

Q20 CAN-DOリストは一般的にどのような妥当性検証を行うのか？

春火 3 限 言語教育基礎 1
世界言語社会専攻 言語文化コース
高景綺 5019030
2020年6月16日

「一連のレベルが存在するということは、ある種のものはむしろ他のレベルに属させた方が良く、あるいは、特別な言語技能の級の記述はむしろ他のレベルにふさわしい」ということが起こりうる。（CEFR日本語訳 p219）

よって、CAN-DO ディスクリプタの割り当てに対して、妥当性検証が必要。

妥当性検証の方法

- 直観的手法
- 質的調査法
- 量的調査法

直観的手法（intuitive method）

- **エキスパート**：個人が資料を集めてスケールを作成。その後、予備調査を行い改善する。
- **委員会**（committee）：コンサルタントのグループがスケール開発の監督を行う。
- **経験主義**（experiential）：機関内部でスケールを試行し、フィードバックを行う。学習者のパフォーマンスを評価したり、逆にパフォーマンスを見ながら定義を見直したりする。

質的調査法 (qualitative method)

- 少人数グループでインフォーマントを交えて行う
- 統計的な解釈を加えない

質的調査法 (qualitative method)

- ・ **スケール設計上の重要概念の形成**

(key concepts: formulation)

a) インフォーマントにスケールの並べ替えをさせる

b) 並べた理由を聞き出す

c) 正解と差違があった場合、なぜそう判断したかを考察する

これにより、スケールの上下関係を判断する際の要因となる重要概念の項目を特定する。

質的調査法 (qualitative method)

- **重要概念とパフォーマンスの照合**

(key concepts: performance)

ディスクリプタと実際の典型的パフォーマンスを照合させてみて、一致しているかを確認する。

例えば、IELTSでは、各レベルの模範となるサンプルスクリプトを選定し、それらを分析した上でディスクリプタに特徴を盛り込むという作業を行なっている。

質的調査法 (qualitative method)

- **主要特性 (primary trait)**

複数のインフォーマントに、作文などのパフォーマンスを順位付けしてもらい、さらに集団で話し合って統一してもらう。どのような原理原則で統一順位が決まったかを記述して、特性として抽出する。

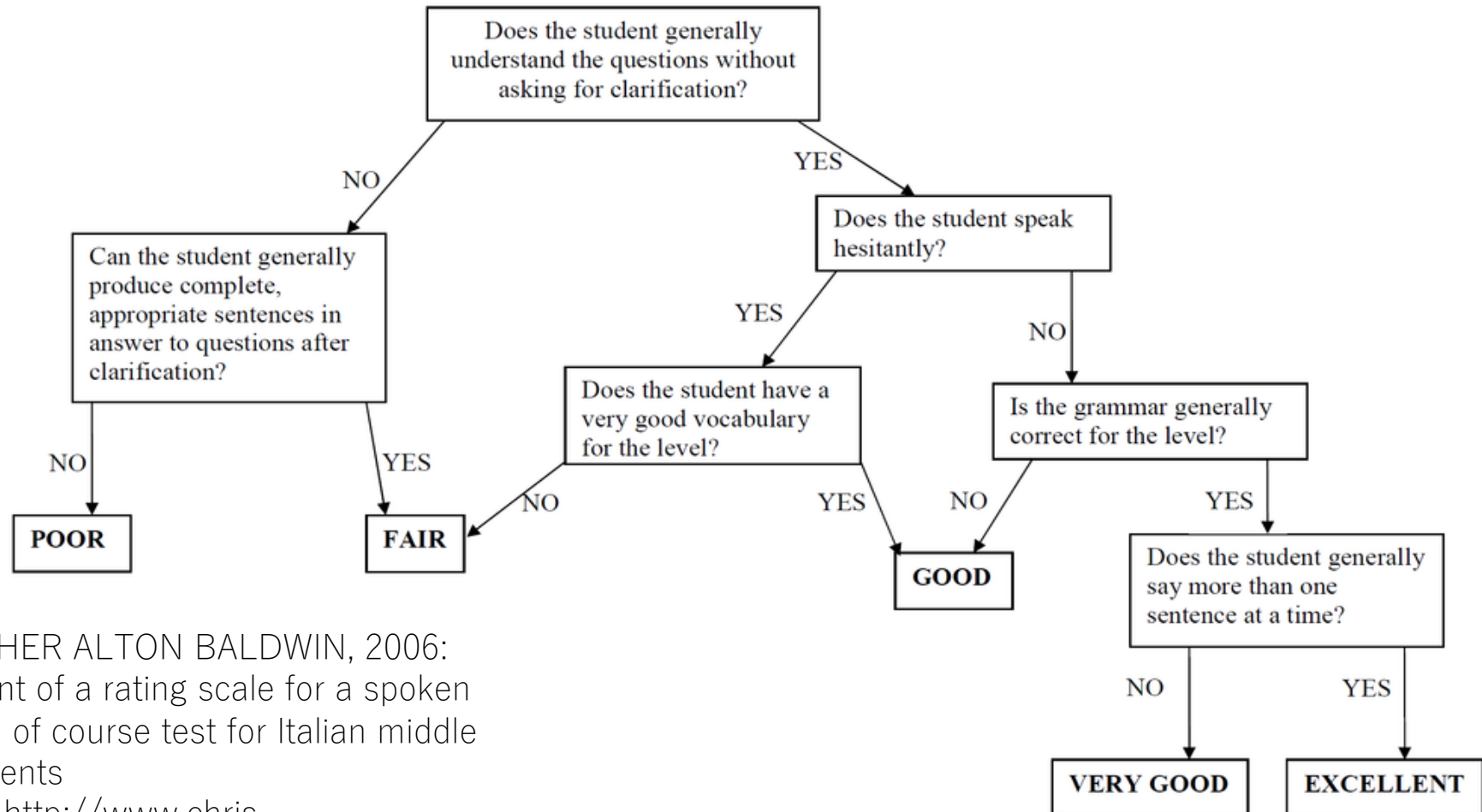
順位付けの仕方は他に、大体同じ作文の山 (pile) をいくつか作っていくなどの方法もある。

質的調査法 (qualitative method)

- **YES/NOによるレベル決定過程の作成 (binary decision)**

前述の特性を、YES/NOで答えられるような判定文に書き換え、決定木のようなツリー構造を作る。これによって、決定プロセスを客観的に示すことができる。

GRADING SCALE FOR ORAL TEST



CHRISTOPHER ALTON BALDWIN, 2006:
Development of a rating scale for a spoken
English end of course test for Italian middle
school students

アクセス：<http://www.chris-baldwin.com/rating-scale.html>

質的調査法 (qualitative method)

- **ペア比較による判断 (comparative judgements)**

異なるパフォーマンスをペアにして、どちらが優れているかを議論し、その過程で使用されたメタ言語をディスクリプタに集約していく。

質的調査法 (qualitative method)

- **並べ替えタスク (sorting tasks)**

ディスクリプタの草稿が出来上がった段階で、インフォーマントに並べ替えをしてもらい、コメント、修正、破棄などを依頼する。目的はディスクリプタの改訂である。

量的調査法 (quantitative method)

- 判別分析
- 多次元尺度法
- 項目反応理論

量的調査法 (quantitative method)

- 判別分析 (**discriminant analysis**)

多変量解析の一種で、学習データから判別モデルを構築し、それを用いて、個体がどのグループに帰属するかを判別する方法である。

説明変数Xにあたる特性変数を選定し、それらがサンプル中（または討議の談話中）にどのくらい現れるかを計量しておく。それから、目的変数Yを該当レベルとし、判別関数を作る。各変数についている判別係数を見ることで、各変数の貢献度がわかるので、貢献度の高い項目をディスクリプタに盛り込んでいく。

量的調査法 (quantitative method)

- **多次元尺度法 (multidimensional scaling)**

学習者の個体間の距離を統計的に求め、それを二次元あるいは三次元に配置する方法である。

パフォーマンスに関して、語彙の豊かさや統語の複雑さなど、様々な言語特徴データがあるとする。その評価データを統合して個体間距離を求め、二次元にプロットする。距離が近い個体の特徴を分析することでグループを分ける主要特性を抽出する。

量的調査法（quantitative method）

- **項目反応理論（IRT, Item Response Theory）または潜在特性分析**

IRTは多変量解析モデルのうち潜在変数モデルの一種である。確率論から発展したもので、主としてテスト課題の難易度を決めるのに使われる。

最もよく用いられるのはラッシュ（Rasch）モデルである。ラッシュモデルの計測方法は、上級者は低いレベルの問題を解けるが、初心者は上級レベルの問題に答えられる可能性が低いという単純な事実から出発して出来上がっており、これによって複数の項目を同一の尺度へと換算できる。

IRTの強みは、項目困難度はサンプルや特定のテスト・アンケートに依存しない点である。

量的調査法 (quantitative method)

- **項目反応理論 (IRT, Item Response Theory) または潜在特性分析**

言語能力尺度への応用：

(a) 質的調査法で得られたデータにラッシュモデルを適用し、項目困難度としてスケールに位置付ける。

(b) ディスクリプタを操作定義して特定のテストの問題項目として複数実装し、そのテスト結果をラッシュモデルで項目困難度を算出し、当該ディスクリプタの相対的困難度スケールとして定義する

(c) ディスクリプタをアンケートの質問項目とし、教師にある学習者について「これができるか否か」を評価させる。その結果を項目反応理論で数値化し、ディスクリプタの相対的困難度を決定する。

CEFR-Jの方法

- アルファ版の作成：チームでの討議（エキスパート、委員会）、専門家による質的フィードバック
- ベータ版の作成：小中高教員200名を対象に並べ替え調査、海外の専門家による質的フィードバック
- Version 1の作成：学生の自己評価アンケート（中高大の学習者5468名に対してベータ版を用いて「CAN-DO調査」）をIRTで分析し修正

ご清聴ありがとうございました。