

# CEFR 基準特性に基づくチェックリスト方式による英作文の採点可能性

## Exploring the Possibility of Assessing English Writing Based on the Checklist of CEFR Criterial Features

根岸雅史

Masashi NEGISHI

東京外国語大学

*Tokyo University of Foreign Studies*

### Abstract

Error-count method, which is often adopted for assessment of EFL writing in Japan, can sometimes produce unfair results to candidates who take risks in using difficult grammatical structures. Research from the English Profile Programme, however, shows that advanced learners tend to use those structures. The purpose of this paper is to explore the possibility of assessing English writing based on such information. 900 Japanese university students participated in this survey. The CEFR levels of their English are assumed to range from A2 to B2+. Three pieces of writing were collected from each participant. Their writing was scored based on the checklist of CEFR criterial features of English drawn from English Profile Programme research with some necessary modifications. The data was analysed by an IRT item analysis programme. The results showed that 42 criterial features were effective in discriminating this level of English learners, and that criterial features of great difficulty are mainly complex sentence structures. The participants' ability (Theta value) correlated moderately with the results of human-rating. This would be partly because the assessment scale includes reference to not only linguistic features but also some other aspects of writing. However, the checklist of CEFR criterial features might become a more accurate measure of linguistic aspects in English writing, if we could include information about the condition of the appearance of each criterial feature. Some implications are also proposed for teaching EFL writing.

### Keywords

CEFR, English Profile Programme, Criterial Features, Writing Assessment

#### 1. 英作文の採点における言語的正確さの評価

英作文は、どのように採点されるであろうか。英作文の採点には、「全体的採点法」「分析的採点法」「減点法」があるが、日本においては、このうち「減点法」が採用されることがよくある。しかしながら、この方法は「和文英訳」の場合には、ある程度機能するかもしれない

が、いわゆる自由作文の採点に当たっては、必ずしもうまく機能しない。「自由作文」では書く分量や内容などが規定されていないために、「減点法」で採点した場合、たくさん書いて誤りが散見されると、わずかな量の英文を正確に書いた作文より低く評価され、時に0点を下回ってしまうことなどもある。

このため、自由作文の採点では、「全体的採点法」または「分析的採点法」が推奨されている。これらの採点法では、様々な観点からそれぞれの作文が評価されることになるが、ほとんどの場合、言語的正確さの要素が、その採点基準に含まれている。言語的正確さは、文法の正確さと語彙の正確さから成ることが多いが、「言語使用 (language use)」として、文法や語法を含むような場合もある。

いくつかの代表的な採点基準の例を見てみる。本稿では、言語的正確さから見た英作文評価の可能性を探ることから、「分析的採点法」の「文法」や「言語使用」などの評価基準に焦点を当てるが、「全体的採点法」においても、ほとんどの場合、これらの要素が含まれる。

次の評価基準は、TEEP (Test of English for Educational Purposes) attribute writing scales (Weir, 1990) から取ったものであるが、ここでは誤りの頻度のみが問題となっている。

#### Grammar

0. Almost all grammatical patterns inaccurate.

1. Frequent grammatical inaccuracies.
2. Some grammatical inaccuracies.
3. Almost no grammatical inaccuracies.

次は Hughes (2003) に紹介されている採点基準である。ここでは誤りの頻度に加え、その質に言及があるが、それは読み手としての採点者の負担の度合いによっており、いわゆる「グローバル・エラー」に対しては、厳しい評価が下されている。

#### Grammar

6. Few (if any) noticeable errors of grammar or word order.

5. Some errors of grammar or word order which do not, however, interfere with comprehension.
4. Errors of grammar or word order fairly frequent; occasional re-reading necessary for full comprehension.
3. Errors of grammar or word order frequent; efforts of interpretation sometimes required on reader's part.
2. Errors of grammar or word order very frequent; reader often has to rely on own interpretation.
1. Errors of grammar or word order so severe as to make comprehension virtually impossible.

英作文の採点においては、Jacobs et al. (1981)の scoring profile(本稿では、スペースの関係で載録はしない)が有名であるが、ここでは、誤りの頻度と通じる度合いのほかに、文構造の複雑さや様々な文法項目に関する誤りの出現頻度への言及がある。ただし、上位の3段階では、agreement, tense, number, word order/function, articles, pronouns, prepositions が繰り返されており、これ以外のレベルに固有な項目は、FAIR TO POOR における negation と fragments, run-ons, deletions だけである。

しかしながら、これらの採点基準の現実的運用には、やっかいな問題がある。それは、「易しい英語を用いて正確に書かれた文章」と、「難しい英語を用いてはいるが、誤りも含む文章」とでは、どちらを高く評価すべきかという問題である。日本では、減点法の使用が多いためか、教室では一般に「簡単な英語を用いて正確に書く」ように指導されることが多い。しかしながら、「難しい英語」を用いて、時に間違えてしまうのと、「易しい英語」しか用いないで、誤りを犯さずに書くのとどちらが評価されるべきかは、簡単には判断できない。

ここでは従来の採点方法の問題点を具体的に考察するために、次の作文を見てみよう。

A.

I enjoyed the school trip. We went to Okinawa. We went there by plane. We visited Shurijo. We stayed in Okinawa for three days.

B.

I enjoyed the school trip to Nagano. Because I had a good time skiing. I want to ski but I have never skied. Our teachers teach us how to ski. I'm a good skier next winter.

Aは易しい英語を用いて正確に書いているのに対して、Bは時制などの誤りが散見される。特に、Bの下線部の箇所は時制に誤りがあるため、文章全体としての意味理解に支障をきたすであろう。

では、これらの作文は、文法的な観点からの採点では、どのような評価を受けるであろうか。上掲のような採点基準であれば、誤りのないAは高く評価され、誤りが頻出しているBは低く評価される。しかしながら、Aは、使っている文型も第1文型と第3文型のみで、そのほかの文法項目の使用も限定的であるのに対して、Bは、第1文型、第2文型、第3文型、第4文型が用いられており、現在完了やhow to ...なども使われている。また、becauseの節を単独ではあるが、用いている。こうして見てみると、どちらが文法能力の発達段階が進んでいるのか、容易には判断しがたい。

## 2. 第2言語習得研究からの示唆

### 2.1 「言語的特徴の難易度」とは何か

本稿では、スペースの関係から、「言語的特徴」のうち「文法的正確さ」を中心に考察する。「英語の学習項目」の中で、何が言語的に難しく、何が易しいかを定めることは容易ではない。学校英語教育では、一般に学習が容易と思われるものから難しいと思われるものへと並んでいると考えられている。現在形は過去形や現在進行形より易しいと考えられるから

こそ、ほとんどの教科書では現在形が最初に配列され、授業でも教えられている。

確かに、先に導入される学習項目は、言語形式上の複雑さは少ない。しかし、だからといって習得が容易とは限らないことは、これまでの第2言語習得研究の結果から明らかである。たとえば、三単現の *s* や定冠詞・不定冠詞は中学1年で教えているが、日本人学習者はなかなか正しく使えるようにならない。はたして、三単現の *s* や定冠詞・不定冠詞は、「易しい学習項目」なのか、「難しい学習項目」なのか、にわかには答えは出せない。

## 2.2 English Profile Programme からの示唆

この点に関して、English Profile Programme から見えてくるものがある。ヨーロッパの共通言語参照枠組みである CEFR は、周知の通り、特定の言語への言及がない。そこで、このプログラムは、ケンブリッジ大学 ESOL を中心に、各 CEFR レベルの英語の基準特性 (criterial features) を見つけていこうとしている。

この目的のために、English Profile Programme では、ケンブリッジ ESOL が実施してきた過去の様々なテスト解答データ(作文データ中心)から成る学習者コーパスを作成し、それぞれの CEFR レベルの学習者の基準特性を同定している。それらの結果は、Hawkins and Buttery (2009), Salamoura and Saville (2009), Salamoura and Saville (2010) などに明らかにされている。ここからは、たとえば、三単現の *s* の習得にはかなり時間がかかり、定冠詞・不定冠詞は、第1言語に冠詞のない学習者は、その習得はかなり遅いことがわかっている。つまり、日本の中学校で指導しているような文法項目も、自ら正しく使えるようになるにはかなりの時間がかかることがある、ということがわかってきたのだ。

## 3. 研究

### 3.1 研究設問

そこで、本稿では、次のような研究設問を設定する。

1. 日本人英語学習者の英語力の弁別に有効な基準特性項目は何か。
2. それらの基準特性項目の難易度はどうなっているか。
3. 基準特性項目の使用の有無をもとにした能力推定値と人間による採点結果には、どのような相関があるか。

### 3.2 研究ツール

English Profile Programme のデータ収集において用いられた3つの作文タスクを用いた。第1のタスクは、外国からの友達を週末自宅に招く E メールを書くもの、第2のタスクは、宇宙からやってきたエイリアンに自分の町を案内することになったとして、何が起こったかを書くもの、第3のタスクは、田舎の生活の長所と短所を書くもの、である。

### 3.3 参加者

日本の国立大学の言語専攻の1年生900名。これらの参加者の英語の CEFR レベルは、

後述するライティング・タスクの採点結果によれば、下記のように分布している。なお、このレベルは、3つのタスクのうちのもっとも高得点の結果がその参加者の CEFR の潜在的レベルを表していると考え、その分布を示す。なお、CEFR レベルの A1, A1+, A2, A2+, B1, B1+, B2, B2+ は、それぞれ1～8に置き換えてある。

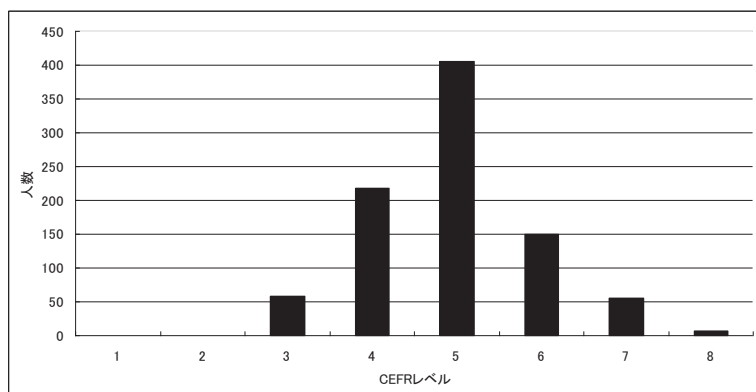


図1 参加者の英語の CEFR レベルの分布

### 3.4 データ

上述のツールを用いて収集された手書きの英作文データを参加者ごとにワープロ入力したものを分析対象とした。綴りや文法の誤りはそのまま入力されている。

なお、これらの英作文には、それぞれ CEFR のレベル判定の結果が付与されている。採点はトレーニングを受けたアドバイザー2名により行われている。1名は中心となって採点基準などの開発を行っている第1採点者、もう1名は第1採点者によるトレーニングを受けた第2採点者で、1レベル以上の不一致の場合は、別の第1採点者が第3採点者として入り、協議により最終判定をした。

採点基準は CEFR に準拠しており、各タスクに応じて具体的に記述をしたものを用い、最初の2008年度パイロットデータのケンブリッジ公式採点官による準公式採点に加えて、昨年度には抜き出しで、CEFR の専門家である英国ベッドフォードシャー大学の Tony Green 博士に確認をしてもらうなどして、タスクごとの採点サンプルの作成を行った。

### 3.5 基準特性とその判断

English Profile Programme において基準特性となっていると考えられる項目は、以下の通りである。

1. 第1文型, 2. 第3文型, 3. I know/think that ..., 4. I want/would like to ... , 5. 直接 wh- 疑問文, 6. 第4文型, 7. 過去分詞による後置修飾, 8. 現在分詞による後置修飾, 9. S V O to 不定詞, 10. S V O ing, 11. it ... that , 12. He said to me (that), 13. 関係代名詞(所有格), 14. 間接 wh- 疑問文, 15. 擬似分裂文タイプ I (wh- 疑問詞が目的語となるタイプ), 16. 不定詞付きの間接 wh- 疑問文, 17. would rather,

18. had better, 19. 主節の後ろに来る分詞構文(現在分詞), 20. 主節の前に来る分詞構文(現在分詞), 21. it ... to, 22. 擬似分裂文タイプ II (wh- 疑問詞が主語となるタイプ), 23. She told me that, 24. 第5文型, 25. may 可能性, 26. may 許可, 27. might 可能性, 28. might 許可, 29. can 許可を暗示する能力, 30. can 可能性, 31. must 義務, 32. must 必要性, 33. should アドバイス, 34. should 可能性, 35. can 能力, 36. could 能力以外, 37. would (would like to 以外), 38. 関係代名詞 制限用法(S-S, 先行詞が文中で「主格(S)」, 関係詞が関係詞節中で「主格(S)」であることを表す。以下, 同様の表記。), 39. 関係代名詞 制限用法(S-O), 40. 関係代名詞 制限用法(O-S), 41. 関係代名詞 制限用法(O-O), 42. 関係代名詞所有格 非制限用法, 43. 関係代名詞 非制限用法(S-S), 44. 関係代名詞 非制限用法(S-O), 45. 関係代名詞 非制限用法(O-S), 46. 関係代名詞 非制限用法(O-O), 47. 関係副詞 制限用法, 48. 関係副詞 非制限用法, 49. 受動態, 50. 完了形, 51. 定冠詞, 52. 助動詞 +have+ 過去分詞, 53. 不定冠詞

これらの項目は, 基本的には English Profile Programme における先行研究をもとに, 日本の英語教育の実態に合わせるために, 若干の修正を行った。なお, もともとの学習者コーパスのレベルが, ケンブリッジ大学 ESOL の試験(KET, PET, FCE, CAE, CPE)によっているために, これらの特性は基本的には A2-C2をカバーしていると考えられる。

英語の CEFR レベルの基準特性の判断は, 筆者の講義及びワークショップを受けた英語教育学を専攻する大学院生8名により行われた。まず, 100名分の英作文データにおける基準特性の有無を全員で判断し, 問題点などを話し合い, 方式を決定した。その後, ワープロ入力された英作文データを1人の受検者につき2人の判断者が, 基準特性の有無をチェックリストに記入していった。特性が正しく使われている場合には, その項目に1を与え, 使われていない, または, 正しく使われていない場合には, 0を与えていった。2人の間で判断が異なった場合は, 問題点を全体で検討し, 統一した判断を最終的には下した。

### 3.6 データの分析方法

データの分析では, それぞれの項目の出現の有無をテストの項目得点として扱い, テスト項目分析ソフト *RASCAL* にかけた。さらに, 基準特性の出現回数に基づく能力値 Theta と CEFR レベルの関係を, グラフで示し, 相関係数を算出した。その際, CEFR レベルの A1, A1+, A2, A2+, B1, B1+, B2, B2+ はそれぞれ1~8に置き換え, 参加者の CEFR レベルは, 欠損値のあるデータを削除したあと, 3つのタスクの評価の合計により求めた。

### 3.7 結果

この手法による, 各項目の項目難易度と標準誤差は, 以下の通りとなった。

表1 RASCAL による基準特性の項目難易度と標準誤差

特性番号	基準特性	項目難易度	標準誤差
2	第3文型	-8.278	0.924
1	第1文型	-7.595	0.659
53	不定冠詞	-5.992	0.303
51	定冠詞	-5.896	0.290
30	can 可能性	-3.910	0.121
5	直接 wh- 疑問文	-3.038	0.090
4	I want/would like to ...	-2.662	0.081
6	第4文型	-2.243	0.075
49	受動態	-2.107	0.073
3	I know/think that ...	-2.070	0.073
14	間接 wh- 疑問文	-1.476	0.070
9	S V O to 不定詞	-1.400	0.069
50	完了形	-1.215	0.070
40	関係代名詞 制限用法(O-S)	-0.847	0.071
21	it ... to	-0.807	0.072
35	can 能力	-0.531	0.074
41	関係代名詞 制限用法(O-O)	-0.488	0.075
16	不定詞付きの間接 wh- 疑問文	-0.465	0.075
36	could 能力以外	-0.115	0.081
38	関係代名詞 制限用法(S-S)	-0.108	0.081
25	may 可能性	-0.036	0.083
37	would (would like to 以外)	-0.009	0.083
24	第5文型	0.004	0.083
23	She told me that	0.088	0.085
33	should アドバイス	0.139	0.086
11	it ... that	0.316	0.091
47	関係副詞 制限用法	0.470	0.095
8	現在分詞による後置修飾	0.488	0.096
7	過去分詞による後置修飾	0.592	0.099
39	関係代名詞 制限用法(S-O)	0.611	0.100
27	might 可能性	0.703	0.103
31	must 義務	0.802	0.107
32	must 必要性	0.908	0.111
19	主節の後ろに来る分詞構文(現在分詞)	0.945	0.113
20	主節の前に来る分詞構文(現在分詞)	1.196	0.124
45	関係代名詞 非制限用法(O-S)	1.275	0.128
29	can 許可を暗示する能力	1.396	0.135
12	He said to me (that)	1.888	0.166
15	擬似分裂文タイプ I	2.007	0.175
48	関係副詞 非制限用法	2.038	0.178
10	S V O ing	2.071	0.181
34	should 可能性	2.071	0.181
43	関係代名詞 非制限用法(S-S)	2.139	0.186
13	関係代名詞(所有格)	2.291	0.200
52	助動詞 +have+ 過去分詞	2.333	0.204
46	関係代名詞 非制限用法(O-O)	2.468	0.217
18	had better	2.808	0.254
22	擬似分裂文タイプ II	2.877	0.263
26	may 許可	3.318	0.325
17	would rather	4.121	0.482
42	関係代名詞所有格 非制限用法	4.121	0.482
44	関係代名詞 非制限用法(S-O)	4.804	0.676
28	might 許可	-Deleted-	

RASCAL の結果から、便宜的に標準誤差が0.250以下の11個の項目を削除した。これらの削除項目のうち、難易度の低いものは、第3文型、第1文型、不定冠詞、定冠詞、難易度の高いものは、had better、擬似分裂文タイプ II、may 許可、would rather、関係代名詞所有格 非制限用法、関係代名詞 非制限用法 (S-O)、であった。また、might 許可は全く使用されなかったために、分析から削除された。

基準特性の出現回数に基づく能力値 Theta と3作文の合計点による作文評価との関係を表すグラフは以下の通り。ピアソンの積率相関は0.36である。

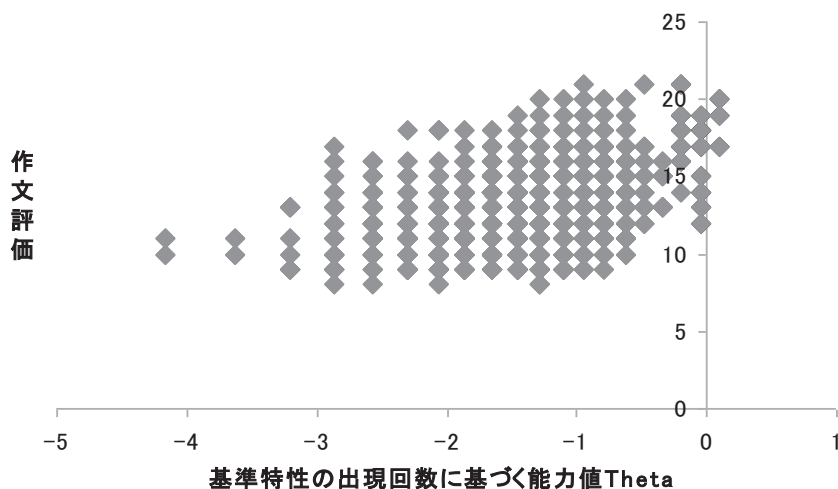


図2 基準特性の出現回数に基づく能力値 Theta と作文評価の関係

作文の CEFR レベルと能力値 Theta との関係は、以下の通りである。

表2 作文の CEFR レベルと能力値 Theta との関係

CEFR レベル	能力値 Theta	対応する基準特性数
A2+ < CEFR ≤ B1	-3.21 ~	4
B1 < CEFR ≤ B1+	-2.87 ~	5
B1+ < CEFR ≤ B2	-1.46 ~	11

#### 4. 考察

RASCAL の結果をもとに、標準誤差が0.250以下の項目を削除した。こうして残った基準特性を見てみると、今回調査対象となった英語学習者にとって弁別力の高い基準特性は、関係詞や分詞による名詞の後置修飾や分詞構文による文修飾に加え、様々な法助動詞である。より詳しく見ると、関係詞節が主語になっている場合 (S-S, S-O タイプ) の方が、目的語に由来している場合 (O-S, O-O タイプ) より、難易度は高い。ただし、関係詞のすべてのタイプが弁別力があるわけではなく、関係詞の非制限用法や擬似分裂文タイプやいくつかの法助動詞はほとんど用いられておらず、弁別力はない。文型で言えば、第1文型や第3文型は、すべての学習者が正しく使ってきており、弁別力は持たないために削除されたが、



第4文型や第5文型やSVO to 不定詞、SVO ing などの発展的な文型は、この集団の下位の学習者の弁別には有効であることがわかる。

難易度の高い項目は、構造が複雑なものである。しかし、見方を変えれば、これらは中学3年から高校にかけて学習した項目であるとも言える。つまり、中学の後半から高校にかけて学習した項目が大学に入って正しく使えるようになっているかどうか、今回の調査対象となったレベルの学習者の能力を弁別している。

文法能力を考える場合、宣言的知識と手続き的知識の違いが問題となる。つまり、一口に文法知識を持っているといっても、それが使えるかどうかでは大きな違いがあるということだ。本稿での示唆は、さらに、一口に「使える」といっても、実際に「使う」ということと必ずしも同義ではないということだ。今回の調査対象者は、高校で学んだような項目も使おうと思えば使えるかもしれない。しかし、問題は、意味に意識を向けて、首尾一貫した文章を書く中でこれらの項目を使ってくるのかどうかである。ここから英作文指導に目を向ければ、基準特性という観点から英作文を見直し、今回明らかになった基準特性の難易度などを参考にしながら、それぞれの学習者が次の使用の目標とするような項目を提示し、積極的な使用を促すという指導があってもいいかもしれない。

これらの基準特性の有無から能力を推定した能力値 Theta と人間の作文評価結果との相関係数は、必ずしも高くない。ただし、表2からわかるように、あるCEFRレベルに達すると一定数の基準特性を使っている(一定の数の基準特性を使えば、必ずあるCEFRレベルに達するとは言えないが)。具体的には、A2+ からB1+ までは、基準特性の数はさほど大きな違いがないが、B1+ を超えると基準特性の数は急激に増加しているのがわかる。

確かに人間の作文評価は、文法的正確さに関する情報だけを用いて行われているわけではない。しかしながら、基準特性に関する情報から能力を推定する精度をもう少し上げることが可能かもしれない。たとえば、今回の分析では、受動態は比較的易しい基準特性となったが、これは **be surprised** といった形を受動態と判断したためかもしれない。このような、ある意味で定型的(formulaic)な言語形式の使用があったからといって、もう少し操作性の高い受動態の使用が可能とは限らないだろう。また、定冠詞や不定冠詞も1度正しく使えたからといって、様々な文脈における使い分けや無冠詞の選択ができていたとは限らない。さらに、助動詞に関する様々な基準特性の出現は、発達段階に影響を受ける部分もなくはないとは言えるものの、タスクの有り様によって、その使用が必ずしも必須であるとは限らない。したがって、これらの観点などを考慮することにより、基準特性の出現の有無に基づく能力値 Theta の精度を上げることも可能かもしれない。

## 5. 結果

今回の調査参加者のCEFRレベルは、作文評価の結果から、B1を中心にA2からB2+まで分布していると推定される。本調査で用いたA2～B2の基準特性項目のうち、A2のうちのいくつかの易しい項目とB2以上のうちのいくつかの難しい項目を除いた、多くの項目はこれらのレベルの学習者の能力の弁別に有効であった。これらの有効な基準特性項目は、文構造が複雑になり、これに伴う言語操作が複雑になると、難易度が高まることがわかった。

これらの基準特性の使用の有無をもとにした能力推定値と人間による採点結果には、ゆるやかな相関があった。人間による作文の採点では、言語的な正確さ以外にも様々な観点が含まれているので、これはある程度当然である。しかしながら、単純な正用法の出現の有無のチェックでなく、それぞれの基準特性の出現条件などに関する情報を加味するなどすることで、それぞれの基準特性項目の精度をさらに上げることができる可能性がある。今後、このような新たな方法を模索していくことが必要となるであろう。

#### 参考文献

- Hawkins, J. A., & Buttery, P. (2009). Using learner language from corpora to profile levels of proficiency: Insights from the English Profile Programme. In L. Taylor & C. J. Weir (Eds.), *Language Testing Matters: Investigating the wider social and educational impact of assessment* (pp. 158-175). Cambridge: Cambridge University Press.
- Hughes, A. (2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Jacobs, H. L., Zingraf, S. A., Wormuth, D. R., Hartfield, V. F., & Hughey, J. B. (1981). *Testing for ESL Composition: A practical approach*. Rowley, Mass: Newbury House.
- RASCAL for Windows95 (Version 3.50) [Computer software]. St. Paul, MN: Assessment Systems Corporation.
- Salamoura, A., & Saville, N. (2009). Criterial features across the CEFR levels: Evidence from the English Profile Programme. *Research Notes*, 37, 34-40.
- Salamoura, A., & Saville, N. (2010). Exemplifying the CEFR: Criterial features of written learner English from the English Profile Programme. In I. Bartning, M. Martin & I. Vedder (Eds.), *Communicative Proficiency and Linguistic Development: Intersections between SLA and language testing research* (pp. 101-132). Eurosla.
- Weir, C. J. (1990). *Communicative Language Testing*. Hemel Hempstead: Prentice Hall.