

第4章 テスト開発 (担当 pp.87-96)

- テスト開発：テストの作成・使用に関わる全過程を指し、それは基本設計・操作化・実施の3段階からなる(Bachman & Palmer, 1996)。
- テスト細目表：基本設計をまとめたもの(テストの目的・対象者・測定したい能力・問題形式・実施形態・解答に必要な知識や技能・難易度の程度 etc.)
- この章は最新の理論的見地も踏まえつつ、より具体的にテスト作成における総合的重要項目とその際に重要とされるテスト細目や構成概念、テスト目的について言及している。

1. テスト開発において念頭に置くべきこと

1.1 Brown (2005)

- テストは状況に縛られるものであり、縛られるべきである。特定の学習者集団に対して有効なテストでも他の状況では役に立たない可能性がある。テストを採用するのか、開発するのか、手直しするのかを考えてから始める。

1.2 McNamara (2000)

- テスト開発者はテスト作成目的やテスト作成に関わる制約(物理的・財政的)などの問題点を明確にした上で詳細なテスト設計をする必要がある(どんなテストを作成しても実施できなければ意味なし)。

1.3 Bachman & Palmer (1996)

- テスト開発は「構想→設計→テスト+テスト結果の保管」といった作成から実施までの様々な作業を含む。状況・目的・利害関係はどうあれ、テストの有用性は常に念頭におくべきであり、どのような状況になっても開発過程を事前に十分検討し、綿密な計画を立てることが重要。このことはテストの有用性を検証する上でも、説明義務を果たす上でも、作成者自身の充実感を得る上でも有効となる。
- テスト開発はテストの作成と使用に関わる全過程を指し、「設計」「操作化」「実施」の3段階からなる。開発は通常、順次段階を経て進むものであるが、時には改訂を繰り返す必要がある場合もある。

2. テスト開発・作成全体に関する記述

2.1 Hughes (2003)：テスト開発の大枠の手順

- ①テストで解決しようとしている問題を詳細かつ明確に述べる
- ②テストの詳細な細目表を書く
- ③問題を作成し、修正する
- ④母語話者を使って試行し、必要に応じて問題の廃棄、修正をする
- ⑤本番の受験者集団のレベルに近い非母語話者で試行する
- ⑥試行の結果を分析し、必要な修正を加える
- ⑦能力尺度を作成する
- ⑧妥当性を確認する
- ⑨受験者、利用者のための手引き書を各
- ⑩必要に応じて面接者や評定者などのスタッフを訓練する

2.2 Downing (2006) : 効果的なテスト開発のための 12 ステップ。これらのステップは循環しており、常に全体を見ながら適切な処理をし、テスト開発を進めていく必要がある。

- ①全体計画：構成概念、解釈方法、形式、妥当性、適用する心理測定モデル、質のコントロールなどが検討される。
- ②内容定義：構成概念妥当性、構成概念の記述を検討
 - テストデータの分析を蓄積しながら言語能力の要素のより正確な分類・抽出を継続していく必要あり
- ③テスト細目表：テストする内容を操作可能な状態で定義するステップ。テスト内容の分野の具体的事項、集団基準準拠テストか目標基準準拠テストか、そして望まれるテスト項目の特性を検討
- ④項目の開発：作成者がテスト項目を実際に作成→いい項目を作るためには常に訓練が必要
- ⑤テスト計画と構築：これまでに作成された項目の中から、適切なテスト項目の選択をする。
 - テストの質をコントロールするためには事前テスト、等化、機密保持が必要
- ⑥テスト産出：これまでのステップを統合し、テストを印刷（PC テストではプログラムのチェック）
- ⑦テストの実施：テスト監督、実施環境の均一化に注意を払う。
- ⑧採点：あらかじめ設定しておいた採点基準・方法に従い、解答を数値化。最初の採点后、テストの項目分析を行い問題がある項目を削除。その後最終的な点数を計算する。
- ⑨分割点の設定：実施後のデータ分析を検討して cut score, performance standard を算出し、基準設定を行う。判断基準の設定はどのような場合も人間が行う。
- ⑩結果のレポート：受験者へのフィードバック。公平性、適時性、適切性が重要。
- ⑪アイテムバンキング：実施したテスト項目について、その項目の特性値を含むすべてのデータを将来の再使用に備えてデータ・ベースに保存する。
- ⑫技術的課題の報告：すべてのステップにおける実践的課題を考察し、改善につなげる。

2.3 Fulcher (2010) : テスト作成の際、テスト項目を書くことをまず考えがちだが実際にはその目的を考えることの方が重要。テスト目的に含まれるものとしてテスト受験者、受験者の能力範囲、言語使用域、知識の範囲などがあり、この内容がテストの構成概念の構造を正当化する際に重要となる。

2.4 日本テスト学会 (2007) : テスト開発者は目的や場面に合わせて測定の内容・形式・方法・手続き・結果の利用法・受験対象者の範囲などを明確に定め、基本設計を行う必要がある。

- ①テストの目的：テスト作成の際には、どのような目的と場面でそのテストが利用できるのかを明確にする（特定の目的があるのか、汎用性を持たせているのか）。それにより、実施者・利用者が目的に合ったテストを適切に選択することができる。
- ②測定対象と測定内容：テストの実施によって目的に合った有益な情報が得られるよう、測定対象となる特性を明確に定義し、それに対応した測定内容を持つテストを作成する必要がある。
- ③受験者の確定：想定される受験者層によって、適切な内容・形式・方法が異なる場合がある。
- ④測定形式：テスト目的に合った形式を選定する。テスト特性は項目の提示方法、解答形式、項目の数、時間、配点など様々な要因によって変化するため、項目が繰り返し利用されることを想定する場合は、毎回のテスト特性の等質性や連続性を確保するため、開発者が決められる要因は制約が許容する範囲内で一定の形に整えておく必要がある。そしてテストの目的や方針の変更でテスト内容や測定形式が変わり、等質性や連続性が保持できない場合、テストは改訂されるべきである。
- ⑤テストシステムとしての配慮：テスト開発にあたっては質問項目の内容のみならず、測定形式・解答時間・実施方法・解答の収集方法・採点方法・質問項目や回答結果の管理及び情報公開まで、全体のシステムを配慮し

た設計を行う。→特に質問項目を繰り返し用いるのか否かは基本設定に大きな影響を与える（用いる場合は細心の配慮のもとでシステム全体の設計を行う必要あり）。

◆同じ目的で繰り返し行われるテスト：内容面だけでなくテストの性能についても各回のテストができるだけ等質となるような尺度化の方法を考慮して作成する。

◆学習や発達の状況を比較するテスト：各回の実施結果に見られる差異が意味のある変化として捉えられるよう、結果が共通尺度として表現されるような構成が必要

⑥結果の利用方法：テスト開発者はテストで得られる結果をどのように利用すべきかについてあらかじめ明記しておく。また、テストに関わる統計的情報を公開し、利用目的を逸脱しないように利用方法を限定することが必要。

2.5 Alderson, Clapham, & Wall (1995)：テスト開発の過程はテストの計画書作成に始まり、報告書作成に至る。その段階は以下表の通り。

1 段階	テスト細目作成	6 段階	合否判定基準の決定
2 段階	テスト問題作成と配列・調整・編集	7 段階	妥当性の検証
3 段階	テストの予備実施と結果の分析	8 段階	テスト得点の報告
4 段階	テスト実施者と採点者の育成	9 段階	テスト改訂
5 段階	テストの信頼性検証		

□また、以上の段階を理解することが優れた言語テストが満たすべき基準の理解につながるとし、テスト細目を中心に以下のように詳述している。

①テスト細目の作成

■**テスト細目(test specifications)**：テストで何をどのような方法で測定するのかを明示した設計書。これはテスト作成者にもテスト使用者にも必要なものである。

■作成者

例) テスト作成：受験者・内容・方法・問題数・時間などの情報を要する。また、予備実施後の調整・編集にもテスト細目が必要となる。

例) 妥当性検証：テストの基盤となる言語理論あるいは言語能力理論が必要→構成要素の決定がいかに行われたかが大切

■使用者

例) クラス編成：受験者を適切なレベルのクラスに振り分けるには、テスト得点が意味するものが明確でなくてはならず、またテスト得点の解釈にはテストの目的が大いに関係する。

例) 受験者の入学許可：大学への入学許可を決定する場合、入学後の言語使用の目的や要求される言語能力とテストの目的や内容が合致している必要がある。

例) テスト準備用教材作成：模擬テストを作成する際には対象受験者のレベル・テスト内容・テスト問題などが本番のテストと合致している必要がある。

②テスト妥当性検証のためのテスト細目

■言語テスト設計を適切に行うためには、言語理論・言語学習・言語習得・言語使用などの総括的な理論が必要不可欠であり、妥当性検証を目的とした細目の内容は使う理論的枠組みによって異なる

例) リーディングテストに必要となるのは→対象測定応力、各測定能力間の関係、テスト方法、テスト方法の測定値への影響、採点基準 など能力の測定に有意な影響を与え得るもの

③テスト使用者のためのテスト細目

- テストの誤用を避けるため、細目にはテストの特徴・使用法・限界・対象受験者を明確に記載する必要がある。
誤用例) 学期の最初と最後に TOEFL などの熟達度テストを行う
- また、テスト細目には問題例、指示なども載せておくとよい。さらに、ある得点を取った受験者が言語を使って実際に何ができるのかといった情報も役立つ。過去に受けた受験者の成績結果、その基準、基準の使用法なども記載すれば教員にも学生にも役立つ。

《授業でのご指摘に対する補足と感想》

2.5 Alderson, Clapham, & Wall (1995)

③テスト使用者のためのテスト細目

- テストの誤用を避けるため、細目にはテストの特徴・使用法・限界・対象受験者を明確に記載する必要がある。誤用例) 学期の最初と最後に TOEFL などの熟達度テストを行う
- また、テスト細目には問題例、指示なども載せておくとよい。さらに、ある得点を取った受験者が言語を使って実際に何ができるのかといった情報も役立つ。過去に受けた受験者の成績結果、その基準、基準の使用法なども記載すれば教員にも学生にも役立つ。

〈アメリカにおける実際の誤用例〉

□ 有名なテスト誤用の例として挙げられる事例は、"wall chart" の State school systems のランキング上に、その他のデータと一緒に the Scholastic Aptitude Test (SAT) のスコアを掲載したというものである。なぜ誤用かという・・・

- ① SAT はさまざまな教育的背景を持つ志望者を、大学の新入生としての個々のパフォーマンスにおいてランク付けするために設計されているテストである。従って志望者が出身高校で受けてきた教科課程や指導や学問的な厳格さは無視されており、志望者が与えられたカリキュラムをどれだけ達成していたかは考慮されていない。従って対象の学校がどれだけ効果的な指導をしているかということの指標にはならない。
- ② SAT を受験しているのは国内の 3 分の 1 の生徒に過ぎない。従って国内すべての生徒に提供する、教育の質の指標としては不十分である。

《感想》

テスト誤用に関する文献を当たってみると、予想していたよりも多くの例が挙がっており誤用例が多いことが分かった。テスト作成者がテスト細目を緻密に作成し、提示することは不可欠なことである。しかし今後は、提示された細目表をしっかりと確認し、テストの目的・対象を把握した上でテストを使用することの重要性を、より広く使用者に浸透させていかなければならないと感じた。

〈参考文献〉

U.S. Congress, Office of Technology Assessment. (1992). Testing in American Schools: Asking the Right Questions. *OTA-SET-519*.