2012/10/24

中村優治 (2010). 「テスト開発」. 石川祥一・西田正・斉田智里 (編著). 『テスティングと評価 4 技能の測定から大学入試まで』 (87–115 頁). 東京: 大修館書店.

第４章　テスト開発

4. テストの目的と種類 (pp. 104-106)

■ テスト細目の作成において, テストの実施目的・理由を記述することが重要になる.

■ ここでは, テストの理由となる「テストの種類」について述べる.

4.1 Alderson, Clapham, and Wall (1995) による「テストの種類」

(1) クラス編成プレイスメントテスト: placement test

■ 受験者の能力を測定し, 各人にもっとも相応しいレベルのクラスに割り当てるためのテスト.

■ テスト内容は, これまで or これから受ける授業のカリキュラムに基づく場合がある.

■ 総合得点でのクラス分け, または技能別得点でのクラス分けができる.

(2) 到達度テスト: achievement test

■ 課程の最終チェックとして, 学習がどれくらい進んだか, 教えたことがどれくらい学ばれたかを調べるために行うテスト.

■ テスト内容は, 課程のシラバス, 教科書の内容, または課程の目的と一致する.

(3) 熟達度テスト: proficiency test

■ 特定の学習カリキュラムを前提とせず, 学習者があるレベルに達しているかどうかを検証するために行われるテスト. あるいは, 特定の領域 (ESP) において十分な言語使用能力を持っているかどうかを確認するために行われることもある.

(4) 診断テスト: diagnostic test

■ 学習者に欠けている能力を特定するために行うテスト. 広義には, ある学習者が特定の言語技能について特別な補習が必要かどうかを確認するために使われる. 狭義には, 特定の領域 (文法のある項目) に弱点があるかを特定することなどを目的とする.

■ 診断を目的としたテストは少なく, (1)〜(3) のテストを診断の目的で使うことも多い.

4.2 Hughes (2003) による「テストの種類」

(1) 熟達度テスト

■ 受験者がそれまで受けてきたトレーニングに関わらず, その時点での言語能力を測定するためにデザインされる. テストの内容は, 授業の内容・目的に基づかない.

(2) 到達度テスト

■ 到達度テストは授業と直接かかわるものであり, 個々の生徒・クラス全体・授業自体がどの程度その目的を達成できたかを見極めることが目的となる.

■ テストには最終到達度テストと中間到達テストの 2 種類がある. 最終到達度テストの内容は (a) 授業の基盤となるシラバスと実際に使用された教科書・教材に直接基づくべき, (b) 授業目標に準拠させるべき

という 2 つの考えがある.

■ (a) の考えでは, 生徒が習ったと思われることしか出題されないという点で「公平だ」という利点があるが, 欠点として, シラバスや教科書が不適切な場合にテスト結果が授業の目的を十分に達成していないということになる.

■ (b) の考えでは, 授業の目的を生徒がどの程度達成できたかが, テスト結果から明らかになるという利点がある. ただし, テスト内容を授業でやったこと以外のものにすると (e.g., 教科書本文を使ってテストを作成しない), 生徒に対して公平ではなくなる.

➔ テストが授業の目標に基づいていれば, 結果からより有益な情報を得られる. そのためテスト内容は授業の目標に基づいたものにする方が良い.

(3) 診断テスト

■ 学習者の得意・不得意な分野を明らかにするものであり, どんな学習が足りないかを確かめることを目的とする. 4 技能の診断であれば, 既存のテストでもこの目的に合うものがある.

■ 純粋に診断を目的とするテストを作成することは非常に難しく, 膨大な量の項目が必要となる. DIALANG のような Web で使えるテストが開発され, 4 技能 + 語彙の診断が可能になっている.

* DIALANG: ランカスター大で開発された診断テスト (http://www.lancs.ac.uk/researchenterprise/dialang/about/)

(4) レベル分けテスト (プレイスメントテスト)

■ 学生をある指導プログラムの中で最も適切なクラスに配置するために役立つ情報を得るためのテスト.

■ 市販のものもあるが, 指導プログラムと市販テストの内容が合致する場合以外は自作すべきである.

5. 構成概念 (pp. 107-112)

■ テストの細目・目的に加え, 測定しようとする能力の構成概念を記述することが必要になる.

■ 構成概念は, 心理言語学的な観点から定義されるよりも, 現在は社会的な部分から定義されるようになってきている.

■ また, 個々の能力よりも状況における相互作用に焦点を当てたダイナミック評価や, 理論的な能力の概念よりも操作的な言語使用に重きを置くパフォーマンスモデルという考え方が広まってきている.

■ そして, 最新の理論を踏まえて作成された典型例が欧州言語共通参照枠 (CEFR) だと言える.

➔ 本稿では構成概念の基本的な考え方について, Thrasher (2002) を参考にする.

(1) 我々の構築した構成概念はその能力を測るのに必要なすべての要素を, また必要とされる要素のみを含んでいるか.

(2) 上記の要素は正しく, 正確に重みづけがなされるのか.

(3) 上記の要素はテストのタスクと現実の行動との対応が確実になされるか.

(4) 採点計画は測定しようとしている能力を測定するのにうまく機能するか.

(5) 受験者が最善の結果が出せるようなテストになっているか.

(6) テスト結果は一般化できるか.

5.1 ライティング能力の構成概念

■ ライティング能力の構成概念は次の 4 つの側面から考えることができる.

(1) ライティングの本質

・ライティング能力の本質は①文字を通して相手に意味を伝える能力, ②自分の考えを表現するために言語

的知識・能力を使用できる能力, ③考えを論理的に構成する能力と特徴づけられる.
・さらに, テクノロジー (パソコン等) を通したコミュニケーションの中で要求されるライティング能力の本質について考える必要がある.

(2) ライティングに関する言語理論 (応用言語学や第二言語習得論)
・テストは言語理論に基づいて開発されるべき (Buck, 1994; Weir, 1993) であり, 様々な言語理論・第二言語習得理論などが関わる (e.g., 言語理論: ジャンルなど, 第二言語習得論: ライティングプロセスなど).

(3) 既存のライティングテストのテスト形式 (フォーマット)
(4) ライティングと他の技能 (リーディングやスピーキング) との関連性
・TOEFL-iBT のテスト形式は, 北米の大学で実際に行われているタスクに近い.
・4 技能を統合した (integrated-task) テスト形式では読んで書いたり, 聞いて書いたりする複合的能力を測定しており, 伝統的なテスト形式では 1 題トピックに関する独立課題 (independent-task) が出される.
・integrated-task の評価は, 構成, 文法, 語彙の正確で適切な使用, および内容の完成度, 正確さを含む.
・independent-task の評価は, 論の展開, 構成, 文法語彙の正確かつ適切な使用を含む.
➔ 日本においては, 学習指導要領と照らし合わせてライティングの構成要素を検討する必要がある.

5.2 スピーキング能力の構成概念
■ スピーキング能力の構成概念は次の３つの側面から考えることができる.
(1) スピーキングの本質
(2) スピーキングに関する言語モデル
(3) 教師の経験
■ スピーキングテストには, (a) 直接, (b) 半直接, (c) 間接という側面からの見方がある. さらに, グループ, 集団討論テストという実施形態でのスピーキング能力の評価も進んでいる.
■ 今後はテストの新しい実施方法, CEFR などの見方を取り入れていく必要がある.

5.3 リスニング能力の構成概念
■ リスニング能力の構成概念は次の３つの側面から考えることができる.
(1) リスニング能力の本質
・聞こえたことを理解する能力
・聞こえたことと知っていることを結びつける能力
・新しい情報を既存の情報の中に取り込む能力
・すでに知っていることと新しい情報を関連付ける能力
・聞こえたものを理解するために, また聞き逃したものを理解しようとして, 言語的知識 (音, 文法, 語彙, 文脈) を活用できる能力
・科学技術の進歩によってもたらされた新しい側面 (コンピュータを用いた新しいリスニングの側面など)

(2) 言語理論的側面
・リスニングにどれだけの要素がどのように関連し合っているのかは明らかにされていない.
・これまでの研究では, メタ認知, 音声区別力, ワーキングメモリ, 既存知識, 読解力がリスニング能力のモデルを構築する上で有効な要因であることが分かっている.

(3) 既存の標準化されたテスト

・複合能力を測定する統合的技能タスクが登場し, authentic なタスクをテストに導入することでコミュニカティブテストの目標に近づきつつある.

・一方, 言語能力の測定には 1 能力 1 タスクという考え方が有力であり, その妥当性を追求すべきかどうか検証する必要がある.

5.4 リーディング能力の構成概念の考え方

■ Nakamura (1998) によるリーディング能力の構成概念.

① 真正性の高い資料の読解能力, ② 日本語への翻訳力, ③ 速やかな状況判断能力, ④ スキミング・スキャニング能力, ⑤ 文と文のつながり (ディスコースレベル) の読解力, ⑥ 英語での要約力

■ Grabe (2000) によるリーディング能力の構成概念

① 言語処理の速さ, ② 言語処理能力の高さ, ③ 言語処理の上手さ, ④ リーディングは相互的に行われる, ⑤ リーディングには目的がある, ⑥ 言語知識の十分さ, ⑦ 広い背景知識, ⑧ 課題遂行のための時間

➔ 実際のテストで測られているもの, 測られていないものを意識する必要がある.

6. 実践テスト作成手順・おわりに (pp. 112-115)

■ レベル分けテストの開発手順の例

1. 問題の所在: 学生を 4 つのレベルに分ける. 授業の目的は読解力の向上であり, 文法・語彙力のチェックも同時に行いたい. テストは 60 分以内で実施し, 採点・解釈が早く容易に実施できるものにしたい.
2. 細目規定:

   <言語操作> 文法力・語彙力を問われ, 速読と精読の両方の能力を問う. メインアイデアの特定・情報の探索・論理構成を把握する能力が問われる. 未知語・照応関係を読み取る能力も問う. **<テキストタイプ>** 母語話者向けに書かれた学問的なもの. **<テキストの長さ>** 350-450 語. **<トピック>** 内容は可能な限り中立なもの. **<リーダビリティ>** 特になし. **<使用文法>** 高校修了時. **<語彙>** 専門的でない学術用語を含む. **<構成>** ３セクション (文法・語彙・読解). **<項目数>** 文法 15, 語彙 15, 読解 20. **<時間>** 50 分. **<回答方式>** ４肢択一方式. **<採点>** 客観方式. **<信頼性>** 全体で α = .70 以上. **<妥当性>** 内容妥当性 (授業目標・シラバスに基づく), 併存妥当性 (他のテストと), 予測妥当性 (授業の評定との比較), 基準関連的妥当性 (教員・学生の判断).

➔ クラス分けの適切性を示す基準関連的妥当性が重要になり, クラス分けのミスが少ないほどこのテストの妥当性は高いということになる.

■ 言語および言語能力をどのように記述するか, どの言語をどのように教えるかは絶えず変化するものであり, その変化に合わせて, 言語能力はどのような方法を用いればもっとも正確に測定できるのかを考える必要がある.

■ テストを変更するには, テストの内容・方法を定期的に改訂し続けなければならない. こうすることでテストの形骸化を防ぐことができる (Alderson, Clapham, & Wall, 1995).

■ テスト開発者・研究者が必要とする原理は, テストの妥当性・信頼性・実用性・実施可能性・波及効果などであり, 理論上の理想と現実的制約との折り合いをつけた開発・改訂が必要になる.

■ しかし, 最も必要な根本原理として, テスト作成の目的・理由を明確にすることを念頭に置かなければならない.