

Title	公開データを用いた心理学卒業研究の提案
Sub Title	A proposal for psychology graduate thesis utilizing open data
Author	平石, 界(Hiraishi, Kai)
Publisher	三田哲學會
Publication year	2024
Jtitle	哲學 (Philosophy). No.152 (2024. 3) ,p.117- 145
JaLC DOI	
Abstract	In the era of the credibility revolution in psychology and the growing adoption of open science practices, an increasing number of articles now openly share the materials, raw data, and analysis scripts utilized in research. This paper introduces a graduate research program leveraging these open resources. Students validate the reproducibility of reported findings using the provided open data and scripts, followed by conducting sensitivity tests to assess the robustness of the findings under varying analysis specifications. I argue that this program meets scientific and educational standards expected of a graduate thesis despite bypassing data acquisition processes typically required in a psychology course in Japan. I delve into the rationale for designing this program, outline its current implementation, and discuss its pros, cons, and future considerations.
Notes	特集：人間科学専攻特集号 寄稿論文
Genre	Journal Article
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00150430-00000152-0117

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

公開データを用いた心理学卒業研究 の提案

平 石 界*

A Proposal for Psychology Graduate Thesis Utilizing Open Data

Kai Hiraishi

In the era of the credibility revolution in psychology and the growing adoption of open science practices, an increasing number of articles now openly share the materials, raw data, and analysis scripts utilized in research. This paper introduces a graduate research program leveraging these open resources. Students validate the reproducibility of reported findings using the provided open data and scripts, followed by conducting sensitivity tests to assess the robustness of the findings under varying analysis specifications. I argue that this program meets scientific and educational standards expected of a graduate thesis despite bypassing data acquisition processes typically required in a psychology course in Japan. I delve into the rationale for designing this program, outline its current implementation, and discuss its pros, cons, and future considerations.

はじめに

本稿は筆者が 2022 年度から慶應義塾大学文学部人間科学専攻のゼミナール活動として開始した、実験や調査によるデータ取得をせずに定量的な心理学卒業論文を執筆するという卒業研究プログラムの現状を報告するものである。

このプログラムでは、公開データ付きの研究論文（ターゲット論文）を

* 慶應義塾大学文学部

「しゃぶり尽くす」ことが卒業研究の目標となる。学生は1本のターゲット論文を選び、その研究の背景（先行研究、著者の研究歴、掲載誌の特徴など）を調べ、付属の公開データから報告された結果が再現されるか確認し、さらに追加分析による頑健性テストを行う。それらの成果をまとめることで、卒業論文は文献研究としての側面と、定量的研究としての側面を持つことになる。卒業研究を通じ、学生が現代の心理学の研究手法について理解を深めるとともに、文献探索、統計、プログラミング、アカデミック・ライティングといったスキルを身につける機会となることを期待したプログラムである。

本プログラム開発の背景には、心理学における信頼性革命を経て明らかとなってきた種々の問題に、学部の卒論指導が如何に対応すべきかという問題意識がある。もちろん、本プログラムこそがそれら問題への唯一の解決策・対応策であるなどと主張するものではないし、他研究室やゼミナールでも同様のプログラムを実施するべきであると勧めるものではない。そもそも本プログラムの設計には、学際的な専攻ゆえ心理実験演習が必修化されていないといった筆者の所属先に固有の事情も大きく影響しており、一般的な心理学専修コースにそのまま適用することは適当と言いがたい。ただ、同様の問題を感じている研究室／ゼミナールの主催者がいたならば、何がしかの参考として貰えるかも知れない。

本稿執筆時点、プログラム一期生にあたる学部4年生が卒業論文の完成を目指して日々励んでいる。まだ卒業生を出していない未完成の段階でこうした原稿を発表するのは、今後のプログラムの改善のためにも、現時点の状況を一度まとめ、そして可能ならば他の研究者・教育者からの批判やコメントを得ることが望ましいと考えたためである。また、慶應義塾で学ぶ学生に、本塾にはこうした学習オプションがあることを伝える機会になることも期待している。

背景

本稿を執筆している 2023 年現在、日本国内で心理学の研究や教育に携わっている多くの人にとって「自らデータを取得することなく、定量的な心理学で卒業研究を行う」というアイデアは、荒唐無稽に思えるかも知れない。筆者も数年前までは、心理学のゼミナールの卒業研究ならば、自分のデータを自分で取ることは必須であると考えていた。実際、日本国内の心理学専修コースのほとんどでは、臨床心理学に完全に特化したコースでもない限り、学生が自らの力で卒業論文のための調査や実験を計画・実行できる水準に至ることを目指して、低学年時からのカリキュラムが構成されている。そのために実験演習や調査演習を通じて、データを取得したり、それを統計的に分析したり、その結果を解釈したり、心理学論文のフォーマットに従ってレポートにまとめたりといった訓練を受けることが標準的なカリキュラムとなっている。

心理学専修コース卒業生の多くは、将来的に心理学研究に携わるわけではないし、企業の調査やデータ分析といった部署に所属するわけでもないだろう。公認心理師として活動する場合にも、必ずしも調査や実験が日々の業務に含まれるわけではなさそうである。それにも関わらず、自らデータを取得することを目標に教育システムが構築されてきたこと背景には、人間に関するデータを自ら取得し、分析し、解釈すること特有の教育的効果があると見込まれてきたからではないだろうか。そしてまた、そうした経験を経て卒業論文を完成させてきた学生たちからの支持もあったからこそ、このシステムが維持されてきたのだと考える。筆者自身も、自らデータを取得するという体験がもたらすものの価値を否定する考えはない。その上で、敢えてデータ取得をしない卒業研究プログラムを実施している理由を説明してみたい。理由は複数あるが、いずれも直接的、間接的に心理学における再現性危機と信頼性革命の影響を受けている。

再現性危機と信頼性革命

再現性危機とは

2010年代初頭から、実験や調査によって集めた定量データを分析する定量的心理学で蓄積されてきた知見の再現可能性が「高くない」ことが大きな問題として注目を受けることとなった。最も有名な例としては、*Psychological Science*, *Journal of Personality and Social Psychology*, *Journal of Experimental Psychology: Learning, Memory and Cognition* という主要三誌に2008年に掲載された研究100本を追試した報告がある (Open Science Collaboration, 2015)。追試のターゲットとされた100の研究のうち、オリジナルの研究で統計的に有意な結果が得られていたものは97あった。しかし追試で統計的に有意な結果が得られたのは、そのうち35であった (35/97=36.1%) であった。この報告に限らず、2010年代前半の心理学界では、教科書に載っているような著名な研究の追試が次々と実施され、その少なからぬ部分が再現に失敗したという報告をもたらした。これが心理学における再現性危機 (replicability crisis) と呼ばれる出来事である。

研究者自由度

追試はなぜ失敗したのだろうか。追試者の技術が低く、正確な実験や調査が実施できなかった可能性もあるだろう。しかし該当するテーマの専門家が加わっても結果を再現できないこともあり、技術的稚拙さに全ての原因を帰することはできない。そうした中、定量的心理学研究における研究上の慣行が原因の一端であると指摘された。例えば、参加者が一定数集まる度に検定をかけて、有意になるまでサンプルを集めることや、関連しそうな多数の心理尺度を測定しておいて、様々な組み合わせで共変量として分析に加えて検定を繰り返すことなどである。こうした行為が偽陽性の確率を高めてしまうことは今や広く知られているが、往時の心理学研究の現場では、多くの人が行っていた (John et al., 2012)。

ここで偽陽性とは、実際には存在しない心理学的な効果が統計的に有意になることである。例えば上述のような慣行を駆使すると、ビートルズの楽曲には参加者の実年齢を若返らせる効果があるという統計的に有意な結果（言うまでもなく偽陽性である）を得ることもできる（Simmons et al., 2011）。元の結果が再現されなかった追試データを、手を替え品を替え分析し直すことで、追試が成功したかのような結果を引き出すこともできる（藤島・樋口, 2016）。研究の過程で研究者が自由に試行錯誤することは、探索的研究のために必要なことである（Nosek et al., 2018）。しかし仮説検証型の研究において、そうした研究者自由度（Researchers Degrees of Freedom）を濫用することは「疑わしい研究慣行」（Questionable Research Practices, QRPs）と言わざるを得ない。QRPs の氾濫が偽陽性の知見を積み上げ、再現性危機をもたらしたと指摘されたのである。

信頼性革命

研究者自由度の濫用という問題を解決し、より厳密な科学として心理学を進めるための提案が提出され実装されていった。これが信頼性革命（credibility revolution）と呼ばれる運動である（Vazire, 2018）。多くの提案・実装の中で最も重要な位置を占めるのが研究計画の事前登録（preregistration）だろう。データ取得もしくは分析より前に、分析対象の変数、測定方法、マテリアル、サンプルサイズ、分析方法、結果の解釈方針などを公的な場所にタイムスタンプ付きの凍結した状態で登録するという提案である。事前登録によって、仮説を検証した部分と、探索的に検討した部分の区別が可能になり、研究者自由度の濫用を防ぐことが期待される（平石・中村 2022）¹。

研究仕様決定の困難性

事前登録の重要性を認めるならば、研究計画の詳細、すなわち研究の仕様（research specifications）を事前に確定させることが必要となる。しかしそれが容易でないことが、信頼性革命の議論の中で明らかになってきた。

先行研究の信頼性

およそ科学研究のほとんどは先行研究の蓄積の上に立つものである。再現性危機とは、しかし、拠って立つ先行研究の信頼性が十分に担保されないという問題である。先行研究の手続きの妥当性や結果の頑健性が保証されないならば、どのような手続きを事前登録すれば良いのか、手がかりが無くなってしまう。具体的にサンプルサイズと剰余変数の統制という問題を検討してみよう。

サンプルサイズと効果量の問題

頻度主義にもとづく帰無仮説有意性検定を行う計画ならば、サンプルサイズを事前に決定する必要がある。しかし先行研究で用いたサンプルサイズをそのまま採用することは、その結果がQRPsによって水増しされている恐れがあることを考えると、リスクがある。十分なサンプルサイズを確保しないと、逆に（水増しされていたとは言え、実際に存在する）効果を見逃すことに繋がりがねないからである。検定力分析によってサンプルサイズを決定しようとしても、先行研究で報告された効果量が水増しされている恐れがあるならば、そのまま用いて必要なサンプルサイズを計算することは出来ない。Cohenの基準（Cohen, 1988）を用いるにしても、心理学理論の多くは効果の方向を論じるだけで、効果の大きさについては無言であるため、大中小いずれを採用するのが適切か、理論的に定まらない（Meehl, 1990）。

剰余変数決定の問題

先行研究の頑健性が保証されないとしたら、研究において統制すべき剰余変数をどのように見繕えば良いかも不明となる。追試が失敗した際に、剰余変数の統制不足が指摘されることは少なくない。例えば有名な仮説の追試が否定的な結果となった際に、講義で仮説を聞いたことのある参加者がいたことの影響が指摘された (Strack, 2016)。事後的な説明や解釈に考慮に値する指摘が含まれていることは稀ではない (Friesse et al., 2019; Coles et al., 2022)。それは単なる追試ですら剰余変数の網羅的な特定が困難であることを意味する。新しい理論や仮説の検証を目的として、データ取得前に統制すべき剰余変数を特定することは更に困難であろう。

一般化可能性の問題

剰余変数決定の困難さは、翻って、知見の一般化可能性の難しさを示している。追試ですら事後的に手続き上の欠点が指摘可能なのだとしたら、いかなる知見であっても、一般化が不可能になってしまう。実験室で得られた知見が日常場面で再現されなかったとしても、剰余変数の統制不足でいくらかでも説明できてしまうし、逆に、そもそものポジティブな結果すら、未測定 of 剰余変数ゆえにたまたま良い結果が得られた可能性が否定できなくなってしまう (Yarkoni, 2020)。言い換えれば、調査や実験からどのような結果が得られても、それは「まさにその調査／実験」において生じた歴史的な記述であり、そこから一般化可能な範囲がほとんど何も保証されないことになる。

信頼性革命は、適切な研究計画を立ててデータを取得することの困難さと、そうして得られた結果の一般化可能性の低さという問題を浮き彫りにした。他方で信頼性革命によって、科学としての心理学研究の手続きが改善してきた面もある。次節はその点を見てみよう。

オープンサイエンスの発展

信頼性革命を通じて、研究者が利用したマテリアルや、取得した生データ、分析に用いたスクリプトなどを公開するオープン・サイエンスの重要性が認識されるようになった（大向，2018）。事前登録もまた、データ取得前に研究者が準備していたものを公開することであり、オープンサイエンスの一環と言える。事前登録抜きでも、マテリアル、生データ、スクリプトが事後的に全て公開されるなら研究者自由度の濫用は抑制できるという議論もある（Rubin, 2020）。加えて、研究者も人間であることを考えれば、研究の過程でエラーを完全に無くすことはできない。

データやスクリプトの公開が進むにつれて、それらを用いた第三者による結果の再生性（reproducibility）の確認が行われるようになってきた（Cruwell et al., 2023）。「結果の再生性」とは、論文に付属して公開されたデータを用いて、論文に記載された分析を実施することで、論文記載の結果が再現されることを示している（国里，2020）。こうした公開データと公開スクリプトの検証を通じて、より適切な分析方法が提案されたケースもある（Silberzahn et al., 2014）。オープンサイエンスが知見の頑健性の確認に寄与したケースと言えるだろう。

データの分析プロセスにおいて、複数の分析方針が同等の妥当性を持ちうる場合がある（Gelman and Loken, 2013）。そこで公開データに対して考えうる限りの妥当な分析方法を網羅的に試すことで、論文で報告された知見が特定の分析仕様に依存したものではないか確認することが提案されている（Steenen et al., 2016; Simonsohn et al., 2020）。マルチバース分析（multiverse analysis）や仕様カーブ分析（specification curve analysis）と呼ばれるこの分析法は、以前からある感受性テストを、システムティックかつ網羅的に行うものと言える。

信頼性革命を通じて、公刊論文に付属して、研究に用いたマテリアル、取得された生データ、分析スクリプトが公開される動きが進んできたこと

を紹介した。そして、第三者によるこれら公開情報を用いた検証が、心理学の進歩に寄与することを論じた。次章では、ここまで紹介してきた信頼性革命の影響を踏まえ、定量的心理学の卒業研究として何が可能かを検討してみよう。

卒業研究は何を目指すのか

信頼性革命は心理学における研究遂行手続きの見直しを迫るものであった。そのことを踏まえて卒論生と共に「研究」を行うことを考えた時に発想されたのが「公開データを用いた心理学卒業研究（以下、公開データ型卒論）」というプログラムである。動機的一方には、信頼性革命を経た現代的な基準を満たしつつ、データ取得型の卒業研究を実施することの困難さという、ネガティブな理由がある。しかし他方には、公開データの存在によって、自らデータを取得せずとも、卒業研究に期待される機能を果たすことが可能な状況が生まれてきたというポジティブな理由がある。また、実験実習などが必修化されていない学際的な専攻で、定量的心理学の卒業研究から「何か」を得たいという固有の事情もある。以下では、卒業研究とは何をを目指すものか筆者の考えを整理しつつ、公開データ型卒論がそれらの目標を満たすものか検討する。

科学への貢献

学部の卒業研究を学生に課すにあたって、心理学という学問への貢献を全く考慮しない指導者は少数だろう。そもそも学問への貢献を全く意図しないのであれば、それは「研究」ではなく「演習」や「実習」と呼ぶのが適切なはずである。研究者（卒論生）の経験も浅く、研究に費やせる時間等のリソースも制限される中であるとは言え、何らかの学問的貢献を目指すことは、結果としてそれが達成されるか否かとは別に、卒業研究にも求められるはずである。

公開データ型卒論は科学に貢献しうるだろうか。既に述べたように、論文に付属の公開データを再分析することは、結果の再生性と頑健性を確認することによって学問に寄与する。「新発見」と呼べるものをもたらす可能性が高くないことは認めざるを得ない。しかし、新奇な結果ばかりを追求するコミュニティの文化が、研究者自由度の濫用と再現性危機を招いたという指摘を踏まえれば（池田・平石、2016）、知見の確実さを地道に確認しつつ進む道を選ぶことも、科学へのアプローチの一つとして認められるべきだろう。

教育効果

卒業研究は教育課程の一環として課されるものであり、そこには一定の教育的効果が期待されることも明らかである。卒業研究を通じて、学生が取得すると期待されるものとは何だろうか。著者は大きく分けて2つを考えている。一つは科学や学問への認識であり、もう一つは具体的・実践的なスキルである。

研究への認識

卒業研究では、学生が自ら「知」を生産する作業に携わることを通じて、新たな「知」を得ることの喜び、研究の楽しさを知ることが期待されているだろう。それは同時に「知」を生産することの困難さを知ることにもつながる。「新しい知」を確立することは容易でないこと知るとは、世間に流布する「発見」の主張を批判的に吟味する態度と技術を養うことにもつながる。研究の楽しさと難しさを体験を通じて学ぶこと、これが卒業研究に期待される教育効果の一つではないかと筆者は考える。

公開データ型卒論は、学生が研究の楽しさと難しさを学ぶ場として機能するだろうか。単純な「新発見の喜び」について言えば、回答は否定的なものとならざるを得ない。しかしそれはデータ取得型卒論にも言えること

かも知れない。実験によって統計的に有意な結果が得られたとして、実験室外へのその知見の一般化可能性が不明ならば、「発見」の程度が減じかねないからである²。しかし学問の楽しさは、ただ発見にあるだけではない。ある研究テーマについて、その学問分野で培われてきた方法論を駆使して、妥当と考える結論を得るという作業そのものに面白さ、楽しさがあることは、研究を生業にしている者であれば同意するところだろう。そして公開データ型卒論も、そうした面白さを得る機会となりうると筆者は考えている。しかしこの点については、実際の学生の声を聞いてみる必要があるとも考えている。

研究の難しさの認識についてはどうであろうか。データ取得という経験を通じてだけ得られる「難しさ」への認識はある。質問項目作成、参加者集め、実験手続きの管理に伴う苦労などは、実際にデータ取得を経験しないと分からないものだろう。公開データ型卒論で得られる「難しさ」の認識は、これらとは少し異なったものとなると思われる。それは、知の確立には研究者コミュニティによる慎重な検証が必要であるという認識からスタートする。座学で学んだそのような認識を、公開データの再分析という実践を通じて確認するのが公開データ型卒論となるだろう。

つまりデータ取得型卒論と、公開データ型卒論では、それを通じて得られる「研究の面白さの認識」「研究の難しさの認識」が質的に異なる部分があると筆者は考えている。プログラム開発においては、公開データ型卒論の持つ特徴をできる限り活かすことを目指すのが効果的だろう。

実践的なスキル

卒業研究にはまた、実践的、実用的なスキルを取得するという機能も期待されていると筆者は考える。どのようなスキル取得が期待できるかに学問分野の特徴が現れる。従来のなデータ取得型の定量的心理学研究であれば、それはプロジェクト管理、文献調査を含む情報収集、英文読解、研究

デザイン（実験や質問紙のデザイン）、実験や調査の実施、統計分析、アカデミック・ライティングといったところであろう。

データ取得型卒論と、公開データ型卒論で最も大きな違いが現れるのが研究デザインにかんするスキル取得だろう。自らの仮説に合った研究デザインを考案する経験は、データ取得型卒論でなければ得られないものである。ただし、心理学研究において、剰余変数やサンプルサイズの設定といった研究デザインの根本部分を理論的に特定することが困難であることは、既に指摘した通りである。研究デザインにまつわる経験は、「正解」が不明な中でも何らかの意思決定をしてプロジェクトを前進させるスキルを学ぶ場として機能している可能性がある。

具体的な実験や調査の実施に必要なスキルの取得も、データ取得型の卒業研究でなければ得られないものである。fMRI や脳波による研究を考えれば分かりやすい。公開データを用いてこれらのデータ分析部分だけを経験することと、実際に機器を操作してデータを取得することには、絶対的な差がある。実験室実験やオンライン調査の実施についても同様である。学生の目標がそうした技術の取得にあるのならば、公開データ型卒論から得られることはかなり限定される。

その他のスキルについては、データ取得型卒論と公開データ型卒論で大きな違いはないと考えられる。公開データ型卒論であってもプロジェクト管理は必要であるし、関連する情報収集、英文読解、統計分析、アカデミック・ライティングといった技術は、論文完成のために必須だからである。データ取得型卒業論文でデータ取得のために用いるリソースを、それ以外のスキルの取得に回すことができるとも言える。演習を通じたデータ取得スキルの養成が不十分なカリキュラム下では、この違いの持つ意味は小さくない。

ここまでの議論から、公開データ型卒論であっても、一般に卒業研究に求められる要件を満たす見通しが得られることを論じてきた。ただし学生

が体験し身につけるものは、データ取得型卒論と全く同じではないことも明らかである。そのため、公開データ型卒論ならではの特徴を活かしたプログラムを開発することが重要となるだろう。次章では、筆者のゼミナールで現在行っているプログラムの概要を紹介する。

プログラムの構成

プログラムの目標

本稿執筆時点（2023年度）における筆者のゼミナールにおける卒業研究の目標を、前章で検討した卒業研究の諸機能に則して整理すると次のようになる。まず科学への貢献としては、既刊論文の結果の再生性と頑健性を、付属の公開データを用いて検討することにある。また、そうした卒業研究が集まることで、現在の心理学研究におけるオープンサイエンスの実態を探り、その促進に貢献することが、ゼミナールのメタな目標となっている。

教育効果のうち、特に学問・研究の認識については、「心」を科学的に研究するという心理学研究の難しさと、そのためのオープンサイエンスの重要性について、実践を通じて理解することが目標である。権威ある雑誌に掲載されたプロの研究者の手になる論文であっても、学部学生がその結果を再検討することは十分に可能であることを体感し、また、そうした再検討が科学の着実な進歩には必須であること、そのための基盤としてオープンサイエンスが不可欠であることを学ぶことを目指している。「研究の楽しさ」については、特定の目標は設けず、学生一人ひとりが自分なりの楽しさを見つけることを期待している。

実践的なスキル取得については、特に情報収集とデータ分析に重点を置いている。卒論では、元の論文で報告された分析の再現だけでなく、同等に妥当と考える分析を追加することを推奨している。そのためにはターゲット論文に類似した他の研究を調べるなど、研究テーマの全体的把握が

求められる。その経験を通じて、1本の論文は孤立して存在するのではなく、先行研究群などによって作られる文脈の中で意味を持つという学問観を得ることも期待している。

ゼミナールへの所属は3年前期から始まる。4年後期までの2年間で上記の目標を達成することを目指してプログラムを配置している。ただし、そこには所属する学部専攻のカリキュラム上の特色と制約が関わっていることは言うまでもない。そのことを断った上で、具体的なプログラムの概要を紹介してみよう。

学部3年時

ゼミ所属1年目となる学部3年時には、1) 英語で書かれた心理学研究論文の読解、2) 心理統計の基本、3) アカデミック・ライティングの基本、4) 信頼性革命を経た現代的な心理学で求められる標準的な手続き、をスキルとして学ぶことを目標としている。第1から第3は一般に心理学専修コースのゼミナールにおいて設定される目標だろう。第4は本稿の冒頭で紹介した信頼性革命の概要に関する知識であり、「心理学研究法」などの科目でカバーされる内容である。しかし筆者の所属先には心理学に特化した方法論の課目が設定されていないため、ゼミナール内で手当している。

具体的な設計は以下の通りである。まず目標2については心理統計の講義(通年)の履修を求めることで対応している。並行して、前期に信頼性革命に関する英語文献を読むことで、目標1と目標4に対応している。これらの文献を読みこなすには統計の基礎的な知識が不可欠であるため、心理統計にかんする知識を定着させることにもつながる。2023年度は再現性危機と信頼性革命から5つのテーマを設定し、関連論文を取り上げた。すなわち、1) 再現性危機の発端となった Bem による超能力論文と、それへのコメント等 (Bem, 2011; Wagenmakers et al., 2011)、2) 研究者自由度に関連する文献 (Simmons et al., 2011; 藤島・樋口, 2016)、3) 大規模追試の

実状 (Hagger and Chatzisarantis, 2016; Baumeister, 2019), 4) 分析仕様の違いによる影響 (Rohrer et al., 2017; Silberzahn et al., 2018), そして, 5) 公開データを用いた結果の再生性チェックの現状である (国里, 2020; Claesen et al., 2021).

3年の後期には, 3から4人のグループで公開データが付属した論文について再分析を行いレポートにまとめることで, 目標3への対応を目指している. 教員指定の論文リストから1本をターゲット論文として選び, その結果の再生性の確認, 追加分析を実施し, 卒業論文の構成(後述)に準ずる形で期末レポートを書くことが学生にとっての課題である. そのプロセスを通じて4つの目標全ての達成度を高めることを目指している. 論文としては英語の短報を選んでいる. 2023年度は, 後期(10月開始)冒頭にターゲット論文の候補リストを提示し, 12月中旬にゼミで分析結果を報告することを第1の課題として設定した. 並行して4年生の手になる卒業論文の草稿を読む機会を設け, そこで得た経験を踏まえて1月末に期末レポートとしてまとめることを第2の課題として設定している. 報告はグループ単位で行うが, 期末レポートは各自での執筆となる.

前期と後期の間の夏期休暇中に一日, 心理学分野で多く用いられる統計分析環境であるRの勉強会を設定している. ターゲット論文に公開スクリプトが付いている場合, その多くはRで書かれている. また, 分析過程はスクリプトとして残す方が公開は容易であり, かつ, フリーで使えるプログラミング言語で書かれていることが望ましい. 所属専攻の統計講義がHAD(清水, 2016)またはSPSSを主として用いているため, ゼミナールとしてRを学ぶ機会を設けている.

学部4年時

学部4年時には, それまでに学んで来た各種スキルを駆使して卒業研究を進め, その過程で各スキルに磨きを掛けることを教育目標としている.

4年時の活動は、大枠を除けば、学生個人ごとに異なるペースで進めている。大枠としては5月末を目標にターゲット論文を確定し、夏期休業中に中間報告を行い、11月末に最初のドラフトを提出することとしている。学部の公的な卒論の締切りは1月中旬である。このスケジュールを基本に、学年開始前の23月に教員を含む少人数のミーティングでテーマ探しの方向性と、進路に向けた活動（i.e., 就職活動）との兼ね合いを念頭におきつつ、各自の48月の大まかなスケジュールを確定することが、卒業研究のスタートとなる。

ターゲット論文の選定においては、自分の興味のある研究テーマであることは当然としつつ、データが公開されているか、分析スクリプトが公開されているかを確認し、用いられている統計手法の難易度についても、教員と相談しつつ、検討することを求めている。ターゲット論文が決定したら、当該論文の詳細をまとめつつ、引用文献などを手がかりに、当該テーマにおける研究状況全体のあらましを把握し、夏期休暇中に行う中間報告会において報告する。その内容を踏まえて、卒業論文の序論（後述）のドラフトを書くことを推奨している。

中間報告を終えたら論文記載の分析結果の再現に入る。しかる後に、論文には記載されていない追加分析を検討・実施する。それらの結果をまとめて11月末に最初のドラフトを作成する。この第1ドラフトを3年生を含めたゼミナール内で互いに読み合い、改善点を話し合う。この時点では個別のドラフトの問題点を挙げるのではなく、一般的な改善点を検討することを目指している。その議論を踏まえた第2ドラフトを12月中旬に作成する。この第2ドラフトには教員が個別にコメントを付ける。個別コメントを参照しつつ最終版を作成し、1月中旬の締切りまでに大学に提出する。その他に1月末にA4で12ページの論文概要を、一般公開することを念頭にまとめる。以上で卒業研究が完成する。

論文の構造

公開データ型卒論で報告が求められる内容は、一般的なデータ取得型の研究のそれとは異なる。「序論・方法・結果・考察」という一般的な心理学論文のフォーマットに合わせようとするとうる覚えが生じる点もあるため、ゼミナール独自のフォーマットを準備し、卒業論文執筆の補助としている。このフォーマットはまだ不十分なところも多いが、現状の報告という観点から、その概要を説明しておく。論文の章立てとしては「序論」「再分析」「追加分析」「総合考察」の4章を設けている。一般的な心理学論文で考えれば、Study 1に相当するのが「再分析」であり、Study 2に相当するのが「追加分析」ということになる。以下、各章に含めるよう指示している内容を紹介する³。

序論

序論では、一般的な心理学論文と同様、研究の目的、理由、意義を含めることを求めている。それに加えてターゲット論文の背景と、その重要性を論じる節を設けるように指導している。具体的には、当該分野の研究史に加え、掲載誌や、著者（著者グループ）についても調査して紹介するように指示している。科学といえども人間の営みであり、そこには社会情勢や研究者間の人間関係など、さまざまな文脈が影響を与える。研究というもの、論文に明示された情報だけで成立する営みでないことを知ることは、研究の難しさ（そして面白さ）を理解することにつながると期待して、通常の研究論文では言及されない、社会的な背景文脈についても論文に含めることを求めている。

背景を述べた上で、ターゲット論文の紹介を行う。特にターゲット論文の方法と結果、そして考察について詳しく紹介することを求めている。付録（supplementary）に記載されている手続きやマテリアルの詳細も確認して、必要ならば説明として加えることを求めている。記述レベルとして

は、専門課程に進んだばかりの専攻の学部2年生が読んで理解できることを目標とするように指示している。つまり「統計的有意性」といった基本概念についても、知らないことを前提に、説明を補いつつ導入するように指示している。

再分析セクション

ターゲット論文で報告されている結果の再生性を確認する。データに加えて分析スクリプトも公開されている場合は、それを用いた再生を試みる。スクリプトが公開されていない場合は、著者にスクリプトの共有を依頼するか、論文の記述に従って自らの手で分析を実行する。

このセクションでは、再分析に当たっての手續きと、再分析の結果ならびにその考察を報告する。手續きとしては、公開データの入手方法、ファイル形式、分析環境（PCのOS、統計ソフトウェア名ならびにバージョン情報など）、分析スクリプトそのものを報告することを求めている。そしてターゲット論文の結果が再現されたか、再現されなかった場合はどのようなレベルで再現されなかったのか（e.g., スクリプトが動かない、出力が少し違う、出力が全く違う）を報告した上で、ターゲット論文の結果の再生性について評価する。

追加分析セクション

ターゲット論文で報告された分析と、同等に妥当と考える他の分析仕様（analysis specifications）による分析を行うことで、ターゲット論文の結論の頑健性をテストする。追加分析セクションでは、目的（どのような追加分析を、何を目的に行うのか）、方法（分析仕様の詳細）、結果、そして考察を記述することを求めている。追加分析とターゲット論文の分析で異なる結論が導かれた場合には、安易にターゲット論文を否定するのではなく、それぞれの分析仕様の妥当性を比較して考察することを求めている。

査読を経て公開された論文と異なる分析仕様を考察するためには、当該テーマ並びに統計手法についての相応に高い専門性が要求される。その意味で本セクションは、多くの卒論生にとって最も高いハードルとなる。ターゲット論文がなぜ特定の分析仕様を選択したのか正確に理解することなしに、それと同等に妥当な別の分析仕様が存在すると主張することはできない。そして何が妥当な分析仕様であるか論じるためには、当該分野における過去の蓄積（先行研究）の理解と、種々の統計手法の理解が求められるからである。研究テーマやデータ、ターゲット論文の性質によっては、適当な追加分析の実行がほとんど不可能な場合もある。例えばターゲット論文そのものにおいて、仕様カーブ分析による分析仕様空間の網羅的探索が行われているような場合に、再分析者（卒論生）が手出しできる隙はほとんど残されていないこともある。これらの難しさを承知の上で、情報収集スキルとデータ分析スキルの涵養という教育目標のために、敢えて高いハードルとして設定しているのが本セクションである。

総合考察

再分析、追加分析の結果と考察を踏まえ、ターゲット論文についての評価を行う。それを踏まえ、より大きな研究テーマについて、自らの卒業研究を踏まえて何が言えるかを考察することを求めている。しかし心理学の知見の一般化可能性の問題に注意し、無理に一般論を述べる必要はなく、むしろその限界を丁寧に検討することを求めている。個々の卒論生の好みで、総合考察の後に「結論」というセクションを設けても良いこととしている。

現状評価

本稿冒頭で述べたように、この公開データ型卒論というプログラムはまだ卒業生を出していない未完成のものであり、その設計にも不十分なとこ

ろが多々ある。ここではプログラム開始（2022年度）から2年弱の時点
で明らかとなってきた問題と、得られつつある成果について紹介しておき
たい。

困難が生じるポイント

公開データ型卒論を進める上では、いくつかの実際上の困難が生じる。
列挙してみたい。

データの公開状況

まず公開データの付属した論文を探すという困難がある。学生が関心
を持った論文がデータを公開しているとは限らない。信頼性革命後の論文で
はデータ公開が進んでいるとは言え、全ての論文に生データが付属してい
るわけではない。それ以前（具体的には2010年度より前）の論文に至って
は、公開データが付属していることは、ほぼ期待できない。公開データが
入手できない場合には、当該論文をターゲットにすることを諦めるか、著
者にデータ共有を依頼するかの二択となる。幸いにして2023年現在、機
械翻訳と生成AI技術の発展により、非ネイティブ話者が英文で依頼メー
ルを書くことのハードルは大きく下がっている。実際、依頼メールを通じ
てデータを入手した学生もいる。オープンサイエンスの理念からすると望
ましくない状況だが、学生の学びの機会としては悪くないかも知れない。

分析スクリプトの公開状況

公開データが入手できたとして、それだけで結果の再生性が容易に確認
できるわけではない。分析スクリプトがセットで公開されていない場合に
は、論文の記述から分析を再現しなければならないが、記述が簡略すぎて
具体的な分析が不明なことがある。例えば分析の単位が参加者なのか、刺
激なのかが不明といった問題である。自由度などの周辺情報から見当をつ

けて候補を絞り込んで分析を行いつつ、ターゲット論文の著者に問い合わせをすることになる。学生の教育機会としては有用と言えるが、オープンサイエンスの観点からは望ましくない状況である。もっとも、そうした問題が実際に生じていることを確認するのが、本プログラムのメタな目的でもあることは、先述した通りである。

計算資源

データとスクリプトの両方が公開されている場合には、結果の再生性確認のハードルは大きく下がる。しかしその場合でも、コードブックが付属していないために、論文で言及される構成概念と、スクリプトに含まれる変数の対応を付けることに苦勞する場合がある。R を用いた分析経験が浅い学部生にとっては、スクリプトの内の各関数がどのような働きをしているのか理解することも容易ではない。もっとも、ChatGPT の助けを借りることで、こうした「解説」の問題は大きく改善されてきている。むしろ問題となるのは、ターゲット論文の分析が SPSS や Stata, Mathematica といった有料ソフトウェアを用いている場合だろう。これらのソフトウェアを使える環境がない場合、結果の再生性確認は困難になる。また少し異なった問題として、ターゲット論文で MCMC 法など計算負荷の高い手法が用いられている場合に、学生私有の計算機環境では計算時間が膨大なものになってしまうケースもある。

統計手法

ここまで述べてきたような問題点がクリアできたとして、ターゲット論文が用いている統計手法が高度に複雑なものであるために、その理解に多大な時間と労力を要する場合も少なくない。例えば近年の心理学論文では、一般化線形混合モデル (GLMM) を用いることがほとんどデフォルトとなっている。一般線形モデル (分散分析と重回帰) をようやく理解した

程度の学生にとって、決して楽なハードルではない。

教育上のメリット

公開データ型卒論で生じる実際的な困難をここまで述べてきた。しかしこれらの困難は、教育的視点から見たときには、必ずしも悪いことばかりではない。ターゲット論文著者にデータやスクリプトの共有を求めて連絡を取る経験は、学生にとって少なからず自信をもたらすであろう。複雑な統計分析モデルを理解しようと奮闘することも、データ分析スキルの獲得につながるだろう。追加分析のために自らの手でスクリプトを書く経験が勉強になることは言うまでもない。言い換えれば、データ取得という一般的な心理学の卒論の一大難関と目されるステップをバイパスしたとしても、学生が様々なスキルを学習する機会は十分に担保されていると考える。

指導する教員側にとってのメリットも少なからずある。個人的に最も大きく感じているのが統計分析にかんする点である。先述のように GLM や GLMM を用いたターゲット論文は少なくない。これらの手法を学部の心理統計の講義で入門的に導入する必要に気づくなど、自分自身の意識をアップデートする機会となっている。今後、例えば機械学習技術の導入が進むなどといった変化があった際に、教育内容をアップデートしていくための良い契機となるのではないかと考えている。

今後の課題

プログラムに現状足りていない課題を、教育面と研究面から挙げておきたい。

教育面の課題

まず、研究を進めるための基本スキル習得の機会（講義）を設ける必要がある。具体的には、文献探索と文献管理スキル（文研管理アプリケーションの活用方法）、効果量と検定力分析、ベイズ統計、プログラミング（プロジェクト管理を含む）、などである。これはプログラム単体の問題というよりも、学部や専攻のカリキュラムの問題と言うべきかもしれない。いずれせよ、学生が卒論を進める上で予め持っている良い知識やスキルをリストアップしなければならない。その上で、それらの知識／スキルを取得できる講義を新たに開講したり、学内外での学習機会（他学部やオンラインの公開講義、教科書など）のリストを整備する必要がある。

統計分析については、「同等に妥当と考えられる分析仕様から構成される空間が存在する」という発想（garden of the forking paths）から、様々な分析手法を整理して示す必要もある。従来の心理統計の講義では、比較的シンプルな問題設定（e.g., コイントス）において最適な分析手法（二項検定）を示す形を取ることが多い。入門の場面ではこうしたアプローチが必要だろう。しかし心理学（特に社会心理学）のデータ分析では、誰もが納得する一つの最適な分析仕様が定まるとは限らないことが明らかとなっている（Silberzahn et al., 2018）。「同等に妥当と考えられる分析仕様」が、どのような次元から構成される空間に配置されているのか整理することが、公開データ型卒論で追加分析を行う際に求められる。

現時点ではサンプル指定（サンプル除外規定、サブグループの設定）、変数の操作的定義（合成変数の作成方法、欠損値の扱い、因子分析モデル）、分析モデル（共変量セット、誤差分布、リンク関数、ランダム効果の設定など）を示している。しかし、それらの概念の意味について、十分な学習機会を提供できているとは言い難い。今後の課題である。また、ターゲット論文の検定力にかんする感度分析によって、ターゲット論文で採用された研究仕様について評価することも重要と考えられるが、検定力分析にか

んする教育も不足している。

研究面の課題

現時点で最も準備が遅れているのが、卒業論文の公開手順の整備である。第三者が公開データを用いて結果の再生性を確認することは、その結果が公開されることによって、初めて学問に貢献するものとなる。他方で、そうした情報の公開は（再分析の結果が元々のターゲット論文の結論に好意的であれ、否定的であれ）、相応の責任を伴うものであり、それを卒論生一人に背負わせるのは不適切であろう。公開に当たって必要と考えているポイントを示しておきたい。

まず再分析者の身分（学部の卒論生）を明示する必要がある。科学の前には誰もが平等であるが、研究者によって技術的な巧拙があることもまた事実である。公開された論文の読者がその価値を判断するための手がかりとして、学部卒論であることを明示することは、あながち間違いではないだろう。ターゲット論文における重大な瑕疵を指摘するような結果が得られた場合には、卒業論文をそのまま公開するのではなく、当該論文の著者に連絡を取る、掲載誌に連絡を取る、Data Colada のようなデータ再分析チームに相談するといった作業をまず行うべきであろう。いずれの場合も、卒論の指導教員名を明示し、論文執筆者と同等の責任を負う必要がある。

公開する内容についても整理が必要である。卒論本文を公開するだけでは不十分であり、再分析、追加分析に用いたスクリプトの公開は必須である。その際には、本文と各スクリプトの対応が明白になるような説明を付与する必要がある。そのためにはプログラミングにおけるプロジェクト管理の知識が有用である。そのような技術を学ぶ機会を用意することは、結果として公開へのハードルを下げることにもなる。

最後になるが、卒論生本人の卒論公開の意思を尊重しなければならない

い、プログラム参加時点で将来的な公開を前提に卒業研究を進めることを明示しつつ、卒業研究が完成した段階で、改めて公開の意思を確認するべきであろう。

まとめ

筆者のゼミナールでこの2年間進めてきた、公開データを用いた定量的心理学卒業論文というプログラムについて、その背景、目標、実装状況、現状評価を述べてきた。冒頭にも書いた通り、他の研究室やゼミナールでもこのプログラムを採用するべきと主張するものではない。実験や調査といったデータ取得を経験しなければ得られないものも確実に存在するからである。他方で、公刊論文に付属する公開データを用いても、学問的・教育的に意味のある卒業研究は十分に可能と考えている。心理学研究において新しいデータを取得することは、シミュレーション研究などの一部を除き、参加者や被験体に何らかの負担を強いることを意味する。特定の時期に卒論を目的とした調査が集中することで「調査公害」とでも呼びうる状況が生じがちであることも、多くの心理学教育担当者が経験するところであろう。今までと異なる形の卒業研究を検討中の方に、本稿が何がしかの参考になれば幸いである。

謝辞

本研究には JSPS 科研費 JP19H01750 の助成を受けた。本プログラムの開発ならびに本稿執筆にあたって、当該科研費のメンバー、ReproducibiliTea Tokyo のメンバーとの議論から多くの示唆を得た。何より、当プログラムに参加して共に研究を進めている、慶應義塾文学部人間科学専攻の平石界研究会に所属の学生たちなくして、本稿が著されることはなかった。記して感謝する。

註

- ¹ 言うまでもなく事前登録をすれば全ての問題が解決するわけではない。
- ² 一般化可能性による「面白さ」への影響は分野によって異なる。統制された環境下での人間の思いがけない行動やその原理を発見する面白さを追求する心理学分野もあるだろう。
- ³ 2023 年度に使用したフォーマットを公開している (<https://osf.io/u5syj>)。

引用文献

- Baumeister, R. (2019). Self-Control, Ego Depletion, and Social Psychology's Replication Crisis. doi: 10.31234/OSF.IO/UF3CN
- Bem, D. J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *J. Pers. Soc. Psychol.* 100, 407–425.
- Claesen, A., Gomes, S., Tuerlinckx, F., and Vanpaemel, W. (2021). Comparing dream to reality: an assessment of adherence of the first generation of preregistered studies. *R. Soc. Open Sci.* 8, 211037.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*(Second Edition). Hillsdale, New Jersey: Laurence Erlbaum Associates.
- Coles, N. A., March, D. S., Marmolejo-Ramos, F., Larsen, J. T., Arinze, N. C., Ndukaihe, I. L. G., et al. (2022). A multi-lab test of the facial feedback hypothesis by the Many Smiles Collaboration. *Nat Hum Behav* 6, 1731–1742.
- Crüwell, S., Apthorp, D., Baker, B. J., Colling, L., Elson, M., Geiger, S. J., et al. (2023). What's in a Badge? A Computational Reproducibility Investigation of the Open Data Badge Policy in One Issue of Psychological Science. *Psychol. Sci.*, 9567976221140828.
- Friese, M., Loschelder, D. D., Gieseler, K., Frankenbach, J., and Inzlicht, M.

- (2019). Is Ego Depletion Real? An Analysis of Arguments. *Pers. Soc. Psychol. Rev.* 23, 107–131.
- 藤島喜嗣・樋口匡貴 (2016). 社会心理学における“*p-hacking*”の実践例. *心理学評論* 59, 84–97.
- Gelman, A., and Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “*p-hacking*” and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University* 348. Available at: <http://stat.columbia.edu/~gelman/research/unpublished/forking.pdf>
- Hagger, M. S., and Chatzisarantis, N. L. D. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspect. Psychol. Sci.* 11, 546–573.
- 平石界・中村大輝 (2022). 心理学における再現性危機の 10 年：危機は克服されたのか、克服され得るのか. *科学哲学* 54, 27–50.
- 池田功毅・平石界 (2016). 心理学における再現可能性危機：問題の構造と解決策. *心理学評論* 59, 3–14.
- John, L. K., Loewenstein, G., and Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychol. Sci.* 23, 524–532.
- 国里愛彦 (2020). 再現可能な心理学研究入門. 専修人間科学論集. *心理学篇* 10, 21–33.
- Meehl, P. E. (1990). Why Summaries of Research on Psychological Theories are Often Uninterpretable. *Psychol. Rep.* 66, 195–244.
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., and Mellor, D. T. (2018). The preregistration revolution. *Proc. Natl. Acad. Sci. U. S. A.* 115, 2600–2606.
- 大向一輝 (2018). オープンサイエンスと研究データ共有. *心理学評論* 61, 13–21. Open Science Collaboration (2015). PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science* 349, aac4716.

- Rohrer, J. M., Egloff, B., and Schmukle, S. C. (2017). Probing Birth-Order Effects on Narrow Traits Using Specification-Curve Analysis. *Psychol. Sci.* 28, 1821–1832.
- Rubin, M. (2020). Does Preregistration Improve the Credibility of Research Findings? *Tutor. Quant. Methods Psychol.* 16, 376–390.
- 清水裕士 (2016). フリーの統計分析ソフト HAD : 機能の紹介と統計学習・教育, 研究実践における利用方法の提案. *メディア・情報・コミュニケーション研究* 1, 59–73.
- Silberzahn, R., Simonsohn, U., and Uhlmann, E. L. (2014). Matched-names analysis reveals no evidence of name-meaning effects: a collaborative commentary on Silberzahn and Uhlmann (2013). *Psychol. Sci.* 25, 1504–1505.
- Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., et al. (2018). Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. *Advances in Methods and Practices in Psychological Science* 1, 337–356.
- Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* 22, 1359–1366.
- Simonsohn, U., Simmons, J. P., and Nelson, L. D. (2020). Specification curve analysis. *Nat Hum Behav* 4, 1208–1214.
- Steege, S., Tuerlinckx, F., Gelman, A., and Vanpaemel, W. (2016). Increasing Transparency Through a Multiverse Analysis. *Perspect. Psychol. Sci.* 11, 702–712.
- Strack, F. (2016). Reflection on the Smiling Registered Replication Report. *Perspect. Psychol. Sci.* 11, 929–930.
- Vazire, S. (2018). Implications of the Credibility Revolution for Productivity,

Creativity, and Progress. *Perspect. Psychol. Sci.* 13, 411–417.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., and van der Maas, H. L. J. (2011).

Why psychologists must change the way they analyze their data: the case
of psi: comment on Bem(2011). *J. Pers. Soc. Psychol.* 100, 426–432.

Yarkoni, T. (2020). The generalizability crisis. *Behav. Brain Sci.* 45, e1.