

令和元年5月28日現在

機関番号：11301

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K04289

研究課題名(和文)素点を利用したDIF指標の開発

研究課題名(英文)Development of DIF index using raw scores

研究代表者

熊谷 龍一(Kumagai, Ryuichi)

東北大学・教育学研究科・准教授

研究者番号：60422622

交付決定額(研究期間全体)：(直接経費) 3,100,000円

研究成果の概要(和文)：本研究では、熊谷(2015)で提案された「素点を利用したDIF(Differential Item Functioning: 特異項目機能)検出方法」について、その理論的精緻化を進め、新たなDIF指標を開発することが目的であった。

具体的には、1)実際の心理尺度データに対してDIF分析を応用し、評定者間の差異に関する分析を行うこと、2)複数のテスト冊子がある状況でのDIF分析の問題点の洗い出しとその解決方法の提案、および実データによる検証に取り組んだ。特に2)については、受検者の群分けについて、項目反応理論(IRT)の潜在特性尺度値を利用する方法が提案された。

研究成果の学術的意義や社会的意義

本研究では、学力テストや心理尺度などにおいてテストの公平性を担保するための分析手法としてのDIFに関して、素点を用いた方法の提案、理論の精緻化、および実データによる検証が進められた。この成果により、実際の教育心理学研究現場で見られる数百件程度のサンプルサイズにおいてもDIF分析が実行可能となり、多くの研究現場でDIF分析が広く活用される契機となることが期待される。

研究成果の概要(英文)：In this study, we theoretically explored "DIF (Differential Item Functioning) detection method using raw score" supposed in Kumagai (2015), and developed a new DIF detection method.

Specifically, 1) we applied DIF analysis for the data of psychological scale and analyzed the difference between raters. 2) we discovered that our DIF detection method included problems in the situation with some test booklets. So we suggested the solution about it, and verified it using real data set. In this solution, we clustered the examinees using latent traits of item response theory.

研究分野：教育測定学

キーワード：DIF 素点 特異項目機能

様式 C-19、F-19-1、Z-19、CK-19（共通）

1. 研究開始当初の背景

心理尺度を含め、テストを用いて人間の内的特性を測定することは、教育心理学の研究において非常に重要な位置を占める。わが国におけるテストの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）では、「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」という「テストの公平性」を担保することが定められている。

このテストの公平性を担保するための分析手法の1つとして、DIF（Differential Item Functioning：特異項目機能）分析がある。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、例えばある知能検査において、男性集団と女性集団とで同一の知的水準であるにもかかわらず、空間把握を測定する項目には男性集団の正答率が高くなるような場合、DIFが検出されたということになる。

このDIF分析は、PISA調査において性差による公平性担保のために実施されるなど、テスト開発・実施場面では必須の分析であるにもかかわらず、我が国のテスト研究・開発場面で実施された事例は非常に少ないのが現状である。この原因については、1）従来のDIF分析では、そのほとんどが正答・誤答の2値型データに対して行われてきた、2）また比較する下位集団の数も2つ（男性と女性、ある外的基準の上位と下位集団など）だけであったこと、などが挙げられる。

研究代表者は、心理測定学的立場からこのDIF研究に取り組んできた。上記の課題を克服するために、3集団以上を同時に比較でき、多値型項目（たとえば5件法のリッカート尺度など）のDIF分析を統一的行なうことができる指標Kを提案し（熊谷、2012）、学術目的であれば自由に利用できる計算プログラム“EasyDIF”の開発および公開を行ってきた。また、この指標Kを心理尺度に対して適用した例としては、和田（1996）のBig Five尺度に対して、年齢および性別を下位分類とした研究を行なった。さらには、現実の研究場面での応用例として、日本語教育のなかでこの指標を用いた研究を進めてきた。

この指標Kを用いたDIF分析を実際の研究場面で適用する中で、次のような課題に直面した。それは、分析に用いることができるデータ数（サンプルサイズ）の問題である。どのような統計手法であれ、データ数の増加が精度の向上（誤差の減少）に寄与するが、特に指標Kは項目反応理論（Item Response Theory、以下IRT）を基盤とした方法であり、安定した母数推定のためには1,000件以上のデータ数が望まれる。しかしながら、教育心理学をはじめ、人間の内的特性を対象とした研究では必ずしも十分なデータ数を確保できるわけではなく、数百件のサンプルサイズで研究が行なわれることも多い。申請者は、このような状況においても用いることができるDIF分析指標の必要性から、指標Kに対して素点（合計得点）を用いる方法への拡張という、新たな研究課題にたどり着き、その端緒となる新たな方法について熊谷（2015）として発表した。この方法については、まだまだ未知である事項が多く、その解明が必須であった。

引用文献

- 熊谷 龍一・野口 裕之（2015）. 素点を用いたDIFの大きさを表す指標の提案, 日本教育心理学会第57回総会発表論文集, PF068.
- 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, 35-43.
- 日本テスト学会（2007）. テスト・スタンダード——日本のテストの将来に向けて—— 金子書房
- 和田 さゆり（1996）. 性格特性用語を用いたBig Five尺度の作成心理学研究, 67, 61-67.

2. 研究の目的

本研究の目的は、指標Kに対して素点（合計得点）を用いる方法へ拡張を行う（熊谷、2015）ことについて、特に実際に得られたデータへの分析において、どのような課題があるのかなどを検討することである。

ここで、指標Kに対して素点を用いた方法への拡張（以下、この方法を指標K2とする）の概略を述べる。指標K2は以下の手続きで計算される。

手続き1 受検者ごとに正答数得点を算出し、その値ごとに層分けを行なう。この時、各下位集団における層の人数が一定数（たとえば10）に満たない場合は、隣接する層と結合を行なう。

手続き2 層ごとに以下の値を計算する。

$$k_{2L} = |P_{LMax} - P_{LMin}| \times W_L$$

ここでLは層を、 P_{LMax} は層Lにおいて下位集団ごとに正答率（多値型の場合は平均値）を算出した時の最大値、および P_{LMin} は最小値である。 W_L は全受検者数に対する層Lの人数の比率を表す。

手続き3 以下の値K2をもって、DIFの大きさを表す指標とする。

$$K2 = \sum_L k_{2L}$$

3. 研究の方法

指標 K2 について、実際に得られたデータを分析する際の様々な課題について検討するために、本研究では、以下の3点を実施した。

- (1) 分担研究者との討議や、海外の教育測定を専門とする機関への調査などを通して、理論的な精緻化を進めた。
- (2) 評定者が複数存在する状況における心理尺度データに対して、DIF 分析を応用することで評定者間の差異に関する分析を行うことができるかどうかを検討した。
- (3) 問題項目が異なる複数のテスト冊子（ただし共通項目は存在する）のデータに対して、素点の合計得点を用いている指標 K2 よりも、さらに多くの情報を用いる形で DIF 分析が行えるかどうかについて検討した。

4. 研究成果

前項の (1) から (3) のそれぞれについて、以下のような成果を得た。

(1) 分担研究者との討議および海外機関調査を通じた理論的精緻化

分担研究者との討議を通じて、(指標 K2 を含んだ) DIF 分析における諸課題を洗い出した。この中で、次項 (2), (3) の研究の方針が見いだされた。また、海外機関の調査を通じて、多くのテストデータ分析場面で DIF 分析が実施されており、DIF 分析自体がテストデータ分析の必要要件となっていることを確認した。ただし、実施されている DIF 分析手法については、正答率の比較等のシンプルな形式のものが利用されていたことから、同じく素点を利用している指標 K2 の有用性が示唆された。

(2) 評定者が複数存在する状況における DIF 分析の応用について

心理学研究においては、自己評定のほかに、他者による評定もよく行われる。このとき、各評定者による評定データを下位集団と見なして DIF 分析を行うことができるのではないかと、ということが本研究の課題の一つとなった。

例えば評定者が4人であるような評定者データの DIF 分析においては、図 1 および図 2 のように2通りのデータ構造が考えられる。図 1 は、各評定者のデータを横に結合した形での DIF 分析を行う方法である。一方、各評定者を下位集団データとみなし、データを縦に結合した図 2 のような DIF 分析の手法も考えることができる。本研究では、この2種類の DIF 分析手法にどのような違いがあるのかを検討することを目的とした。



図 1. 横に連結したデータ

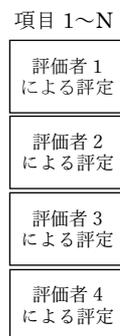


図 2. 縦に連結したデータ

本研究では、岩田・熊谷・佐伯 (2018) で用いられた実際の心理尺度データを分析対象として検討を行なった。調査データの概要は以下の通りである。

調査協力者： 小学 5, 6 年生およびその保護者、担任教師に対して、児童 541 名 (男子 283, 女子 258) を評価対象者として回答を得た。

心理尺度： 子どもの内在化・外在化問題を測定する Strength and Difficulties Questionnaire (SDQ; Goodman, 1997) を用いた。「問題行動」「多動・不注意」「情緒不安定」「仲間関係問題」「向社会的行動」の 5 下位尺度、それぞれ 5 項目ずつ、計 25 項目から構成されている。本研究では、このうち、「問題行動」「多動・不注意」「情緒不安定」の 3 尺度を分析対象とした。各項目は 3 件法で実施した。

分析手法： 本研究では、その性質がある程度明らかになっている指標 K を用いて、2 通りのデータに対して DIF 分析を行なった。

表 1 に、横に連結したデータ形式での指標 K と、縦に連結したデータでの指標 K の数値を示す。形式間の指標 K の相関は .61 であった。

表 1. データ形式ごとの指標 K の比較

	問題行動		多動・不注意		情緒不安定	
	データ形式		データ形式		データ形式	
	横	縦	横	縦	横	縦
項目 1	0.49	0.46	0.37	0.25	0.41	0.17
項目 2	0.51	0.36	0.21	0.25	0.33	0.38
項目 3	0.24	---	0.39	0.14	0.40	0.15
項目 4	0.32	0.32	0.50	0.26	0.83	0.44
項目 5	0.15	0.09	0.41	0.17	0.45	0.36

*「問題行動」における項 3 は、適切な項目母数の推定値が得られなかった

両形式間での指標 K の相関は中程度であったものの、数値にはややずれが見られ、また横に連結したデータのほうが、指標が大きくなる傾向が見られた。

また、以下の事項が課題となった。

- ・横に連結したデータの分析においては、評定者が異なる評定結果を一人の受検者データのように取り扱っていること。
- ・縦に連結したデータでは、総項目数が 5 と少なくなることから、適切な項目母数の推定が得られない項目があった。

このうち、2 点目の適切な項目母数の推定が得られない点については、指標 K2 を用いることでその解決が図られる可能性が見いだされ、今後の検討課題となった。

この研究成果については、熊谷・岩田・佐伯 (2018) として日本心理学会第 82 回大会にて発表を行なった。

引用文献

Goodman, R. (1997). "The Strengths and Difficulties Questionnaire: a research note." *Journal of Child Psychology and Psychiatry*, 38(5), 581-586.

岩田 昇・熊谷 龍一・佐伯 いずみ (2018). Can fathers and mothers properly evaluate their children's problems? -An IRT investigation on the Strengths and Difficulties Questionnaire- 第 34 回日本行動科学会ウィンターカンファレンス.

(3) 複数のテスト冊子がある状況における DIF 分析について

分析対象としたい項目数が多く、一つのテスト内に含めることができる項目数が少ない場合などにおいて、共通項目を含めた複数冊子 (版) による分析を行うことがある。このような場合において、上記の指標 K2 を用いた DIF 分析 (この時、冊子と下位集団は 1 対 1 対応しているとは限らない) を行うとき、手続き 1 の「正答数得点」を算出する場合には、全項目の正答数 (総得点) ではなく、共通項目の部分のみの正答数を用いなければならない。これは、等しい能力を持っている受検者を各層に付置するためであり、各冊子に含まれる項目数や項目の困難度などが異なる場合には、「総得点」ではその機能を果たせなくなるためである (この点については、同じく正答数得点を利用する DIF 分析手法の Mantel-Haenszel 法においても同様に問題となる)。共通項目の数が非常に少ないような場合には、得点段階も少なくなることが問題となる。

この問題について、全冊子の反応パターンを同一にしたデータ行列 (各冊子の未提示部部分については、欠測値とする) について、項目反応理論 (Item Response Theory; 以下 IRT とする) の分析による潜在特性尺度値 (以下 θ とする) を用いることを提案した。 θ を用いることにより、共通項目部分以外の情報も「能力値」に組み込むことができ、得点段階の減少も防ぐことができると考えられる。この方法について、実際に収集されたデータを用いた分析例を示す。

調査協力者： ベトナム、中国、オーストラリア、日本国内在住の、日本語を母語としない日本語学習者、556 名。

調査内容： 欧州言語共通参照枠 (CEFR) における Can do descriptors (言語活動で何ができるかを記述したもの) の「読む」、「聞く」から 70 項目を抽出し、4 件法の自己評価形式とした質問紙データ。

冊子構成： ベトナム語 4 版、中国語 3 版、英語 3 版、それぞれに共通項目が含まれている。

下位集団： 国内在住 (88 名)、国外在住 (468 名) の 2 集団とした。

上記データの 70 項目について、Rasch モデルにより θ を算出し、それにより受検者を 4 層に分けた上で、指標 K2 を算出したところ、DIF 検出の基準となる $K2 > 0.3$ を示した項目は 2 項目であった。図 3 は $K2 = 0.40$ となった項目 24 「(日本語で) 毎日使っている機器・設備の取扱説明のような、簡単な専門的情報を理解することができる」の DIF 状況である。

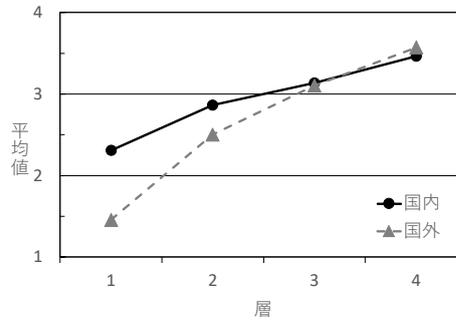


図3：項目24のDIF状況

この結果から、指標 K2 の産出手続きにおける層分けのための「正答数得点」の代わりに、IRT の θ を用いることが可能であることが示唆された。なお、この研究結果については、日本教育心理学会第 61 回総会にて公表を行なう予定である。

5. 主な発表論文等

〔雑誌論文〕 (計 0 件)

〔学会発表〕 (計 3 件)

- ①岩田 昇・熊谷 龍一・佐伯 いずみ (2018). Can fathers and mothers properly evaluate their children's problems? -An IRT investigation on the Strengths and Difficulties Questionnaire- 第 34 回日本行動科学会ウインターカンファレンス.
- ②岩田 昇・熊谷 龍一・佐伯 いずみ (2018). SDQ 評定における男女間 DIF の検討～評定者による相違, 日本心理学会第 82 回大会発表論文集, 865.
- ③熊谷 龍一・岩田 昇・佐伯 いずみ (2018). 複数評定者データの DIF 分析に関する方法的検討, 日本心理学会第 82 回大会発表論文集, 949.

〔図書〕 (計 0 件)

〔産業財産権〕

○出願状況 (計 0 件)

名称：
 発明者：
 権利者：
 種類：
 番号：
 出願年：
 国内外の別：

○取得状況 (計 0 件)

名称：
 発明者：
 権利者：
 種類：
 番号：
 取得年：
 国内外の別：

〔その他〕

ホームページ等
<http://irtanalysis.main.jp/>

6. 研究組織

(1) 研究分担者

研究分担者氏名：野口 裕之

ローマ字氏名：NOGUCHI Hiroyuki

所属研究機関名：名古屋大学

部局名：教育発達科学研究科

職名：名誉教授

研究者番号（8桁）：60114815

(2) 研究協力者

研究協力者氏名：

ローマ字氏名：

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。