

# 第15章

## 統計的な分析と分析結果の解釈

英語教育実践は、複数の要因が複雑に絡み合う交互作用の中でその成果が定まる。実践のための指針を有効に得るためには、英語教育学研究の成果が適切に蓄積されなければならない。しかし、コンピュータやソフトウェアの急速な発展に伴い、統計的な分析に関する誤用や誤解釈も散見される。そのため、差の分析を解釈するという観点からは、効果量、検定力、信頼区間、そして図示といった手法を主軸に、研究成果をより有効に伝え、より有効に蓄積する方法を示した。また、複数の変数を使った分析と解釈という観点からは、相関分析、多変量分散分析、重回帰分析、探索的因子分析、そして共分散構造分析といった手法をとりあげ、多くの観測変数や、場合によっては潜在変数をも扱う多変量解析において、留意しなければならない点について指摘した。

### キーワード

- 第1節 統計的仮説検定, 探索的因子分析, 重回帰分析, 適性処遇交互作用
- 第2節 測定された差の分析, 効果量, 信頼区間, 追試, データの可視化
- 第3節 複数の変数の分析, 妥当性, 信頼性, メタ分析

## 15 統計的な分析と分析結果の解釈

### 1 統計的な分析と分析結果の解釈の現状と展望

#### 1.1 統計的な分析と分析結果の解釈の現状と展望

1990年代半ば以降のインターネットの発達と計算機の普及は、英語教育学研究における統計的な分析を後押しした。それまでは、たとえば等分散を仮定した独立2平均の差の検定（いわゆる  $t$  検定）を行う際には電卓を用い、各群の平均と標準偏差などを計算し、公式にあてはめて  $t$  値を計算し、統計の専門書の巻末の数表でサンプルサイズから求めた自由度と特定の有意水準の場合の臨界値を見つけて  $t$  値と比べ、統計的に有意か有意ではないかの判断をしていた（つまり、正確な有意確率を求めるという考え自体が希薄だった）。それが、多くのパーソナル・コンピュータに出荷時からインストールされているような汎用表計算ソフトの機能で有意確率までもが求められたり、データを入力すれば同様の計算を行える無償のウェブサイトも存在したりする時代となった。また、ソフトウェアについては無償のものも含め、それまでのコマンドを入力して操作するタイプのもののほか、マウスで操作するのが前提となったものも普及したため、統計的な分析を始めるための障壁が低くなったといえよう。

統計的な分析が増える中で、いわゆる誤用や誤解釈が見受けられることも多くなった。研究において、先行研究をふまえることは当然のふるまいであるものの、統計的な分析の行いかたが誤っていたり、分析結果の解釈が誤っていたりすると、その再生産をしてしまうという危険が伴う。統計的な分析の専門的な訓練を受けたわけではない場合には、先行研究の論文に示されている手法名や図表がすべてであり、掲載論文に示されたものだけを行えばよいと、実際には論文化に至るまでのさまざまな検討作業（たとえば次節以降で述べる作図等）があったということを見過ごしてしまいがちである。

統計的な分析の方法や解釈にまつわる問題については、拠って立つ立場の違いなどによってどちらとも言えないものもあるが、適切・不適切が比較的はつきりしたものも多い。たとえば前田（2000）は、探索的因子分析に焦点を当て、1995年から1999年に発行された国内の英語教育学関連の学会誌である、*ARELE (Annual Review of English Language Education in Japan)* の第6巻から第10巻、*JACET Bulletin* の第26巻から第30巻、*JALT Journal* の第17巻から

第21巻、*JLTA Journal*の第1巻から第2巻、*Language Laboratory*の第32巻から第36巻に掲載された212本の論文のうち、15回の探索的因子分析について検討を行った。その結果、満足できるものはなかったと報告している。また、前田（2004a）は、重回帰分析に焦点を当て、2000年から2002年に発行された前田（2000）が対象にしたものと同じ学会誌の172本の論文のうち、11件の重回帰分析を取り上げた。その結果、相関係数行列を示さずに標準偏回帰係数の相対的な強弱を解釈しているという点においては8件に問題があったことなどを報告している。

そのような留意すべき点については、前述した誤りの再生産の問題が付きまとう。英語教育学の分野においても、竹内・水本（2012）や平井（2012）のように、分野特有の事例を伴うため学習しやすい書籍が出版されている。統計的な分析や解釈に対する研究者の理解も全般的には向上しているとも言える。しかし、すでに刊行された論文は先行研究として残り続ける。

英語教育実践の成果は、さまざまな要因に左右される。教室には、教師と、教材と、ひとりひとり別々の学習者が居る。それぞれの学習者にとってぴったりの指導は理想ではあるものの不可能であることや、教師や教材が変わると別の成果が得られることは、簡単に想像できる。このことについて並木（1997）は「すべての人にとって最も効果的な教え方は果たしてあり得るのだろうかという問い」が適性処遇交互作用（ATI: Aptitude Treatment Interaction）の研究の出発点だとしている。

個々の学習者の適性（個人差）と、指導（教師や教材）とが交互に作用しあった結果が、教育実践の成果である。学習者の全員が積極的な成果（たとえばテスト得点の向上）を収めても、その成果の幅に差があった場合には、幅が小さい学習者は大きい学習者に比べると、他の手立てがあったのではないかと考えられる。時間以外すべての要因を固定したままとしても、ある時点で積極的な成果（たとえば言語活動に活発に参加してくれた）があっても、別の時点では消極的な成果となる可能性もある。時がうつろい環境が変われば好みや感じかたに変化が出るという事実を織り込み、それらも含んだ多くの要因の交互作用の中で、あらゆる営みは行われている。

英語教育学の研究成果をよりよいかたちで蓄積することは、研究者の責務である。研究で明らかにした（教育実践から見るとほんの一部の）側面を、教育実践の指針として教室にフィードバックするためには、ただ良い研究をすれば

よいというだけではない。他者がそれを読めば同じことができるように詳しく、誤用や誤解釈がないように十分な情報を示しつつ、論文化することが必要である。そのため、本章では、本節でこれまで述べたように1990年代半ばからの時勢を振り返った後、次節以降において、差の分析と解釈や、多くの変数を用いた分析について、資料となる文献を引用しつつ、現状や、これから分析を行ったり先行研究を読むときに留意すべき点を指摘する。 [前田啓朗]

## 2 測定された差の分析と解釈

### 2.1 ARELE掲載論文に見る測定された差の分析

応用言語学系の国際ジャーナルでは、 $t$  検定や分散分析 (ANCOVA や MANOVA を含む) のような平均値に基づいた分析 (観測された差の分析) が大半であるということが報告されている (Gass, 2009; Lazaraton, 2005)。Plonsky (2013) は、1990年から2010年までの20年間で *Language Learning* と *Studies in Second Language Acquisition* に掲載された論文のうち、統計的手法を用いた論文606本における分析手法の傾向と報告内容 (reporting practices) を調査した。その結果、半数以上 (56%) の論文で分散分析が使われていたという事実からも、グループ間の平均値差に焦点を当てた論文がいかに多いかがわかる。

Mizumoto, Urano, and Maeda (2014) は、ARELE の第1号から第24号 (1990-2013) に掲載され、CiNii で公開されている論文450本のうち、76% (341本) が量的研究であることを示した。国際ジャーナルと同じように量的研究がより多く掲載されている ARELE でも、観測された差の分析を行っている論文が多いことが予想できる。そのため、以下では、ARELE 第13号から第24号までの過去12号分における「測定された差の分析」の使用傾向を報告する。

ARELE 第13号から第24号までで、CiNii に公開されている論文290本のうち、統計的手法を用いているものは255本であった。そのうち、観測された差の分析を行っている論文が186本 (73%) であり、その他の69本 (27%) は相関係数を基にした分析 (重回帰分析やその他の多変量解析など) であった。この結果からも、ARELE に掲載されている論文の多くで、観測された差の分析が行われていることがわかる。

表 15.1 は観測された差の分析を行っている論文186本における分析方法と

報告内容を集計したものである。分析方法では、 $t$  検定のみが 30 本 (16%)、分散分析のみが 115 本 (62%)、その両方が 38 本 (20%) となっており、分散分析は合計で 186 本中 153 本 (82%) において使用されており、*ARELE* 掲載論文は分散分析の割合が高いといえる(対応するノンパラメトリック検定も集計に含む)。「検定なし」の 2 本は、効果量とその信頼区間で解釈を行っているもの(Koizumi, 2011) と、記述統計と論旨に合った図示方法で議論を行っているもの(前田, 2008) の 2 本であるが、統計的検定の限界を理解した上で、必要な情報は十分に提示し、適切な解釈を加えているものなので、意図的に統計的検定を用いていない良い例であるといえる。「記載なし」の 1 本については、「有意な差 (significant difference) があった」とのみ記載されており、行っている分析を見ればおそらく分散分析であるということはわかったが、本文中に分析方法に関する記述は見られなかった。

表 15.1 観測された差の分析方法と報告内容 ( $K = 186$ )

	$K$	%
分析方法		
$t$ 検定	30	16
分散分析	115	62
両方	38	20
検定なし	2	1
記載なし	1	1
報告内容		
サンプルサイズ	184	99
平均 (Mean)	174	94
標準偏差 ( $SD$ )	155	83
統計量	177	95
$p$ 値 (*のみも含む)	182	98
$p$ 値の明記	86	46
信頼性を示す指標	74	40
効果量	57	31
効果量の信頼区間	3	2

報告内容については、標準偏差 (standard deviation:  $SD$ ) や信頼性を示す指標、 $p$  値の明記、そして効果量やその信頼区間が記載されていない論文が多く見られ、*ARELE* 掲載論文のうち、少なくとも、測定された差の分析を行っているものについては報告内容の改善が望まれるといえるだろう。

## 2.2 測定された差の分析と解釈における注意事項

高度な分析が可能な統計解析ソフトウェアの普及と、竹内・水本 (2012)、平井 (2012)、前田・山森 (2004) などのように、英語教育学研究で用いられることの多い統計解析をわかりやすく説明している入門書が、ここ 10 年で発行されている状況が相まって、量的研究の中でも測定された差の分析が非常に多い現状であるが、前述のように報告内容については改善の余地がある。以下では特に報告内容に注目し、測定された差の分析と解釈における注意事項を挙げていく。(分析のための前提条件などの詳細は上記の入門書を参照されたい。)

### (1) 再現に必要な情報を必ず書く

測定された差の分析を行っている論文において、標準偏差 (*SD*) やその他の記述統計情報が報告されていないことがある。そのような場合、元データの再現や結果の検証が不可能である。そして、そのような論文は、量的研究において先行研究の効果量を統合して、その効果を検討するメタ分析で必要な効果量が計算できないために、メタ分析の対象から除外されやすい (Plonsky, 2013)。また、次節の「複数の変数を使った分析と解釈」にもあるように、変数間の相関係数も再現 (二次分析) には必要になるため報告したい。さらに、*p* 値の報告についても正確な記述を心がけたい。*APA Publication Manual* 第 6 版 (2010) では、(図表でのわかりやすさを重視する場合を除いて)  $p < .05$  やアステリスク (\*) での表記ではなく、 $p = .031$  というように明記することが求められており、0.001 以下の値を取る場合のみ  $p < .001$  と表記する (p. 114)。

### (2) 有意差至上主義からの脱却

統計的仮説検定の結果得られる *p* 値は、サンプルサイズに依存し、再現性に乏しい (Cumming, 2012)、そのため、水本・竹内 (2008) は、実質的な差を検討するために効果量を用いる重要性を唱えており、*ARELE* 掲載論文でも 2011 年 (第 22 号) 以降は効果量の報告が増えてきている。 $p < .05$  やアステリスク (\*) での表記では、「有意差あり・なし」の二分法だけに注目してしまい、多くの研究者・実践者が本当に関心のあるはずの「どの程度効果があるか」ということまではわからない。*ARELE* 掲載論文でも、報告すべき内容が書かれておらず、 $p < .05$  やアステリスク (\*) のみの記載が優先されているものがあるという状況は、「有意差至上主義」のマインドセットのせいであるといえるだろう。

また、 $p$  値は再現性が低いため、再現性を重視するのであれば、追試(replication)によって結果を蓄積し、吟味していくメタ分析的アプローチを優先させるべきである。そのためには、効果量とその信頼区間を報告し、それに基づいた解釈を行うのが適切である(Cumming, 2012; 大久保・岡田, 2012)。ARELE掲載論文のうち、効果量と信頼区間による報告・解釈を行っている論文はKoizumi (2011)のみであるが、今後、そのような報告方法のさらなる普及が期待される。

### (3) 実質的重要性 (practical significance) を考える

効果量や信頼区間をもってしても、実際に測定された差が、学習者や英語教育学研究にとって、どのような意味を持つのかという実質的重要性 (practical significance) はわからない。この解釈については、より多くの研究から得られた効果量とその信頼区間を用い、研究の文脈を理解した分析者自身が考えていく必要がある。

### (4) 測定の信頼性を示す指標を報告する

ARELE掲載論文では、表 15.1 からわかるように、測定の信頼性を示す指標の報告が少ない (186 本中 40%)。いくら高度で複雑な分析を行っていたとしても、そもそもの測定の信頼性が低ければ、本当に差がある場合に有意差を検出できる力である検定力も低くなってしまい、結果の再現性が低くなる (検定力については、水本・竹内, 2011 を参照)。信頼性係数を使って各種の補正を行う二次分析もあることから、測定の信頼性を示す指標は報告すべきである。

### (5) データの可視化を心がける

前田 (2008) に見られるような論旨に合った図示方法は、“A picture is worth a thousand  $p$  values” (Loftus, 1993) という言葉からもわかるように、 $p$  値のみでの結果提示よりも価値があるといえる (統計的検定は行っていないでも必要な情報があれば再現し検定もできる)。また、上位群と下位群に分けてから分析を行うような場合も、それぞれの群の分布がどのようなものであるかを見せるという目的から、生データが可視化できるような方法が好ましい。

### (6) 要因は増やすよりも減らす努力を

一般的に、英語教育学のみならず心理学のような関連分野でも、先行研究の

結果から得られる検定力は低い。これはサンプルサイズが小さいことに起因するが、複雑なデザインで二元配置以上の分散分析を行う場合には、検定力も低くなりがちである。そのため、いたずらに要因を増やすよりも、先行研究をしっかり概観することで、できるだけ少ない要因で仮説検証型の研究に落とし込む努力が必要であるだろう (Mizumoto, Urano, & Maeda, 2014)。

〔水本 篤〕

### 3 複数の変数を使った分析と解釈

#### 3.1 ARELEの過去10年の動向

英語教育学研究には学習者・教員・学習環境などの多くの要因が複雑に関わる。そのため、研究目的に応じ、適切な分析手法を使うことが必要である。適切な分析手法を用いて初めて、より妥当なデータ解釈につながるため、研究のデザインを考える際に、どの分析手法を用いれば、どのような解釈ができるかを吟味する必要がある(水本, 2012)。

このような背景を受け、ARELEの収録論文でも複数の変数を対象にした分析が多く行われてきた。2004年発行の第15号から現時点で最新号の2013年発行の第24号までの過去10年を概観すると、相関係数を用いた論文が29件(例 Ushiro, Hijikata, Shimizu, Nakagawa, Koga, Ohno, & Umehara, 2007; 山西・廣森, 2008), 多変量分散分析(MANOVA)を用いた論文が2件(Adachi, 2013; Koga & Sato, 2013), 重回帰分析を用いた研究が6件(例 Ohno, 2004; Sakuma, 2011), 探索的因子分析を用いた研究が13件(例 Fujita & Noro, 2009; Yamanaka & Takeuchi, 2009), 共分散構造分析(structural equation modeling)を用いた研究が11件(例 Katagiri, 2010; 齊田, 2012)ある。特に相関分析は基本的な分析手法であり、広く使われていることが分かる。また、これらの分析をすべて包括する共分散構造分析が、件数は少ないが継続的に使われている。適切な分析と解釈を行い、英語教育学研究をより良いものにするため、これら複数の変数を扱う分析手法を使う際に特に注意すべきことを7点述べる。

#### 3.2 複数の変数を使った分析と解釈の注意事項

第1に、変数間の分析結果を、因果関係と直に解釈することはできない。因果関係と解釈できるのは、(a)原因が結果に時間的に先行していること、(b)原



因に変化があると結果にも変化が起こること、(c)原因と結果の間にその他の変数がないこと、(d)原因と結果の因果関係の方向性が正しいこと(例:原因  $X \rightarrow$  結果  $Y$  であり、 $X \leftarrow Y$  や、 $X \Leftrightarrow Y$  でないこと)、(e)各変数の統計的分布が正しいことの5つの条件を満たした場合である(Kline, 2011)。例えば、アンケート調査を1度実施し分析する場合、すべての変数を同時に測っている。そうすると、原因と結果を同時に測ることになり、どの変数が原因でどの変数が結果であるかをモデル化するには理論的根拠が必要である。特に共分散構造分析を用いた研究では多くの変数をモデルに入れることがあるが、変数間の関連や因果関係のモデル化は理論に基づくべきであり、統計的なモデル指標のみに基づくべきではない。研究のデザインを考える際に上記5点をできるだけ検討することが望ましいが、単独の研究で満たすことは難しい。そのため、単独の研究では、変数間に関連があることの記述にとどめ、因果関係を含意する解釈を避けるのが無難かもしれない。同一のテーマについて研究が増えるにつれ、比較考察を加え、徐々に因果関係を構築するのがよいだろう。

第2に、重回帰分析を用いる際には、独立変数の投入法に気をつける必要がある。ステップワイズ法は、多くの独立変数の中から従属変数の説明に役立つ変数を統計的に選ぶ投入法である。先行研究が少ないため非常に多くの独立変数を検討する必要がある場合や、理論的にも独立変数を精選できない場合は、ステップワイズ法の使用はやむを得ない。ただし、変数の選択結果はデータに大きく依存するため(Tabachnick & Fidell, 2007)、新たな研究でも同じ変数が重要と選択されるとは限らず一般化可能性が低くなる。その結果、研究間での比較検討が難しくなる一因となる。もし先行研究や理論に基づき独立変数の投入順を指定できる場合、階層的重回帰分析(hierarchical/sequential multiple regression)の使用が好ましいだろう。最も重要と考えられる独立変数を最初に投入し、説明率が累積的にどのように変わるかを検討することで、各独立変数の重要性を調べることができる。

第3に、アンケートやテストを使い探索的因子分析を行い、各因子を構成する項目における回答の合計値や平均値、または因子得点(今野, 2012)を求め、重回帰分析を行う事例が見られる。この方法は、先行研究や理論的背景が少ないため因子構造の探索が必要である場合はよいかもしれない。一方、先行研究や理論的背景が十分にある場合、モデルを作り、共分散構造分析を用いれば、以上の分析をすべてまとめて行うことができ、さらには項目の測定誤差を考慮

することもできる。また、共分散構造分析では、モデルの適切さを様々な視点から評価することができる。従って、可能であれば共分散構造分析の使用を推奨する。

第4に、第3点目と関連するが、アンケートやテストをモデル化する際に、各項目をそのままモデルに入れるか、各項目を合計した値をモデルに入れるかを検討する必要がある。例えば、扱う変数の一つに中間テストの得点があるとすると、1問1点で35問からなり、内訳は語彙(5問)、文法(5問)、読解(10問)、リスニング(10問)、ライティング(5問)セクションからなる。特に共分散構造分析を用いてこの中間テストをモデル化する場合、3つの方法がある。第1に、各問を各変数と捉え、35の観測変数としてモデル化する。第2に、35問全体で一つの変数と捉え、35問の合計点を出し、1つの観測変数としてモデル化する。第3に、5つのセクションからなると捉え、各セクションの合計点を出し、5つの観測変数の背後に中間テストで測る力、例えば到達度の因子があるとモデル化する。後者2つのように得点を足し合わせモデル化することをパーセルング(parceling)と言う。いずれの方法をとるかは、理論的根拠が最優先されることに加え、分析上の前提を満たすかどうかや、統計パッケージの機能にも依存する。詳細はBrown (2006)を参照のこと。

第5に、複数の変数を使う分析ではアンケートやテストを用いることが多いが、アンケート(項目)やテストの選定・作成は慎重に行いたい。先行研究や理論的根拠に加え妥当性・信頼性による裏付けがあれば、先行研究で使用されたアンケートやテストをそのまま用いてよいだろう。ただし、先行研究で妥当性・信頼性ともに適切なアンケートやテストという結果が示されていたとしても、対象や実施条件が異なれば妥当性・信頼性が低くなることもあることは知っておきたい。さらに、時間的制約などのため既存アンケートやテストの一部分を実施する場合、本来の意図とは異なる特性を測ってしまう恐れがある。その結果、データは分析できたとしても、先行研究との比較が困難になる。測定したい能力をより短いアンケートやテストで測ることは吟味すれば可能であり、Smith, Combs, & Pearson (2012)が詳しい。

また、各先行研究では少数の概念しか測定していないが、新たに行う研究内でより多くの概念を比較する場合は注意が必要だろう。例えば、先行研究Aは「不安」がテスト得点に与える影響を扱い、先行研究Bは「動機づけ」がテスト得点に与える影響を扱い、新たに行う研究では両者がテスト得点に与える影

響を調べるとする。面白い研究テーマであるが、先行研究では少数の概念のみを対象にしたことを考えると、おそらく少数の概念を測るだけでもかなりのアンケート項目数が必要であり、より多くの概念を一度に対象にするのは難しいかもしれない。一般的に、各概念は複数の下位要素から構成され、下位要素の数だけ、アンケート項目数が増えてしまう。そして、各概念で別のアンケートを作成したとしても、1つ目のアンケートには誠実に回答しても、2つ目のアンケートで回答者は疲れてしまい、回答が揺らぐかもしれない。不安を測る項目と動機づけを測る項目を同一のアンケート内にランダムに配置する場合と、各概念で別のアンケートに分けた場合では回答の質は変わってしまう。より多くの概念を単一の研究内で扱う場合は予備調査を行い、測定時にかかわる影響を検討することが必要だろう。

第6に、分析の詳細が報告されていない場合は、分析結果および解釈の信憑性に疑問が残る。例えば、共分散構造分析を使う際には、かなりの受験者数に加え、観測変数が正規分布しているか(外れ値はないか)、変数間にはU字形ではなく線形の関係があるか、欠損値はないか、パラメーターの推定法に何を用了か、どのパラメーターを固定したか、分散が負の値になるなど分析中に問題が生じなかったか、などを報告する必要がある。外れ値(outlier)があればその理由を検討し、除去するかを決める。観測変数が正規分布でなければ、Satorra-Bentlerの方法を用いパラメーター推定の過程で修正を行う必要がある。たとえ最も単純な分析である2つの変数間の相関係数であっても、外れ値が大きく影響することは決して少なくない。散布図を描く(前田, 2004b, 2008; 竹内・水本, 2012)などの方法で、外れ値は見つけることができる。詳細はBrown (2006), Byrne (2010), Kline (2011)が詳しい。

第7に、すべての分析において、平均値・標準偏差・相関係数行列を報告することが好ましい。これらの値があれば、再分析が可能になることが多い。効果量が報告されていない場合、計算することもできる。これらは査読の過程で、査読者が分析の適切さを確認する際にも役立つ。また、類似テーマの研究を集め、メタ分析を用い統計的に統合することも可能になる。具体的な手順は、水本・竹内(2011)や、Zientek & Thompson(2009)に記載されている。メタ分析では、元の研究では扱わなかった分析も可能になる。例えば、元の研究Aでは日本人、研究Bでは韓国人、研究Cでは中国人が対象であった場合、元の研究では各国の学習者への影響に焦点があるが、メタ分析では国間で影響を

比較考察でき、一般化可能性の程度が分かる。共分散構造分析を用いて分析すれば、新たなモデル構築も可能である(手順は In'nami & Koizumi, 2012 を参照)。このように平均値・標準偏差・相関係数行列があれば、多くの分析が可能になる。そのため、これらの情報は紙幅が足りなければオンライン上で公開したり、問い合わせがあった際には積極的に共有するなど、何らかの形で利用可能にしたい。その結果、研究間で知識の再検証・蓄積が進み、英語教育学研究での学習者・教員・学習環境などの多くの要因間の関係がより一層解明されるだろう。

【印南 洋】

### 【引用文献】

- Adachi, R. (2013). Pupils' changes in communicative attitudes toward English activities—A case study at a Japanese elementary school. *Annual Review of English Language Education in Japan*, 24, 221-233.
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford.
- Byrne, B. M. (2010). *Structural equation modeling with Amos: Basic concepts, applications, and programming* (2nd ed.). New York, NY: Routledge.
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.
- Fujita, K., & Noro, T. (2009). The effects of 10-minute extensive reading on the reading speed, comprehension and motivation of Japanese high school EFL learners. *Annual Review of English Language Education in Japan*, 20, 21-30.
- Gass, S. (2009). A survey of SLA research. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 3-28). Bingley, UK: Emerald.
- 平井明代 (編著) (2012). 『教育・心理系研究のためのデータ分析入門：理論と実践から学ぶ SPSS 活用法』東京：東京書籍
- In'nami, Y., & Koizumi, R. (2012). Reproduction of structural equation models in second language testing and learning research. *Reports of 2011 Studies in Japan Association for Language Education and Technology, Kansai Chapter, Methodology Special Interest Groups (SIG)* (pp. 15-40).
- Katagiri, K. (2010). Three-year longitudinal study of the progress in listening and reading proficiency of Japanese senior high school students and their developmental patterns. *Annual Review of English Language*

- Education in Japan*, 21, 221-230.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford.
- Koga, T., & Sato, R. (2013). Effects of a debate task on changes of communication variables. *Annual Review of English Language Education in Japan*, 24, 295-306.
- Koizumi, R. (2011). Test-taking processes of the Lexical Organisation Test: Comparing it with the Word Associates Test. *Annual Review of English Language Education in Japan*, 22, 153-168.
- 今野勝幸 (2012). 「因子分析」 In 平井明代 (編著), 『教育・心理系研究のためのデータ分析入門: 理論と実践から学ぶSPSS活用法』 (pp. 181-203). 東京: 東京書籍
- Lazaraton, A. (2005). Quantitative research methods. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 109-224). Mahwah, NJ: Erlbaum.
- Loftus, G. R. (1993). A picture is worth a thousand *p* values: On the irrelevance of hypothesis testing in the microcomputer age. *Behavior Research Methods, Instruments, & Computers*, 25, 250-256.
- 前田啓朗 (2000). 「構成概念の妥当性の検証—日本の英語教育学研究における傾向と展望—」 *JLTA Journal*, 3, 119-126.
- 前田啓朗 (2004a). 「因果分析の妥当性の検証—日本の英語教育学研究における傾向と展望—」 *JLTA Journal*, 6, 140-147.
- 前田啓朗 (2004b). 「テスト得点間の関係の検討—相関分析」 In 前田啓朗, 山森光陽 (編著), 『英語教師のための教育データ分析入門: 授業が変わるテスト・評価・研究』 (pp. 64-72). 東京: 大修館書店
- 前田啓朗 (2008). 「WBT を援用した授業で成功した学習者・成功しなかった学習者」 *Annual Review of English Language Education in Japan*, 19, 253-262.
- 前田啓朗・山森光陽 (編著) (2004). 『英語教師のための教育データ分析入門: 授業が変わるテスト・評価・研究』 東京: 大修館書店
- 水本 篤 (2012). 「研究をはじめるとは: 研究計画をたてるには」 In 竹内 理・水本 篤 (編著), 『外国語教育研究ハンドブック—研究手法のより良い理解のために』 (pp. 3-16). 東京: 松柏社
- 水本 篤・竹内 理 (2008). 「研究論文における効果量の報告のために—基礎的概念と注意点—」 『関西英語教育学会紀要 英語教育研究』 31, 57-66. Retrieved from [http://www.mizumot.com/files/EffectSize\\_KELES31.pdf](http://www.mizumot.com/files/EffectSize_KELES31.pdf)
- 水本 篤・竹内 理 (2011). 「効果量と検定力分析入門—統計的検定を正しく使うために—」 『より良い外国語教育研究のための方法: 外国語教育メディア学会 (LET) 関西支部 メソドロジー研究部会 2010 年度報告論集』 47-73. Retrieved from <http://www.mizumot.com/files/metho2010.pdf>
- Mizumoto, A., Urano, K., & Maeda, H. (2014). A Systematic review of published

## 第15章 統計的な分析と分析結果の解釈

- articles in *ARELE* 1-24: Focusing on their themes, methods, and outcomes. *Annual Review of English Language Education in Japan*, 25, 33-48.
- 並木 博 (1997). 『個性と教育環境の交互作用：教育心理学の課題』東京：培風館
- 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』東京：勁草書房
- Ohno, T. (2004). What first-sight oral reading performed by Japanese EFL learners implies: general English proficiency or working memory capacity? *Annual Review of English Language Education in Japan*, 15, 139-148.
- Plonsky, L. (2013). An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, 35, 655-687. doi: 10.1017/S0272263113000399
- 斉田智里 (2012). 「授業満足度と成績に影響を及ぼす授業評価要因の検討—大学英語教育プログラム改善の観点から—」 *Annual Review of English Language Education in Japan*, 23, 389-403.
- Sakuma, Y. (2011). Cognitive features of working memory in elementary school students participating in foreign language activities. *Annual Review of English Language Education in Japan*, 22, 233-248.
- Smith, G. T., Combs, J. L., & Pearson, C. M. (2012). Brief instruments and short forms. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology: Vol. 1: Foundations, planning, measures, and psychometrics* (pp. 395-409). Washington, D.C.: American Psychological Association.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn and Bacon.
- 竹内 理・水本 篤 (編著) (2012). 『外国語教育研究ハンドブック—研究手法のより良い理解のために』東京：松柏社
- Ushiro, Y., Hijikata, Y., Shimizu, M., Nakagawa, C., Koga, T., Ohno, M., & Umehara, C. (2007). Relationships between cue types in recall tests and L2 reading proficiency. *Annual Review of English Language Education in Japan*, 18, 31-40.
- Yamanaka, Y., & Takeuchi, O. (2009). Why some Japanese students have trouble learning English: A factor analytic study. *Annual Review of English Language Education in Japan*, 20, 201-210.
- 山西博之・廣森友人 (2008). 「適切な指導と評価を目指した、愛媛大学共通教育「英語」カリキュラム開発への取り組み：英語運用能力判断基準(Can-Do リスト)の開発とその意義」 *Annual Review of English Language Education in Japan*, 19, 263-272.
- Zientek, L. R., & Thompson, B. (2009). Matrix summaries improve research reports: Secondary analyses using published literature. *Educational Researcher*, 38, 343-352.