

7. The politics of language testing

■ 教育とテストは政治的方策を実現させる上で重要な役割を果たしている。

- ・ Mcnamara & Roever (2006) は、「調査(assessment)の政策化 (政治性) は言語テストの最も著しい特徴である」としている。
- ・ また、テストは政策者にとって判断の説明責任を果たすための道具として用いられる。

■ テストに政治性が伴う 2 つの理由

① individual learners の育成

= 学習者が最も良い教育を自ら選択できるように、テストによって測定される各機関の業績・能力を公開し、学習者に **informed choice** を持たせる。

② ビジネスに求められる人的資源の生産

= 潜在的な経済の有効性は経済に貢献できる人口のレディネスであり、経済の効果的な運営のために教育は重要な役割を持つ。

■ 新しい言語教育の目的 - 世界経済における ビジネスの観点から見る

- ・ ビジネスの視点では経済の観点で言語教育における教師・学習者の成功が判断される。→例) ヨーロッパの雑誌に見られる教育の新たな見方
- ・ OECD の PISA 調査のデータ等も経済のための教育の指標として判断に用いられる。

■ 学習者の評価

- ・ 経済活動における言語能力の水準は、スケール評価 (i.e., rubrics) によって表される。
(例: can-do list)

■ 言語テスト・言語教育の課題

- ・ **high-stakes** を必要とする職業における最低限の言語到達度を設定すること
- ・ 標準ベースのシステムが学習者・教師・教育機関の社会基盤を制御する根拠となる。

8. Historical interlude II

■ テストの経済的役割

- ・ テストを経済の向上のための道具とする見方は従来から存在した(e.g., プラトン・全体主義) ⇒テストにより人材の適材適所の配置が可能となる

■ Max Weber の労働者の経済的能力を計測するための3つの観点

- (a) 機能に対する最適な適性
- (b) 実践を通して習得が行われる最適条件
- (c) 仕事のための inclination の最適条件

■ テスト理論研究

(旧) Spearman の主張: 「テストによってもたらされるもの = intellectual index」 (p.16 引
用参照)

⇒ 当時能力測定テストは社会開発と優生学の潮流と強く結びついていた (e.g., Cattell, 1937)。

⇒ 1930 年代以降、心理学界とテスト作成者が経済のニーズに応えテスト理論の検証が急速に進展する。

(現在) しかし、現在でも経済的優勢に関連するテストは有能さを示すものとしてエリート主義に寄与している (例: 弁護士や医師など有能さを求められる職業人に対する価値観)。

- ・ 以上のように、政治的な視点に立つと教育と望む社会の実現・維持は強い関係を持っている。
- ・ そのため教育内容やテスト内容は政治視点や教師の視点など異なる立場の間で議論の的となる。

9. Professionalising language education and testing

・ Consequential validity が検討されるように、テスト (特に high-stakes test) は washback effect として社会へ良い影響も悪い影響も与える可能性がある。

⇒ テストは、理想の社会・経済の実現のために教師や教育体制を制御するために使用されてきた (e.g., 学習指導要領の変更 ⇒ 教育内容, high-stakes test の内容の変更)

⇒ テストは意図的でない結果を引き起こす有害なものになりうる (e.g., high stakes test の失敗によって受験者の心身の健康を損なわれる)。

■ John Stuart Mill 見方: 「テストは民主主義における能力主義を実現するために必須である。」

- ・ 国家的教育体制により、自己判断できる個人の育成と選挙権の普遍化への布石がで

きる。

- ・ テストにより教育体制の制御と個人の洗脳が可能になる。

■ Mill の唱えるテストの 3 原則

- ① テストは政府によって制御されるべきではなく、一つの独立した機関によって実施されるべきである。
- ② テストは受験者に複数観点が存在する問題を問うたり真偽が問われている問題に関して問うべきではない(e.g., 死刑制度やジェンダー問題など価値観にとって”正解”が異なる問いを出すことは公平ではない)
- ③ 国はどの資格が社会的認知されていると表明するようなタスクをとるべきでない
⇒これらの原則から義務テスト使用の分析に示唆がある

- ・ Fulcher and Davidson (2007): テスト開発者はそのテストが対象とする人と狙う目的を明らかにするべきである。

- ・ 言語テスト作成者によって、テストに関する倫理規約と実践のガイドラインを作成され、倫理的責任の問題が検討されている。

例) ILTA(International Language Testing Association) によって発行されている *Standards for Educational and Psychological Testing* (AERA, 1999)

10. Validity

- ・ テストを作成する際に重要なのは「妥当性」の観点である。

- ・ “validity”の定義

(旧) 1989 年まで: 「テストが意図する対象を測定しているか」という観点 (Ruch, 1924)

(新) Messick (1989) 以降: 多数の異なる側面を持つ一つの概念

- ① ①関係性と実用性の側面: テスト得点は受験者の知識・技能・能力を推定し、判断の基準とする
- ② ②構造的側面: テストにより多数の能力・技能に関する情報を得て、焦点となる能力。技能を得点化・構造化する
- ③ ③内容的側面: テストの内容はその分野の内容を代表する(領域代表性)
- ④ ④一般化可能性の側面: テスト得点はそのテストを超えて一般的な能力をあらわす
- ⑤ ⑤外的妥当性の側面: 他の同様の能力を測定するテストと(相関)関係がある

- ⇒正確な推測と公正な判断のためのテストの作成へ
- ⇒教室でのテストによって指導・学習に示唆を与える

【Activities】

- ・ 具体的事例を考慮し、「どのようにテストを作成すべきか」を検討する。

1.1 Why do you test?

- ・ 一般的なテスト使用の目的は、①到達度(achievement), ②適性(aptitude), ③診断(diagnostic), ④配置(placement), ⑤能力(proficiency) の評価である。
- ・ しかし、実際のテスト使用にはこの他に多くの動機があるため以下を検討すべきである。
 - (a) テスト作成の際にはそのテストを実施する理由をリストアップする
 - (b) 外部テストを受けさせる際にも同様に理由を検討する

⇒ (a)解答例: 上記5つの他 motivation を高めること (e.g., 到達度テストに基礎的な問題から発展的な問題まで含めることで達成感と向上心の motivation を高める、あるいは〇〇模試のように目標に specific なテストにより、目標を明確化して motivation を高める) や reward / punishment を与えること (e.g., TOEIC で満点をとると会社からボーナスが出る / 追試に合格するまで部活を禁止する) など。

(b)解答例: 各県の実力テスト=到達度, や能力、知能検査=診断・能力・発達

1.2 The grounds for selection (省略)

- ・ 教育やある種のキャリアを受容する際、多くの基準から人々は選定される。

※p.21 表 参照

Q. この他にも基準があるか?

Q. これらの基準は適当か? 不適当だと思われる基準はどのような価値観に基づいているか?

1.3 Identifying unintended consequences

- ・ テスト実施は必ず何らかの影響を与える。

T. 指導・学習の状況を想定し、テスト実施される場合に行うこと、もしくはテストがなかったら行わないことをリスティングし、それらが positive (+) /negative (-)な影響を持つか検討する。

例) a) 行うこと: (専門)知識の獲得 (+), output 練習 (+), 過去問対策 (+), 受験費支払い (-), 徹夜 (-), フィードバックから自分の到達度把握 (+), test-taking 技術を磨くことへの努力 (+)

⇒能力の向上や自分の診断、motivation を高めるなど

b) 行わないこと: テスト勉強 (-), 結果から自分の到達度把握 (-), high-stakes test のプレッシャーの減少 (+), 徹夜 (+), 他の活動に時間を割ける (+),

1.4 Mitigating unintended consequence

T. p.22 の記事を読んで、どのような行動が意図しない結果に通じてしまう事を防ぐかを検討する。

・ 記事内容＝中国で大学入学試験において最低限のスコアを満たせなかった 16 歳の少女が自殺をした。

⇒例) 「勉強と試験からくるプレッシャー」が生徒の心身に害を及ぼす大きな原因といわれていることから、一つの判断に対し一つの試験を唯一の指標をしない、またテストを受ける機会を増やすことがある。他にも、high-stakes テスト結果が及ぼす選定にセーフティネットを設けること等の解決策の方向が考えられる。

(具体的には、大学入試で推薦入試、AO 入試、一般前期・後期入試などの多様で複数のテストを設けること、もしくは一つのテストに対して 3 回機会がありその平均点や最高点を用いる等の対策が挙げられる。また、資格が必要な職業に就くためにも多様な道を用意することなどが有効である。)

1.5 The Big Brother debate

T. p.23 の記事を読んで個人情報のデータベース (国民全員の詳細データや試験結果の記録) を作成する事は論理的と言えるか? 賛成/反対 の意見を検討する。

・ 賛成: 個人が人生の経緯を簡単に把握できるため、個人の現在と未来を検討する参考に活用できる。

- ・ 反対: 過去を参照されることにより、現在の個人を判断する際に **bias** がかかる。過去の失敗を挽回できなくなる可能性がある。そのため、今以上にテストへのプレッシャーが増大する。

1.6 Finding the right level?

- ・ 移民の要件となる標準言語能力をとして採用される **Common European Framework of Reference for Language: Learning Teaching Assessment** の level. B1 (p. 24) を検討する。

Q.1 新しい移民の受け入れ要件として十分な記述になっているか?

Q.2 言語要件は移民受け入れの基準として適切か?

Q.3 移民がこの言語条件に達しているか測定するタスクはどんなものか?

⇒例) 最低限必要なコミュニケーションのための4技能について言及している。しかし、ある条件において対処できるなどの曖昧な記述も有り、テストを作成するにはより詳細な **can-do list** を作成する必要がある。その条件を測定するためには、日常生活の具体的場面を想定し、4技能について、また4技能を統合的に活用する能力を測定するタスクが求められる。具体的な場面とタスクの例としては、入国後の役所での住民手続きにおける **reading, listening, writing, speaking** のペーパーテストで行うことや **roll play** を行うこと、他にも教育の場面を想定したグループディスカッションなどが挙げられる。

1.7 The efficiency drive ~Yoakum and Yerkes (1920) の精神工学

- ・ 各職業人の **Army test** の得点を表す p.25 のグラフから (効率的な国家の再生のために)個人の仕事適性への指針が得られるとする彼らの主張を検討する。

Q.1 本データと彼らの主張は妥当か?

Q.2 そうでない場合、彼らの理由付け・調査・データの解釈に関してどのような問題点があるか?

⇒例) 効率の良い職業の選択には、能力・適性・意欲など多くの要因が関わってくるが、**Army** テストで本当にそれら全ての構成要素が考慮されていないと考えられる。このことから、彼らの主張は能力の一側面のみから判断してしまっている可能性がある。具体的には、**cook** や **mechanic** などは職業特有の専門技能が大きく関わるもので、知能テストの結果とは大きく違う要因が関わりと想定される。また、この結果を採用することは、倫理的に職業に優劣をつけることに

つながりかねないという問題を持ち、さらに同じ社会の再生産につながる可能性がある。