

Flucher, G. (2010). *Practical Language Testing*. London: Hodder Education.

Chapter 6.1-6.2

6.1 Investigating usefulness and usability

- (1) テストの仕様書とサンプル項目は、テストの設計に関する協議を通して精査される。
 - (2) ここで精選された項目が *prototyping* とよばれる次のステージに進む。このステージでは2~3のグループによってタスクがどう機能するかが検証される。
 - (3) タスクが想定通り想定したら、パイロット調査に十分な量の項目が作られる。ここでは統計的な調査のために多人数での検証が行われる。
 - (4) タスク、項目、そして作成の仕様が上手く機能するようならテストを実際使用が可能な状態 (*operational*) に仕上げる段階へと移行する。テスト実施の **GO** サインを出すか否かを決定するのがこの段階である。
 - (5) **GO** サインが出たら、テストの大量生産が始められるよう、タスクや項目のまとまりが作成される。それぞれの項目は生産されるたびにチェックされる。プールに使用可能な項目が加われば、それはテストフォームを作るために使用される。
- このプロセスを踏むことは面倒だが、テストの妥当性の証左を与えてくれる点で非常に重要なものである。教室環境においては、ここに含まれる多くのステップが省略されがちだが、最初のステップは教師にとって特に重要なものである。よって以下ではテスト仕様を決める協議で、サンプル項目と仕様書がどのように評価されていくかをみる。

6.2 Evaluating items, tasks, and specifications

- 協議の目的は、測定したい構成概念に関する有益な情報を与える項目を選定することである。例として実際のテスト仕様協議をあげる (pp. 162-170)。これはコンピュータベースのプレイスメントテストの作成に関するもので、以下のステージに分割される。
- (1) 教師がグループを作り、アイスブレイキングの活動を行う。
 - (2) 構成概念を見直し、タスクや項目仕様書をサンプルアイテムとともに作成する。
 - (3) グループでサンプル項目を交換し、逆光分析を行う。
 - (4) (3) でみた項目を仕様書に照らして批判的に検討する。項目を作成したグループにフィードバックを与えるため、それぞれの項目が仕様書から合理的に作成されているか否かを検討する。
 - (5) 各項目作成グループが仕様書のフィードバックを受け、コメントに返答する。
- 中でも重要なものが (3) の逆行分析と、(4) の項目と仕様書の一貫性チェックである。後者のステージにおいて評価グループは、項目が仕様書と合致しているか、両者が共に機能するかを確認する。
 - では、実際に項目の協議をみる。以下に示すものは作成グループが作ったテスト項目である。対象とされる構成概念は「簡明かつ単純なメッセージやアナウンスで述べられる事柄を特定する能力」である。コンピュータによるテストなのでマウスとキーボード操作が必要であることを留意すること。

スクリプト: p.161 参照

- これを留守番電話のメッセージにしたいと思います。
- 各行で正しくない語や数字をクリックしなさい。
- 正しいものを 1~6 に入力しなさい。
- 音声は 2 回流れます。
- 答えは好きなタイミングで入力できます。
- 1 度音声 flowed 後、答えを入力するために 60 秒が与えられます。

City Ticket Agency

0	Event	Jazz concert
1	Place	Regent Cinema
2	Address	Bank Road
3	Time	8:30

Ticket bought

4	Price	10 euros
5	Number	5
6	Seat (s) row	front

1 () 2 () 3 () 4 () 5 () 6 ()

- まず自分たちでこのテストを逆光分析してみましょう。
- 次に逆光分析についての協議のダイアログを見てみる。彼らはこれが何のテストなのか、どのような構成概念を対象としているのかを知らない点に注意すること。

(1) 第一段階 (pp.162-164)

- 口を揃えて **complicated, complex, difficult** と評されている。協議の初期段階では、構成概念を特定するどころか、受験者が項目に対してどう回答するのかという次元で論が停滞している。この点に関して、彼らはコンピュータを使うことだけでも受験者にとっては難しいと感じている。

(2) 第二段階 (pp.164-165)

- スクリプトを読み上げて実際にテストをやってみようとして試みている。結果、重大な問題が 2 点浮上した。
- (a) 生徒の答えをタイプさせる回答形式では、スペリングのテストになってしまう可能性があること。リスニングの構成概念から逸脱してしまふ可能性がある。
- (b) 受験者が正しくないものをマークし修正するためには、正しいものをまず聞き取る必要があること。つまり、1 度スクリプトを聞いてからでないと回答は不可能な仕様になっており、記憶力によって成績が左右されるおそれがある。

(3) 第三段階 (pp.165-166)

- ここで興味深い点は、評価グループ内で「留守番電話のメッセージ」が何を意味するのか、という点について意見が分かれていることである。招待だととらえるものもいれば、映画館からのメッセージと考えるものもいる (p.165)

(4) 第四段階 (pp.166-167)

- ・評価グループは以下の問題点に言及している。

(a) リスニングテキストの長さ (p.166)。

(b) 2度目に音声流れる際、受験者は何をすべきなのか (p.167)。

→①答えを書く

→②書いた答えを確認する

→③理解内容を確認する (この後に与えられる 60 秒の間で答えを書くため)

→最後の場合、やはり記憶力が問題になる。

(c) レイアウト上の問題

→答えを入力するテキストボックスが、設問箇所と離れているのはなぜ？

- ・また、1st Part (間違っている語や数をクリック) と 2nd Part (正しい答えを入力) の目的も考察している。

→1st Part は 2nd Part のあくまで前提として位置付けられるものなのか？もしそうなら、1st Part は採点対象から外されるべきでは？

- ここで彼らは項目仕様書を与えられ (pp.168-169)、これを参照しながら項目を再検討するステージ 4 に移行した (pp.169-170)。

リスニング訂正タスク
総合記述
リスニング、リーディング、スキミングが統合された項目。特定の情報を聞き取る能力。音声を聞き、テキストと音声の差異を特定、訂正する。
プロンプト特性
チュートリアルと例が与えられる。音声は時間、数、イベント、場所、値段、電話番号などの情報を伝える。各行に誤りがあり、生徒は誤りをクリックし、テキストの下部にあるスペースに正しい情報を入力する。6つの誤りのうち、3つが数字で、残りが語である。レベルは音声速度のよって変更可能である、
回答特性
生徒は音声を聞き、テキスト内の正しくない情報をクリックする。そして正しい語や数字を入力する。音声は2度流れ、解答の入力に1分与えられる。
補足
サンプル問題は中間レベルに合わせてある。複数の解答パターンを用意する必要がある。スペリングのミスもある程度は許容する。数字はアラビア数字で書いてもスペルアウトしても良い。

■ ステージ 4 の概要

- ・評価グループは項目に関する問題をまとめていた
 - ・項目の難易度が調整できるのかについて意見が分かれていた
 - ・回答特性に2つのステップが含まれることが、デザインの大きな問題点であることを合意した
 - ・スペリングの問題に関して、どの程度まで許容するのかの線引きが難しいとされた。
- 結果、評価グループは問題形式を多肢選択式に変更することを提案した。

- このように、グループで協議してテストデザインを考えることによって、デザインした者が自分では気付けないような問題点を明るみに出すことができる。現にデザイングループはこの後、多肢選択式の提案を快諾した。
- 注意すべき点は、この例のテストはデザイングループが「創造力」を働かせて、新規性のある問題を作成しようとした結果生まれたものだという点である。「創造力」はデザイン初期の段階では確かに重要だが、思いつきから新しいものを生むことの難しさも認識しなければならない。
- 対して評価グループは、作成された項目を分析、「創造的」からは程遠い多肢選択式問題を提案した。そして最終的にはこの提案が優位を得たのである。
- この例と Fulcher and Davidson (2007: 56-57) が示す通り、教師はテストやタスクの「典型」を念頭に置くべきである。多肢選択式問題のような「典型」が長く用いられてきたことには理由があるのである (e.g., 作成や採点が容易、目標準拠基準テストに応用できる)。このような「創造力に反する力」に意識を向けることは、テスト作成の初期段階で不可欠な作業である。

☆討議した内容

- 多肢選択式の問題点はなにか？
 - ①ストラテジックに正解できてしまう場合がある。
 - ②錯乱肢がヒントになる可能性がある。
 - ③弁別力のある錯乱肢を作るのは難しい
- 以上のような問題点があるものの、作成・採点が容易な点を考慮すると、多肢選択式問題の使用を避けるよりも、どう上手く使用するかを考える方が賢明だといえる。つまり、問題の質をあげることが重要だといえる。
- しかしながら、現状の英語の試験における多肢選択式問題は、必ずしも良質なものばかりとはいえないのではないだろうか。英検の読解セクションの選択肢を読解理論に基づいて分析した高木 (2011) は、理論的に記憶に残りやすいといえる錯乱肢はほとんど見られなかったと指摘している。
- では、良い選択肢を作成するには何が必要なのであろうか。まず正解の選択肢に関しては、正解となる記号に偏りがなく (i.e., c,c,c,などとならない)、唯一の正解だと特定できることが条件としてあげられる (e.g., Haladya & Dowing, 1989)。より大きな重要性が置かれる錯乱肢に関しては、(a) 非論理的な文言を避けること、(b) 受験者が頻出する誤りに基づいて作成すること、(c) 専門用語を避けること、(d) よく認知されているが正解にはならない語句を用いること、(e) 真実だが不正解につながる事象を使用すること、(f) ユーモアを避けること、等に留意すべきとされる。
- 十分に練られた多肢選択式問題は、受験者の処理プロセスを自由回答式問題と同程度に深めることができることが示唆されている (e.g., Ozuru, Best, Bell, Whitherspoon, & McNamara, 2007)。良質な多肢選択式問題は妥当性の高い測定を可能にするだけでなく、受験を通して学びを深める「テストからの学習」にもつながることが期待できる。