【2013/11/27　異文化言語教育評価論】　　PLT Ch. 6 Sec. 6 ~ 8　Y.M.

6. Field testing

■ field test とは、サブテストの数や各サブテストのテスト項目数など assembly model が決定されると、完全な形のテストが大規模に試されることをいう。サンプルサイズは大きく、テスト受験人口を反映する。

■ field test により項目分析で安定した推定ができるが、複雑な統計分析も可能である。

・ 母語や性別、障害などによってテスト得点に偏りがないか分析する。

・ 異なる構成概念を測定するようデザインされたサブテストがデザイン通りに機能しているか、テストの内部構造を因子分析で分析する（手計算では不可能なので、SPSS などで行われる)。

→ 収束的妥当性（同じ構成概念を測るテスト成績は高い相関がある）

→ 弁別的妥当性（異なる構成概念を測るテスト成績の傾向は異なる）

■ computer based などの場合に、テスト受験者がテスト機材を使用できるかを確認する（特に弱視など障害を持つ受験者)

■ テスト全体やその構成要素に割り当てられた時間が受験者にとって十分かを確認する。時間が不十分だと、意図した構成概念ではなくスピードを測るテストになってしまう。逆に時間が長すぎると、テスト作成者のコストが高くなり、受験者のモチベーションを下げることにもつながる。

7. Item shells

■ 大規模テストでは field test 終了時に、テストに含める / 除外すべきもの、受験形態、制限時間などが決定される。この際、テスト仕様書に従って多くの項目やタスクを発展させる必要がある。そのためにテスト作成機関が作成者に、item shell を提供する。

■ item shell とは、項目作成者が新しい内容を挿入できる電子テンプレートのことである。これは変更できない標準テストの指示とレイアウトだけを最小限含んでいる。Word などのソフトで簡単に作成できる。

■ pp. 186-187 の例では、項目作成者は難易度に影響するテキストの特徴を含む何かのサービスや施設に対する不満を述べる記事を用意するよう、仕様書で支持される。

■ item shell の別の利点として、item pool に入れる前に新たな項目を確認する際、編集作業が少なくなる。

8. Operational item review and pre-testing

■ 項目やタスクから pool を作成する際、確認プロセスが行われる。問題があればテスト作成者が修正し、再度確認プロセスを行う。このプロセスは各項目が全てを通過するまで行われる。その後その項目は item pool に加えられ、最終的にテストに使用される。

■ 本章で紹介した手順は、learning for assessment の場合は全ては必要にならないだろう。しかし high-stake test で結果の解釈を保証しなければならす、本章で説明した手順が全て必要になってくる（信頼性・妥当性の確保)。

(1) Content review

・ 項目やテキストなどのマテリアルが仕様書に合っているかを確認する。

(2) Key check

・ 意図した答えが本当に正解になるかを確認する。多肢選択問題では、錯乱肢が本当にもっともらしくならないように注意しなければならない。記述問題の場合、採点用紙に正解になる範囲を明記し、予期される正答・誤答が全て含まれるようにしなければならない。

コメント [MY1]: 錯乱しとして機能させるにはもっともらしくする必要があるが、もっともらしくしすぎてしまうと正答になってしまうので注意が必要。

(3) Bias / sensitivity review

・ ある一部の受験人口を差別する可能性の高い引用やマテリアルが含まれていないかを確認する。これには文化的感受性も含まれ、high-stakes test では様々な文化に詳しい人物が確認を行う。Hambleton and Rodgers (1995) では designated subgroups of interest (DSIs) を特定し、教育的・文化的経験を超えるためにテスト受験者が不利益を被ることがないかを確認することを提案している。

(4) Editorial review

・ 最後にスペルや文法のミスがないか、フォーマットが適切かを確認する。

Activities 6.6 Item review

■ item 1

・ Key check: 錯乱肢 (a), (d) も正解になりうる?

・ Editorial review: Tony の発言は"It is going to take forever to pack for our holiday at this pace."の方が適切?

□ 解答例 (appendix 6)

・ 問題は Linda の応答を問う問題なのに、その手掛かりは Tom の発言の中にある。

・ Linda の発言に注目すると、Tom の発言で"at this pace"は入れる必要がなかった。

・ 正解の選択肢が一番長くなっている。 ・ 正解の方が錯乱肢より複雑な構文を使っている。

・ 錯乱肢 (c), (d) も正解になりうる。

■ item 2

・ Content review: 2 人の発言が載っているが、このテキストの形式は会話といえるのか? また presenter が少女の飲酒に関する問題提起をしているのに、それに対する Tom の発言は容認するものとなっている (一貫性がない?)。

・ Key check: 錯乱肢 (b) も正解になりうる?

□ 解答例

・ 10 代の若者に関するトピックとしては適切ではない。

・ 一部のテスト受験者に不快感を与える恐れがある。

・ イギリスの 10 代のライフスタイルや飲酒に関する価値判断が含まれてしまう。

・ Tom の応答がトピックと関係していないように思われる。

・ Tom の発言が口語的すぎで、照応関係が非常に複雑になっている。

■ item 3

・ Key check: 錯乱肢 (a) も正解になりうる?

・ Editorial review: × "I've invite" → ○ "I've invited" × "loft" → "left"

□ 解答例

・ 上記の文法やスペルの誤り ・ 錯乱肢 (a) も正解になりうる

・ 錯乱肢 (c) だけ否定文が使われている ・ 解答がイントネーションに依存し過ぎている

・ リスニングで登場人物が両方女性だと区別が難しい ・ 会話内容が不自然である

■ item 4

・ Content review: テキストが会話とはいえない ・ 錯乱肢 (a), (b) も正答になりうる?

□ 解答例

・ 語彙、稿文、長さの面からテキストが受験者にとって難しすぎる ・ 背景知識を多く必要とする

・ "excessively tighter"が非文法的である ・ "forcing more small"の方が適切 (and は不要)

・ 錯乱肢 (b), (d) はテキストと関係が薄く、錯乱肢として機能しない

コメント [MY2]: excessively「(ある基準と比較して) 過度に」と比較級で意味が重複するため?

コメント [MY3]: more が small-to-medium-sized を修飾する