

Chapter 5 Designing test specifications

1. What are test specifications?

- 本章では項目デザイン (item design) と試験仕様書の作成 (writing test specification) を詳細に見ていく (図 4.1 参照)。
 - これらはテストの「青写真」とよばれ、建築や工学のそれと類似する。

 - 試験仕様書には多くの形式があり、テストの目的、構成概念、項目タイプ、項目数の statement を説明する記録である。
 - Mislevy, Almond & Lukas (2003) によるテストに必要であり得る試験仕様
- ① Item/task specifications:
 - ・ 目標とする学習者の能力についての推論の根拠となる証拠を引き出すために設計された prompt について書かれる。
 - ・ テスト受験者がどのような input マテリアルに出会うか、また指導がどのようなものであるか、その他重要な prompt の特徴
 - ・ この仕様書はタスクを変更する方法や修正の許容範囲の特徴を含み、項目作成者がどの程度自由に作成できるかを説明する (そのためにテストの意図を説明する試験仕様を伴った項目のサンプル等が通常与えられる)。

 - ② Evidence specification:
 - ・ テスト受験者のそれぞれのタスクにおける期待される反応 (response attribute) とそれらの反応をどのように得点化するかについて書かれる。
 - ・ closed-response 項目については正解・不正解で採点されるが、performance タスクについてはより複雑なスケールや採点のためのガイドが必要。

 - ③ Test assembly specification:
 - ・ テスト全体がどのように構成されるかについての指導書
 - ・ 現実世界での基準の特徴を適切に代表する項目の幅広さと項目数
 - ・ 例) いくつの項目タイプがテストには必要か、文脈やトピックによって項目が符号化されている場合の反応に必要な語用論的知識の程度、基準となる領域の中で、それぞれのカテゴリーからいくつの項目が必要か)

 - ④ Presentation specification:
 - ・ 項目と支援マテリアルがテスト受験者にどのように提示されるかを test production team に伝える。

- ・ ペーパーテスト →余白、フォント、文字サイズ、スペース、ページ番号の位置等を規定
- ・ コンピュータ →色、スクロール、一画面あたりのテキスト量等を規定

⇒多様さや変動を限定することにより、無関係であるはずの要因からスコアにゆらぎが生じるのを防ぐ

⑤ Delivery specification:

- ・ テスト運営やセキュリティーやタイミングなどの詳細について規定する。
- ・ 例) 机と PC の間隔や監視員の人数、テスト中の使用可/不可の道具、テスト・テスト付随する活動の時間配分

■ 上記の仕様書によって具体的なテスト運用における「テスト形式(test form)」の作成が可能となる。

■ 試験仕様書を実際にテストにした場合、それらは test version ではなく、test form にあたる。

※ test forms →同じ項目数や構成概念、困難度、タイミング等条件を揃えられていて、parallel な性質を持つため、何度も同じ結果が得られる

※ test versions →テストによる測定の改良を目指して、更新されていく (例: gender の要因が反映されないように改善される)。新しく作り出された test version から parallel な性質を持つ test forms が作り出される (図 5.1 参照)。

■ 初期の試験仕様書に関する議論として、Ruch (1924) では上記の①-⑤の仕様を含むべきと主張されている。

■ また、Yoakum & Yerkes (1920) の army test では内容と統計的に parallel な複数のテスト形式が用意されていた。

■ Burt (1922) のタスク説明の例 (Understanding Simple Commands) と Burt (1923) のオープン・タスク説明の例 (Describing Pictures) に示される試験仕様書の特徴

- ① テストが何を意図しているかが述べられている。
- ② 手順が述べられている。
- ③ どのような反応が期待されるか、またどのような反応まで理解の証拠としてカウントするかが述べられている。
- ④ Presentation specification に関する情報が述べられている (e.g., 写真の大きさ)。

→提示法に関する研究は近年の TOEFL のコンピュータ化の際に、リスニングテスト時に visual image を画面に提示しておく等のテスト開発に示唆を与えている。

■ まとめとして、試験仕様初期からある考え方であり、テストが測定法として信頼性を持つ (つまり、テスト形式が parallel な性質を持ち、同じ結果を得ることができる) ために必要である。試験仕様が進歩してきたように、テストに使用する instrument も進歩しており、さらに

妥当性の高い (見たい構成概念を検証できる) テストへ向けて更新されている。

2. Specification for testing and teaching

- 第2章では基準準拠テストの重要な原則として、テストとそれによって予測される現実世界のその領域(の能力) のつながりについて見た。
- Glaser (1963)においても、テスト結果の説明・測定のために、テストと指導のつながりが考慮され始めている。
- Hambleton (1994: 23) によると、
基準準拠テストの実施は、意図していた指導の結果を説明することである。教師やテスト開発者が測定される知識やスキルを明確に説明することで、妥当なテスト項目の作成、評価項目の作成 (item-object congruence¹)、テスト得点の解釈のために必要な枠組みがもたらされる。

※1. item-object congruence はテストで測定される項目やタスクと学習事項の関係を表す。

項目の specification によって学習事項は定義されるため、これは item-spec congruence や item/task fit to spec と呼ばれる。

- specification を作成することによって、言語テスト作成者は、その項目の使用やその項目が何を測定するよう意図されているかを明確にしなければならなくなる。
- 適切な specification は、その項目やタスクの成功から構成概念を、また構成概念から基準を予測するための論理的根拠の明確な記述を持つ。
- Popham and Huseck (1969: 3) は、基準準拠テストの妥当性を調査する上で重要なのが、specification によって表される学習事項・項目・結果などのつながりであるとしている。
- 図 5.2 (p.135) は Popham (1978) のテストの specification のためのフォーマットである。specification に必要な要素として、主に以下の5つを挙げている。

(フォーマットの最初には、タイトルや整理のための番号などが記載される)

(1) General Description (GD):

- テストで測定しようとする意図している構成概念や目標とする言動についての記述
- 教室内テスト/タスクで使用する場合には、タスクの形式をシラバスに関連づける GD が必要である。この記述は特に螺旋型シラバスを複数の教師で行う場合に役立つ。

(2) Prompt Attributes (PA):

- テスト受験者の目標である反応を導き出すために、どのような説明を受け、どのようなインプットを受け取ることが必要かを定義する。
- PA は使用マテリアルの出典や難しさに関する情報を持つ (e.g., テキストタイプ・ジャンル)。

(3) Response Attributes (RA): a) Selected Response or b) Constructed Response

- テスト受験者による期待される反応がどのようなことか明確に記述する (=正答)

(4) Sample Item (SI)

- 上記の specification に基づく答えの具体例についての情報
- anti-item の提示
- anti-item によって、項目作成者がつくるべきでない問題に示唆が与えられる。適切でない問題が作られる毎に、それを防ぐため SI の記述が見直される。

(5) Specification Supplement (SS)

- 項目/タスク作成者が parallel な項目/タスクを作成できるように、示唆を与えるその他の情報
- 例 1) Speaking test では面接官が持つべきトピックや質問の枠組みがここで提示される。

- このようなテンプレートに基づいた specification により、テストの目的の明確化が可能である。

→例: CEFR A1 に基づき作成された specification の例 (pp.137-138)

CEFR A1 Guiding language (一部例)

- (1) GD: シンプルな処理が習得される →Can ask people for things and give people things. 等
- (2) PA: オーラルの刺激は録音により行われる。
- (3) RA: 受験者は 4 つの選択肢から正答を選択する
- (4) SI: 複雑な統語構造の文を使用することは避けること

- まとめとして、specification によって、多数あるテスト形式において、多数の項目/タスク作成者による、parallel な問題作成が可能になる。specification を詳細に作成することで、テスト項目と構成概念のつながりを明確になる。
- また、教室内テスト/タスクにおいても、教師が協同して螺旋型の指導/テストを行う際の、重要な指針となる。