

Hughes, R. (2011). 'Chapter 4 Issues in assessing speaking. *Teaching and Researching: Speaking. 2nd ed.* United Kingdom: Pearson ESL. pp.80-112.

(後半 pp. 97-112)

4.3 A comparison of contrasting test paradigms for oral assessment in three high-stakes tests

Concept 4.5 High-stakes testing

High-stakes なテストとは受験者の生活に大きな影響を与えるようなテストのこと。

どのようなテストでも受験者には少なからず影響があると考えられ、テストの失敗は人生の大きな障害となる。

そうしたテストの例として医師国家試験や司法試験などが挙げられ、合格不合格で受験者に線引きがされる。

言語テストでも同様に、例えばイギリスでは移民が市民権を得るための試験として **ESOL** が使われ始めた。また、大学入学の際にも語学力を保証する証拠が必要になる。このようなテストが **High-stakes** なテストである。

- 対照的な3つのスピーキングテストについて本項では調査を行なっている。

Quote 4.16

学習者・教師のニーズと大規模テストのコンテキストは合致しない。大規模テストでは、受験者を差別化、カテゴライズすることを目的としており、また一般性、公共性、グループの識別ができ、測定可能、複製可能、予測可能な一貫した特有のものが作成されるのに対し、教室テストではそれとは逆の性質を持つ。

4.3.1 Internet-based Teaching English as a Foreign Language (TOEFL) speaking test

- TOEFL は世界的に 7300 以上の団体と組織に受け入れられている代表的な **Hi-stakes** なテストである。
- p.99 の図 4.2 は TOEFL を取り囲む独特かつ、明確に定義されたそれぞれの領域に関するイメージであるが、このイメージと ETS のテスト開発の哲学は合致している。
- TOEFL のフォーマットと評価プロセスは意図的に客観的に作成されており、これは対面のテストによってバイアスが生じることを避けるための考え方によるものである。
- TOEFL のスピーキングセクションは、受験者自身に馴染みの深いトピックについて意見を発表する 2 つの **independent** タスクとライティング・リスニングによる情報を統合して回答する 4 つの **integrated** タスクが設けられている。
- 基準の細部は異なるが、これらの 2 つのタスクは根本的に同じ側面を持つ。そしてこれらのすべてに受験者はモノログ形式で回答し、それらは録音が行われる。

- 6つの異なったタスクがオンラインスコアリングネットワークを経由して、訓練を受けた ETS の評価者チームに送られる。少なくとも 3 人の採点者が同じ受験者を評価し、信頼性検討のために 2 人の採点者のスコアがチェックされる。
- スコアは 0–4 点の間で付けられ、それらの平均を利用して 0–30 点に最終的に換算が行われる。
- 6つのすべてのタスクで、発話がなめらかで、エラーが少なく、話の筋が通ったものが高評価を与えられ、内容に不足があり、hesitation を含むもの、反復的なもの、基本的なアイデアのみを含むものが低い評価を与えられる。
- 幾分「非人間的」な哲学がこのスピーキングテストの理論を支えているのとは対照的に、ETS は極めて受験者に優しいインターフェイスを利用している。
- 話し言葉の、コミュニケーションスタイルについてと、詳細なテスト戦略のアドバイスがフィードバックとして受験者に与えられる(see Quote4.17)。
- そこで与えられるフィードバックは、このテストが同じ枠組みの中で行われるため、良いアドバイスになる。

4.3.2 The IELTS speaking test

- TOEFL iBT とは対照的に、IELTS は対面式で、対話者（試験官）と一緒に試験が行われる。テストは 10–15 分程度続き、それが録音される。
- テストの流れは 3 段階であり、最初の 4 分間は受験者にも一般的で、慣れ親しんだトピックについてのインタビューが行われる。第二段階として、セリフ付きのテストが使用され、約 1 分間の準備時間を経て、2 分間モノログで話すように指示される。第 3 段階へ移行すると、より抽象的な内容について試験官と対話を行うタスクが課せられる。
- IELTS のスピーキングテストは 2001 年に改訂が行われ、そこではフォーマットには幾つかの変更が会ったが、構成概念の観点に顕著な変更が与えられた。それによると、パフォーマンスは以下の 4 エリアで分析される様になった。①流暢さと一貫性、②語彙の使用範囲と量、③文法の使用範囲と正確さ、④発音。
- 記述子は IELTS ウェブサイトにおいて入手可能で'IELTS speaking band descriptors'と検索すれば出てくるようになっている。テストの外観については Quote 4.18 で、評価尺度については最高点と最低点のものが Table 4.2 に提示されている。
- これらのテストに影響力があることは顕著であるにもかかわらず、インタラクションに焦点を当てた IELTS のスピーキングテストと、能力に強い焦点を当てた非人間的なデザインの TOEFL はそれぞれスピーキングの正と負の側面を含むものになっている。
- これらの著名な 2 つのテストは、「良いスピーキング」の考え方を明らかにしてくれる。良いパフォーマンスでは、一般的に意味が通じ、話の筋が通った、語彙が多様な長い発話を指し、逆にシンプルでためらいや自己修正を含み、不完全でチャンクの繰り返しがあるような発話は望ましいもので

はない。

- 優れた発話を作り出すことは誰にとっても容易なもの、とは言えず、母語話者の会話はこうしたテストの基準とはならない。
- それに関連して、試験官が受験者の自己修正の瞬間をどう見なすか、受験者の発話と母語話者のパフォーマンスの瞬間を比較するのがこれまでに行われてきた研究の特徴になる。
- 言語学習者は目標言語でこうした処理が自然にできることを最高の目標として掲げるが、少ない語彙を使い、繰り返しを使うことは通常の会話の局面にも似たものであり、そういった発話は聞き手にとって聴きやすいこともある。

4.3.3 UK Border Agency Knowledge of Language and Life assessments

- 2005年にイギリス政府は多国籍の市民に新しい基準を適用し始めた。その一つとして Knowledge of Language and Life in the UK test (KOL test)が挙げられる。
- 移住の出願者は UK National Qualifications Framework で Entry Level3 に達していることが認可の条件となっている (Entry Level3 は CEFR の B1 と同程度)。
- このテストは一般知識と言語の評価を混ぜあわせたもので、その評価基準は直接 KOL テストのために設計されたものではないがイギリスのリテラシーカリキュラムにつながるものになっている (see Quote 4.19)。
- KOL テストはイギリスの生活と文化の一般的知識についての 24 問の多肢選択肢問題になっている。このテストを合格できなかった受験者はこの試験を何度も受け直すか、'ESOL with citizenship materials'の資格を得る必要がある。
- 移住関連のテストがコミュニケーションと話し言葉の機能特性に焦点を当てる傾向があり、Cambridge ESOL のようなテストはその観点から CEFR に基づいて築かれたものである。
- 以降、ESOL と IELTS・TOEFL との対照を行う。ESOL with citizenship materials では CEFR の Can do statements とインタラクティブ活動の評価基準の要素を含んだベンチマークが描かれている。
- Quote 4.20 は受験者が何を達成することを期待しているかということに関して、初級 3 レベルを概観している。
- それによると E3 では話し言葉を聞いて直接的で narrative な情報を含め、対面でも電話でも応答が可能なレベルが要求されており、馴染みのあるトピックであれば複数の人と対話ができることなども基準に入っている。合格条件として、言語機能に注目が置かれており、planning や narrating といった文言だけでなく、過去や未来の出来事について尋ねることができる、などといった言語機能も必要とされている。
- ESOL のスピーキング・リスニングフォーマットは、上述されたテストとは著しく異なる。受験者はペアで評価が行われ、対話者と外部の評価官が試験に参加し、モノログと対話の両方を含めて広範囲のタスクに参加することになる (ただし E1~E3 のレベルでは評定者は試験に参加しない)

- 合格基準が **can do statements** に基づいて全体的、かつ相互的に関連付けられる。ESOL の評定者は **CEFR** の異なるレベルのサンプルを基に訓練が行われる。
- 「インタラクション」を含んだすべての基準が先述のテストと同様、言語使用の範囲、正確さ、流暢さ、一貫性を含んでいる。
- ESOL では複数の受験者が一緒に受験するため、採点システムに対話を含めることが適当であるが、対照的に **iBT** のスピーキングはモノログであるため、インタラクションの記述は余分なものと言える。
- テスト全体のデザインとスピーキングの構成概念がぴったり適合することはない。
- 移住を目的とした言語評価では **High-stakes** なテストであり、スピーキングを含むことに異論がないことは興味深いことである。
- テストの哲学やテスト文化として、政治的に **KOL** はデリケートな問題を含む。これは国内にウェールズ語やゲール語母語話者がいるためである。そうしたマイノリティ言語についても選択肢やリーディングマテリアルを通じて情報が提示されることもある。
- 他の国では、移住に際してより厳格な言語の必要条件を設けている場合もある。オーストラリアの場合、4 技能のすべてを満たすことが必要とされるし、カナダの場合は英語か仏語でリスニングとスピーキングができることを要件として設定している。
- このように **High-stakes** なテストでは、倫理上の問題や社会的責任についての懸念を含むものになっている。