

回顧的プレテストの2つの効用

——授業による変化の把握と受講生の自己認識への影響——

The Dual Role of the Retrospective Pretest: Measuring Subjective Gains in Learning Content and Cultivating Perceived Growth

浦上昌則

Masanori URAKAMI

要 旨

本研究は、心理学の調査法にかかる授業を対象として、プレテストーポストテストに回顧的プレテストを組み入れることの効用について検討することを目的とした。4月から7月にかけて開講された授業を対象とし、学習内容である調査研究を行う力に関する自己報告等について、プレテストと第1クォーター、第2クォーター終了時それぞれで回顧的プレテスト、ポストテストへの回答を求めた。分析の結果、今回の調査項目においては、レスポンスシフトが生じないとはいえないことや、授業内容との対応関係から、回顧的プレテストーポストテストの方が妥当な変化を把握できていることが示唆された。また回顧的プレテストを採用し、授業開始時を回顧し、現在と比較することに関する受講生の感想を分析したところ、頻繁に用いられる言葉に「成長」があり、回顧や比較することで自分の成長感に意識が向くことが示唆された。得られた結果を踏まえ、回顧的プレテストを使用した研究、実践について議論した。

問題と目的

通例、授業の成績は、授業終了時における受講者のパフォーマンスがどの程度であるか、どの程度授業の目的に近づいているかの指標に基づいて付与されよう。しかしこういった指標は、授業の効果を反映しているとはいえない。なぜなら授業の効果は、授業前後での受講生の変化に表れるものだからである。そこで受講による変化の程度を把握しようとする場合、プレテストーポストテストデザインが利用されることが多いだろう。ところが、以前よりこの方法に対しては問題が指摘されており、対応策として回顧的プレテストが提案されている。そこで本研究では、プレテストーポストテストに回顧的プレテストを加えたデザインを用いて、ある授業を通しての受講生の変化を把握することについて検討する。加えて、受講生が回顧的プレテストに答えることで、現在の自分と以

前の自分を比較することの影響についても検討することを目的とする。

準実験計画である1群プレテスト→ポストテストデザインは、教育現場のように、適切な対照群を設定しての実験計画が実施困難な場合に利用されることが多いといえよう（山田，2020）。このデザインでは、同一対象に、同一の測定を繰り返して実施することで、その差が実施時点間での回答者の変化を反映するとみなされている（Cronbach & Furby, 1970）。ところが、以前よりこの方法によるテスト間の得点差は、変化を正しくあらわさない場合もあることが指摘されている。

その代表的なものとして、Howard, Ralph et al. (1979) を指摘できよう。この論文は5つの研究から構成されるが、その第1研究で、独断性の減少を目的としたスキル研修ワークショップの効果を、自己報告式のプレテスト→ポストテストデザインを用いて検討した結果、予測と異なり、参加者の半数以上で独断性が高くなり、平均値も統計的に有意に高くなったことが明らかになった。そこで参加者にインタビューを行ったところ、参加者はワークショップを通して、参加前の自身の独断性に対する認識を変化させたことが見いだされた。ワークショップ前は、自分の独断性を過小評価しており、ワークショップを通してそれに気づいたという参加者が少なくなかったのである。ここから、プレテスト→ポストテスト間で回答者の判断基準が同一でないことによって、その差が変化を正しくあらわさないという問題点が明らかにされた。

介入によってプレテスト→ポストテスト間で判断の基準が変化すること（再校正）を、Howard, Ralph et al. (1979) はレスポンスシフト（response-shift）とよんだ。そして、かれらはレスポンスシフトの問題を回避する方法を模索し、Campbell & Stanley (1963)などを参考に、ポストテスト時に回顧的プレテスト（retrospective pretest）を用いることを提案した。回顧的プレテストとは、ポストテスト時にプログラム参加前を回顧して回答を求める方法である。通常のポストテストと回顧的プレテストを合わせて実施することで、判断の基準が統一されレスポンスシフトの影響が排除できるとされる。このような、回顧的プレテストを用いた評価研究、レスポンスシフトに注目した研究は、その後も継続的に行われてきており、Hill (2020) や Nimon (2014), Nimon & Allen (2007) などによってレビューされているように、知見が蓄積されてきている。

ところが、この回顧的プレテストについては現在でも賛否両論がある。それぞれの立場の大きな違いは、バイアスやエラーを、主にプレテストにあると考えるか、回顧的プレテストにあると考えるかにあると指摘される（Hill, 2020; Hill & Betz, 2005, Geldhof et al., 2018）。回顧的プレテストを推奨する立場は、レスポンスシフトに焦点をあて、レスポンスシフトが生じるならばプレテストが信頼できないものとなるとし、回顧的プレテストのメリットを強調する。回顧的プレテストに疑問を投げ掛ける立場は、記憶の正確さ、自分を有利に見せたいという欲求（社会的望ましさ）、プログラムに費やした努力の正当化（effort justification bias）、他者に有能であるように見せたいという欲求（印象管理）、暗黙理論ほか、回顧的な方法における各種バイアスの影響を問題視し、回顧的プレテストのデメリットを強調するのである。現状では、どちらが正しいかという判断は定まっておらず、それぞれのメリット、デメリットを考慮し、方法の選択や結果の解釈に慎重さをもってあたるのが生産的といえるだろう。

このように問題点も指摘されている回顧的プレテストであるが、レスポンスシフトへの対応以外にもそれを実施するメリットへの言及もある。たとえば Hill & Betz (2005) は、ポストテストで回顧的に評価をする機会は、以前と現在の比較することから対象者の効力感を強化する可能性を指摘する。また Hill (2020) は、プログラムの効果や個人の成長について、参加者自身がどのように感じているかを理解することを目的とする場合、従来のプレテストよりも、回顧的なテストの方がよ

り直接的な情報を与えてくれると示唆している。

以上のような議論がある回顧的プレテストを、本研究では心理学の調査法にかかる授業に適用し、この授業による受講生の変化について把握し、検討する。この授業は、いわゆる方法論に関する科目であり、それゆえ単に調査法に関する知識を記憶するだけでなく、研究を行うためのスキルを身につける側面もある。知識についてもスキルについても、それを客観的に測定することは可能であろうが、授業の中で、プレテスト、ポストテストを実施し、それぞれでさまざまな知識、スキルを客観的に把握することは現実的には難しいといわざるを得ない。特にスキルに関しては、客観的な測定には、多大な時間と労力が必要になる。

そのため、現実的な方法として自己報告式の調査が利用されるが、上述の議論を踏まえると、特にプレテスト時の回答が問題となろう。Howard, Dailey et al. (1979)などは、介入前には、対象者は測定対象となる構成概念に関する十分な情報を持っておらず、またその構成概念に関する自分のレベルが低いことから、プレテスト時点で有効な自己報告をすることが妨げられると指摘する。Lam & Bengo (2003)はレスポンスシフトの影響を受けやすい概念として、知識やスキル、態度をあげているが、介入で得た新しい情報に基づいて調整されるならば、それらについては参加者が事前の状態を回顧的に判断したものの方がより妥当なものになると考えられる (Norman, 2003)。

このような指摘は、調査法という授業と、そこで扱われる調査法に関するスキルや自信の自己報告にもあてはまるだろう。たとえば、「研究目的に沿った調査項目を準備する」ことを考えてみても、そのようなことをしたことがない段階で、「研究目的に沿った調査項目を準備する」ことができるかと問うても、判断材料に乏しい中での回答しか得られないかもしれない。もしプレテストでの自己報告が有効なものでなければ、プレテストーポストテストデザインでは変化を正しく把握することは困難となり、回顧的プレテストーポストテストデザインを採用すべきかもしれない。類似した先行研究としては、Drennan & Hyde (2008)が、看護学の修士課程に在籍する学生を対象として、論文の準備のための高度な量的・質的研究の手法に関する授業の効果の検討を行っている。この研究では、21の自己評価項目の内、14項目でプレテストよりも回顧的プレテストの得点が低いという統計的に有意なレスポンスシフトが認められている。

このように、本研究で対象とする調査法のような授業では、授業を受けることによってレスポンスシフトが生じ、プレテストーポストテストデザインでは受講生の変化を正しく把握できない可能性が考えられる。そこで本研究では、調査法を用いた研究を行うために必要な力（本研究ではこれを便宜的に研究力とよぶ）に関する自己報告に、レスポンスシフトが生じるのかどうかを検討することを第1の目的とする。

さらに、授業期間中に中間的なポストテストを行うことで、授業内容とその期間における測定された変化の対応関係という点から検討する。合理的に考えれば、ある期間の学習内容と、その内容の研究力の伸長は対応するはずである。プレテストーポストテストデザイン、回顧的プレテストーポストテストデザイン、またレスポンスシフトにみられる変化の様相が、学習内容との対応関係をどのように反映するのかを検討する。

最後に、授業開始時を回顧し、現在と比較することを、対象者がどのように感じているのかという点について検討する。Hill & Betz (2005)が指摘するように、ポストテスト時に回顧的に評価をする機会は、以前と現在の比較することになり、それが対象者の効力感を強化する可能性が考えられる。

これまでの研究でも、プレテストと回顧的プレテストの差など、回顧的プレテストに回答する

ことについて、インタビューや自由記述による調査を行っている研究がある（たとえば Cantrell, 2003; Manthei, 1997; Sprangers, 1989）。これらから回顧的プレテストに回答することや、そのために現在と過去を比較することによって生じる意識について把握することができる。たとえば Sprangers (1989) は、対象者がプレテストとポストテストの得点の違いに気づいているかという点について検討している。病院での実習を前にした医学生を対象とする、「アナムネーゼとコミュニケーションのスキル」に関するプログラムがとりあげられ、対象者は自己報告式のプレテスト、ポストテスト、回顧的プレテストへ回答し、あわせてインタビューを受けている。インタビューへの回答の中には、プレテストと回顧的プレテストの違いへの言及があり、「当時は何も判断材料がなかった。今と以前の違いは、自分がどうであるか (what I am) をより明確に把握できるようになったこと…自分に対する判断がより意識的になったこと」、「当初は、どちらともいいがたかった (toss-up) …プログラムのおかげで自分に何が求められているのか、より理解できるようになった」といった回答が紹介されている（訳は筆者による）。

こういった自分の変化に関する回答は、授業の評価に加え、Hill & Betz (2005) が指摘する回顧的に評価する機会の有用性についても有用な情報を与えてくれると考えられる。そこで第3の目的として、回顧的プレテストによって、授業開始時を回顧し、現在と比較することで対象者は何を感じるのかという点について検討する。

方法

対象とした授業の概要

2023年4月から7月にかけて開講された心理調査法にかかる授業をとりあげた。当該大学ではクォーター制が採用されており、この授業は第1クォーター（以下、Q1と略記する）、第2クォーター（以下、Q2と略記する）の連続で開講されている。ただしQ1は週2コマ（連続）、Q2は1コマである。

この授業は「質問紙調査法についての理解を深め、それを使った研究計画を立案できるようになる」、「適切な質問紙を作成できるようになる」、「統計処理パッケージRを用いてデータを適切に分析できるようになる」、「論文、研究要旨として適切に研究をまとめられるようになる」という4点を到達目標としてかかげている。3年生時以降で受講できる科目であり、心理測定法に関する科目の単位取得が履修の条件になっている。この心理測定法に関する科目を含むカリキュラム構成から、受講生は研究や調査法に関わる基本的な知識、たとえば実証的研究に関する基礎的知識、頻繁に利用される統計手法に関する知識、および統計処理パッケージRの基礎的な操作の経験を持っている者がほとんどであると推察される。

授業は講義と演習・実習を織り交ぜて展開された。4、5名からなるグループを構成し、調査法を用いた研究を立案し、論文としてまとめるまでの過程を経験する。概ねQ1では計画立案から調査用紙の作成まで、Q2はデータ収集から分析、論文作成までの一連の作業を行った。Q1の最終課題には、自分たちが行おうとしている研究について、「問題」、「目的」、「方法」までを論文化し、分析計画にも触れたものを提出するように求められた。Q2では、完成した論文と要旨の提出が求められた。なおQ2は、授業内の講義において論文や要旨の作成における留意点などについて触れたが、作業の進捗状況という点では、データ分析の完了程度までであり、論文の完成以前にQ2の最終回授業を迎えている。

こういった授業内容に加え、論文を読む課題（Q1は多読課題として6編、Q2は精読課題として英語論文1編）も課された。

調査内容と調査計画

調査は以下の3回の時点で行った。なお、いずれの回でも、他の調査内容とともに実施された。

・1回目（以下T1）：初回授業時にweb調査の形式で実施した。

研究力に関する自己査定 授業の到達目標や授業内容、さらにDrennan & Hyde (2008)で用いられている研究能力に関連する調査項目などを参考に、「調査研究の一連の手順・流れをイメージすること」、「研究に必要な論文を探すこと」など10項目を準備した。また上述のような各クォーターの授業内容との整合性から、Q1での学習内容と関わりの深い内容6項目をI群、Q2での学習内容と関わりの深い内容2項目をII群、両クォーターでの学習内容と関わりの深い内容2項目をIII群と分類した（Table 1参照）。なお、いずれもこの授業で学習する内容であるが、当該学科のカリキュラムや受講生が3年生以上であることから、回答の手がかりさえないような項目は含まれていないと考えられる。回答は、「まったくできない」の0から、「うまくできる」の5までの間で評定を求めた。選択肢は両極のみに示し、その間は等間隔に並ぶ数値とした。以下、これをプレテストとよぶ。

・2回目（以下T2）：Q1最終回の授業時にweb調査の形式で実施した。

研究力に関する自己査定 T1の調査で用いた10項目について、今現在、あなたがどう思っているかと、授業のスタート時の自分を思い返し、その時に同じことを聞かれていたらと当時を回顧して回答するように求めた。また「なお、この授業のスタート時に同様の調査を行いました、その時の回答を思い出して回答することを求めるものではありません。今、当時を振り返ってどうか、という観点から回答してください」という説明を加えた。回答は、「まったくできない」の0から、「うまくできる」の5までの間で評定を求めた。選択肢は両極のみに示し、その間は等間隔に並ぶ数値とした。

調査項目のレイアウトは、Howard, Ralph et al. (1979)に準じて、項目1の提示、現在の査定（ポストテスト）、スタート時の査定（回顧的プレテスト）、項目2の提示、現在の査定、スタート時の査定、と続くレイアウトとし、項目ごとに現在の査定、スタート時の査定のそれぞれが一目で把握できる形式を採用した。この形式は、バイアスにさらされやすいとも指摘されるが（Hill, 2020; Hoogstraten, 1982）、先述した本研究の第3の目的に照らして、回答者の意識が自身の変化に向きやすい形式と判断して採用した。なお、Howard, Ralph et al. (1979)では、両テストの回答が同一視点であることを確認するため、回顧的な回答はポストテストの回答と関連させて行うように指示しているが、今回の調査では、この指示は行っていない。

・3回目（以下T3）：Q2最終回の授業時にweb調査の形式で実施した。

研究力に関する自己査定 T1の調査で用いた10項目について、今現在、どう思っているかと、授業のスタート時（Q2の初めではなく、Q1の初め）の自分を思い返し、その時に同じことを聞かれていたらと考えて回答するように求めた。また、T1やT2時点の回答を思い出して回答するのではなく、今、当時を振り返ってどうか、という観点から回答してほしい旨も説明した。回答方法等はT2調査時と同様とした。

授業開始時を回顧し比較することによる感想 「今回の授業で協力してもらった調査では、授業スタート時と現在を比較するような質問を行いました。こういう学習の振り返り方について何か感じることを、考えることがあれば自由に書いてください」という教示を用い、自由記述で回答を求めた。

倫理的配慮

毎回の調査時には、個人を特定して分析をするものでも、成績に反映されることもないことを説明した。研究のためのデータとすることの可否についての意思表示を求め、同意が得られた対象者のみのデータを分析に用いている。また3回の調査間での回答者の同定は、回答者のみを知るIDの記入を求めることによって行った。

結果

3回の調査すべてで研究力の項目に回答があり、回答者の同定ができ、また研究への同意が得られた対象者は30名（登録者数の約70%）であった。分析には、R（4.2.2）およびpsych（2.2.9）パッケージ、anovakun（4.8.7）を用いた。

3度、5つの測定それぞれで、「まったくできない」を0、「うまくできる」を5とする0から5を付与し、項目それぞれの分布の様相を確認し、基礎統計量を算出した。平均値および標準偏差をTable 1に示す。

Table 1
基礎統計量と分散分析結果

群	T1		T2		T3		分散分析結果		
	プレ	ポスト	回顧プレ	ポスト	回顧プレ	F値と自由度	p	η^2	
	平均 (SD)	平均 (SD)	平均 (SD)	平均 (SD)	平均 (SD)				
研究に必要な論文を探すこと	I	2.60 (0.97)	3.47 (0.97)	2.13 (0.86)	3.47 (0.73)	2.13 (0.82)	30.2 (4, 116)	**	0.51
論文を正確に読解すること	I	2.30 (1.12)	3.37 (0.72)	1.70 (0.92)	3.47 (0.73)	2.10 (0.92)	39.4 (3, 90)	**	0.58
論文を批判的に評価すること	I	1.77 (0.90)	2.73 (1.14)	1.60 (1.00)	3.17 (0.91)	1.87 (1.04)	23.1 (4, 116)	**	0.44
研究の対象となる事象・ 現象を見つけること	I	1.93 (1.20)	3.10 (1.16)	1.60 (1.10)	3.40 (1.04)	1.77 (1.17)	28.7 (4, 116)	**	0.50
適切な仮説を導くこと	I	1.97 (0.89)	2.77 (0.94)	1.83 (0.99)	3.10 (0.80)	1.77 (1.04)	29.4 (3, 90)	**	0.50
研究目的に合った調査用具(質 問紙など)を作成すること	I	2.03 (1.07)	3.23 (0.97)	1.83 (1.18)	3.47 (1.01)	1.77 (0.94)	25.9 (3, 92)	**	0.47
収集したデータを統計的 に分析すること	II	1.70 (1.06)	2.23 (1.22)	1.47 (0.97)	3.33 (1.03)	1.23 (1.07)	39.1 (4, 116)	**	0.57
Rを操作すること	II	1.43 (1.14)	1.53 (1.04)	1.23 (0.86)	3.33 (1.21)	1.10 (0.92)	48.6 (3, 75)	**	0.63
研究を論文にまとめること	III	2.20 (1.19)	2.63 (0.96)	1.80 (0.96)	3.17 (0.83)	1.90 (1.03)	23.2 (3, 83)	**	0.44
調査研究の一連の手順・ 流れをイメージすること	III	2.30 (1.06)	3.40 (0.86)	1.37 (0.89)	3.60 (0.77)	1.67 (0.96)	74.0 (3, 91)	**	0.72

** $p < .01$

統計的検定結果

5つの測定値間での平均値の差異について検定を行った。1要因分散分析の結果、すべての項目で有意な結果となった。なお、Mendozaの多標本球面性検定を行い、有意である場合は、Greenhouse-Geisserのεによる自由度の調整を行った。多重比較はShafferの方法(5%水準)で行った。その結果をTable 1およびTable 2に示す。

まず、T1からT2の間では、「Rを操作すること」以外のすべての項目で有意な得点の上昇が認められた。これは、プレテスト→ポストテストでも、回顧的プレテスト→ポストテストでも同様であった。

次に、T1からT3の間では、全ての項目で有意な得点の上昇が認められた。なお、プレテスト→ポストテストでも、回顧的プレテスト→ポストテストでも同様の結果であった。

さらにT2からT3の間について、T2のポストテストとT3のポストテスト得点の差を検討した。その結果、「適切な仮説を導くこと」、「収集したデータを統計的に分析すること」、「Rを操作すること」、「研究を論文にまとめること」で有意な得点の上昇が認められた。先のT1からT2の間の結果を考慮すると、Ⅱ群の「Rを操作すること」はQ2で、その他の「適切な仮説を導くこと」(Ⅰ群)、「収集したデータを統計的に分析すること」(Ⅱ群)、「研究を論文にまとめること」(Ⅲ群)は、Q1、Q2全体を通して伸長するといえよう。

次に、レスポンスシフトについて確認した。T1のプレテスト得点、T2の回顧的プレテスト得点、

Table 2
多重比較の結果

	群	変化 プレーポスト		変化 回顧プレーポスト		T2-T3 のポスト ト変化	レスポンスシフト		回顧プレ の安定性	その他	
		プレ	プレ	T2 回 顧プレ	T3 回 顧プレ	T2 ポスト	プレ	プレ	T2 回 顧プレ	T2 ポスト	T2 回 顧プレ
		T2 ポスト	T3 ポスト	T2 ポスト	T3 ポスト	T3 ポスト	T2 回 顧プレ	T3 回 顧プレ	T3 回 顧プレ	T3 回 顧プレ	T3 ポスト
研究に必要な論文を探すこと	I	*	*	*	*	n.s.	n.s.	n.s.	n.s.	*	*
論文を正確に読解すること	I	*	*	*	*	n.s.	*	n.s.	*	*	*
論文を批判的に評価すること	I	*	*	*	*	n.s.	n.s.	n.s.	n.s.	*	*
研究の対象となる事象・ 現象を見つけること	I	*	*	*	*	n.s.	n.s.	n.s.	n.s.	*	*
適切な仮説を導くこと	I	*	*	*	*	*	n.s.	n.s.	n.s.	*	*
研究目的に合った調査用具(質 問紙など)を作成すること	I	*	*	*	*	n.s.	n.s.	n.s.	n.s.	*	*
収集したデータを統計 的に分析すること	Ⅱ	*	*	*	*	*	n.s.	n.s.	n.s.	*	*
Rを操作すること	Ⅱ	n.s.	*	n.s.	*	*	n.s.	n.s.	n.s.	n.s.	*
研究を論文にまとめること	Ⅲ	*	*	*	*	*	n.s.	n.s.	n.s.	*	*
調査研究の一連の手順・ 流れをイメージすること	Ⅲ	*	*	*	*	n.s.	*	*	n.s.	*	*

*は多重比較の結果が有意であることを示す

T3の回顧的プレテスト得点の平均値の比較から、「調査研究の一連の手順・流れをイメージすること」は、T1のプレテスト得点に比べT2、T3の回顧的プレテスト得点が有意に低いことが明らかになった。また「論文を正確に読解すること」は、T2の回顧的プレテスト得点が、T1とT3に比べて有意に低かった。これら2項目以外では、有意な差は認められなかった。

効果量

5つの測定値間の変化の大きさを検討するために、それぞれの差得点から効果量として標本標準化平均値差 d_b を算出した（橋本・荘島，2016；大久保・岡田，2012など参照）。差得点の平均値および標準偏差はTable 3に、効果量をまとめたものをTable 4に示す。なお、Table 2においてその他とした、T2のポストテストとT3の回顧的プレテスト、T2の回顧的プレテストとT3のポストテストについては除外している。

T1からT2の間では、プレテスト→ポストテストでみると、平均はすべて上昇しており、その効果量は「Rを操作すること」0.08から、「論文を正確に読解すること」の1.05の間にあった。これら以外の項目では「研究を論文にまとめること」、「収集したデータを統計的に分析すること」が0.50

Table 3
差得点の平均と標準偏差

群	変化 プレーポスト		変化 回顧プレーポスト		T2-T3 のポスト 変化	レスポンスシフト		回顧プレ の安定性	
	プレ	プレ	T2 回 顧プレ	T3 回 顧プレ	T2 ポスト	プレ	プレ	T2 回 顧プレ	
	T2 ポスト	T3 ポスト	T2 ポスト	T3 ポスト	T3 ポスト	T2 回 顧プレ	T3 回 顧プレ	T3 回 顧プレ	
研究に必要な論文を探すこと	I	0.87 (0.94)	0.87 (1.01)	1.33 (0.88)	1.33 (0.76)	0.00 (0.91)	-0.47 (1.04)	-0.47 (1.07)	0.00 (0.83)
論文を正確に読解すること	I	1.07 (1.01)	1.17 (1.09)	1.67 (0.96)	1.37 (0.81)	0.10 (0.71)	-0.60 (1.19)	-0.20 (1.24)	0.40 (0.81)
論文を批判的に評価すること	I	0.97 (1.03)	1.40 (1.00)	1.13 (1.31)	1.30 (0.95)	0.43 (0.94)	-0.17 (1.32)	0.10 (1.16)	0.27 (0.94)
研究の対象となる事象・ 現象を見つけること	I	1.17 (1.21)	1.47 (1.33)	1.50 (1.25)	1.63 (1.16)	0.30 (1.06)	-0.33 (0.99)	-0.17 (1.39)	0.17 (1.15)
適切な仮説を導くこと	I	0.80 (0.89)	1.13 (0.78)	0.93 (0.94)	1.33 (0.84)	0.33 (0.61)	-0.13 (0.94)	-0.20 (1.00)	-0.07 (0.69)
研究目的に合った調査用具（質 問紙など）を作成すること	I	1.20 (1.30)	1.43 (1.41)	1.40 (1.35)	1.70 (1.26)	0.23 (1.01)	-0.20 (1.21)	-0.27 (1.01)	-0.07 (1.05)
収集したデータを統計的 に分析すること	II	0.53 (1.01)	1.63 (1.22)	0.77 (0.97)	2.10 (0.96)	1.10 (0.92)	-0.23 (1.01)	-0.47 (1.14)	-0.23 (0.94)
Rを操作すること	II	0.10 (1.27)	1.90 (1.37)	0.30 (0.95)	2.23 (0.97)	1.80 (1.03)	-0.20 (0.66)	-0.33 (0.84)	-0.13 (0.68)
研究を論文にまとめること	III	0.43 (0.86)	0.97 (1.16)	0.83 (0.65)	1.27 (0.78)	0.53 (0.73)	-0.40 (1.07)	-0.30 (1.15)	0.10 (0.76)
調査研究の一連の手順・ 流れをイメージすること	III	1.10 (1.16)	1.30 (1.02)	2.03 (0.76)	1.93 (0.78)	0.20 (0.71)	-0.93 (0.94)	-0.63 (0.96)	0.30 (0.79)

Table 4
効果量 d_p

	群	変化 プレーポスト		変化 回顧プレーポスト		T2-T3 のポスト 変化	レスポンスシフト		回顧プレ の安定性
		プレ	プレ	T2 回 顧プレ	T3 回 顧プレ	T2 ポスト	プレ	プレ	T2 回 顧プレ
		T2 ポスト	T3 ポスト	T2 ポスト	T3 ポスト	T3 ポスト	T2 回 顧プレ	T3 回 顧プレ	T3 回 顧プレ
研究に必要な論文を探すこと	I	0.92	0.86	1.51	1.76	0.00	-0.45	-0.43	0.00
論文を正確に読解すること	I	1.05	1.07	1.74	1.69	0.14	-0.50	-0.16	0.49
論文を批判的に評価すること	I	0.94	1.40	0.87	1.37	0.46	-0.13	0.09	0.28
研究の対象となる事象・ 現象を見つけること	I	0.97	1.10	1.20	1.41	0.28	-0.34	-0.12	0.15
適切な仮説を導くこと	I	0.90	1.46	0.99	1.58	0.55	-0.14	-0.20	-0.10
研究目的に合った調査用具(質 問紙など)を作成すること	I	0.93	1.02	1.03	1.35	0.23	-0.16	-0.26	-0.06
収集したデータを統計的 に分析すること	II	0.53	1.34	0.79	2.19	1.19	-0.23	-0.41	-0.25
Rを操作すること	II	0.08	1.38	0.32	2.30	1.75	-0.30	-0.39	-0.20
研究を論文にまとめること	III	0.50	0.83	1.29	1.61	0.73	-0.37	-0.26	0.13
調査研究の一連の手順・ 流れをイメージすること	III	0.95	1.27	2.66	2.46	0.28	-0.99	-0.66	0.38

程度であり、他の項目は0.90を超えていた。項目分類の観点からすれば、II群の2項目、III群の1項目が相対的に効果量が小さく、I群のすべて、およびIII群の1項目が相対的に効果量大きいことが示されたといえよう。

他方、回顧的プレテスト→ポストテストでみると、平均はすべて上昇しており、その効果量は「Rを操作すること」0.32から、「論文を正確に読解すること」の2.66の間にあった。「Rを操作すること」が、他の項目に比べて大きく離れて低い位置にある。その他の項目は、「収集したデータを統計的に分析すること」の0.79から、「論文を正確に読解すること」の1.74までの間に位置していた。すなわち、II群の2項目が、他に比べて低い位置にあることが認められる。

プレテスト→ポストテストと回顧的プレテスト→ポストテストの効果量を比較すると、「論文を批判的に評価すること」でプレテスト→ポストテストの方が回顧的プレテスト→ポストテストよりわずかに大きかった(0.07の差)が、その他の項目では回顧的プレテスト→ポストテストの効果量の方が大きい。また効果量の差(絶対値)は0.07から1.71であったが、その大小から項目を2群に大別できそうである。大きい群にはI群の2項目(「研究に必要な論文を探すこと」, 「論文を正確に読解すること」)とIII群の2項目が該当する。

続いて、T1からT3の間では、プレテスト→ポストテストでみると、平均はすべて上昇しており、その効果量は「研究を論文にまとめること」0.83から、「適切な仮説を導くこと」の1.46の間にあった。T1からT2の間に比べ効果量の項目間差異は小さく、すべての項目で大きな差が認められるといえよう。項目分類の観点からすれば、T1からT2の間では効果量が小さかったII群の2項目で1.3程度の値が認められ相対的位置が上位に変化している点が特徴的といえるだろう。

回顧的プレテスト→ポストテストでみると、平均はすべて上昇しており、効果量は「研究目的に合った調査用具（質問紙など）を作成すること」1.35から、「調査研究の一連の手順・流れをイメージすること」の2.46の間にあった。プレテスト→ポストテストでもみられたように、T1からT2の間に比べて効果量の項目間差異は小さくなっているといえよう。またT1からT2の間では他の項目よりも大きく離れて低い位置にあった「Rを操作すること」や、「収集したデータを統計的に分析すること」のⅡ群の項目は、Ⅲ群の「調査研究の一連の手順・流れをイメージすること」とともに2.00を超える最上位に位置していた。

プレテスト→ポストテストと回顧的プレテスト→ポストテストの効果量を比較すると、T1からT2の間と同様に、「論文を批判的に評価すること」でプレテスト→ポストテストの方が回顧的プレテスト→ポストテストよりわずかに大きかった（-0.03）が、その他の項目では回顧的プレテスト→ポストテストの効果量の方が大きかった。その大小から、T1からT2の間のように項目を大別することは困難であるが、Ⅰ群の項目の多くは効果量の差が小さい傾向、Ⅱ群やⅢ群の方が大きい傾向がうかがえる。

T2からT3の間では、「研究に必要な論文を探すこと」の0.00から、「Rを操作すること」の1.75までの効果量が認められた。その上位から「Rを操作すること」、「収集したデータを統計的に分析すること」というⅡ群の2項目、続いてⅢ群の「研究を論文にまとめること」と続いており、T2からT3の間で学習、経験する内容が上位に位置していた。

レスポンスシフトの効果量は、小さい効果量とされる0.2（絶対値）を判断基準とすると、T1とT2、T1とT3のいずれでも0.2に届かなかった項目は「論文を批判的に評価すること」のみであった。しかし、基準を中程度とされる0.5とすれば、いずれの場合もそれに届かなかった項目は8項目となる。中程度を超える効果量が認められた項目は、統計的検定で有意と判断されたものと同一である。すなわち、全体的には、小から中程度の効果量のレスポンスシフトが生じており、プレテスト時の自己報告がやや高いといえよう。

授業開始時を回顧し比較することによる感想

この設問に対しては無回答者も多く、得られたデータは18件（60%）であった。記入された回答すべてをTable 5に示す。なお、記述内容の類似性によって並び替え、便宜的に番号を割り振った。

1から8および10、11の10件の内容には、すべて「成長」という言葉が含まれている。回答数の半数以上であり、全対象者の1/3でもあることから、「成長」は回顧し比較することによって意識化される代表的な概念と考えられるだろう。また、3「どの部分においてどの程度成長」、4「どれだけ成長」などといった内容があったり、1「小さな成長」、5「少しでも成長」などといった表現があることから、漠然とした成長感ではなく、その領域や程度などが特定できるような具体性のある「成長」が実感できているのではないだろうか。また「うれしい／嬉しい」という言葉が5、6、8の回答で用いられており、成長の実感と共に肯定的感情が生起していることもうかがえる。さらに、6「意外に成長」、7「自分では分からなくても思い返せば成長」など、授業開始時を回顧すること、回顧的プレテストに取り組むことが気づきにつながることを認識できたことを示す内容もみられた。

9から12の4件は、「できない」「変わっていない」というような内容が含まれる回答である。それぞれの回答の意味内容はやや異なっていると推察されるが、11、12などから、回顧し比較することは、成長した部分にのみ意識を向かわせるわけではなく、成長していない部分にも焦点が当てられることを示唆しよう。

Table 5
自由記述内容

1	ふりかえることで小さな成長も気づくことが出来る。
2	以前よりも成長をしているように感じる。
3	自分がどの部分においてどの程度成長したかが可視化されて良かった。
4	自分がどれだけ成長したか知るいい機会だと思う。
5	少しでも成長したと感じられてうれしい。
6	意外に成長していることが分かって自分もうれしい
7	自分では分からなくても思い返せば成長したと思える点があることに気がついた
8	自分のできなかったことが少しできるようになっていて嬉しく、頑張ったんだと思える。
9	前は全然できてなかったなど感じる
10	自分がどのくらい成長したのか変わっていないのかかがはっきりわかるようになりました。
11	自分の成長した部分、または今後の課題が明確になったと思う。
12	勝手にわかったからこそ、できていないと感じることが多かった。
13	前回の自分の回答を完全に覚えているわけではないため、過去の記憶や感情が現在の自分のものに引張られて回答が変動しているのではないかと感じました。
14	授業スタート時の自分の回答をあまり覚えておらず、現在との比較が難しく感じるので、授業スタート時の回答を読み返したいと思った。
15	授業スタート時の自分の力が正確に思い出せなかった。
16	以前よりもできるようになっていると思いたい気持ちが生じながらの回答であった。比較することでこれから先も項目にあったことを意識しながら研究を進めていきたいという考えをもった。
17	授業の始まりと終わりだけだと、途中で意識するのを忘れてしまうので中間でもう一回ぐらい振り返りがあっていいと思った。
18	授業の終わりをイメージすることで、自分の行動を振り返る習慣ができた。

13 から 15 の 3 件は、回顧した内容の正確さ、記憶の正確さに言及したものである。T2、T3 のいずれの調査時でも、T1 での回答を思い出して回答することを求めるものではなく、今、当時を振り返ってどうか、という観点から回答するよう教示文で説明している。これは自己査定時の判断の基準を、現在の時点で統一するための教示であるが、これがあっても、回顧することが求められると、当時の回答を正確に思い出すことに意識が向く対象者が含まれることを示す結果であろう。

16 は、13 などとともに、回顧的な方法におけるバイアスの存在を自覚しているような内容である。17 は、おそらく、回顧、比較することを肯定的にとらえている内容と推測できようが、研究力として設定した 10 項目を目標として意識していること、それを確認する機会をもっと増やすことを期待した内容ともいえよう。こういう傾向は、18 や 16 の回答にも共通していると推測でき、回顧し比較する機会は、目標の再確認の機会ともなるのではないかと考えられる。

考察

本研究は、心理学の調査法にかかる授業を対象として、プレテストーポストテストに回顧的プレテストを組み入れた調査を行い、以下の 3 点について検討することを目的とした。第 1 に、調査法を用いた研究を行うための研究力に関する自己報告に、レスポンスシフトが生じるのかどうかを検

討すること、第2に、ポストテストを2回行うことで、授業内容とその期間における測定された変化の対応関係という点から調査デザインやレスポンスシフトを検討することである。そして第3に、回顧的プレテストによって、授業開始時を回顧し、現在と比較することを対象者がどのように感じるのかという点について検討することである。

まず、調査法を用いる研究力に関する自己報告に、レスポンスシフトが生じるのかという点に関しては、統計的に有意な差が確認されたのは、「調査研究の一連の手順・流れをイメージすること」と「論文を正確に読解すること」の2項目だけであった。有意差が認められなかった項目については、効果量の観点から、小さい効果量とされる0.2よりは大きいものの、中程度とされる0.5には届かない項目が7項目、0.2より小さい効果量しか認められない、ほとんどレスポンスシフトが生じないと考えられる項目は「論文を批判的に評価すること」の1項目であった。そして、いずれの項目でも、最も高い平均値はプレテスト時の自己報告であった。

以上のことから、研究力の自己報告においては、消極的な表現ではあるが、レスポンスシフトは生じないとはいえないと結論できるだろう。類似した先行研究である Drennan & Hyde (2008) では、レスポンスシフトの効果量の多くは小から中程度であり、プレテスト時には過大評価気味であることが報告されている。この研究とは対象者数に違いがあり (Drennan & Hyde は $n=96$)、それが統計的検定の結果に表れている可能性があるが、プレテスト時は過大評価気味であることや、またレスポンスシフトの大きさは、本研究結果と類似しているといえよう。また授業開始時の研究力に関する自己査定は過大評価となる傾向は、他領域の教育プログラムを対象とした先行研究と同様である (e.g., Cantrell, 2003, Drennan & Hyde, 2008; Hoogstraten, 1982; Sprangers, 1988)。

先に述べたように、今回の受講生は、調査法に関わる基礎的な知識は得ていると考えられるものの、実際に調査法を用いた研究を行うことははじめてという場合が多いと推察される。それゆえ受講を通してレスポンスシフトが生じやすいと考えたが、関連する先行研究と類似した結果が得られたとはいえ、認められたレスポンスシフトは、それほど大きなものではなかった。ここから、授業開始時において受講生はそれほど自分の力を過大評価しているわけではなく、プレテストの回答は、かなり正確な見積もりといえるかもしれない。とはいえ本研究結果は、今回のようなレスポンスシフトが生じるかもしれない授業を対象にした研究では、プレテストーポストテストデザインを用いて変化を把握することにはリスクがあることを示唆する結果といえるだろう。

次に、3時点でのテストを実施したことから、授業内容とその期間における測定された変化の対応関係という点から調査デザインやレスポンスシフトを検討する。既述のように、研究力として用意した10項目に関して、各クォーターの授業内容との整合性から、Q1での学習内容と関わりの深い内容6項目をⅠ群、Q2での学習内容と関わりの深い2項目をⅡ群、両クォーターでの学習内容と関わりの深い2項目をⅢ群とした。そのため、Ⅰ群はT1とT2の間で変化が大きく、Ⅱ群はT1とT3の間で、Ⅲ群はその両方でプレ／回顧的プレテストーポストテスト間の変化、またレスポンスシフトが大きくなると予測するのが合理的であろう。

検定結果から、「Rを操作すること」というⅡ群の項目は、T1とT2の間では変化がなく、T1とT3の間では有意に平均が高くなるという予測に従った差が明らかになったが、プレテストーポストテストでも回顧的プレテストーポストテストでも同じ結果であった。他方、効果量を参照すると、3つの項目群のT1とT2、T1とT3の間での変化の効果量、レスポンスシフトの効果量の大小を踏まえると、回顧的プレテストーポストテストの方が合理的な予測と合致した変化を示す傾向にあるといえるだろう。

先の検討で、研究力に関しては、レスポンスシフトは生じないとはいえないと指摘したこととも合わせ、今回の場合はプレテスト→ポストテストよりも、回顧的プレテスト→ポストテストの方が妥当な変化を把握できていると判断してよいのではないだろうか。そして、プレテスト→ポストテストは、特にQ2や両クォーターを通して学習する内容に関する研究力の変化を過小評価する傾向

Figure 1
プレテスト→ポストテストにおける効果量

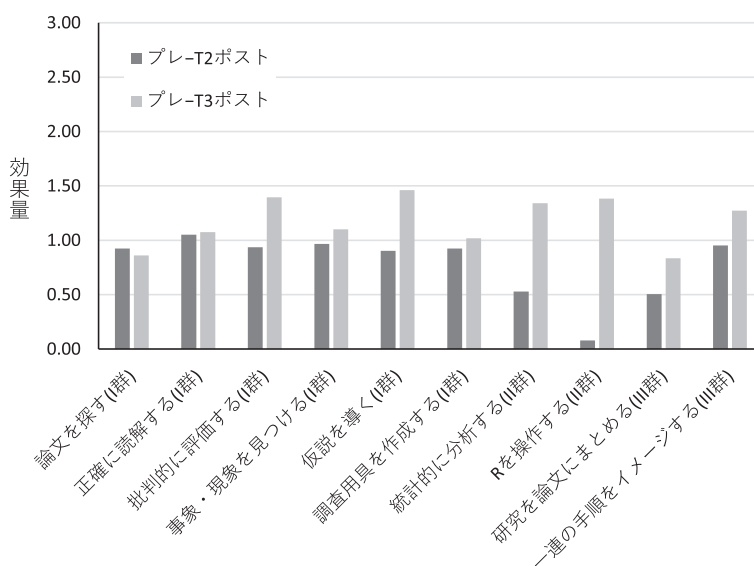
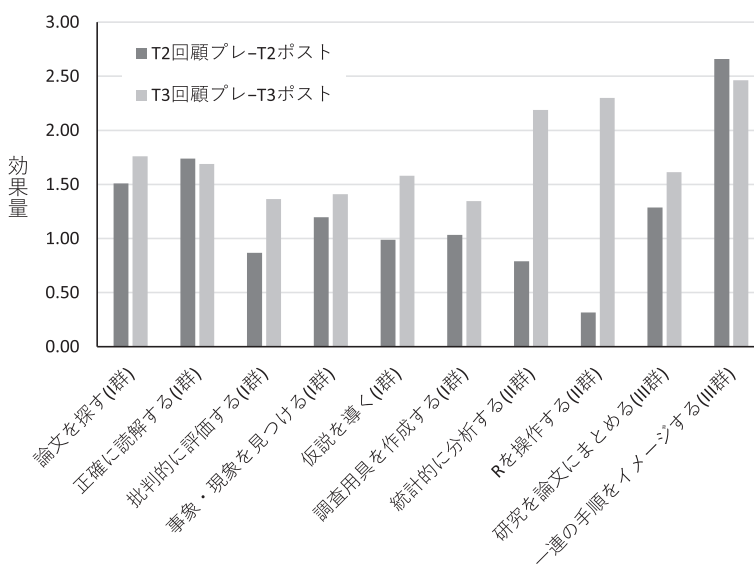


Figure 2
回顧的プレテスト→ポストテストにおける効果量



があるといえよう。

これらの結果をもとに、授業の効果について触れておく。Table 4をもとに、プレテストーポストテストによる効果量、回顧的プレテストーポストテストによる効果量をグラフ化したものがFigure 1およびFigure 2である。これらの図に示された様相の比較から、プレテストと回顧的プレテストを利用する場合には、授業の効果について異なる結論を導けそうである。たとえば、プレテストーポストテストでは、この授業全体(Q1とQ2)を通して、まんべんなく研究力が上昇しているといえようが、回顧的プレテストを参照する場合は、「調査研究の一連の手順・流れをイメージすること」や、Ⅱ群の2項目での上昇が顕著であること、「研究の対象となる事象・現象を見つけること」、「論文を批判的に評価すること」、「研究目的に合った調査用具(質問紙など)を作成すること」などでの上昇が小さいと評価できよう。この上昇が小さい項目とも重複するが、Q1で主として扱われるⅠ群の項目の中に、当該期間、すなわちT2の回顧的プレテストーポストテストで認められる差が相対的に小さいものが認められ、こういった点は授業の改善ポイントとなろう。なおこのような特徴は、プレテストーポストテストの結果だけからは見出しにくいといえよう。

研究力の伸長における相対的に差の小さい内容は、研究対象となる現象を見つけ、それを説明する仮説を考え、その仮説について検討するための調査案を検討するという、研究指導を行う際に最も難しい部分と重なるといえるのではないだろうか。全体的にみれば、この授業の参加者は、授業を通して研究力に関してかなり大きな伸長を報告している。しかしながら、特に回顧的プレテストーポストテストを参照すると、改善を検討すべき点も残されているといえよう。

ただし、これらの結果だけをもって、授業の効果、影響を論ずるには慎重であるべきである。これを論ずるには一層明確に因果関係を特定できるデザインを利用すべきであるし、また今回は客観的なパフォーマンスを指標としていない。さらに、回顧的プレテストの実施にあたり、バイアスにさらされやすいと指摘される形式(Hill, 2020; Hoogstraten, 1982)を採用している。そのため、授業の効果、影響を論ずるには今後の更なる検討が必要であろう。

最後に、回顧的プレテストによって、授業開始時を回顧し、現在と比較することを対象者がどのように感じるのかという点については、自由記述を求めた。その内容から、意識化される代表的な概念として、「成長」を見出すことができた。またそれは漠然とした成長感ではなく、より具体性を伴った感覚であり、成長していない部分にも目が向くとも推察できた。さらに、回顧し比較する機会は、目標の再確認の機会ともなることが示唆される。こういった回答の特徴は、Hill & Betz (2005)が指摘する、回顧的に評価をする機会が対象者の効力感を強化する可能性を、部分的にではあるが支持する結果といえるだろう。

このような結果を踏まえると、本研究では、第2の目標との関連でT2、T3という2回の回顧的プレテスト、ポストテストを実施する計画を採用したが、授業の工夫という点で、授業開始時を回顧し比較する機会を、一連の授業の途中で取り入れることも有益と考えられる。Table 5にある17の回答内容から推察すれば、授業の目標が意識から外れそうになった時期に行うことがひとつの案として浮上しようが、動機づけ研究などの枠組みから回顧的プレテストを検討する有用性を示唆するものといえよう。

本研究では、以上のようにプレテストーポストテストデザインと、回顧的プレテストーポストテストデザインでは、回顧的プレテストーポストテストデザインの方が変化をより妥当に把握できると考えられた。さらに、回顧的プレテストーポストテストに取り組み、授業開始時を回顧し現在と比較することが、自分の成長の確認に寄与するであろうことも明らかになった。授業終了時のパ

パフォーマンスは重要であるが、授業を通しての変化に対する自己評価や、その過程をふりかえる機会を保証することもまた重要である（和栗，2010）。ここから、回顧的プレテストの有用性を指摘できるが、本研究はサンプル数も少なく、結果の一般化には今後の研究を待たざるを得ない。

しかしながら、我が国においては、介入の効果を検討する際には、プレテスト―ポストテストを用いる場合がほとんどであり（山田，2020）、回顧的プレテストが利用されることは極めて少ないという現状がある（浦上，2023）。そのため、介入によるレスポンスシフトの影響は考慮されないまま、介入の効果が議論されている場合もあるといえよう。確かに回顧的プレテストが抱える問題は小さくないが、Hill & Betz (2005) や Hill (2020) が指摘する、また本研究からも示唆されるメリットも一層考慮されるべきであろう。授業による変化を測定する研究はもちろん、授業を構成する一部として回顧的プレテストを採用するような実践においても、回顧的プレテストの利用をめぐった研究、議論が活性化することが期待される。

引用文献

- Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171-246). Chicago: Rand McNally.
- Cantrell, P. (2003). Traditional vs. retrospective pretests for measuring science teaching efficacy beliefs in preservice teachers. *School Science and Mathematics, 103*, 177-185.
- Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we? *Psychological bulletin, 74*, 68-80.
- Drennan, J., & Hyde, A. (2008). Controlling response shift bias: the use of the retrospective pre-test design in the evaluation of a master's programme. *Assessment & Evaluation in Higher Education, 33*, 699-709.
- Geldhof, G. J., Warner, D. A., Finders, J. K., Thogmartin, A. A., Clark, A., & Longway, K. A. (2018). Revisiting the utility of retrospective pre-post designs: the need for mixed-method pilot data. *Evaluation and Program Planning, 70*, 83-89.
- 橋本貴充・荘島宏二郎 (2016). 実験心理学のための統計学——*t*検定と分散分析—— 誠信書房
- Hill, L. G. (2020). Back to the future: Considerations in use and reporting of the retrospective pretest. *International Journal of Behavioral Development, 44*, 184-191.
- Hill, L. G., & Betz, D. L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation, 26*, 501-517.
- Hoogstraten, J. (1982). The retrospective pretest in an educational training context. *The Journal of Experimental Education, 50*, 200-204.
- Howard, G. S., Dailey, P. R., & Gulanick, N. A. (1979). The feasibility of informed pretests in attenuating response-shift bias. *Applied Psychological Measurement, 3*, 481-494.
- Howard, G. S., Ralph, K. M., Gulanick, N. A., Maxwell, S. E., Nance, D. W., & Gerber, S. K. (1979). Internal invalidity in pretest-posttest self-report evaluations and a re-evaluation of retrospective pretests. *Applied Psychological Measurement, 3*, 1-23.
- Lam, T. C., & Bengo, P. (2003). A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation, 24*, 65-80.
- Manthei, R. J. (1997). The response-shift bias in a counsellor education programme. *British Journal of Guidance and Counselling, 25*, 229-237.
- Nimon, K. (2014). Explaining differences between retrospective and traditional pretest self-assessments: competing theories and empirical evidence. *International Journal of Research & Method in Education, 37*, 256-269.
- Nimon, K., & Allen, J. (2007). A review of the retrospective pretest: Implications for performance improvement evaluation and research. *Workforce Education Forum, 44*, 36-55.

- Norman, G. (2003). Hi! How are you? Response shift, implicit theories and differing epistemologies. *Quality of Life Research, 12*, 239-249.
- 大久保街亜・岡田謙介 (2012). 伝えるための心理統計——効果量・信頼区間・検定力—— 勁草書房
- Sprangers, M. (1989). Response-shift bias in program evaluation. *Impact Assessment, 7*, 153-166.
- 浦上昌則 (2023). キャリア講演会への参加と意識変化——回顧的ブレテストを組み込んだ変化を把握するアプローチ—— アカデミア人文・自然科学編, 26, 23-36.
- 和栗百恵 (2010). 「ふりかえり」と学習 国立教育政策研究所紀要 139, 85-100.
- 山田洋平 (2020). 教育実践研究における研究法のあり方についての一考察——協同学習の実践を通じた検討—— 人間と文化, 3, 163-174.