

## Chapter 8 Aligning tests to standards

### 8.1 It's as old as the hills

- standards-based assessment はテストを学習者のパフォーマンスと達成度を絶対尺度に基づいて評価するものである。ゆえに standards-based assessment はしばしば基準準拠テストの発展形とされる。
- しかしながら、standards を設定する難しさは、教育評価の歴史と同じくらい古くから主張されている。例えば Latham (1877) はカットスコアの決定に関する困難性を指摘している。

### 8.2 The definition of 'standards'

- standards の議論に入る前に、この語の明確な定義が必要である。本章では Devies et al. (1999) による「求められるパフォーマンスのレベル」という定義を用いる。  
→(例) その大学入学に求められる standards は英語で A を取得していることである。
- 採点に関する前章では、各レベルの受験者が知っている、出来ると考えられる記述を含む尺度をみた。政策立案者が、学位取得や、入国に必要なパフォーマンスレベルを定める場合でも、このレベルが満たされるべき standards となる。

### 8.3 The uses of standards

- standards を設定し、standards-based assessment を活用することは教師にとって便利な場合がある。さらに、業務上必要な standards を定めることは一般的に大変重要である。
- 業務の成功に特定の言語能力が必要ならば、個々人がその standards に達しているか否かをテストすることが大切である。  
→例えば飛行機のメカニックには、英語で書かれたマニュアルを順序立てて理解したり、問題が起こればレポートを書いたりする能力が要請される。これらを満足にできない者が作った飛行機に乗りたいとは思わないだろう。  
→しかしながら、業務上の standards を設定したり、個人が standards を満たす能力があるかどうかを評価するテストを作ることは大変難しいことが、これまでの章から示されている。
- standards が high-stakes testing によって課されるとしばしば議論を巻き起こす。  
→No Child Left Behind (NCLB) という政策下で、アメリカの学校の多くは、生徒の英語力に関し 'adequate yearly progress (AYP)' を報告することを要請された。  
→この AYP を測るテストは教授内容に甚大な影響を与える言語政策のような様相を呈し、多くの州では、この standards と繋がりのある high-stakes test を実施する必要にみまわれた。  
→standards は達成度の指標としてだけでなく、政治的権威者が教育システムをコントロールするための手段としても機能し得る。
- さらに物議を醸しだすのは、standards-based assessment を移民政策に利用することである。  
→移住の是非をテストで問うことに関し、倫理的な観点から反対する者もいれば、移民が社会に馴染むには、言語的 standard の設定は順当だという者もいる (Bishop, 2004)。  
→どちらにせよ明らかなことは、移民に求められる standards は経済的状况によって変化するという事実である。
- たとえばオーストラリアでは、移民に対し IELTS Bounds 4~6 の言語能力を求めている。

→その一方、優秀な人材を求めするために **Bounds 6** 以上の移民のみを受け入れるべきだと主張する団体もある。  
→言語能力と業務遂行能力は関係がありそうだが、この文脈で **IELTS Bounds 6** が最適な **standard** だという証拠はどこにもない。この例は政策に関するテストの **standard** の基準を決めるむずかしさを表している。

- さらに **standards** の記述と受験者の能力との関係性を示す証拠がないため、テストスコアと受験者の能力とがリンクしているという前提が破たんしている。このように移民におけるテストは、能力の測定ではなく、単に移民の数を減らすことが目的となっているのが実状である。

#### 8.4 Unintended consequences revisited

- **standards** が政治目的で使用は、しばしば想定外の結果をもたらす。
  - NCLB では、**standards** に達した生徒は、「英語学習者」の区分から外れる仕組みになっていた。
  - standards** に達した者が「英語学習者」から除かれるために、学習者のスコアは常に **standards** よりも低く、ゆえに生徒が **AYP** であることを示すのは非常に難しい、という構造的な問題が発生した。
- 様々な教科のテストで英語の使用を義務付けることも、想定外の結果をもたらす。
  - アラスカの学校では、日々の指導・学習は **L1** のユピク語で行われるが、生徒はテストを外国語で受験し、成果を上げることが期待されている。
  - L1** なら十分に得点できるテストでも、外国語であるために、はるかに低い点数しかとれないという事態が発生している。この例は、思慮なく政治目的にテストを利用することが、少数言語を脅かし得ることも示唆している。
- 想定外の結果は、言語テストを移民政策に利用する文脈でより大きくなる。
  - テストが新たな人生や栄光へのゲートキーパーとして機能するものだと、それに関する詐欺が多く発生する。
  - テストが移民の大学入学審査として使われると、テストに合格できないインドの経済移民の中には、合格する能力を持っていたり、学習を経済的にサポートしてくれたりする相手と偽装結婚を行う者が大勢いた。さらにこの偽装結婚の発覚を防ぐために暗躍する弁護士も多く見受けられた。

#### 8.5 Using standards for harmonization and identity

- 社会的な同調 (**harmonization**) を作るために **standards** を用いることは古くからあった。この目的は、同調を強いることで政治的統一性を維持することにある。
- 例えばカロリング朝 (発表者注: 古代フランス、フランク王国の王朝) でも、統一性を強いる政治目的で **standards** とテストが利用された。このテストに正解はなく、同調に都合良く設定された **standards** が認める回答が「正しい」とみなされる格好であった。
- 今日でもこのような例はしばしばみられる。**standards** のディフェンスとしてよく用いられる論理が、**standards** は様々な人々が学習者のレベルについて議論する「共通言語」を提供するというものである。
  - 最たる例が **CEFR** である。**CEFR** は言語の性質それ自体には大した関係がないものの、わかりやすく、便利な基準を提供している。そのため、学校や教材開発機関はこぞって自分たちのテストが **CEFR** と関連していることを主張している。
- このような同調は多くの問題を発生させ得る。
  - ・①教師から創造性を奪うことになる
    - 状況によって、最適な能力レベルや **standards** は異なるが、**standards** による同調圧力はこの柔軟性を奪いかねない。

## ・②妥当性の問題

- 同調論者は、複数のテスト結果が CEFR と関連して報告されれば、それらは同じスコアを意味すると主張する。
- 今後の研究によって様々なテストと CEFR の関連性が実証されれば、テスト機関は妥当性を叫ぶだろうが、目的、構造、内容が異なるテストを、CEFR とのつながりによって、一括りに妥当とは言えない。
- しかし、このような傾向は経済的利便性から CEFR 以外の文脈でも良く起こり得ることをしておくべきである。

## ・③究極的に、同調は支配への動きを強める

- 1 つの standards によって全てにニーズを満たすことが不可能なことは歴史から明らかである。
- しかし、権威ある機関によって standards が設定されると教師はそれに従う必要性を感じずにはいられなくなってしまう。
- わたしたちはテストや教育機関が、言語や教育に関したった 1 つの standards を押し付けさせることのないよう注意すべきである。

## 8.6 How many standards can we afford?

- standards の記述を作成する際に、最初にやるべきことはパフォーマンスレベルの数を設定することである。
  - パフォーマンスレベルは最少に定めることが大切である。多くのテストのパフォーマンスレベルは合格・不合格の 2 つである。
- ただ、パフォーマンスレベルを規定するカットスコアを決めるのは難しい。隣接するグループの違いはグラデュエーションのような連続性をもつからである。
- パフォーマンスレベルの数を最少に抑えるべき理由として、テストの信頼性に関するものもある。
  - (パフォーマンスレベルによって区分される) 群の数が増えれば、得点の小さな差 (small differences) が起きる危険性が大きくなる。この差はテストの信頼性をしばしば損なうことが知られている (2 章参照)。信頼性の低いテストではスコアが偶然によるものである可能性が大きくなり、カットスコア付近の受験者はどちらの群にも入ってしまうという問題がある。
- Wright (1996) はパフォーマンスレベルを分けるカットスコア数に、どのくらいの信頼性が必要かを推定するために、分類指標 (Index of Separation) を使用することを推奨している。公式は以下の通りである。

$$G = \sqrt{\frac{R}{1-R}}$$

G = 分類指標, R = テストの信頼性

- この公式によれば、1 つのカットスコア (i.e., パフォーマンスレベル数) を設定するために必要な信頼性は .50 である。この値自体は低いが、カットスコア付近の信頼区間非常に大きくなってしまっだろう。
- 求めるカットスコアの数が増えれば、必要な信頼性はどんどん高くなる。例えばカットスコアを 2 つ設定したいなら (i.e., パフォーマンスレベル数 3) に必要な信頼性は、.8 から .9 で、3 つではさらに厳しい値が要求される。
- ここまでみたように、パフォーマンスレベルは出来るだけ少ない方が好ましいが、教師は時に意図してレベルの数を増やす必要性にかられることがある。
  - 大学における言語クラスなどがその例で、パフォーマンスレベルが少ないと、学習者が進歩を感じにくい場合があるのである。この事態はモチベーションの低下にもつながる可能性がある。

→このような状況では、スモールステップによるモチベーションを与えるために、理論的には多すぎるパフォーマンスレベルの設定が必要な場合もある。しかしながら、これらは、あくまで教育上の「フィクション」なのであって、第二言語習得における「現実」ととらえてはいけない。

☆移民法に定められるテストに関して

■移民を受け入れている諸外国では、入国を希望する者に対し何らかのテストを課している国がほとんどである。たとえば、カナダは移民法によって公用語である英語もしくはフランス語能力を示すテスト (e.g., IELTS バンド 6 以上) の提出を義務付けている。また、アメリカは英語のテストに加え、アメリカの歴史や政治に関する基礎知識を問うテストの受験を求めている。

■このような帰化テストに最も頻繁に用いられている英語能力テストが IELTS である。具体的には、4 技能を対象とした受験者の成績は 0~9 のバンドスコアで算出され、ジェネラル・アカデミック 2 種類のモジュールからなる。帰化テストに利用されるのはふつうジェネラル・モジュールで、バンドスコア 6 (Competent User) 前後が standards として定められることが多い。

■まず帰化テストが必要か否かに関して、私個人の意見としては必要であると考え。言語も文化も異なる国で社会生活を営むためには、周囲との意思疎通を図ったり、常識に則った言動を身に着けたりするためにも、相手国に関する最低限の知識は求められて然るべきだろう。

■しかし、気になる点として、受験者が国で生活、就業していくために必要な言語力を IELTS 等の単体のテストによって、画一的に評価できるのかということがある。同じ英語圏であっても、アメリカとオーストラリアでは用いられる英語は発音や語法といった表面的な部分でも異なるし (たとえ IELTS が様々な英語を対象としていても)、さらに国が違えばディスコースの運び方や、コミュニケーション方略の機微も異なってくるだろう。

■帰化テストを IELTS や CEFR に関連付けたとしても、このような微妙な差異を考慮に入れない限り、真に「相手国で生活するための」言語能力を測定することは難しいと感じる。さらに同じ国であっても、移住する目的の違い (e.g., 留学目, 就業目的) によっても求められる言語使用は変わってくる。また、テストを課すならアメリカのように歴史・文化や政治に関することも問うべきだろう。

■各国の移民法を調べていると、それぞれに努力はみられるものの、現実の移民がもつ多様性、そして移民の数量的な問題には到底追いつけていないという印象を抱かされた。確かに、個々の移民に対応可能な帰化テストの設計は、現実的には殆ど不可能なのかもしれない。しかし、それでも IELTS や CEFR で一様な評価を行うだけでは、本書の著者の指摘通り、必要な能力を反映したテストとしては働かず、移民の数を増減させることが関の山であろう。我が国では移民の受け入れがないため、今まであまり意識に上らなかったことであつたが、移民に関するテストは最も high stakes な問題として、検討されるべき課題であると考え。