

少数カテゴリの併合がパラメタ推定に及ぼす影響 —パフォーマンステストデータに多相ラッシュモデルを適用した場合—

○光永悠彦 (名古屋大学)
神澤克徳 # (京都工芸繊維大学)

羽藤由美 # (京都工芸繊維大学)

キーワード：項目反応理論，言語テスト，パフォーマンス評価

背景と目的

京都工芸繊維大学において行われている，英語発話能力を測定するためのテスト (KIT test, Hato et al., 2018 参照) では，受験者の発話内容が 2 名の評価者によって，二つの評価観点 (Task Achievement 及び Task Delivery) から評価される。発話パフォーマンスはルーブリックにしたがって複数の評価者によって評価され，多相ラッシュモデル (many-facet Rasch model, MFRM; Linacre, 1994) により尺度化される。KIT test では 6 カテゴリからなるルーブリックが用いられているが，(1) 最下位カテゴリに入る受験者数が少数 (10 名程度) になった場合，MFRM の推定が不安定になる恐れがある，(2) 最下位カテゴリに対応するカテゴリ境界パラメタの標準誤差が他のカテゴリにおけるパラメタよりも顕著に大きくなる (最大 10 倍程度) という理由により，最下位カテゴリをその一つ上のカテゴリと併合し，5 カテゴリのデータを用いて分析を行ってきた。本研究では，6 カテゴリのデータと 5 カテゴリのデータの結果推定されたパラメタの点推定値を比較し，カテゴリを併合することにより項目困難度や受験者の能力母数の推定結果がどのように変動するかを，最下位カテゴリと判断された頻度との関連に着目して検討する。最下位カテゴリと判断された受験者は能力が相対的に低い受験者が多いため，最下位カテゴリと判断された頻度が大きな項目において，困難度パラメタの点推定値も変動することが予想される。

方 法

2018 年 12 月に実施された KIT test による 661 名のデータを用いた。9 問からなる問題冊子を 3 種類用意し，受験者 (大学 1 年次生) には 3 種類の冊子のうちランダムに一つを割り付けた。また 66 名のモニター受験者に対し，3 種類の問題冊子すべてを解答させた。解答はすべて英語発話で行わせ，その内容を録音しておき，一つの解答につき評価者 (英語ネイティブ・ノンネイティブ) を 2 名割り当て，独立に二つの評価観点 (Task Achievement [TA] 及び Task Delivery [TD]) ごとに 6 件法で評価させた。

分析については (1) 「評価者」「評価観点」「項目」「受験者」の 4 相を仮定した MFRM (Overall) 及び (2) TA のみの評価で，3 相を仮定した MFRM (TA)，(3) TD のみで 3 相を仮定した MFRM (TD) の三通り

で行った。(1) から (3) のいずれについても，最下位カテゴリを一つ上のカテゴリに併合した場合 (NCAT5) と，元の 6 カテゴリの場合 (NCAT6) の場合で分析し，結果を比較した。いずれの分析も Facets 3.71.4 を用いて行った。ただし，評価者の属性を考慮し，評価者と評価観点との間に交互作用項を仮定した。

結果と考察

TA, TD それぞれで 14, 274 回の評定が行われ，そのうち TA で 78, TD で 47 個の最下位カテゴリの評定が下された。全体に占める割合はそれぞれ 0.55%, 0.33% であった。

項目困難度の推定結果を比較すると，Overall において最も最下位カテゴリと評定された頻度が大きかった (頻度=25) 項目においては，推定値が 0.01 の差であった。一方で，推定値の絶対値差は最大 0.04 であり，最下位カテゴリ頻度は 5 であった。また TA や TD に限定した場合の推定値の絶対値差は最大 0.04 であったが，最下位カテゴリ頻度が 1 の項目であっても絶対値差が最大となる場合もみられ，最下位カテゴリ頻度が直線的に推定値差と関連する傾向はみられなかった。

Table 1 に，受験者ごとに推定された期待カテゴリが，カテゴリ数 5 と 6 の場合で異なった受験者数を示した。わずかながら，期待カテゴリが変化する受験者が存在することがわかった。最下位カテゴリ数が多い場合の検討が今後望まれる。

Table 1 期待カテゴリが異なった受験者の人数。

	Difference of category (CAT5 - CAT6)			
	-1	0	1	
Overall	5	656	0	
Task Achievement	5	650	6	
Task Delivery	8	653	0	

引用文献

- Linacre, J.M. (1994). *Many-facet Rasch measurement*. Chicago: MESA Press.
Hato, Y., Kanzawa, K., Mitsunaga, H. & Healy, S. (2018). Developing a computer-based speaking test of English as a lingua franca: Preliminary results and remaining challenges. *Waseda Working Papers in ELF*, 7, 87-99.

付記 本研究は科学研究費補助金 (基盤研究 (B))，課題番号 16H03448 の助成を受けて行われた。