Karien Hattingh & Johann L. van der Walt

Centre for Language Practice, Potchefstroom Campus, North West University

An index formula for measuring development in second language writing

A B S T R A C T The measurement of second language development is an essential aspect of second language research. An index of development is an objective measure with which to measure second language development. This article focuses on syntactic accuracy in writing and describes the process followed in establishing an index formula that can measure development in writing. Criteria for an index formula are formulated, and the unit of analysis, the T-unit, is discussed. Grade 12 compositions were analysed and the statistical procedures used in the analysis (correlation, stepwise regression and discriminant analysis) are described. The proposed index formula is Length x EFT/T. This formula can be used in future to determine developmental progress in second language acquisition, and whether standards are improving or not. Pedagogical implications are also pointed out.

Keywords: Second language development, syntax, accuracy, index of development, T-unit, index formula.

1. Introduction

The accurate measurement of development in a second language remains problematic. Various methods have been used, and criticised, such as morpheme studies, standardised tests, and error analysis (cf. Gass & Selinker, 2001). Level of development is a central issue in second language research, but has practical implications as well – for example in the assessment of learners in high-stake examinations, such as the school-leaving examination in South Africa, where the standards achieved are of national concern.

One way of establishing level of development in a second language is by means of an index. Gass and Selinker (2001: 49), however, point out that the field of Second Language Acquisition has not yet come up with an index of development. This article is an attempt to make a contribution in this regard. Our aim is to describe the development of an index formula that can be used to measure second language writing objectively. We focus on writing as it is generally assumed

that the characteristic patterns of advanced learners (the objects here) are best studied in written production (Lorenz, 1999: 11). The developmental level reached by a learner can be determined by measuring those characteristics of a learner's production that reveal some point or stage along the interlanguage continuum (cf. Wolfe-Quintero et al., 1998: 2). It implies a developmental measure that can trace this progress in a linear fashion as the second language develops. Developmental measures are usually based on accuracy, fluency or complexity measures. Here we focus on accuracy only, which can be defined as 'freedom from error', or comparison with target-like use.

2. An index of development

Gass and Selinker (2001: 49) point out that, unlike in first language acquisition research, there is no easy way of determining whether a given second language learner is more or less proficient than another. An index of development could provide such a measure. It should be an independent yardstick by which the development of a learner can be gauged.

The idea of an index of language development is not new. As long ago as 1964, Brooks noted that it was necessary to establish a systematic programme of tests to measure success in second language development. Labels such as 'beginner', 'advanced' and 'intermediate' are often used to describe someone's level of language development. These as very subjective terms and they are too vague to express the actual level of development of a particular group of students. An index of development would provide a more precise and objective description. Thomas (1994: 309) argues that it is important to understand how knowledge and language use may change during the course of acquisition and as long as research results are not connected to learners at identifiable stages in L2 development, it will remain difficult to investigate these issues.

3. Criteria for an index formula

An index formula that indicates developmental level should satisfy a number of criteria. These include:
- It should be based on a formula according to which a numerical score, representing the level of development, can be calculated;
- It should be calculated in an objective way;
- It should be a global measure and not based on any specific language structure;
- The measure should be readily available and applicable *post hoc* to previously collected data;
- Factors such as the first language background, educational background and the age of learners should not influence the index (cf. Unsworth, 2002);
- Its value should increase uniformly and linearly as learners proceed towards full acquisition of the second language.

4. Unit of analysis in establishing an index formula

An index formula should be based on a valid and reliable unit of analysis. The question is: Which unit is the best for measuring syntactic development in writing?

There seems to be general agreement that the most reliable unit for measuring second language development is the T-unit. Gass and Selinker (2001: 50) describe it as an "exact means of measuring syntactic development", and it has been employed as basis for a number of measures of development.

The T-unit was first proposed by Hunt in 1965 in a study of first language acquisition entitled *Grammatical Structures Written at Three Grade Levels*. The purpose of his study was 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.

After investigating sentence length as a possible measure on which to base an index, Hunt (1965) concluded that it was not an adequate measure of language development, since learners tend to write run-on sentences connected with numerous *ands*. Furthermore, the definition of 'sentence' posed a problem. Hunt (1965: 20) found that the mean length of clauses and subordination ratio both increased from one grade to another. Since both these lengths increase in successively older learners, he concluded that the complete length of a unit that contains one main clause with all the subordinate clauses attached to it might be a good index of development. Hunt (1965: 21) named such a unit – containing one main clause, together with all the subordinate clauses attached to it – a 'minimal terminal unit' or 'T-unit'. It preserves all co-ordination between words and phrases, as well as subordination expressed by the learner. A T-unit is the shortest possible grammatically allowable sentence into which a passage can be divided, without creating fragments. Most units contain only one main clause, but multi-clause units, like complex sentences, may also occur. Units like compound sentences or compound-complex sentences are divided into separate T-units. A conjunction such as *and* 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). Sentences are divided as follows:

- *I know an attractive girl* / *and her name is Jane* / – two T-units.
- *John woke up, although he was very tired* / – one T-unit.
- *I told him everything of course* / *because I could see that I did not have to hide anything from him* / – two T-units.

Hunt (1965) calculated sentence length, clause length, T-unit length, mean number of clauses per T-unit and mean number of T-units per punctuated sentence. The results of his statistical analyses showed that T-unit length was 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 indicator. Hunt therefore determined that the T-unit was the best measure of syntactic development.

The T-unit was adapted for use with non-native speakers by modifying its definition to incorporate *error-free* T-units rather than just T-units (Gass & Seliger, 2001: 50). The large number of second language studies that have used the T-unit since the 1970s attests to its validity as a measure of linguistic development and maturity. Scott and Tucker (1974) were the first to use error-free T-unit ratio for analysing English written by Arab students. They found that the number of error-free T-units that learners produced increased as learners became more developed, both orally and in writing. Larsen-Freeman and Strom (1977) found that, even without time limitations, more advanced students produced more words and more error-free T-units per composition than younger ones. Farhady (1978) found a significant correlation between the length of T-units and five independent measures of learners' writing ability – organisation, logical development, grammar, writing mechanics and style – that were based on the subjective evaluation of three judges of students' compositions. Larsen-Freeman (1978) continued the

search for an index of development to serve as a yardstick that researchers could use to "expediently and reliably gauge a learner's proficiency in a second language" (Larsen-Freeman, 1978: 439). She investigated the writing of a group of EFL students and determined the average composition length, the average number of T-units, the percentage of error-free T-units and the number of words written per error-free T-unit. Her results showed that all four measures distinguished between groups of learners with different levels of development. Larsen-Freeman (1978: 446) concluded that at least some of the measures worked well to discriminate proficiency differences among groups.

Bardovi-Harlig and Bofman (1989) examined the relationship between syntactic complexity and overall accuracy of the writing of advanced adult foreign language learners. They wanted to describe three aspects, namely morphology, syntax and errors, or "overall accuracy with regard to the distribution of errors" (Bardovi-Harlig & Bofman, 1989: 18). They concluded that any index must take into account both length and accuracy to measure movement toward target-like production. Wolfe-Quintero et al. (1998) report twelve studies that have found a significant relationship between error-free T-units per T-unit and development. Lee (2005: 116) investigated the differences between first and second drafts of written products, and showed that there were significant differences in number of T-units and T-unit length between the two drafts.

All the studies cited above found T-unit length to discriminate among second language learners at different levels of proficiency, and that the mean length of T-units was a useful measure of development towards second language maturity. As learners develop, their writing changes, and they produce more words, add more sentences, use longer T-units, and produce more error-free T-units. The error-free T-unit measure is not always able to take into account the severity of the errors, and the definitions and types of 'error' and the methods of counting errors have varied considerably from one researcher to the next, but T-units remain popular in second language research, because they are easy to identify and are relatively low-inference categories (Mackey & Gass, 2005: 232).

5. Formulas for an accuracy index

A limited number of accuracy indices have been proposed. These date from the 1970s to the mid-1990s. After this, the search for an accuracy index seems to have been abandoned, probably as a result of the growth in qualitative approaches to second language research. We briefly review the major studies here (cf. also Wolfe-Quintero et al., 1998).

Gipps and Ewen (1974) used an Intelligibility Index to measure the intelligibility of the writing of child second language learners. A scale of intelligibility (IntellIndex = sum of points for the intelligibility of each T-unit) was used to analyse T-units, with unintelligible T-units rated 0, partly intelligible ones 1, completely intelligible ones 2, and completely accurate ones 3. The final score was the sum of all the T-unit scores. Intelligibility was correlated with how long the children had been using the L2 in school. A moderate correlation was found.

Evola, Mamer and Lentz (1980) proposed an Error Formula that calculates the correctness of single grammatical morphemes (e.g. conjunctions, pronouns or articles). 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 correlation was found between the error formula and holistic ratings.

Kaczmarek (1980) was interested in establishing whether raters are influenced by the number of errors in an ESL composition when rating a composition. She developed an Error Index that considers all types of errors produced by learners. The index was defined as the number of error-free words, minus the number of errors, divided by the number of error-free words (EFW_E÷EF). The error index correlated highly with holistic ratings of compositions.

Arnaud (1992) proposed an index based on the use of lexical items, and determined the lexical quality of second language writing. 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), and then subtracting twice the number of erroneous lexical words on same length samples (LWT+RWT÷2LWE). Arnaud found a moderate correlation between the lexical formula and scores on discrete item vocabulary tests.

Engber (1995) developed a Lexical Accuracy Index, based on lexical errors (such as word choice, collocation, morphemes and spelling) and variation. His assumption was that ESL writers with a higher level of proficiency would produce more lexical variation with fewer lexical errors, which would result in a higher score. 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). Engber found a moderate correlation between the lexical index and holistic ratings of writing.

These indices are based on various units of analysis (T-units, language errors, specific grammatical morphemes, or lexical items), reflecting their different approaches and purposes. They all involve the calculation of frequencies, ratios and correlations, as well as some kind of judgement, such as the identification of errors. It is essential that this identification be reliable. Only one index formula, that of Gipps and Ewen, is based on the T-unit, which we regard as the most reliable indicator of development in writing. Gipps and Ewen's criterion of intelligibility is not a very reliable one, and we propose to use grammatical accuracy instead.

6. Establishing a new index formula

An index formula can only be validated if there is a criterion (an independent variable). For our study, Grade 12 ESL compositions that had been assessed in the final 2003 matriculation examination were randomly collected and the mark used as independent variable. 216 creative writing compositions were collected from 6 provinces in South Africa. (These provinces formed the study population because we had convenient access to them.) Of these, 112 were Higher Grade and 104 Standard Grade. A composition of 250-300 words had been required. Learners were from different native language backgrounds. Different native languages do not play any significant role in the analysis of T-units (cf. Larsen-Freeman, 1978). These compositions were assessed by experienced markers, and the assessment was also moderated.

The compositions were first divided into five groups – poor, fair, average, good and excellent – according to the mark allocated by the marker. Poor compositions (0%-19%) were divided into Group 1 (N=29), fair ones (20%-39%) into Group 2 (N=57), average ones (40-59%) into Group 3 (N=64), good ones (60%-79%) into Group 4 (N=48), and excellent ones (80%-100%) into Group 5 (N=18).

Each essay was divided into T-units by two researchers to ensure validity and reliability. Each T-unit was analysed for grammatical accuracy. Spelling errors were ignored in this particular

analysis. There were so many spelling errors in the compositions that we had to make a decision in this regard. Spelling is not regarded as part of syntax (cf. Huddleston & Pullum, 2005: 6), and we decided to focus on syntax only.

The dependent variables were: the average length of the compositions, the average number of T-units per composition, the average number of words per T-unit, the average percentage error-free T-units per composition and the average percentage of words per error-free T-unit (defined as the total number of words in error-free T-units units divided by the length of the composition times 100). An Analysis of Variance test (ANOVA) (cf. Hatch & Farhady, 1982: 151; Seliger & Shohamy, 1989: 234) was done to compare the mean characteristics of the compositions. This was followed by a multiple comparison using the method proposed by Tukey (cf. Miller, 1981: 37-48) in order to determine which of the five groups differed significantly for each variable/characteristic. Both these tests were conducted at a 5% level of significance.

The best predictors of the mark were then selected by means of three steps: a correlation procedure, a stepwise regression analysis and a discriminant analysis. The correlations were drawn to determine the relationship between the variables. The stepwise regression was done in order to identify the most important variables and explain to what extent the data could predict the mark (cf. Hatch & Farhady, 1982: 234-235). The discriminant analysis was then done in order to determine which characteristics would make up the best index formula (cf. Seliger & Shohamy, 1989: 226)

7. Results

Table 1 reports the results of each of the dependent variables for the five groups as well as the mean for the whole sample. The ANOVA indicated significant F values for the length of the compositions, the number of words per T-unit, the percentage error-free T-units and the percentage words per error-free T-unit.

Variable	Group 1	Group 2	Group 3	Group 4	Group 5	Mean	ANOVA F value
Composition length in words	222.14	279.86	287.25	266.27	343.83	276.61	6.44 $p < 0.0001$
Number of T-units per composition	23.76	29.91	29.29	26.85	30.83	28.27	Not significant
Number of words per T-unit	9.64	9.86	10.18	10.36	11.85	10.18	2.92 $p < 0.0024$
% Error-free T-units	21.06	34.30	40.10	51.06	69.75	40.91	21.76 $p < 0.0001$
% words per error-free T-unit	17.02	29.63	34.33	46.71	63.41	35.94	21.17 $p < 0.0001$

Table 1: Results of the T-unit measures for the five groups.

Tukey's Studentized test was used to determine which of the measures clearly distinguished between the five groups. The results are reported in Table 2. There is no clear distinction between groups with the same superscript letter.

Variable	Group 1	Group 2	Group 3	Group 4	Group 5	Mean	ANOVA F value
Composition length in words	222.14	279.86	287.25	266.27	343.83	276.61	6.44 p< 0.0001
Number of T-units per composition	23.76	29.91	29.29	26.85	30.83	28.27	Not significant
Number of words per T-unit	9.64	9.86	10.18	10.36	11.85	10.18	2.92 p< 0.0024
% Error-free T-units	21.06	34.30	40.10	51.06	69.75	40.91	21.76 p< 0.0001
% words per error-free T-unit	17.02	29.63	34.33	46.71	63.41	35.94	21.17 p< 0.0001

Table 1: Results of the T-unit measures for the five groups.

In Table 2 the group means with different superscript letters differ mutually; for example, composition length and average number of word per T-unit clearly distinguish learners in Group 5 from all other groups.

Variable	Group 1 Poor	Group 2 Fair	Group 3 Average	Group 4 Good	Group 5 Excellent
Length	222.14^c	279.86^b	287.25^b	266.27^{bc}	343.83^a
Words/T-unit	9.63^b	9.78^b	10.18^b	10.28^b	12.0^a
% EFT	21.065^d	34.30^c	40.09^{bc}	51.06^b	69.75^a
% words/EFT	15.6^d	25.7^{cd}	35.7^{bc}	47.0^b	62.8^a

Table 2: Results of the Tukey tests conducted for each variable and for the five groups.

8. Selecting the best predictor of development

A correlation procedure was conducted to see how the dependent variables correlate with the mark (the independent variable). These correlations are reported in Table 3.

	Number of T-units	Length	Number of words per T-unit	% words per EFT	%EFT
Mark	0.1004	0.2603	0.1929	0.5153	0.5229
	$p=0.1424$	$p<0.0001$	$p=0.0045$	$p<0.0001$	$p<0.0001$

Table 3: Correlation between the mark and the dependent variables.

The correlations with the mark range from moderate to high (cf. Wolfe-Quintero et al., 1998: 38), except for the number of T-units.

In order to establish an index, one needs to determine which variable or combination of variables would best predict the mark (the independent variable or criterion), i.e. the level of syntactic development. The Stepwise Regression Analysis and the Discriminant Analysis were therefore conducted.

The Stepwise Regression included the STEPWISE and REG procedures of the SAS program. 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 (%EFT) per composition may be the best predictors of the independent variable. The percentage EFTs proved to be the most significant of the two predictors with $F(2; 211) = 99.63$ with $p < 0.0001$, with the value for length being $F(4,211) = 59.96$ with $p < 0.0001$. The coefficient of determination (R^2) supports these results, with $R^2 = 0.363$ for the product of both variables, whereas $R^2 = 0.320$ for the percentage EFTs, which was the more significant of the two variables taken separately.

The Stepwise Regression results identified length of composition and the percentage EFTs as the most important variables. The Stepwise Regression showed that these two characteristics were good predictors of the mark. The combination of the two predictors, (Length) x (%EFT), correlated with the mark at 0.589. According Wolfe-Quintero et al. (1998: 38), this is a moderate to high correlation. This combination, (Length) x (%EFT), correlates better with the mark than the two variables taken separately. One could therefore base an index of development on these two predictors. However, if the percentage of EFTs is used, the values would become very big. The accuracy ratio of EFTs is calculated by dividing the number of EFTs by the total number of T-units. By using this ratio (i.e. EFT/T) instead of the percentage EFTs, the values are more practical to work with. Therefore, the error-free T-unit ratio instead of the percentage EFTs was used.

To verify the selection of these two variables, all possible subsets of regression were performed. The process produced essentially the same results.

Length of composition and the EFT/T ratio therefore seem to be the best predictors of the level of development in writing. In other words, the product of these two characteristics would constitute the best index to determine the level of syntactic development that this particular group of subjects has achieved in writing.

The Discriminant Analysis entailed determining the accuracy with which compositions were classified into the five 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 new classification, using cross validation (cf. Huberty, 1994), was based on length and %EFT, and the results were then compared with the original division to determine the accuracy with which the compositions were initially grouped.

The results of the DISCRIM procedure appear in Table 4.

Classification by means of Discriminant Function

From Group	Good	Poor	Total
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%

(Original Classification)

Table 4: Number of observations and percentage classified into groups.

Table 4 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 results of the Discriminant Analysis supported the result of the Stepwise Regression.

9. The proposed index formula

The results of the Stepwise Regression and the Discriminant Analysis show that the best indicator of the level of development is a combination of the length of the composition and the error-free T-units ratio (EFT/T). The length of compositions and the percentage of error-free T-units (EFT ratio) both increase for the five consecutive groups (cf. Table 1). In other words, subjects who scored higher marks both wrote longer essays and had a higher percentage of error-free T-units.

The multiple correlation from the Stepwise Regression produced a correlation of 0.589, with $R2 = 0.364$. The results of the Discriminant Analysis reinforce the suggestion that the product of length and accuracy ratio 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 (Length) x (EFT/T), resulted in a division similar to the original classification. The two characteristics combined, (Length) x (EFT/T), produced a moderate to high correlation (cf. Wolfe-Quintero et al., 1998: 38) with the mark of 0.602 with $R^2 = 0.363$.

The two values for the squares (R^2) are also very similar.

Thus, the suggested index formula, (Length) x (EFT/T), is supported for two reasons:
- Length and the EFT accuracy ratio are good predictors of the mark (cf. Stepwise Regression and Discriminant Analysis);
- The product of the two characteristics correlates with the mark at 0.602.

10. Application of the index formula

We used this index formula to calculate the level of development in writing of the group of Grade 12 ESL learners. This resulted in a numerical value that can be regarded as the level of syntactic development of this group of subjects.

For Higher Grade, an index value of 130.26 was calculated; the index value for Standard Grade was 100.41, and the level of the group as a whole was 115.89. Index values for each of the six provinces were also calculated. The results are presented in Table 5.

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. Province D produced the highest value, 139.57, which is higher than that of the Higher Grade group.

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 5: Index values for six provinces.

11. Conclusion

The index formula proposed in this article can be used in second language research to determine syntactic development in writing. The process depends on a valid and reliable identification of T-units and error-free T-units, but results in an objective measure of development that can be expressed as a numerical value. The formula can be used to keep track of standards, such as those achieved in the matriculation examination, to establish whether they are improving or not. The index values obtained in this study can be used as benchmark with which to compare subsequent analyses of Grade 12 learners' writing. The formula can also be used to trace the syntactic development of learners over a period of time (e.g. in secondary school).

In addition, the T-unit has pedagogical value, and can be used as a core concept in the teaching of writing. Learners should be taught to produce longer T-units. This will involve them in the production of subordinate clauses, and make their writing more mature and sophisticated. As Lee (2005) indicates, there can be a marked difference between the number of T-units and T-unit length in a first draft and those in a second draft. Of course, length does not automatically imply accuracy, and teachers must pay particular attention to the development of accuracy in writing. Writing demands higher degrees of accuracy than oral production, and remains an essential component of competence in a second language. Judging from the quality of the compositions analysed for this study, accuracy in writing seems to be an area that is neglected at present. But it remains an essential part of competence in a second language.

REFERENCES

Arnaud, P.J.L. 1992. Objective lexical and grammatical characteristics of L2 written compositions and the validity of separate-component tests. Pp. 133-145 in Arnaud, P.J.L & Bejoint, H. (eds.), *Vocabulary and Applied Linguistics*. London: Macmillan.

Bardovi-Harlig, K &Bofman, T. 1989. The attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition* 11: 7 - 34.

Brooks, N. 1964. *Language and language learning: Theory and practice*. New York: Harcourt, Brace and World.

Engber, C.A. 1995. The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing* 4: 139 - 156.

Evola, J., Mamer, E. & Lentz, B. 1980. Discrete point versus global scoring for cohesive devices. Pp. 177-181 in Oller, J. W. & 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. Unpublished master's dissertation, University of California, Los Angeles.

Gass, S.M. & Selinker, L. 2001. *Second language acquisition: An introductory course*. (2nd ed.). Mahwah, N.J./London: Lawrence Erlbaum.

Gipps, C. & Ewen, E. 1974. Scoring written work in English as a second language. *Educational Research* 16:121-125.

Hatch, E. & Farhady, H. 1982. *Research design and statistics for Applied Linguistics*. Rowley, Mass.: Newbury House.

Hunt, K. 1965. *Grammatical structures written at three grade levels*. Urbana: National Council of Teachers of English.

Huberty, C.J. 1994. *Applied discriminant analysis*. New York: Wiley and Sons.

Huddleston, R. & Pullum, G.K. 2005. *A student's introduction to English grammar*. Cambridge: Cambridge University Press.

Kaczmarek, C. M. 1980. Scoring and rating 'essay tasks'. Pp. 151-159 in Oller, J.W. & Perkins, K. (eds.) *Research in language testing*. Rowley, Massachusetts: Newbury House.

Larsen-Freeman, D.E. 1978. An ESL index of development. *TESOL Quarterly* 12: 439-448.

Larsen-Freeman, D.E. & Strom, V. 1977. The construction of a second language acquisition index of development. *Language Learning* 27: 123-134.

Lee, Y.J. 2005. A summary of construct validation of an English for Academic Purposes placement test. *Spaan Fellow Working Papers in Second or Foreign Language Assessment* 3: 113-131.

Lorenz, G. 1999. *Adjective intensification – learners versus native speakers*. Amsterdam: Rodopi.

Mackey, A. & Gass, S. 2005. *Second language research: Methodology and design*. Mahwah, N.J.: Lawrence Erlbaum.

Miller, R.G. 1981. *Simultaneous statistical inference*. New York: McGraw Hill.

Seliger, H.W. & Shohamy, E. 1989. *Second language research methods*. Oxford: Oxford University Press.

Scott, M. & Tucker, G.R. 1974. Error analysis and English language strategies of Arab Students. *Language Learning*, 24: 69 - 97.

Thomas, M. 1994. Assessment of L2 proficiency in second language acquisition research. *Language Learning* 44(2): 307 - 336.

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.].

Wolfe-Quintero, K., Inagaki, S. & 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.

ABOUT THE AUTHORS

Karien Hattingh *is an EFL tutor at the Potchefstroom campus of North-West University and specialises in second language acquisition and methodology.*

Ms K. Hattingh
School of Languages
North-West University
Private Bag X6001
Potchefstroom
2520

Johann van der Walt *is Head of the Centre for Language Practice and Professor of English at the Potchefstroom campus of North-West University. He specialises in second language acquisition, methodology and assessment.*

Prof. J.L. van der Walt
Centre for Language Practice
School of Languages
North-West University
Private Bag X6001
Potchefstroom
2520