

<資料>

英語の口頭技能に関する自己評価の開発

酒井英樹 信州大学教育学部言語教育講座

和田順一 聖ヨゼフ学園中学・高等学校

Development of a Tool for Self-Assessment of L2 Oral Skills

SAKAI Hideki: Language Education, Faculty of Education, Shinshu University

WADA Junichi: St. Joseph's Junior and Senior High School

This study examined the reliability and validity of a questionnaire as a tool for self-assessment of L2 oral skills. The authors developed a 6-point Likert questionnaire consisting of 30 items covering five areas (listening comprehension, fluency, vocabulary, pronunciation, and grammar) based on the FLOSEM by Padilla and Sung (1999). The results showed high reliability coefficients and high intercorrelations for the scores of those subskills. This study also examined the reliability and validity of an abbreviated set consisting of 15 items and found similar results for the shorter questionnaire. This validates the usefulness of the 15-item version. Finally, this study discussed the face validity of the questionnaire.

【キーワード】 英語口頭技能 自己評価 質問紙票 信頼性 妥当性

1. はじめに

直接的に能力や技能を測定する伝統的な評価方法に対して、ポートフォリオ評価、自己評価、相互評価などの方法を代替評価 (alternative assessments) と呼ぶ (e.g., Bulter & Lee, 2006, p. 506). 本研究では代替評価の1つとして自己評価に焦点を当てる。

自己評価の利点として第1に教育的意義を挙げることができる。中央教育審議会初等中等教育分科会教育課程部会が平成22年3月24日に出した「児童生徒の学習評価の在り方について(報告)」の中で、「児童生徒が行う自己評価や相互評価は、児童生徒の学習活動であり、教師が行う評価活動ではないが、児童生徒が自身のよい点や可能性について気付くことを通じ、主体的に学ぶ意欲を高めること等学習の在り方を改善していくことに役立つことから、積極的に取り組んでいくことも重要である。また、児童生徒の自己評価を学校評価においても反映し、学校運営の改善につなげていくことも考えられる。」(p. 12)と、自己評価の教育的意義も確認されている。英語力に関しては、Blanche & Merino (1989)がこの意義について指摘している。また、自己評価の利点として、コミュニケーション能

力など複合的な能力や技能で、直接的に測定することが難しい能力や技能の評価に利用することができる点が挙げられる。

本研究では、The Stanford Foreign Language Oral Skills Evaluation Matrix (FLOSEM) という口頭技能を評価するための評価基準 (Padilla & Sung, 1999) に基づいて自己評価項目を作成し、その信頼性及び妥当性を検討した結果を報告する。

2. 研究の背景

Padilla & Sung (1999) は、外国語による口頭技能を評価する基準表として、FLOSEM を作成し、その信頼性と妥当性を検討している。FLOSEM は、聴解、流暢さ、語彙、発音、文法の領域について、Level 1 (初級) から Level 6 (母語話者レベル) まで設定されている。例えば、聴解は次のように記述されている (pp. 37-38)。

Level 1: Learner can understand a limited number of high frequency words in isolation and short, common conversational formulaic expressions (e.g., "How are you?", "My name is...").

Level 2: Learner can understand short questions and simple non-formulaic statements when they are embedded in a short dialogue or passage. However, the entire dialogue or passage must be repeated at less-than-normal speed for learner to understand.

Level 3: Learner can comprehend the main point(s) of a short dialogue or passage which contains some statements with embedded structures heard at less-than-normal speed, though it is likely that details will be lost. Even at this speed, some repetition may be necessary.

Level 4: Learner understands most of what is said (all main points and most details) in both short and longer dialogues and passages which contain abstract information heard at almost-normal speed. Some repetition may be necessary, usually of abstract information.

Level 5: Learner understands nearly everything at normal speed, although occasional repetition may be necessary.

Level 6: Learner understands everything at normal speed like a native speaker.

評価者 (教師もしくは学習者自身) は、学習者の熟達度がどの程度かをこの基準に基づいて判断する。Padilla & Sung (1999) は、教師が中国語、日本語、韓国語を学ぶ高校生 (564 名) の能力を FLOSEM を使って評価し、その結果、学年の最初と最後で上昇が見られたことを示した。また、同じ参加者が自分自身の第二言語能力を評価し、自己評価と教師による評価との間に高い相関 ($r = .70, p < .0001$) が見られることを示した。さらに、132 名の高校生が外部指標として Classroom Oral Competency Interview (COCI) というインタビューテストを受験し、COCI と FLOSEM の間に高い相関 ($r = .829, p < .0001$) が

見られることを示した。これらの結果に基づき、FLOSEM を自己評価や教師による評価として用いることができることを述べている。

FLOSEM は、Butler (2004) や渡部 (2010a, 2010b) によって、小学校英語の教員の英語力を測定するために利用されている。Butler (2004) は、FLOSEM の項目にリーディングとライティングの領域を加えて基準表を作成し、小学校において英語を教える日本、韓国、台湾の教師に対して調査を実施した。渡部 (2010a) は、Butler (2004) が使用した基準表 (7 領域) を用いて、4 つの公立小学校の教員 54 人に対して調査を実施した。渡部 (2010b) は、Butler (2004) 及び渡部 (2010a) で用いられた質問紙票のうち、ライティングとリーディングの領域を除いた 5 領域に、「教室内英語力」という領域 (8 項目) を加え、さらに回答方法をリカート式に改良して 38 項目の質問紙票を作成した。公立小学校 12 校 169 人の教師を対象に調査を実施した。リカート式に改良したため、信頼性係数が計算できるようになった。6 つのカテゴリーにおいて .806 から .896 と高い信頼性係数が得られたことがわかった。また、渡部 (2010b) では、予備調査として、大学生を対象にして、自己評価と TOEIC、英検などの外部指標との相関を報告し、並存的妥当性が十分高いことを主張している。

質問紙票の回答方法として、リカート式を採用したものは渡部 (2010b) しかない。しかし、渡部 (2010b) では小学校教師を対象にしており、中学生や高校生に対して妥当かどうかの検討がなされていない。一方で、高校生を対象にした Padilla & Sung (1999) ではリカート式は検討されていない。そこで、リカート式の自己評価を中学生と高校生を対象に実施し、信頼性と妥当性を検討することが必要であると考え、本研究を実施した。

3. 方法

3.1 参加者と調査実施校の英語指導の状況

ある私立女子中学校・高等学校の 230 名が質問紙票に回答した。欠損値及び重複回答を行った者を除いた結果、199 名 (中学 2 年生 72 名、高校 1 年生 72 名、高校 3 年生 55 名) となった。

各学年の授業の状況と学習歴について簡略に述べる。高校 3 年生の必修教科目は、リーディングが 4 時間、英語 II が 1 時間となっており、その他選択科目として、英語研究 I、英語研究 II、英語特講がある。英語研究 I は基本的文法事項について行っており、英語研究 II は大学入試センター試験の問題演習を行っている。また、英語特講は理系以外の生徒が受講し、大学入試センター試験の英語リスニングに焦点を置き、リスニングを中心に授業が行われている。また、この学年は、高校 1 年生後期より、語彙に関しては教科書と単語帳の単語テストを頻繁に行っている。テストされる単語に関しては、高校 2 年生までは、教師のモデル発音の後、生徒が続いて発音した。教科書の Q&A 活動では、生徒の発話における心理的負担を減らすため、あまり長く沈黙を作らず、教師がなるべく助言を入れた。Q&A 活動においては解答が単語のみで行われるということもある。

高校1年生では選択科目はなく、必修科目だけで授業が構成されている。内訳は英語Iが4時間、ライティングが2時間、オーラルコミュニケーションIが1時間である。この内英語Iの1時間は英語の絵本を読む活動（50分中約30分）と2人の英語母語話者教師による活動（英語によるゲーム形式の活動20分、活動内容は日替わり）に充てられている。ライティングは分割授業で、約25名と15名のクラスに分けられている。オーラルコミュニケーションIは2人の英語母語話者教師が担当をし、2分割（1クラス約20名）で行われている。この学年は高校1年生になってから毎週英語の絵本を読む活動と英語母語話者教師による活動を行っている。この活動の例として、生徒同士でのカード当てクイズ（グループで活動し、1人の生徒が自分のカードを見ずに他の生徒に見せ、他の生徒がそのカードのヒントを出し、カードを見せている生徒がそのカードに書かれているものを推測する）や、語彙を増やすためのジェスチャーゲームなどを行っている。

中学2年生は必修科目のみで、英語5時間、英会話1時間で授業が構成されている。この学年は中学1年生の頃より、極力英語による授業が行われている。教師の質問に対しては1文で答えず、その答えに何か新しい情報を付け加えて話させ、なるべく多く発話するように求めている。語彙に関しては発音ができ、意味が分かることに重点を置き、スペリングの学習は後に回している。また新しい語彙を使用できるように、なるべくその語彙を含む文を聞かせ、さらに生徒にその語彙を使用させるようにしている。文法に関しては、生徒が文の構造に意識を向けられるように促し、発話を求めている。教室英語の使用も多く、生徒は先生に話しかけるときは英語で話させるようにしたり、また使用した表現を、文字を通して確認させたりしている。

本研究の参加者の特徴をまとめる。3つの学年の中では、中学2年生が口頭技能においては集中的に指導されており、実際に英語に触れている量も多いと思われる。また、中学1年生から高校2年生まで共通していることは、英会話もしくはオーラルコミュニケーションIの授業が週に1回あり、この授業は英語母語話者教師によって行われており、1クラスを2つに分け指導されていることである。その他、今年度より、多読指導を行い、学年により時間数は違うが、できる限り絵本やその他の本に触れる時間を確保している。

3.2 測定具

Padilla & Sung (1999) の評価基準表を基にして、渡部 (2010) の訳を参考にしながら、筆者らが改めて日本語訳を作成した。主な修正点は、以下の2点である。Padilla & Sung (1999) の表は、学習者のレベルを特定するための基準が書かれているが、本研究では「まったくあてはまらない」から「とてもあてはまる」までの6件法のリカート尺度で回答してもらうように変更した。また、Padilla & Sung (1999) の評価基準には複数の内容が含まれるものがあつたが、本研究ではその項目を複数項目に分割して記述した。その結果、5領域30項目となった（資料A参照）。具体的には、聴解6項目、流暢さ6項目、語彙7項目、発音5項目、文法6項目となった。

3.3 分析方法

本研究の自己評価の信頼性を検討するために、内的一貫性の指標であるクロンバック α を計算した。また、本研究では、妥当性のうち、構成概念的妥当性 (construct validity) と、内的妥当性 (internal validity) の1つである表面的妥当性 (face validity) を検討した (e.g., Alderson, Clapham, & Wall, 1995)。まず、構成概念的妥当性を検討するために、相関分析及び因子分析を実施した (Alderson, Clapham, & Wall, 1995, pp. 183-186)。つまり、本自己評価の5領域が口頭技能を測定するものであるならば、相互の相関係数が高く、因子分析では因子解が1つであることが予測される。また、表面的妥当性を検討するために、中学校・高等学校の教員である著者の一人が項目について分析を行った。

4. 結果と考察

4.1 すべての項目の場合—30項目分析—

表1によると、各領域の信頼性係数は十分高かった (.842 から .893)。また、30項目では、.967 であった。十分な内的一貫性があったといえる。

表1. 領域ごとの信頼性係数 (α)

聴解	流暢さ	語彙	発音	文法	合計
6項目	6項目	7項目	5項目	6項目	30項目
.873	.866	.893	.842	.868	.967

表2は、各領域の尺度得点の間の相関係数を示している。聴解と発音の間の相関係数が最も低かったが、 $r = .693$ と高く、また統計的に有意であった ($p = .000$)。その他の相関係数は .700 を超える高い値を示した。

表2. 領域間の相関係数

	聴解	流暢	語彙	発音	文法
聴解		.793	.800	.693	.773
流暢さ	.000		.893	.755	.886
語彙	.000	.000		.701	.882
発音	.000	.000	.000		.721
文法	.000	.000	.000	.000	

注. 右上は相関係数、左下は確率を示している。

各領域の尺度得点に対して、主因子法による因子分析を実施した。その結果、一因子解が適当であると判断された。固有値が 4.166 で、分散の 83.3% を占めた。信頼性係数、領域間の相関係数及び因子分析の結果から、5領域の合計得点を口頭技能得点として計算しても妥当であると考えた。

表3は口頭技能得点の記述統計量を示している。Kolmogorov-Smirnov と Shapiro-Wilk

の検定の結果、正規性からの逸脱は見られなかった。しかし、図1をみると、分布の左側で床効果が見られることがわかる。

表3. 30項目の合計得点の記述統計量 ($N = 199$)

	M	SD	Skewness	SES	Kurtosis	SEK
30項目	79.598	22.400	0.056	0.172	-0.399	0.343

注. SES = 歪度の標準誤差；SEK = 尖度の標準誤差。

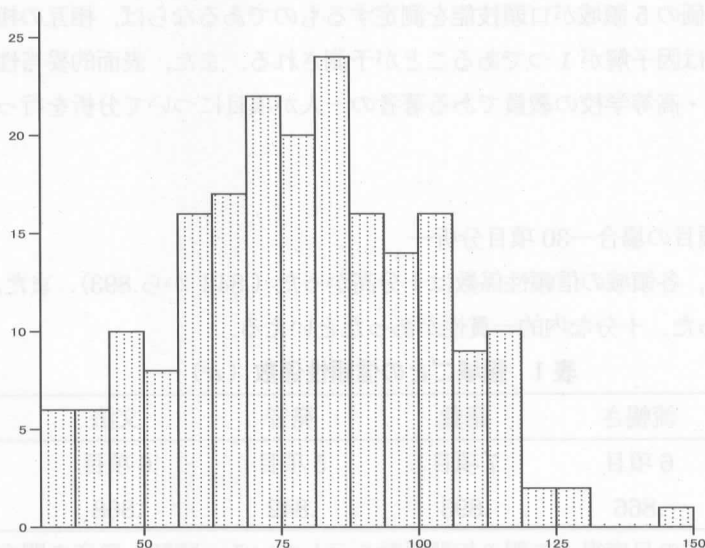


図1. 30項目の合計得点の分布

4.2 項目数を減らした場合—15項目分析—

実施可能性を高めるために、30項目の質問紙票よりも項目の少ない質問紙票の信頼性及び妥当性の検討を行った。本分析では、各領域において3項目選び、合計15項目の縮小版質問紙票の検討を行った。項目の選択基準は、(a)最も得点が低かった項目、(b)最も得点の高かった項目、(c)中間程度の得点の項目、である。(a)と(b)の基準は、できるだけ幅広い能力を測定するために採用した。(c)の選択にあたっては、項目内容も検討した。その結果、資料Aで挙げられた15項目となった。

表4は、領域ごとの信頼性係数を示している。項目数が少なくなったために、表1の結果と比べて低くなっているが、15項目全体では $\alpha = .930$ という高い信頼性が得られた。

表4. 領域ごとの信頼性係数 (α)

聴解	流暢さ	語彙	発音	文法	合計
3項目	3項目	3項目	3項目	3項目	15項目
.699	.720	.729	.764	.746	.930

表5は、各領域の尺度得点の間の相関係数を示している。語彙と発音の間の相関係数が最も低かったが、 $r = .628$ と高く、また統計的に有意であった ($p = .000$)。

表 5. 領域間の相関係数

	聴解	流暢	語彙	発音	文法
聴解		.762	.754	.682	.725
流暢	.000		.763	.690	.790
語彙	.000	.000		.628	.733
発音	.000	.000	.000		.671
文法	.000	.000	.000	.000	

注. 右上は相関係数, 左下は確率を示している.

主因子法による因子分析の結果, 一因子解が適当であると判断された. 固有値が 3.833 で, 分散の 77.7% を占めた. 信頼性係数, 領域間の相関及び因子分析の結果より, 縮小版においても 15 項目を合計し, 口頭技能得点とすることが妥当であることが示唆された.

表 6 は合計得点の記述統計量を示している. Kolmogorov-Smirnov の検定と Shapiro-Wilk の検定の結果, 正規性からの逸脱は見られなかった. 図 1 でみられた床効果が, 図 2 では解消されている.

表 6. 15 項目の合計得点の記述統計量 ($N = 199$)

	<i>M</i>	<i>SD</i>	Skewness	SES	Kurtosis	SEK
15 項目	38.899	10.986	0.162	0.172	-0.416	0.343

注. SES = 歪度の標準誤差; SEK = 尖度の標準誤差.

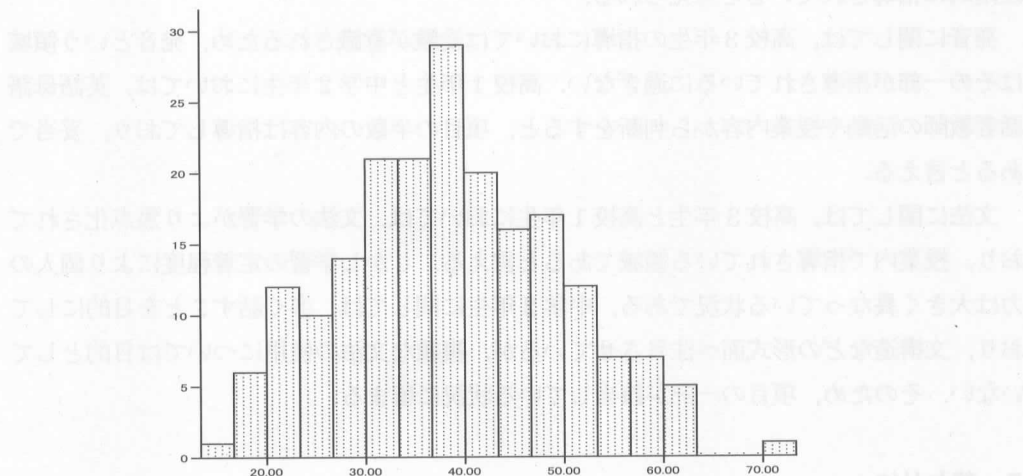


図 2. 15 項目の合計得点の分布

4.3 表面的妥当性

本節では, 本研究の自己評価の領域・項目の内容が, 中学生・高校生の英語による口頭技能の測定として妥当であるか検討をする. 特に, 指導内容と自己評価の内容の整合性という点から考察する.

先に記述した各学年の特徴を鑑みると、聴解領域に関しては、高校3年生に関しては、大学入学センター試験対策のため、文系の生徒（英語特講選択の生徒）にとっては妥当なものであると言える。また高校1年生に関しては英語会話や英語活動の時間における英語によるゲームや内容の説明を考えると妥当である。中学2年生に関しては普段からの授業（英会話を含む）で、英語母語話者教師がゆっくりと話をしたり、普通のスピードで話をしたりすることを考えるとこれも妥当といえる。つまり聴解領域に関しては全学年を通し妥当な項目であると言える。

流暢さに関しては、高校3年生と高校1年生に関しては、重点を置いて授業において指導されているものではない。しかし一部の生徒に関しては両学年とも、個別に実用英語技能検定試験などを受験することにより、口頭による練習を行うこともあり、この項目の一部は指導されていると言える。一方で中学2年生に関しては、定型表現などを多く使用し、また日常生活の一部（英語の教師に用事があり職員室に来る時は、できる限り英語で話すようにしている）としても使用しているため、流暢さに関する項目の大部分が、授業の中や、教師との関わりがある学校生活でサポートされている。

語彙に関しては、各学年とも工夫がされており、単語テストという形式でチェックが実施されている。単語テストの形式は、教師が読み上げる単語を聞き、書き取り、単語の意味を書くというものや、スペリングを見て意味を記述していくものなどがある。このように語彙を増やしていくことは行われているが、学年の発達段階に応じで単語の抽象度は異なっている（高学年にいくほど抽象度の高いものを扱う）。それゆえこの領域は学年により段階的に指導されていると考えられる。

発音に関しては、高校3年生の指導においては受験が意識されるため、発音という領域はその一部が指導されているに過ぎない。高校1年生と中学2年生においては、英語母語話者教師の活動や授業内容から判断をすると、項目の半数の内容は指導しており、妥当であると言える。

文法に関しては、高校3年生と高校1年生においては、文法の学習がより焦点化されており、授業内で指導されている領域であると言える。しかし学習の定着程度により個人の力は大きく異なっている状況である。中学2年生に関しては、多く話すことを目的にしており、文構造などの形式面へ注目させているが、複雑な文法の使用については目的としていない。そのため、項目の一部が該当している状況に留まる。

5. 終わりに

本研究では、FLOSEM に基づいて作成したリカート式の自己評価の信頼性と妥当性の検討を行った。その結果、かなり高い信頼性及び妥当性が見られることが示された。しかし、本研究では今後追求すべき課題も残されている。本研究では並存的妥当などの外的妥当性（external validity）を検証しなかった。適切な外部指標を選び、検討することが必要である。また、今回は私立学校の英語学習者を分析対象とした。公立の学校や他の校種

の学習者についても同様の結果が得られるかどうかという一般化の可能性を検討する必要がある。

引用文献

- Alderson, J. C., Clapham, C., Wall, D., 1995, *Language test construction and evaluation*, Cambridge University Press.
- Blanche, P., Merino, B. J., 1989, Self-assessment of foreign-language skills: Implications for teachers and researchers, *Language Learning*, 39, pp. 313-340.
- Butler, G. Y., 2004, What level of English proficiency do elementary school teachers need to attain to teach EFL? Case studies from Korea, Taiwan, and Japan, *TESOL Quarterly*, 38(2), pp. 245-278
- Butler, G. Y., Lee, J., 2006, On-task versus off-task self-assessments among Korean elementary school students studying English, *The Modern Language Journal*, 90(4), pp. 506-518
- Padilla, A. M., Sung, H., 1999, *The Stanford foreign language oral skills evaluation matrix (FLOSEM): A rating scale for assessing communicative proficiency*, ERIC ED445438.
- 渡部健介, 2010a, 小学校教員の英語力に関する自己評価, 『中部地区英語教育学会紀要』, 39, pp. 63-70.
- 渡部健介, 2010b, 『自己評価に基づく小学校教師の英語力に関する調査研究』, 信州大学教育学研究科提出修士論文.

(2010年7月28日 受付)

資料A. 質問項目

領域	項目
聴解	1 基本的な単語や、日常的で定型的な会話表現を聞きとることができる。
	2* 普通より遅いスピードで繰り返しのある、短い会話文や文章における、簡単な表現や短い質問文を聞きとることができる。
	3 普通より遅いスピードで話されたり繰り返しが時々ある短い会話文や文章の重要な部分を聞きとることができる。
	4 抽象的な情報を含む対話文や文章が短くても長くても、普通に近いスピードで話されることのほとんど（重要な部分のすべてと、詳細部分のほとんど）を聞きとることができる。
	5* 普通のスピードで言われたことをほとんどすべて聞きとることができる。
	6* 普通のスピードで話されたことをすべて、英語のネイティブスピーカーのように聞きとることができる。
流暢	1* 決まりきった質問・応答のやりとりや、短い質問に短い応答をするようなやりとりに参加できる。

さ	2	日常生活に関する身近なトピックについて、定型でない表現や質問をして、簡単な会話を行うことができる。
	3*	会話や議論に参加しているとき、簡単な文構造を使って自分の考えを表現できる。
	4	ためらったりつまったりせずに、複雑な考えを表現したり、あまり使わない単語や表現を使うことができる。
	5	自分の話す速さやリズムは、コミュニケーションの流れの目立った妨げにはならない。
	6*	自分の話す英語の流暢さは英語のネイティブスピーカーのようである。
語彙	1	日常生活で使われる物や行為に関する基本的な単語や、会話の定型表現や連語表現の単語を知っている。
	2	簡単な会話の中で、具体物について簡単に話したり、質問したりするための基本的な単語や基本的な連語表現を知っている。
	3	同義語や、簡単な考えを表現するいくつかの別の言い方を知っている。
	4*	コミュニケーションするときに困らない程度の適切な単語を知っている。
	5	抽象的な考えを含むような会話を理解したり、参加したりするための十分な単語を知っている。
	6*	トピックが具体的であろうと抽象的であろうと、様々な議論に参加することのできる幅広い単語を十分知っている。
	7*	ネイティブスピーカーのような幅広い単語を知っている。
発音	1	基本的な単語や定型的な表現の発音やイントネーションを正しくできる。
	2*	個別の単語や表現を、正しい発音やイントネーションで言うことができる。
	3*	相手に理解してもらえるように発音することができる。
	4*	自分の発音やイントネーションは、英語のネイティブスピーカーの力に近い。
	5	自分の発音やイントネーションは、英語のネイティブスピーカーのようである。
文法	1*	英語を話すとき、基本的な単語や短くて定型的な会話表現を正確に使用できる。
	2	英語を話すとき、基本的な文構造や疑問文構造を正しく使うことができる。
	3	英語を話すとき、誤解や不理解を生じさせるような文法的な間違いをおかさない。
	4*	英語を話すとき、比較的複雑な文型や文法規則のいくつかを正しく使うことができる。
	5	英語を話すとき、比較的複雑な文型や文法規則のほとんどを使いこなせる。
	6*	英語を話すとき、複雑な文型や文法規則を英語のネイティブスピーカーのように使いこなせる。

注. * は 4.2 の分析で使われた 15 項目を示している。