

上で表わした。その結果、高校入学時の高校生の英語学力平均値が年々低下していること（8年間で偏差値換算で約3.8の低下）、特に1998年度と2002年度の落ち込みが大きいことが示された。

学力テストデータの差異項目機能（DIF）分析

孫 媛

1 DIF とは？一項目バイアスと DIF

1960年代、1970年以降米国では「項目バイアス」研究が重視された。それは、テストが測定しようとしている能力とは別の要因によって項目正答率に集団差があることが問題となったことによる。そして、集団差を探す統計的方法の研究が活発化した。その中で「バイアス」という用語の使用については慎重であるべきで、テスト項目における集団差を指すには「DIF」という中立的な用語を用いることが望ましいという主張が出されてきた。すなわち、集団間で違いのあることが、直ちにバイアスであるとは限らないということを用語にも反映させようというものである。

比較する2つの集団について、基準となる集団を「参照集団」、DIFの影響を受ける可能性に関心のある集団を「焦点集団」と呼ぶ。例えば米国社会において、白人集団は前者、黒人・ヒスパニック集団は後者にあたることが多い。

また、DIFは「均一DIF」と「不均一DIF」に分類される。

前者は、一方の集団の項目正答率が能力尺度のどの水準においても他方の集団より高くなるようなDIFをいい、後者は、2集団の項目正答率の差がすべての尺度水準で等しくないようなDIFをいう。

2 IRTの観点から見たDIF

基本的には、1) 参照集団、焦点集団の別に項目パラメタを推定、2) 2群における θ の分布の違いについての調整（等化）を行なう、3) 2群における項目パラメタを比較する（=ICCを比較）、という手順を採る。

IRTモデルを適用して、DIF検出のための統計量を計算する方法は、1) 項目特性曲線間の面積を用いた方法、2) 項目パラメタ値の集団間の差を検定する方法に大別される。後者には、Lordの χ^2 法（Lord, 1980）、尤度比検定法（Thissen, Steinberg, & Wainer, 1993）、BILOG-MGのDIFコマンドによる方法、などがある。

なお、IRTモデルを用いることなくDIF検出のための統計量を導く、Mantel-Haenszel法、SIBTEST（Simultaneous Item Bias Test）などの方法もある。

3 学力テストデータのDIF分析

3.1 学力テストデータ

全国の公立共学高校15校の高校3年生に2001年6、7月に実施され、受験者総数は4,718名であった。

テストの構成は、国語23項目、数学17項目、英語33項目で、すべて多肢選択形式で正答か誤答かの2値で採点された。

3.2 参照集団と焦点集団

性別と学校により下位集団を設定した。性別に関しては、男性を参照集団、女性を焦点集団とした。また、学校に関しては、15校のうち、学力が平均水準に近く、受験者数も比較的多い学校Jを焦点集団として選び、その他の14校をまとめて参照集団とした。

3.3 DIF検出法

IRTに基づく方法は、1) 項目パラメタ値を正確に推定するために多くの受験者を必要とするが、今回は学校Jを焦点集団とする分析では受験者数が300名程度と少ない、2) 今回のデータは一次元性が満たされない、3) 計算手順が複雑でコストがかかる（特別なプログラムを必要とする）、などの理由から、DIF検出法にはMantel-Haenszel法を用いた。

このM-H法は、Holland & Thayer (1988) が医学・生物統計学の分野で古くから用いられてきたM-H統計量をDIF探索に適用したもので、1) 概念的に単純、2) 有意性検定ができる、3) 均一DIFに対する検出力が高い、4) 標本数が比較的少なくてもよい、5) 利用が容易（SPSSなどを使って簡単に実施でき、特別なプログラムを必要としない）、などの長所がある。

3.4 DIF分析からの示唆

各教科別に基本統計量（略）の計算を実施したのち、DIF分析を実施した。

DIF分析の結果から、1) 各科目とも性別について、数は少ないが比較的大きなDIFを示す項目が存在した。したがって、重要な決定を伴う大規模テストについては、性別DIFに関心を向けるべきである、2) 国語における「古文」「漢文」、数学における「数と式」「三角比」などは、領域として男性あるいは女性に有利になっているように見える。検証的に確認するにはSIBTESTによりDBF（Differential Bundle Functioning）を分析することが有益かと思われる、3) 学校別について、性別よりも大きなDIFが見い出される傾向がある、4) 数学や英語（特に数学）において領域により学校Jの有利・不利が目立つ、5) そのような傾向があるのならば、教育上の重点の違いなど、学校の特徴をDIF分析で見い出せる、6) カリキュラムを検討する際の重要なヒントを与えてくれる、などの示唆が得られた。

引用文献

- Holland, P. W., & Thayer, D. T. 1988 Differential item functioning and the Mantel-Haenszel procedure. In H. Wainer & H. Braun (Eds.), *Test validity*. Hillsdale, NJ : Lawrence Erlbaum. Pp. 129-145.
- Lord, F. M. 1980 *Applications of item response theory to practical testing problems*. Hillsdale, NJ : Lawrence Erlbaum.
- Thissen, D., Steinberg, L., & Wainer, H. 1993 Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning*. Hillsdale, NJ : Lawrence Erlbaum. Pp. 67-113.