

別紙1-1

論文審査の結果の要旨および担当者

報告番号	※ 乙 第	号
------	-------	---

氏 名 並 川 努

論 文 題 目

心理尺度短縮版作成におけるIRTの活用に関する研究

論文審査担当者

主 査 名古屋大学大学院教育発達科学研究科教授 野口 裕之
名古屋大学大学院教育発達科学研究科教授 高井 次郎
名古屋大学大学院教育発達科学研究科准教授 石井 秀宗

論文審査の結果の要旨

別紙 1 - 2

本論文の全体を通しての目的は、心理尺度短縮版の作成方法について、具体的な事例をもとに検討することである。

心理尺度を用いた測定は様々な場で行われている。しかし、調査には時間的な制約があったり、回答者にかかる負担を少しでも減らす必要があったりするため、項目を減らした短縮版尺度が作成され、利用されることも多い。しかし、短縮版尺度をどのような情報をもとに作成すべきかについては、十分な検討がなされているわけではないのが現状である。

そこで本研究では、まず心理尺度短縮版作成の国内外の状況について検討した上で、実際に複数の尺度を取り上げ、短縮版を作成する試みを行うことを通して、短縮版の作成方法について検討を行った。なお、本研究では、従来の古典的テスト理論に比べ、尺度や項目に関する精緻な情報が得られると考えられる **Item Response Theory** (以下、IRT) に注目をし、短縮版作成を行った。また、IRT を用いた方法と、因子分析を用いた方法の比較を行い、それらの違いについて検討を行った上で、短縮版作成方法についての今後の課題について指摘を行った。各章で行った検討の具体的内容は以下の通りであった。

まず第2章では、心理尺度の短縮版がどのように作成されているのかについて、レビューを行った。心理尺度短縮版は、調査や研究において多く利用されているが、その作成方法については必ずしも十分な検討がなされているとは言えない。実際に短縮版作成においてどのような方法が多く用いられているかについては、Goetz et al. (2013) によって英語論文を対象にしたレビューがなされており、因子分析を用いた方法などが多く用いられていることなどは示されているものの、日本国内での状況等は明らかになっていない。そこで、データベース (CiNii, J-Stage) を用いて論文を抽出し、日本における短縮版尺度の短縮版作成研究の数や短縮化による項目数の削減率、短縮版の作成方法等についてレビューを行った。その結果、近年になるほど、短縮版作成の研究数が増えていることや、85.7%にあたる論文において、因子分析の結果を利用した項目選択がなされていることなどが示された。また、先行研究と比較し、因子分析が多く用いられていることや、IRT を用いた研究が少ないことなどが指摘された。

論文審査の結果の要旨

第 3 章から第 5 章では、IRT を用いて心理尺度の短縮版を実際に作成することを通して、具体的な課題の検討を行った。まず第 3 章では、子ども用の抑うつ尺度 (DSRS-C) を取り上げた。小学校 3 年生から中学校 2 年生までの 4683 名を対象に行った調査データをもとに、IRT のパラメタを推定し 18 項目から 9 項目を短縮版として選択した。その上で、作成された短縮版の信頼性および妥当性について検討し、実用水準で十分な信頼性・妥当性を備えていることを確認した。

また、第 4 章では国内で用いられるパーソナリティに関する尺度の中でも利用頻度の高い尺度の一つとして和田 (1996) の Big Five 尺度を取り上げ、短縮版作成を行った。2099 名のデータをもとに推定した slope パラメタと location パラメタをもとに 29 項目 5 件法の尺度が作成された。また、外在基準として同じ 5 因子モデルの尺度である NEO-FFI との相関を検討することで作成された短縮版の妥当性の検討を行った。その結果、作成された短縮版はオリジナル版と同様の相関等を示し、十分な信頼性・妥当性を備えていることが示された。

第 5 章では、新たな尺度として「過去の自己」と「現在の自己」とを比較する程度を測定する継時的比較志向性尺度を作成するとともに、その短縮版を作成し、検討を行った。まず、社会的比較に関する先行研究をもとに、継時的比較志向性尺度を作成し、信頼性・妥当性の確認を行った。その後、934 名から得た回答をもとに、IRT を適用した分析を行い、オリジナルの 11 項目に対して、5 項目からなる短縮版が作成されるとともに、当該尺度が短縮版として十分な信頼性、妥当性を持つことが示された。

第 6 章では、3 章から 5 章までの間で取り上げた尺度を用いて、因子分析をもとに短縮版の項目選択を行う方法と、IRT を用いた方法との比較を行った。その結果、いずれの方法で作成された場合においても短縮版として選択される項目は共通するものが多く、ほぼ同程度の信頼性や妥当性を備えた尺度が作成されることが示された。また、得点算出方法についても検討を行った。具体的には、従来多く用いられているように各項目の合計点を用いる場合と、IRT の θ を用いる場合との比較を行った。いずれの方法においても、得られた得点間の相関は非常に高く、同様の値が得られることが示唆された。これらの研究結果から、IRT を用いた方法も短縮版作成に有用であることが示された。しかしながら、本論文で行ったものは基礎的な検討にとどまっているため、IRT のパラメタ推定のプロセ

論文審査の結果の要旨

別紙 1 - 2

スや IRT のモデルの選択等においても課題が残っている。また、目的に応じた柔軟な短縮版作成や、短縮版作成方法のガイドラインの作成などの検討も今後必要であることが指摘された。

本論文に対して審査委員から以下の疑問点や示唆が示された。

- 1) 項目の DIF 分析を実施して、DIF 分析の結果に基づいて、「よい」項目と「悪い」項目とを分けて短縮版を構成することができるのではないかと。
- 2) テストが測定する内容を統一することなく各種の分析を行ったのには何か意図があったのか。
- 3) 短縮版開発の因子分析に探索的因子分析を用いているが、確認的因子分析は用いないのか。
- 4) 継時的比較を取り上げているが、社会的比較の方が研究も多いし、実際に個人が行なう比較でも多いのではないかと。
- 5) 心理臨床や医療などの現場では精度が低くても短時間で実施できるものが必要と云うが、それを使って何を知らうとするのか、また、結果はどのように利用するのか。
- 6) 少数項目の尺度が複数あると、個人の回答傾向など系統誤差の大きさが相対的に大きくなる。それが研究結果にどの程度影響するか考えているか。
- 7) 因子分析による短縮版作成の方がテスト情報量が高い。IRT 版の項目選択に問題があるのではないかと？
- 8) 短縮版の妥当性をどう担保するか。精度だけ追及すると妥当性が下がるのではないかと？

などの疑問に対して、

- 1) 第 3 章のデータで別の研究として実施している。小学生と中学生とで回答傾向が異なる項目を検出するなどは発達尺度を構成する際に重要な分析で、短縮版を開発するのにそういう分析も重要だと考える
- 2) 今回は特定の領域・概念に絞らずに、広い範囲を対象として応用可能性を示したかったため、テストが測定する内容を統一しなかった。
- 3) 項目選択の時は探索的因子分析、開発後に確認的因子分析を実施することになると考える。

論文審査の結果の要旨

別紙 1 - 2

- 4) 他者との比較、自分の過去との比較、自己評価では両方あり得ると考える。上の世代の方が若い世代よりも過去との比較をする傾向があると考えられるし、欧米と日本との間では文化差があると考えられる。
- 5) スクリーニングを目的として、短時間で多くのヒトを判別する。また、複数の尺度を使って多数の概念の関係を見たい研究などで重要な役割を果たす、と考える。
- 6) 確かに、項目数を少なくすればいいというものではない。研究でも臨床でもどの程度の項目数でいいのかは、原版開発の際には考えられていない。尺度毎に1セットの項目群として開発されるが、それを崩さない方がいい場合もあると考える。
- 7) 短縮版は、むしろ項目困難度が中程度に集中して、原版は項目困難度を散らすように開発する、という発想もあるかもしれない。そうすれば、指摘されたような問題も生じないと考える。
- 8) 医療現場で現実的な問題（スクリーニング）に応えるにはいいが、心理学研究で安易に短縮版を使って研究する場合には妥当性の問題が生ずると考える。

など適切に回答した。

また、何らかの処遇の事前と事後とで同じテストを用いて測定したい場合に、折半して2つの平行テストとして実施してもいいのではないか、そのような場合に IRT が特に有効に機能するのではないか、という示唆があり、今後研究を展開するにあたっての重要な課題と考える、という回答があった。

以上のように、博士学位請求者は審査委員からの疑問や示唆に対して、具体的、かつ適切に回答した。

論文の内容および口述試験の内容を総合して、審査委員は全員一致して、本論文を博士（心理学）の学位に値するものと判断し、論文審査の結果を「可」と判定した。