Englewood Cliffs, NJ: Prentice Hall Regents.

BRUTSCH. S. 1979. Convergent-Discriminant Validation of Prospective Teacher Proficiency in Oral and Written French by means of the MLA Cooperative Language Proficiency Tests for Teachers (TOP and TWP), and Self-ratings. Minneapolis, MN: University of Minnesota. (Dissertation – Ph.D.).

BURT, M. and DULAY, H. 1980. On acquisition orders. (*In* Felix, S *ed.* Second language development. Gunter Narr, Tubingen.)

BURT, K. and KIPARSKY, C. 1972. The Gooficon: a Repair Manual for English. Rowley, MA: Newbury House.

CANALE, J. A. 1994. Linking Language Assessment to Classroom Practices. (*In* R. Rodriguez, Ramos, N and Ruiz-Escalante, J.A. *eds.* Compendium of readings in bilingual education: issues and practices. Austin, Texas: Texas Association for Bilingual Education.)

CANALE, M. and SWAIN, M. 1980. Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, 1(1): 1-47.

CANCINO, H., ROSANSKY, E. and SCHUMANN, J. 1978. The Acquisition of English Negatives and Interrogatives by Native Spanish Speakers. (*In* Hatch, E. *ed.* Second Language Acquisition. Rowley, Mass: Newbury House.)

CARLSEN, G, R. 1965. National Council of Teachers of English, Research Report No.3 (*In* Hunt, K. Grammatical Structures Written at Three Grade Levels).

CHENGAPPA, S., BHAT, S. and HIWARALE, J. 2002. Mean Length of Utterance and Syntactic Complexity in the Speech of Mentally Retarded. *Language in India: Strength for today and bright hope for tomorrow,* 2(6). [Web:] http://www.languageinindia.com/sep2002/minimumlengthutterance.html [Date of Access: 3 June, 2004].

CHOMSKY, N. 1965. Aspects of the Theory of Syntax. Cambridge, Mass: MIT Press.

CLARK, J. L. D. 1985. Language Testing: Past and Current Status - Directions for the future. *Language Testing*, 64(4): 431-443.

CLIFFORD, R. T. 1978. Reliability and Validity of Language Aspects Contributing to Oral Proficiency of Prospective Teachers of German. (*In* Clark, J.L. *ed.* Direct testing of speaking proficiency: theory and application. Princeton, NJ: Educational Testing Service.)

CLIFFORD, R. T. 1981. Convergent and Discriminant Validation of Integrated and Unitary Language Skills: the need for a research model. (*In* Palmer, A.S., Groot, P.J. & Trosper, G. A. *eds*. The construct validation of tests of communicative competence. Washington, DC: TESOL.)

COETZEE- VAN ROOYEN, A. S. and VERHOEF, M. M. 2000. Perception of English Proficiency: views from Southern Sotho speakers. *South African Journal of Linguistics,* 38: 163 – 185.

COOK, V. 1993. Linguistics and Second Language Acquisition. London: The Macmillan Press.

COOPER, T. C. 1976. Measuring Written Syntactic Patterns of Second Language Learners of German. *Journal of Educational Research,* 69 (5): 176-183.

CORDER, S. P. 1967. The Significance of Learner's Errors. *IRAL*, 5: 161-170.

CORDER, S.P. 1971. Idiosyncratic Errors and Error Analysis. *IRAL*, 9(2): 147 – 59.

CORDER, S.P. 1978. Language Learner Language. (*In* Richards, J.C., *ed.* Understanding Second and Foreign Language Learning: Issues and Approaches. Rowley, Mass: Newbury House.)

CORDER, S. P. 1981. Error Analysis and Interlanguage. London: Oxford University Press.

CRYMES, R. 1971. The Relation of Study about Language to Language Performance: with special reference to nominalization. *TESOL Quarterly,* 5(3): 217 – 230.

CUMMING, A. 1997. The testing of writing in a second language. (*In* Clapham, C. and Corson, D., *eds*. Encyclopedia of language and education, 7. Language testing and assessment. Dordrecht, Netherlands: Kluwer.)

DLAMINI, C. 1998. The Thorny Question of Language and Education. *The Sunday Times.* 3 March.

DREYER C. and VAN DER WALT, J.L. 1994. The Interpretation and Evaluation of Quantitative Research Studies In Second Language Acquisition. *Literator,* 15(3): 121-137.

DULAY, H.C. and BURT, M.K. 1973. Should We Teach Children Syntax? *Language Learning*, 23(2): 245 – 258.

DULAY, H.C. and BURT, M.K. 1974. Natural Sequences in Child Second Language Acquisition. *Language Learning,* 24: 37 – 53.

DULAY, H.C. and BURT, M.K. 1975. Creative Construction in Second Language Learning and Teaching (*In* Burt, M and Dulay, H. *eds*. On TESOL '75: New Directions in Second Language Learning, Teaching and Bilingual Education. Washington, DC: Teachers of English to Speakers of Other Languages.)

DULAY, H.C. and BURT, M.K. 1977. Remarks on Creativity in Language Acquisition. (In: Burt, M., Dulay, H. and Finnochiaro, M. *eds*. Viewpoints on English as a second language. New York: Regents.)

DULAY, H.C. and BURT, M.K. 1980. On Acquisition Orders. (*In* Felix, S. *ed.* Second Language Development: Trends and Issues. Tubingen: Gunter Narr.)

DULAY, H., BURT, M and KRASHEN, S. 1982. Language Two. New York: Oxford University Press.

ELLIS, R. 1985. Understanding Second Language Acquisition. London: Oxford University Press.

ELLIS, R. 1994. The Study of Second Language Acquisition. Oxford: Oxford University Press.

ELLIS, R. 1995. Interpretation Tasks for Grammar Teaching. *TESOL Quarterly,* 29:87 – 105.

ELLIS, R. 1996. Understanding Second Language Acquisition. Oxford: Oxford University Press.

ENGBER, C.A. 1995. The Relationship of Lexical Proficiency to the Quality of ESL Compositions. *Journal of Second Language Writing,* 4: 139 – 156.

EUBANK, L. 1987. The Acquisition of German Negation by Formal Language Learners. (*In* Van Patten, B., Dvorak, T. and Lee, J. *eds*. Foreign Language Learning: A Research Perspective. Rowley, Mass: Newbury House.)

EVOLA, J., MAMER, E. and LENTZ, B. 1980. Discrete Point versus Global Scoring for Cohesive Devices. (*In* Oller, J. W. and Perkins, K. *eds*. Research in Language Testing. Rowley, MA: Newbury House.)

FARHADY, H. 1978. The Differential Performance of Foreign Students on Discrete Point and Integrative Tests. Los Angeles: University of California (Thesis - M.A).

FELIX, S. 1980. Second Language Development: Trends and Issues. Tubingen: Gunter Narr.

FELIX, S. 1984. Maturational Aspects of Universal Grammar. (*In* Davies, a., Ciper, C. and Howatt, A.P.R. *eds*. Interlanguage. Edinburgh: Edinburgh University Press.)

FELIX, S. 1987. Cognition and Language Growth. Dordrecht: Foris.

FINN, S. M. 1986. The Abuse of English by Vista Students. *English Usage in Southern Africa,* 17 (1): 1 – 18.

FLAHIVE, E. D. and SNOW, B. G. 1980. Measures of Syntactic Complexity in Evaluating ESL Compositions. (*In* Oller, J.W. and Perkins, K. *eds*. Research in Language Testing. Rowley, Massachusetts: Newbury House.)

FOSTER, P. and SKEHAN, P. 1996. The Influence of Planning and Task Type on Second Language performance. *Studies in Second Language Acquisition*, 18: 299-323.

FOTOS, S.S. and ELLIS, R. 1991. Communicating about Grammar: a task-based approach. *TESOL Quarterly,* 25: 605 – 628.

FOTOS, S.S. 1994. Integrating Grammar Instruction and Communicative Language Use through Grammar Consciousness-raising tasks. *TESOL Quarterly,* 28: 323 – 351.

FULCHER, G. 1987. Tests of Oral Performance: the need for data-based criteria. *ELT Journal,* 41: 287 – 291.

GAIES, S.J. 1976. Sentence-Combining: a technique for assessing proficiency in a second language. Paper presented at the conference on Perspectives on Language, University of Louisville, May 6-8. Louisville, Kentucky.

GAIES, S. J. 1977. The Nature of Linguistic Input in Formal Second Language Learning and Communicative Strategies in ESL Teachers' Classroom Language. (*In* Brown, H. D., Yorio, C. A. and Crymes, R. H. eds. On TESOL '77. Teaching and learning English as a second language: trends in research and practice. Washington, D.C.: Teachers of English to Speakers of Other Languages, 204-212.)

GAIES, S. 1980. T-unit Analysis in Second Language Research: applications, problems and limitations. *TESOL Quarterly,* 14: 53 – 60.

GASS, S.M. & SELINKER, L. 2001. Second Language Acquisition – an introductory course. 2nd ed. Mahwah, New Jersey, London: Lawrence Erlbaum Associates.

GIPPS, C. & EWEN, E. 1974. Scoring Written Work in English as a Second Language. *Educational Research,* 16:121 – 125.

GREENBAUM, L. & MBALI, C. 2002. An Analysis of Language Problems Identified in Writing by Low-Achieving First Year Students, with some Suggestions for Remediation. *Southern African Linguistics and Applied Language Studies,* 20 (4): 233 – 244.

GRUENEWALDT, J.T. 1999. Achieving Academic Literacy in a Second Language: South Africa's Educational Predicament. *South African Journal for Higher Education,* 13(1): 205 – 211.

HAKUTA, K. 1974. Prefabricated patterns and the Emergence of Structure in Second Language Learning. *Language Learning,* 24: 287 – 297.

HAKUTA, K. 1975. Becoming Bilingual at Age Five: the story of Uguisu. Harvard University (Thesis).

HAMMOND, R.M. 1988. Accuracy versus Communicative Competency: the acquisition of grammar in the second language classroom. *Hispania,* 7: 408 – 416.

HAMP-LYONS, L. 1990. Second Language Writing: assessment issues (*In* Kroll, B. *ed.* Second Language Writing: Research insights for the classroom. Cambridge: Cambridge University Press).

HARGREAVES, P. 1987. Royal Society of Arts: Examinations in Communicative Use of English as a Foreign Language. (*In* Alderson, J.C., Krahnke, K.J. and Stansfield, C.W. *eds.* Reviews of English Language Proficiency Tests. Washington, DC: Teachers of English to Speakers of Other Languages.)

HARLEY, B., ALLEN, P., CUMMINS, J. and SWAIN, M. 1993. The development of Second Language Proficiency. Cambridge: Cambridge University Press.

HARMER, J. 2004. The Practice of English Language Teaching. Edinburgh: Longman.

HATCH, E. and FARHADY, H. 1982. Research Design and Statistics for Applied Linguistics. Rowley, Massachusetts: Newbury House.

HENNING, G. 1976. A Developmental Analysis of Errors of Adult Iranian Students of English as a Foreign Language. Paper presented at the UCLA-USC Second Language Acquisition Forum, Los Angeles.

HIGGS, T. & CLIFFORD, R. 1982. The Push toward Communication. (*In* Higgs, T. *ed.* Curriculum, competence, and the foreign language teacher. Skokie, IL: National Textbook Company.)

HIRANO, K. 1991. The Effect of Audience on the Efficacy of Objective Measures of EFL Proficiency in Japanese University Students. *Annual Review of English Language Education in Japan,* 2: 21 – 30.

HOMBURG, T. J. 1984. Holistic Evaluation of ESL Compositions: can it be validated objectively? *TESOL Quarterly,* 18: 87 – 107.

HUBERTY, C.J. 1994. Applied Discriminant Analysis. New York: Wiley & Sons.

HUNT, K. 1965. Grammatical Structures Written at Three Grade Levels. Illinois: National Council of Teachers of English, Research Report No.3.

HUNT, K. 1970. Recent Measures in Syntactic Development. (*In* Lester, M. *ed.* Readings in Applied Transformational Grammar. New York: Holt Rinehart.)

HUOT, B. A 1993. The Influence of Holistic Scoring Procedures on Reading and Rating Student Essays. (*In* Williamson, M. and Huot, B., *eds.* Validating holistic scoring for writing assessment: theoretical and empirical foundations. Cresskill, NJ: Hampton Press.)

HYLTENSTAM, K. 1977. Implicational Patterns in Interlanguage Syntax Variation. *Language Learning,* 27(2): 383 – 411.

HYMES, D. 1971. Competence and Performance in Linguistic Theory. (*In* Huxley, R. & Ingram, E. *eds.* Language Acquisition: models and methods. London: Academic Press.)

INGRAM, D. 1989. First Language Acquisition. Cambridge: Cambridge University Press.

ISHIKAWA, S. 1995. Objective Measurement of Low-Proficiency EFL Narrative Writing. *Journal of Second Language Writing,* 4: 51 – 70.

JACOBS, H.L., ZINGRAF, S.A., WORMUTH, D.R., HART, el. V.F. and HUGHEY, J.B. 1981. Testing ESL Composition: a practical approach. Rowley, MA: Newbury House.

JOUBERT, J. 2004. Goeie Engels = goeie uitslae. *Beeld,* September 7.

KACZMAREK, C. M. 1980. Scoring and Rating 'essay tasks'. (*In* Oller, J.W. and Perkins, K. *eds.* Research in Language Testing. Rowley, Massachusetts: Newbury House.)

KAUR, S. 2000. Problems in Assessing Proficiency in English among Postgraduate Students. University of Malaysia. Original ultiBASE publication. Paper presented at the "Sixth International Literacy and Education Research Network Conference on Learning", Bayview Beach Resort, Penang, Malaysia 27 - 30 September 1999. [Web:] http://ultibase.rmit.edu.au/Articles/may00/ kaur1.htm {Date of Access: 4 April, 2004].

KAWATA, K. 1992. Evaluation of Free English Compositions. *CASELE Research Bulletin,* 22: 49 – 53.

KOBAYASHI, M. 2002. Method effects on reading comprehension test performance: text organisation and response format. *Language Testing,* 19(2): 193 – 220.

KONDO-BROWN, K. 2002. A FACETS Analysis of Rater Bias in Measuring Japanese Second Language Writing Performance. *Language Testing*, 19 (1) 1 – 29.

KRASHEN, S.D. 1979. The Monitor Model for Second Language Acquisition. (*In* Gingras, R., *ed.* Second Language Acquisition and Foreign Language Teaching. Washington, DC: Center for Applied Linguistics.)

KRASHEN, S.D. 1982. Principles and Practices in Second Language Acquisition. Oxford, New York: Pergamon Press.

KRASHEN, S.D., BUTLER, J., BIRNBAUM, R. and ROBERTSON, J. 1978. Two Studies in Language Acquisition and Language Learning. *ITL: Review of Applied Linguistics,* 39 – 40: 73 – 92.

KRASHEN S.D. and TERRELL, T.D. 1995. The Natural Approach: Language Acquisition in the Classroom. New York: Prentice Hall International.

KROLL, B. 1998. Assessing Writing Abilities. *Annual Review of Applied Linguistics*, 18: 219-240.

KUO, H. 2003. The Nature and Causes of Interlanguage Fossilization. *The Source- an online journal of education,* 4(1). [Web:]
http://www.usc.edu/dept/education/TheSource/issue_2003_spring/kuo.htm
[Date of access: 10 May, 2004].

LADO, R. 1957. Linguistics across cultures. University of Michigan: Ann Arbor.

LARSEN-FREEMAN, D. 1976. An Explanation for the Morpheme Acquisition Order of Second Language Learners. *Language Learning,* 26(1): 125 – 134.

LARSEN-FREEMAN, D.E. 1976. Evidence of the Need for a Second Language Acquisition Index of Development. (*In* Ritchie, W.C. *ed.* Second Language

Acquisition Research – issues and implications. London, New York: Academic Press.)

LARSEN-FREEMAN, D.E. 1978. An ESL Index of Development. *TESOL Quarterly*, 12: 439- 448.

LARSEN-FREEMAN, D.E. 1983. Assessing Global Second Language Proficiency. (*In* Seliger, H.W. and Long, M. *eds*. Classroom Oriented Research in Second Language Acquisition. Rowley, MA: Newbury House.)

LARSEN-FREEMAN, D.E. and STROM, V. 1977. The Construction of a Second Language Acquisition Index of Development. *Language Learning*, 27: 123-134.

LARSEN-FREEMAN, D.E. and LONG, M. 1992. An Introduction to Second Language Research. London, New York: Longman.

LENNEBERG, E. 1967. Biological Foundations of Language. New York: Wiley and Sons.

LENNON, P. 1990. Investigating Fluency in EFL: A Quantitative Approach. *Language Learning*, 40: 387-417.

LIGHTBOWN, P. 2002. The Role of SLA Research in L2 Teaching: reply to Sheen. *Applied Linguistics,* 23(4): 529-535.

LIGHTBOWN, P.M. and SPADA, N. 1990. Focus-on-Form and Corrective Feedback in Communicative Language Teaching: effects on second language learning. *Studies in Second Language Acquisition,* 12: 429 – 448.

LOBAN, W. 1976. Language Development: kindergarten through grade twelve. NCTE Research Report No. 18. Urbana, Ill: National Council of Teachers of English.

LUMLEY, T. 2002. Assessment Criteria in a Large-Scale Writing Test: what do they really mean to the raters? *Language Testing:* 19(3): 246 – 276.

LYONS, J. 1968. Introduction to Theoretical Linguistics. Cambridge: Cambridge University Press.

MACKAY, D. G. 1982. The Problems of Flexibility, Fluency, and Speed-Accuracy Trade-off in Skilled Behaviour. *Psychological Review,* 89: 483 – 506.

MATTHEWS, M. 1990. The Measurement of Productive Skills. Doubts Concerning the Assessment Criteria of certain Public Examinations. *ELT Journal,* 44: 117-120.

MAWASHA, A. L. 1982. A Study of Common Errors in Written English with Special Reference to North-Sotho-speaking Children. *English Language Teaching Information Centre (ELTIC) Reporter,* 7 (2): 28 – 34.

MELLON, J. C. 1969. Transformational Sentence Combining: a method for enhancing the development of syntactical fluency in English composition. NCTE Research Report No. 10. Urbana, Ill: National Council of Teachers of English.

MILLER, R.G. 1981. Simultaneous Statistical Inference. New York: McGraw Hill.

MILLS, C. 1990. Syntax and the Evaluation of College Writing: a blind alley. (*In* Arena, L. A. *ed*. Language Proficiency: defining, teaching and testing. New York, London: Plenum Press.)

MONROE, J. H. 1975. Measuring and Enhancing Syntactic Fluency in French. *The French Review,* XLVIII(6): 1023- 1031.

MULLEN, K. A. 1980. Evaluating Writing Proficiency. (*In* Oller, J.W. and Perkins, K. *eds*. Research in Language Testing. Rowley, Massachusetts: Newbury House.)

MULLEN, K. A. 1987. Test of Ability to Subordinate. (*In* Alderson, J.C., Krahnke, K.J. and Stansfield, C.W. *eds*. Reviews of English Language Proficiency Tests. Washington, DC: Teachers of English to Speakers of Other Languages.)

NAS, G. 1975. Determining the Communicative Value of Written Discourse Produced by L2 Learners. Utrecht, The Netherlands: Institute of Applied Linguistics.

NEMSER, W. 1971. Approximative Systems of Foreign Language Learners. *International Review of Applied Linguistics*, 9: 115 – 123.

NEUMAN, R. 1977. An Attempt to Define through Error Analysis the Intermediate ESL level at UCLA. Los Angeles: UCLA. (Thesis – Ph.D.).

NEY, J. W. 1966. Review of Grammatical Structures Written at three Grade Levels, Kellogg W. Hunt. *Language Learning,* 16 (2): 230 – 235.

NKANI, M.D.S. 1998. An Analysis of the Errors in Grammar made by Xhosa Learners of English Second Language. Potchefstroom: PU for CHE. (Mini-dissertation – M.A.)

NORRISH, J. 1983. Language Learners and Their Errors. London: Macmillan Press.

NORTH, B. and SCHNEIDER, G. 1998. Scaling Descriptors for Language Proficiency Scales. *Language Testing*, 15(2): 217-263.

NWAILA, C. 1993. English in South Africa: the dilemma facing teachers. *ELTIC Reporter,* 17: 3 – 10.

O' DONNELL, R. C. 1976. A Critique of Some Indices of Syntactic Maturity. *Research in the Teaching of English,* 10: 33 – 38.

O' DONNELL, R. C., GRIFFIN, W. and NORRIS, R. C. 1967. Syntax of Kindergarten and Elementary School Children: a transformational analysis. NCTE Research Report No. 8. Champaign, Ill: National Council of Teachers of English.

O' HARE, F. 1973. Sentence Combining: improving student writing without formal grammar instruction. NCTE Research Report No. 15. Urbana, Illinois: National Council of Teachers of English.

OLLER, J. W., Jr. 1979. Language Tests at School: a pragmatic approach. New York: Longman.

PICA, T. 1983. Adult Acquisition of English as a Second Language under Different Conditions of Exposure. *Language Learning*, 33: 465 – 497.

PIENEMANN, M. and JOHNSON, M. 1987. Factors Influencing the Development of Language Proficiency. (*In* Nunan, D. *ed.* Applying Second Language Acquisition Research. National Curriculum Resource Centre: Adult Migrant Education Program Australia.

PIKE, L. W. 1973. An Evaluation of Present and Alternative Items Formats for use in the Test of English as a Foreign Language. Research Report. Princeton, New Jersey: Education Testing Service.

ROSANSKY, E. 1976. Methods and Morphemes in Second Language Acquisition Research. *Language Learning,* 26: 409 – 425.

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

SAVIGNON, S.J. 1983. Communicative Competence: theory and classroom practice. Reading, MA: Addison Wesley.

SCHACHTER, J. and CELCE-MURCIA, M. 1983. Some Reservations Concerning Error Analysis. (*In* Robinett, W. and Schachter, J., *eds*. Second Language Learning: Contrastive Analysis, Error Analysis, and Related Aspects. Michigan: University of Michigan Press.)

SCOTT, C. M. and STOKES, S. L. 1995. Measures of Syntax in School-Age Children and Adolescents. *Language, Speech and Hearing Services in School,* 26: 309 – 317.

SCOTT, M. and TUCKER, G.R. 1974. Error Analysis and English Language Strategies of Arab Students. *Language Learning,* 24: 69 – 97.

SELINKER, L. 1970. Grammar and Technical English. (In: Robert C. Lugton ed. English as a Second Language: Current Issues. Philadelphia: Center for Curriculum Development.

SELINKER, L. 1972. Interlanguage. *IRAL*, 10(3).

SELINKER, L. 1992. Rediscovering Interlanguage. London: Longman.

SELINKER, L. and LAMENDELLA, T. 1978. Two Perspectives on Fossilization in Interlanguage Learning. *Interlanguage Studies Bulletin,* 3(2): 144-191.

SELINKER, L. and LAMENDELLA, T. 1979. The Role of Extrinsic Feedback in Interlanguage Fossilization. *Language Learning,* 29: 363 – 375.

SELIGER, H.W. and SHOHAMY, E. 1989. Second Language Research Methods. Oxford: Oxford University Press.

SHARWOOD-SMITH, M. 1994. Second Language Learning: theoretical foundations. London: Longman.

STEVENSON, D.K. 1987. Test of English as a Foreign Language. (*In* Alderson, J.C., Krahnke, K.J. and Stansfield, C.W. 1987 *eds*. Reviews of English Language Proficiency Tests. Washington, DC: Teachers of English to Speakers of Other Languages.)

STEWARD, M.F. 1978. Syntactic Maturity from High School to University: a first look. *Research in the Teaching of English,* 12: 37 – 46.

THOMAS, M. 1994. Assessment of L2 Proficiency in Second Language Acquisition Research. *Language Learning*, 44(2): 307 – 336.

THORNHILL, D. E. 1969. A Quantitative Analysis of the Development of Syntactical Fluency of Four Young Adult Spanish Speakers learning English. Florida: Florida State University (Dissertation – Ph.D.)

TOMITA, Y. 1990. Assessing the Writing Ability of High School Students with the use of T-units. *Step Bulletin,* 2: 14 – 28.

TURNER, C.E. 2000. Listening to the Voices of Rating Scale Developers: identifying salient features for second language performance assessment. *Canadian Modern Language Review,* 56(4): 556.

UNSWORTH, S. 2002. Measuring L2 Proficiency in Theoretical Second Language Acquisition Research. Why and How? Utrecht Institute of Linguistics. [Web:] http://www.let.uu.nl/~Sharon.Unsworth/personal/Unsworth%202002%20NOTT%20handout.PDF [Date of access: 20 January, 2004.]..

UPSHUR, J. A. and TURNER, C.E. 1995. Constructing Rating Scales for Second Language Tests. ELT Journal, 49.

UPSHUR, J. A. and TURNER, C. E. 1999. Systematic Effects in the Rating of Second-Language Speaking Ability: test methods and learner discourse. *Language Testing,* 16(1): 82- 111.

VAN DER WALT, J.L. 1981. The Practical English Course: A Proposed Syllabus for an Afrikaans University.

VAN DER WALT, J.L. and VAN ROOY, B. 2002. Towards a Norm in South African Englishes. *World Englishes,* 21(1): 113 – 128.

VAN PATTEN, B. 1984. Processing Strategies and Morpheme Acquisition. (*In* Eckman, F.R., Bell, L.H. and Nelson, D. *eds*. Universals of Language Acquisition. Rowley, Mass: Newbury House.)

VAUGHEN, C. 1991. Holistic Assessment: What Goes on in the Rater's Mind? (*In* Hamp-Lyons, L. *ed.* Assessing Second Language Writing in Academic Contexts. Norwood, New Jersey: Ablex Publishing Corporation.)

VAVRA, E. 2000. A Theory of Natural Syntactic Development. *Teaching Grammar as a Liberating Art.* [Web:]
http://nweb.pct.edu/homepage/staff/evavra/TGLA/TGLA04.htm
[Date of Access: 20 Jan, 2004.]

VIGIL, F. and OLLER, J. 1976. Rule Fossilization: a tentative model. *Language Learning,* 26(2): 281 – 295.

WASANASOMSITHI, P. 1998. Focus-on Form or Focus-on Meaning: how far should ESL/EFL teachers go? *Thai TESOL Bulletin,* 11(1). [Web:] http://www.thaitesol.org/bulletin/1101/110101.html
[Date of Access: 2 March, 2005].

WEIDEMANN, A. and VAN RENSBURG, C. 2003. Language Proficiency: current strategies, future remedies. *Journal for Language Teaching,* 36(1&2): 152 – 163.

WEINREICH, U. 1953. Languages in Contact. Mouton: The Hague.

WHITE, L. 1982. Grammatical Theory and Language Acquisition. Dordrecht: Floris.

WHITE, L. 2003. Second Language Acquisition and Universal Grammar. Cambridge: Cambridge University Press.

WIDDOWSON, H.G. 1978. Teaching Language as Communication. Oxford: Oxford University Press.

WIDDOWSON, H.G. 1979. Explorations in Applied Linguistics. Oxford: Oxford University Press. Oxford: Oxford University Press.

WIDDOWSON, H.G. 1983. Learning Purpose and Language Use. Oxford: Oxford University Press.

WODE, H. 1981. Learning a Second Language. Tubingen: Gunter Narr.

WOLFE-QUINTERO, K., INAGAKI, S. and KIM, H. 1998. Second Language Development in Writing: measures of fluency, accuracy and complexity. Second Language Teaching and Curriculum Centre. Manoa: University of Hawai'i.

ZOBLE, H. 1984. Cross-language Generalisations and the Contrastive Dimension of the Interlanguage Hypothesis. (*In* Davies, A., Criper, C. and Howatt, A. *eds*. Interlanguage. Edinburgh: Edinburgh University Press.)