様式Ｆ－１９

科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成２５年　５月１７日現在

機関番号：１１３０１
研究種目：若手研究(B)
研究期間：2011～2012
課題番号：23730602
研究課題名（和文）　ＤＩＦを利用したテストの公平性に関する研究

研究課題名（英文）　Investigation of the fairness of tests using DIF analysis.

研究代表者
熊谷　龍一（KUMAGAI RYUICHI）
東北大学・大学院教育学研究科・准教授
研究者番号：60422622

研究成果の概要（和文）：
本研究では，テストの公平性を担保するアプローチの一つとして，DIF（differential item functioning：特異項目機能）分析を利用した方法について検討を進めた。従来の DIF 分析に対して，対象集団数や順序付き多値型項目への対応といった理論的拡張が進められ，分析を実行するためのコンピュータプログラムの開発，公開を行った。さらに，性格検査やうつ評価尺度といった心理尺度，英語教育や日本語教育の現場で広く利用されている can do statement 尺度への適用例を示し，その方法の妥当性を示すことができた。

研究成果の概要（英文）：
In this research, the fairness of tests was investigated using DIF analysis. I developed a new approach of DIF, which is applicable when there are more than three groups and polytomous items. A computer program (EasyDIF) was developed to calculate. To investigate the validity of this approach, surveys using psychological scales and can-do-statements scales were conducted.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
交付決定額	1,600,000	480,000	2,080,000

研究分野：社会科学
科研費の分科・細目：心理学・教育心理学
キーワード：教育評価，教育測定

１．研究開始当初の背景

人間の能力・特性を測定するために用いられる「テスト」に関して，信頼性や妥当性の検討と同様に，テストの公平性に関しても客観的に保証する必要がある。テスト品質保証のためのガイドラインである「テスト・スタンダード」（日本テスト学会，2007）においても，「受検者は，テストの全ての過程において年齢，性，国籍，障害の有無などによって差別されてはならない」と定められていることからも明らかである。この公平性を担保するための有力な手法の一つとして，テスト分析の一つである DIF（Differential Item Functioning：特異項目機能）を用いた方法が存在する。DIF は「テストが測定しようとしている特性・能力が等しいにもかかわらず，所属する下位集団によって正答率が異なる」と定義され，国際的には，例えば PISA 調査においても性差による公平性を担保するために実施されている。しかしながら，我が国において実際のテスト場面でそのような現象が生じるということがあまり認識されておらず，DIF 分析が行われることはこれまでほとんど無かった。

２．研究の目的

本研究の目的は，熊谷ほか（2005）で提案された「母集団分布を考慮した項目反応モデ

ルによるDIF検出方法」を基軸とし，
（1）方法の理論的整備を行い，従来の DIF 検出指標との比較検討を行う，
（2）実際のテスト開発・分析場面における DIF 分析に際して，分析対象母集団数・多値型項目などの拡張による本手法の有効性，妥当性を提示する，
（3）分析ソフトウェアを開発し，HP 等で公開することにより，その成果を実用に供する形で公表する，
の3点である。

3．研究の方法

上記3点の研究目的を達成するための方法は，以下のようになる。
（1）熊谷ほか（2005）で提案された「母集団分布を考慮した項目反応モデルによる DIF 検出方法」について，対象母集団が3つ以上の場合への拡張（熊谷，2007）や，順序付き多値型項目への拡張（熊谷，2010）といった理論的拡張を整理･統合して学術雑誌へ投稿する。
（2）（1）で整理・統合された方法について，コンピュータによるシミュレーション・データおよび実際の心理尺度におけるデータに対して適用し，従来用いられてきた DIF 検出方法との比較検討を行いながら，その特徴や有効性・妥当性を検証する。
（3）（1）で整理・統合された方法を実際に計算するためのコンピュータ・プログラムの作成を行い，フリーソフトとして公開する。
（4）実際に実施された様々なテストデータ（心理尺度や can do statement 調査など）に（1）で整理・統合された方法を適用し，その，心理学分野で広く DIF 分析が活用されるための端緒を示す。

4．研究成果

（1）理論の整理・統合

本研究で提案され，理論の整理・統合が行われた DIF 検出方法の詳細は以下のようになる。

①2 値型項目・下位集団数が 2 つの場合

手続き 1　受検者数 N, 項目数 n の項目反応行列における項目 k について，DIF の存在を検討する状況を考える。

この時受検者集団は，下位集団 A（N_A名）と下位集団 B（N_B名）から構成されるものとする。この項目反応行列の項目 k について，Thissen, Steinberg, & Wainer (1993) による尤度比を用いた方法と同様に，項目 k_A と k_B という二つの項目に分割する。すなわち，項目 k_A については，下位集団 A の項目反応をそのまま用いるが，下位集団 B の項目反応部分については欠測値とする。同様に項目 k_B については，下位集団 A の項目反応部分を欠測値とし，下位集団 B の項目反応はそのまま用いる。

手続き 2　手続き 1 で得られた $N \times (n+1)$ の項目反応行列に対して，項目反応理論（Item Response Theory：以下 IRT とする）に基づく項目母数の推定を行う。項目母数推定においては，多母集団モデルを適用し，下位集団 A および B の母集団分布についても推定を行う。これにより，項目 k については，k_A および k_B という 2 組の項目母数推定値が得られ，また下位集団 A の母集団分布 $g_A(\theta)$ と下位集団 B の母集団分布 $g_B(\theta)$ が得られる。

手続き 3　SIBTEST 法 (Shealy & Stout, 1993) と同様に，以下の式で項目 k において，指標 K を算出する。

$$K = \int_{-\infty}^{\infty} \left| P_A(\theta) - P_B(\theta) \right| g_T(\theta) d\theta. \quad (1)$$

ここで $P_A(\theta)$，$P_B(\theta)$ は，それぞれ k_A および k_B の項目母数を用いた項目特性関数であり，$g_T(\theta)$ は母集団分布 $g_A(\theta)$ および $g_B(\theta)$ に対して N_A と N_B の比率を乗じて足し合わせた混合母集団分布である。指標 K は 2 組の項目特性曲線に挟まれた領域の面積に対し，全体の母集団分布で重みづけをしたものであり，項目 k について下位集団間での正答率差の期待値を表すものである。

②下位集団数が 3 つ以上の場合

下位集団数が 3 つ以上の場合において，①の下位集団数が 2 つの場合との相違点は以下の通りである。

手続き 1　項目 k について，下位集団数 L だけ分割を行う。

手続き 2　$(N \times (n+L-1))$ の項目反応行列に対して，項目母数の推定を行い，項目 k について L 組の項目母数の推定値を得る。

手続き 3　(1) 式を以下の式に変更する。

$$K = \int_{-\infty}^{\infty} \left[P_{\max}(\theta) - P_{\min}(\theta) \right] g_T(\theta) d\theta. \quad (2)$$

ここで $P_{max}(\theta)$ は，潜在特性尺度値上のある点 θ において，L 組の項目特性関数の中の最大値であり，同様に $P_{min}(\theta)$ は，L 組の項目特性関数の最小値を示す。

③順序付き多値型項目の場合

DIF の存在を検討する項目 k が，5 件法リッカート尺度のように順序付き多値型項目の場合の，指標 K の計算手続きは以下のようになる。

手続き 1　②と同様に DIF を検討する項目 k について，下位集団数 L だけ分割を行う。

手続き 2　$(N \times (n+L-1))$ の項目反応行

列に対して，項目母数の推定を行う。使用する項目反応モデルについて，順序付き多値型項目に対応したものとしては，Graded Response Model （Samejima, 1969） や Generalized Partial Credit Model (Muraki, 1992) などがあるが，ここではその別を問わない。

手続き 3　手続き 2 により，L 組の項目母数推定値および項目反応カテゴリ特性関数が得られる。各項目反応カテゴリ特性関数について，

$$E_l(\theta)=\sum_{c=1}^{C} P_{lc}(\theta)\cdot c \qquad (3)$$

により，項目期待カテゴリ特性関数を算出する。ここで C は項目 k のカテゴリ数，$P_{lc}(\theta)$ は L 組中 l 番目のカテゴリ c に対する項目反応カテゴリ特性関数である。

手続き 4　以下の式により，指標 K を算出する。

$$K=\int_{-\infty}^{\infty}\left[E_{\max}(\theta)-E_{\min}(\theta)\right]g_T(\theta)d\theta. \qquad (4)$$

ここで $E_{max}(\theta)$は，潜在特性尺度値上のある点 θ において，L 組の項目期待カテゴリ特性関数の中の最大値であり，同様に $E_{min}(\theta)$は，L 組の項目期待カテゴリ特性関数の最小値を示す。

（2）方法の特徴，および有効性・妥当性の検証

①シミュレーション・データを用いた検討

様々な状況設定のもとでシミュレーション・データを発生し，指標 K と従来の DIF 検出方法（Mantel-Haenszel 統計量および SIBTEST 法）との比較検討を行った。これにより，

・Mantel-Haenszel 統計量における Δ_{MH} および SIBTEST 法における β 推定量と高い相関を有する，
・顕著に DIF が見られるとする $\Delta_{MH}\geqq 1.5$ に対応する指標 K の数値として（カテゴリ数－1）×0.1 が妥当である，
・指標 K は識別力母数の違いによる影響を強く反映する，

といったことが示された。

②実際の心理尺度データを用いた検討

和田（1996）の Big Five 尺度の中の“外向性因子”12 項目を用いて，指標 K の検討を行った。調査協力者は 15 歳から 83 歳の男女 3612 名（男性 1627 名，女性 1985 名）であった。性別および年齢により，4 つの下位集団を設定し DIF 分析を行なった。指標 K を用いた DIF 検出においては，項目期待カテゴリ特性曲線を表示することで，DIF の特徴を視覚的に捉えることが可能となることが示された。

（3）分析ソフトウェアの開発・公開

指標 K の計算には繰り返し計算が必要であり，非常に煩雑な手続きが必要となるため，これを容易に実行できる計算ソフトウェア“EasyDIF”の開発を試みた。開発にあたっては，既に初学者向けの IRT 分析ソフトウェアとして開発されている EasyEstimation（熊谷，2009）と同様のユーザー・インターフェースを持ち，簡易なマウス操作のみで分析を行うことができるように設計が行われた（図 1 参照）。

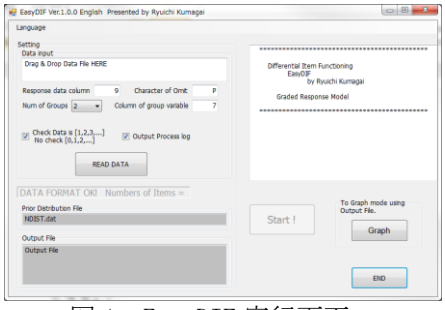

図 1：EasyDIF 実行画面

積分計算については，θ を Q 個の離散点に分割し，そのときの重みを $A(\theta_q)$ とする離散近似による計算を採用した（ここで，q=1,2,…,Q）。

EasyDIF は，学術研究目的での利用に限るフリーソフトとして，http://irtanalysis.main.jp/にて公開された。

（4）実際のデータに対する適用

①Birleson 自己記入式抑うつ評価尺度に対する適用

子どもを対象とした抑うつ測定尺度である Birleson 自己記入式抑うつ評価尺度（DSRS-C）について，小学生集団，中学生集団を下位集団として指標 K による DIF 分析を行なった。

調査協力者　小学校 3～6 学年の児童および中学校 1～2 学年の生徒 4638 名（男性 2414 名，女性 2269 名）。ただし回答に不備のあったものを除いた 4630 名を対象とした。

尺度構成　DSRS-C（村田ほか，1996）を用いた。回答選択枝は 3 件法。谷ほか（2010）に従い 2 因子構造を採用し，「活動性および楽しみの減衰」尺度に関して分析を行なった。

結果　「家族と話すのが好きだ」という項目において，中学生の方が「いつもそうだ」と回答しにくい傾向が示された（指標 K=0.23）。

②Big Five 尺度に対する適用

（２）方法の特徴，および有効性・妥当性の検証でもちいた Big Five 尺度のデータに対して，5 つの下位尺度全てに対して指標 *K* による DIF 分析を行なった。調査協力者は（２）の時と同様であり，下位集団の設定は男性・女性の 2 つとした。

結果　下位集団を男性・女性の 2 集団にした時の DIF 分析については，尺度に含まれる全 60 項目において，DIF 項目は検出されなかった。

③英語 Can-do リストに対する適用

日本人の英語学習者を対象に，ある一定レベルの英語力があると認定された人が，実際にどのようなことが英語を使ってできるかを調査するための英語 Can-do リスト（日本英語検定協会，2006）に対して，指標 *K* による DIF 分析を行なった。

調査協力者　大学 1 年生 1370 名。下位集団の設定は文系男子，文系女子，理系男子，理系女子の 4 集団とした。

尺度構成　英語 Can-do リストのうち，3 級から準 1 級の全 91 項目（読む、聞く、書く、話す）を含む 134 項目。回答方式は 4 件法。

結果　134 項目中，20 項目で DIF が検出された。身近な話題については，男子より女子のほうが，社会的な話題については，女子より男子のほうが，よりできると回答する傾向が見られた。文理で差がでる項目は聞くこと，読むことに多かった。

④日本語 Can-do-statements に対する適用

日本語の運用能力を測る自己評定尺度として開発された日本語 Can-do-statements（三枝，2004）に対して，指標 *K* による DIF 分析を行なった。

調査協力者　国内の日本語学校および大学に所属する外国人学生 868 名。下位集団は，中国語母語話者 322 名，韓国語母語話者 407 名，その他の母語話者 139 名の 3 集団とした。

尺度構成　読む」，「書く」，「話す」，「聞く」の 4 技能について，各 15 項目（合計 60 項目）。7 段階評定であったが，回答度数の関係で 5 段階に変換した。

結果　21 項目で DIF が検出された。「その他」集団は「読む」に自信が無く（自己評定による日本語能力の水準が同程度であっても「できない」に回答する傾向），「話す」は自信がある（「できる」に回答する傾向）ことが示された。

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計１件）

１．熊谷龍一、2012、統合的 DIF 検出方法の提案－“EasyDIF”の開発－、心理学研究、83(1)、35-43、査読有、URL：https://www.jstage.jst.go.jp/article/jjpsy/83/1/83_35/_pdf

〔学会発表〕（計４件）

１．熊谷龍一・野口裕之、日本語 Can-do-statements に対する拡張型 DIF 分析の試み、2012 年度日本語教育学会秋季大会予稿集、251-252、2012 年 10 月 13 日～10 月 14 日、北海道

２．斉田智里・熊谷龍一、英語 Can-do リストの DIF 分析の試み、第 38 回全国英語教育学会愛知研究大会発表予稿集、316-317、2012 年 9 月 4 日～9 月 5 日、愛知

３．熊谷龍一・野口裕之・谷伊織、心理尺度における多値型項目応答モデルの適用事例（２）、日本教育心理学会第 53 回総会発表論文集、537、2011 年 7 月 24 日～7 月 26 日、北海道

４．並川努・谷伊織・熊谷龍一・脇田貴文・中根愛・野口裕、Birleson 自己記入式抑うつ評価尺度における特異項目機能の検討、日本教育心理学会第 53 回総会発表論文集、379、2011 年 7 月 24～7 月 26 日、北海道

〔その他〕

ホームページ等

http://irtanalysis.main.jp/

６．研究組織

(1)研究代表者

熊谷　龍一（KUMAGAI RYUICHI）

東北大学・大学院教育学研究科・准教授

研究者番号：60422622

(2)研究分担者

（　　）

研究者番号：

(3)連携研究者

（　　）

研究者番号：