

## 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—

水本 篤 (流通科学大学)  
竹内 理 (関西大学)

## Basics and Considerations for Reporting Effect Sizes in Research Papers

Atsushi MIZUMOTO (University of Marketing and Distribution Sciences)  
Osamu TAKEUCHI (Kansai University)

キーワード: 効果量, 統計的検定, 検定力, サンプルサイズ, メタ分析

Keywords: effect size, statistical testing, power, sample size, meta-analysis

### SUMMARY

Reporting effect sizes is now mandatory when submitting research papers to professional journals in our field. In this paper, therefore, we introduce and discuss the fundamental concepts and some considerations of the effect sizes along with other related issues in order to facilitate using and reporting effect sizes in EFL research in Japan.

#### 1. はじめに

心理学をはじめとする諸分野の従来の研究論文においては, 何らかの検定を行った際には, 対象となった比較グループの平均値, 標準偏差の他に, サンプル数, 検定量, 有意値という情報を中心に記載する形がほとんどであった。しかし, 国内外の論文誌が論文執筆の際の基準として用いている APA Publication Manual 第5版 (2001) (以下, APA 第5版と略記) では「効果量(effect size)」(効果サイズ, 効果の大きさ)の報告が必要であると以下のように記されている。

For the reader to fully understand the importance of your findings, it is almost always necessary to include some index of effect size or strength of relationship in your Results section. (p. 25)

効果量自体はそれほど最近の考え方ではなく、たとえば、海外における応用言語学研究の統計的分析の入門書としてよく引用されている Hatch and Lazaraton (1991) の *The Research Manual* でも、効果量の概念はすでに紹介されていた (p.265 など)。また、最近では、Dörnyei (2007) の応用言語学における研究手法を包括的に扱っている専門書でも、効果量の報告が必要であると記載されている (p.212)。しかし、現状としては、効果量が報告されているものは、特に日本国内の論文誌に掲載されている論文では数が少ない。たとえば、全国英語教育学会の発行している、*ARELE* の2007年度版に掲載されている24本の研究論文のうち、記述統計や相関分析以外の統計的手法が用いられている22本では、効果量の報告が行われているのはわずか3本であった。また、わが国の外国語教育において効果量の考えが浸透していないことは、2003年に発行された『応用言語学事典』(研究社)においても、まったく取り上げられていないことから明らかである。小野寺・菱村 (2005) は、心理学、教育、カウンセリングなどの論文雑誌での効果量について、「日本ではこのような流れはまだみえてきていないが、方向性は定まっており、いずれ数年のうちに効果サイズを無視した研究報告はできなくなると考えられる。」(p.135) としている。もし、外国語教育学が人文系の一分野として他の研究領域と同じような水準の研究が求められるのならば、効果量の報告があまりにも少ない現状は変えていかなければならない。しかし、日本においては、統計関係の書物でも効果量が断片的に取り扱われているか、もしくはまったく取り上げられていない場合が多い。また、APA 第5版 (2001, p.25) に例として挙げられている、 $r^2$ ,  $\eta^2$ ,  $\omega^2$ ,  $R^2$ ,  $\phi^2$ , Cramer's  $V$ , Kendall's  $W$ , Cohen's  $d$  のように、いくつかある効果量の指標のうち、どれをどの分析に選択すれば良いかが十分に説明されている書籍も皆無に近い。そこで、本稿は効果量の報告が外国語教育の論文でも活用しやすくなるように、効果量の基礎的概念をわかりやすく紹介し、その使用の目的と注意点を解説する。

## 2. 効果量とは

### 2.1 効果量の基本的な考え方

これまでの研究では、いくつかのグループの間の平均値(平均点)に統計的な違いが存在するかを証明するためには、 $t$  検定や分散分析(ANOVA)などの分析を行い、基準として  $p$  値が 0.05 以下 ( $p < .05$ ) の場合には、「平均値に有意な差があった」という報告をするのが慣例であった。このような検定は、手元にある、サンプル(標本, sample)のデータから、母集団(population)の平均値差を推定する推測統計の考え方を利用しており、サンプルサイズ(サンプル数,あるいは被験者数)が大きくなればなるほど、統計的に有意であるという結果になりやすいという大きな問題がある。このため、ある検定を行ったところ、20人では有意ではなく、200人のデータの場合には有意になるということも十分にあり得る。<sup>1</sup> 一般的に、どの分野の論文誌でも有意な結果 ( $p < .05$ ) が出た研究論文を掲載する傾向があるため、研究者はサンプルサイズをどうにか増やして、有意な結果を出しやすくなるようなケースも考えられる。<sup>2</sup>

効果量とは、「効果の大きさ」のことを指し、実験的操作(experimental manipulation)の

効果や変数間の関係の強さ(strength of association)を表す指標である (Field & Hole, 2003, p.152)。 $p$  値は前述のように、サンプルサイズによって変わるものなので、実質的効果が大きいか小さいかについての情報は何も与えてくれない。そこで、サンプルサイズによって変化することのない、標準化された指標である効果量が用いられることとなった。芝・南風原 (1990) によると、効果量は、「測定単位にたよらない指標となっている。そのため、効果量を用いれば、単位の異なる変数を用いた研究の間でも、実験条件の効果の大きさを互いに比較することができる」(p.118) と定義されている。実験の条件によっては、有意差があっても ( $p < .05$ )、実質的効果があまりない(効果量が小さい)場合もあれば、有意差がなくても ( $p > .05$ )、効果量が大きい場合も考えられるため、有意差があろうがなかろうが、どちらにしても効果量は報告すべきである (Field, 2005; Kline, 2004 など)。このような効果量の特性は後述するメタ分析(meta-analysis)に適していて、いくつかの研究結果を統合する目的で使用されることも多い。

## 2.2 効果量と検定力, サンプルサイズ, 有意水準の関係

効果量を使用するにあたって、「検定力」の考え方をある程度理解しておく必要がある。検定力(power)とは、「有意差を見つける力」(磯田, 2004, p.48),あるいは「母集団において差があるとき、サンプルにおいて有意な結果が得られる確率」(南風原, 2002, p.143)である。サンプルサイズを大きくすれば検定力は高まるが、逆に、検定力が強すぎる場合には、前節の説明にあるとおり、実質的には有意差がなくても、有意な差を検出する可能性がある。

有意差検定では、「実際には有意差がないのに有意差あり」としてしまう第 1 種の誤り (Type I error) を犯す確率 ( $\alpha$ ) と、「実際には有意差があるのに有意差なし」としてしまう第 2 種の誤り (Type II error) を犯す確率 ( $\beta$ ) の可能性を常に考えなければならない。有意水準や危険率とも呼ばれる第 1 種の誤り (Type I error) を犯す確率 ( $\alpha$ ) は、(心理学などの分野で) 統計的検定の慣例として、 $\alpha = 0.05$  に決められており、第 2 種の誤り (Type II error) を犯す確率 ( $\beta$ ) は、 $\beta = 0.2$  が望ましいと提案されている (Cohen, 1988)。<sup>3</sup> そして検定力は  $1 - \beta$  で定義されるため、 $\beta = 0.2$  の場合、 $1 - 0.2$  で 0.8 になる。検定力が 0.8 ということは、80% の確率で実際に有意差があるときには、それを検出できることを意味している。また、Cohen (1992) では、「0.80 以下の検定力の場合には、第 2 種の誤りを犯す可能性が高くなる」(p. 156) としている。このように、第 2 種の誤りは検定力の計算に直接関わってくる問題である。

Field and Hole (2003) や 村井 (2006) によると、サンプルサイズ、有意水準 ( $\alpha$ )、検定力 ( $1 - \beta$ )、効果量の 4 つは、他の 3 つが決まれば残りの 1 つが決まるという関係であるとされている。上述のように、検定力 ( $1 - \beta$ ) は 0.8、有意水準 ( $\alpha$ ) は 0.05 と決まっているので、研究をデザインする際に実際に考慮しなければならないのは、サンプルサイズと効果量の関係になってくる。それでは、研究においてサンプルサイズはどう決めれば良いのだろうか。「20 人～30 人ぐらいからデータを取れば正規分布となるので良いのではないか」という従来の考

え方の場合、実用性は高いが、サンプルサイズが小さすぎると検定力が下がってしまう可能性がある。また、「できるだけ多くの実験参加者からデータを集める」というのは、検定力が高まる反面、微妙な差でも有意差ありという結果になってしまうので、これも薦めることはできない。このような観点から、村井（2006）は検定力分析(power analysis)を利用し、サンプルサイズを決定するいくつかの方法を提示している。検定力分析では、(1) 実験を実施した後に、サンプルサイズ、効果量、有意水準( $\alpha$ )から、検定力( $1 - \beta$ )を確認する方法と、(2) 実験を実施する前に、これまでの先行研究からわかっている(推測される)効果量、<sup>4</sup>有意水準( $\alpha$ )、目指している検定力( $1 - \beta$ )からサンプルサイズを決定する目的のものが多い (Field & Hole, 2003, p.154)。

このように、検定力分析によって、検定力を確認したり、サンプルサイズを決定したりする際にも、効果量は重要な位置を占めているため、非常に重要な概念であるといえる。なお、検定力分析は、フリーソフトである G\*Power (Faul et al., 2007) を使って簡単に確認できるため、一度、実際のデータを使って試してみることをお薦めする。

### 2.3 効果量の種類と注意すべき点

効果量を表す指標は数多く存在するため、どの指標を用いれば良いのかが非常にわかりにくい、大きく 2 種類に分けることができる (Kline, 2004, p.97)。まず、(1) *d family* と分類される、「グループごとの平均値の差を標準化した効果量」(Cohen's *d*, *f* など)、そして、(2) *r family* と分類される、「変数間の関係の強さ(strength of association)を示す効果量<sup>5</sup>」( $r^2$ ,  $\eta^2$ , partial  $\eta^2$ ,  $\omega^2$ ,  $R^2$ ,  $\phi^2$  など)である。どちらの種類も、絶対値が大きくなればなるほど効果は大きく、小さくなればなるほど効果が小さいことを表す。

(1) の *d family* と分類される、「グループごとの平均値の差を標準化した効果量」の代表的な指標である Cohen's *d* は、*t* 検定のような 2 グループの平均値の差を比較するとき使用し、平均値の差の効果量を以下のような式で求めることができる (小野寺・菱村, 2005, p.136)。

$$d = \frac{(\text{実験群の平均} - \text{統制群の平均})}{\sqrt{\frac{\text{実験群の標準偏差}^2 + \text{統制群の標準偏差}^2}{2}}}$$

この計算から得られる値はグループごとの平均値の差を標準化したもの (standardized mean difference) になっている。算出される数値は、標準偏差を単位として平均値がどれだけ離れているかを表しており、たとえば、 $d = 1$  なら、1 *SD* 離れていることを意味する (吉田, 1998)。このような種類の効果量は、値が 1 を超えることもある(理論的には上限と下限は無制限である)ため、解釈がわかりづらいという一面もある。しかし、*d* を用いる方法も広く行われているため、指標としてどのようなものであるかは理解しておくべきである。

一方、(2) の  $r$  family と分類される、「変数間の関係の強さを示す効果量」は、相関係数 (correlation coefficient) に基づいた効果量である。特に、ピアソンの積率相関係数 ( $r$ ) は、相関係数の指標としてどのような種類の研究でも使用することが多い指標であり、また効果量の指標として使用する際には 0 ~ 1 の範囲に収まるため、直感的にも理解しやすい (Field, 2005, p. 32)。さらに、Cohen's  $d$  を  $r$  に変換する式が Cohen (1988) などでも紹介されていることから、<sup>6</sup> Field (2005) は、効果量の指標として  $r$  の使用を推奨している。 $t$  検定における効果量  $r$  の計算は  $t$  値と自由度 (df) を使って以下のように計算される。

$$r = \sqrt{\frac{t^2}{t^2 + df}}$$

ここで求められる  $r$  は、たとえば、実験群と統制群の 2 つのグループの点数を並べたものの相関係数ではなく、実験群と統制群のグループを表す名義尺度 (0 を実験群とし、1 を統制群とするような 2 値データ) と点数の間の点双列相関係数 (point-biserial correlation coefficient) であるため、<sup>7</sup> 実験群と統制群の 2 つのグループの得点同士の相関係数とは違う種類のものであるということに注意しなければならない。また、2 変数のピアソンの積率相関係数ではマイナスの値になることもあるが、効果量を表す  $r$  ではマイナスの値が算出されることはなく、常にプラスの値となる (Field, 2005, p. 32)。

相関係数を使った効果量のうち、分散分析 (ANOVA) を使用している研究では、相関比 (correlation ratio,  $\eta$ ) を用い、 $\eta^2$  (eta squared, イータ 2 乗) で効果量が表されることが多い。 $\eta^2$  は以下のように定義される。<sup>8</sup>

$$\eta^2 = \frac{\text{ある要因の平方和 (SS}_{effect})}{\text{全体平方和 (SS}_{total})}$$

この式から明らかなように、 $\eta^2$  では、全体における、ある要因の占める割合 (分散説明率) が計算されている。ちなみに、 $t$  検定の効果量として説明した  $r$  は相関比の特殊な形であるため、 $\eta^2$  は、分散説明率と呼ばれる相関係数 ( $r$ ) を 2 乗した  $r^2$  や  $R^2$  (回帰分析の決定係数, coefficient of determination) と同じものであると考えてよい (Field, 2005, p.357)。このように、 $\eta^2$  は  $r^2$  と同じものであるが、分散分析のときには  $\eta^2$  が慣例として使われるということに注意しなければならない。

表 1 は、研究で使われることが多い検定 (分析) と、それぞれに対する代表的な効果量の指標をまとめたものである。注意すべき点を表 1 に対応する形で以下にまとめておく。

#### (1) 相関分析

相関分析の場合は、算出された数値がそのまま効果量として使われる。



## (2) 重回帰分析

$R^2$  は重回帰分析を行った際に SPSS などの統計ソフトではアウトプットで表示される。

## (3) $t$ 検定

効果量算出方法の種類が違う  $r$  と  $d$  では効果量の目安も異なる。 $t$  検定には繰り返しありと繰り返しなしのパターンがあるが、 $r$  と  $d$  ともに計算式は同じ形で効果量を求めることができる。また、Hatch and Lazaraton (1991, p. 266) では  $\eta^2$  が紹介されているが、前述のように  $\eta^2$  は  $r$  を 2 乗したものとまったく同じであり、呼び名が違うだけである。ちなみに、 $d$  と同じように、グループごとの平均値の差を標準化した効果量として、Hedge's  $g$  や、Glass's  $\Delta$  も使われることがある<sup>9</sup> (詳細は Kline, 2004 を参照)。

## (4) 一元配置の分散分析 (ANOVA)

$\eta^2$  は  $SS_{\text{effect}} / SS_{\text{total}}$  で表されるが、他の独立変数の個数や、その独立変数の統計的有意性の影響を受けるという欠点がある (Tabachnick & Fidell, 2006, p. 54)。具体的には、すべての  $\eta^2$  の合計が 1.0 (100%) になるため、独立変数 (要因) の数が増えれば、その分、1 つの独立変数が占める割合も少なくなってしまう、どれほど効果があるのかが分かりづらくなってしまう。そのような欠点を補うために、他の変数の要因を統制した上で、ある 1 つの独立変数の影響の効果量を計算するのが、partial  $\eta^2$  (偏イータ 2 乗, 偏相関比) である。partial  $\eta^2$  ( $\eta_p^2$  と書く) は、 $SS_{\text{effect}} / (SS_{\text{effect}} + SS_{\text{error}})$  で求められるので、 $\eta^2 (SS_{\text{effect}} / SS_{\text{total}})$  との違いを認識しておくべきである。しかし、SPSS のヘルプでは、partial  $\eta^2$  が  $\eta^2$  であるかのような説明がされており、実際に出力される値も partial  $\eta^2$  であるため (堀, 2006)、多元配置の分散分析では研究者が誤って partial  $\eta^2$  を  $\eta^2$  と報告する例があとを絶たない<sup>10</sup> (Levine & Hullett, 2002)。ゆえに、partial  $\eta^2$  を効果量として使用した場合には、必ずそのように明記すべきである (Tabachnick & Fidell, 2006, p. 55)。また、partial  $\eta^2$  から  $\eta^2$  への変換式はないため、後から他の研究者が効果量を計算できるように、有意であっても有意でなくてもすべての変数に対して、平方和や平均平方などの情報を記載すべきである。なお、 $\eta^2$  は母集団における推定値が不正確であり一般化できないため、より正確な母集団推定値として、Field (2005) や Hatch and Lazaraton (1991) は  $\omega^2$  を使うべきであるとしている。ただし、 $\omega^2$  はグループごとの人数が等しいときにしか使用できず (Field, 2005, p.384; Hatch & Lazaraton, 1991, p. 331)、繰り返しありの場合 (反復測定, repeated measures) には計算式が異なる (Field, 2005, p.452)。また、分散分析を使った研究ではほとんどが  $\eta^2$  (もしくは partial  $\eta^2$ ) を報告しているので、 $\eta^2$  で統一する方法を選ぶか、あるいは  $\omega^2$  も同時に提示することが考えられる。大切なのは、Cortina and Nouri (2000) が分散分析における結果を報告する際のガイドラインを提示しているように (p. 62)、後から (メタ分析等で) 他の研究者が必要になるかもしれない値はすべて論文中に提示しておくことであろう。

- (5) 二元配置分散分析 (Two-way ANOVA), 多元配置分散分析 (Multi-way ANOVA)  
主効果, 交互作用, 多重比較, それぞれに対して効果量の報告が必要である。
- (6) 共分散分析 (ANCOVA)  
共変量 (covariate) の影響を取り除いて分析し, 主効果, 交互作用, 多重比較, それぞれに対して効果量の報告を行う。
- (7) 多変量分散分析 (MANOVA), 多変量共分散分析 (MANCOVA)  
多変量検定の統計量 (Wilks's  $\Lambda$  など) を基にして,  $1-\Lambda$  の形で multivariate  $\eta^2$  が計算できる。また, 変数ごとの分散分析でも効果量の報告を行う。 multivariate partial  $\eta^2$  は multivariate  $\eta^2$  を補正したものである (Tabachnick & Fidell, 2006, p. 261)。
- (8) カイ2乗検定 ( $\chi^2$  test, chi-square test)  
 $2 \times 2$  の分割表 (クロス表) の場合には, 相関係数の一種である  $\phi$  (ファイ) 係数を用い, それ以外の場合は, Cramer's  $V$  を利用する。これらの値は SPSS などの統計ソフトの場合は結果にアウトプットされる。またカイ2乗検定の効果量はオッズ比 (odds ratio) で提示されることもある (Field, 2005, p. 693)。
- (9) ノンパラメトリック検定  
データに正規性が確認できない場合に使う, ノンパラメトリック検定も効果量が算出できる。SPSS では, 検定統計量が  $Z$  に変換されたものが出力されるので, それを用い,  $r = Z/\sqrt{N}$  で計算すればよい。3つ以上のグループを比較するクラスカル・ウォリスの順位和検定などではグループの組み合わせごとに効果量の計算を行う。

それぞれの効果量の計算方法については, 自ら数式を見ながら計算してみることが理解のためには重要であるが, 論文執筆の際の効果量の報告をできるだけ容易にするために, Excel による計算シートを作成したので (<http://www.mizumot.com/stats/effectsize.xls>), 具体例や計算式の詳細はこちらで確認いただきたい。

## 2.4 メタ分析

メタ分析 (meta-analysis) は, いくつか複数の量的な研究結果を統合して, 全体としての効果を検討する統計的分析手法である。検定の結果得られた  $p$  値を基準として比較するのは, サンプルサイズの影響を受けるためふさわしくない。ゆえに, いくつかの先行研究を  $p$  値以外の基準, すなわち効果量で比べるものがメタ分析である。

メタ分析を行うときに使用する効果量は同じ指標に統一しなければならない。グループごとの平均値の差を標準化した効果量を用いたメタ分析では,  $d$  が使われる場合が多いが, 最近では相関係数  $r$  が使用されることも多く (小野寺・菱村, 2005, p.134), その使用を推奨する統計解説書も増えている (Field, 2005; Cortina & Nouri, 2000, p.60; Howitt & Cramer, 2003, p.384)。たとえば, Howitt and Cramer (2003, p.384) では  $r$  に変換してメタ分析を行えば, 必ずしも同じ検定を使用した研究でない場合でも, 効果量の比較が可能であることを具体例として示している。一般的に, メタ分析は (1) 研究の対象を決める, (2)



データベースの検索を行う,<sup>11</sup> (3) 統計分析の報告部分からデータを抽出し, 1 つの効果量指標にコーディングする,<sup>12</sup> (4) 複数の研究の効果量を変換したり, 平均したりして) 統合し, 結果を検証する, という流れで行う (Howitt & Cramer, 2003; Kline, 2004)。

外国語教育学の分野においてメタ分析が用いられている研究は, 海外では, L2 指導の効果 (Norris & Ortega, 2000), 言語学習におけるテクノロジー利用の効果 (Zhao, 2003), 学習への態度と動機づけ, そして学習達成度との関係を調査したもの (Masgoret & Gardner, 2003) など, 近年増加しつつある。日本国内でも, 印南・小泉 (2007) のように, メタ分析を活用する研究がすでに行われている。このような観点からも, 効果量の理解は, 外国語教育学の今後の発展のために不可欠になってくるであろう。

### 3. おわりに

本稿では, 日本の外国語教育研究論文でも, 今後, 報告が必要になってくると考えられる効果量の基礎的な概念を紹介し, 同時に注意点を議論した。「有意差至上主義」ともいえる, これまでの研究から脱却し, これから外国語教育学が研究分野として発達していく上で, 他の研究者の行った研究結果を, 効果量を用いて比較し, メタ分析で使用していくような継続的な調査を行う必要がある。このためにも, 効果量の報告は必要不可欠になるといえる。本稿により, 外国語教育研究に携わる研究者諸氏が, 効果量報告の必要性を理解していただき, 今後の研究に役立てていただければ幸いである。

### 謝辞

執筆にあたり, 印南洋 (神田外語大学・日本学術振興会特別研究員), 小泉利恵 (常磐大学), 前田忠彦 (統計数理研究所) の各先生から貴重なご意見をいただきました。ここに記して感謝いたします。

### 注

- 1 前田 (2004, p. 66) では, 相関係数の検定 (無相関検定) を例にあげて, サンプルサイズが大きければ, .14 の相関係数でも有意であるという結果になることを指摘しており, 「相関係数の強さにどのような意味があるかを解釈する際には, (相関係数の検定は) ほとんど意味がない」と述べている。
- 2 村井 (2006, p. 116) は, 視聴率調査のように母数をできるだけ正確に推定する必要がある場合には, 「多ければ多いほうがよい」の原則があてはまるとしている。
- 3 あるいは研究者が自分の研究目的に応じて, 適切な値に定める。
- 4 先行研究がなければ, 予備調査の結果によって, おおよその効果量の目安をつけて, 本調査のサンプルサイズを決定するという方法も考えられる。
- 5 「従属変数における各独立変数の割合を示す効果量」ともいえる。
- 6 もちろん, 逆の数式変換も可能である。 $d$  (Cohen's  $d$ ) は効果量を計算してくれるサイト (<http://web.uccs.edu/lbecker/Psy590/escalc3.htm>) を使えば簡単に確認できる。
- 7 片方の変数が 2 値データである以外, 点双列相関係数とピアソンの積率相関係数はまったく同じのものであると考えてよい。
- 8 換言すると, 「独立変数の変動 (グループ間平方和) / 従属変数の全変動 (全平方和)」となる。
- 9 有効な使用例は Koizumi & Katagiri (2007) を参照。
- 10  $\eta^2$  は報告されているすべての値を合計しても 1.0 を超えることはないが, partial  $\eta^2$  の場合には合計で 1.0 を超えることがあるため, 報告の間違いを確認しやすい。一元配置分散分析のときには partial  $\eta^2 = \eta^2$  である。
- 11 有意差がないため, 専門誌などへの投稿・掲載が見送られた論文も含めてメタ分析すべきであるた

- め、このステップにおいては出版されていない論文や紀要論文、HP に掲載されている Working Paper のようなものも対象にして検索を行う必要がある。
- 12 増井 (2003) のようなメタ分析ソフトを使えば、サンプルサイズ(人数)、平均値、標準偏差、相関係数などを用いて、メタ分析を行うことができる場合もある。

## 参考文献

- American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: American Psychological Association.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.
- Cortina, J. M., & Nouri, H. (2000). *Effect size for ANOVA designs*. Thousand Oaks, CA: Sage Publications.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191. Retrieved September 21, 2007, from <http://www.psych.uni-duesseldorf.de/aap/projects/gpower/>
- Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage Publications.
- Field, A., & Hole, G. (2003). *How to design and report experiments*. London: Sage Publications.
- 南風原朝和. (2002). 『心理統計学の基礎－統合的理解のために』. 東京：有斐閣.
- Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Boston: Heinle & Heinle.
- 堀 啓造. (2006). 「分散分析 偏イータ 2 乗」『SPSS ときど記 (242)』. 2007 年 10 月 4 日検索. <http://www.ec.kagawa-u.ac.jp/~hori/spss/tokidoki24.html#242>
- Howitt, D., & Cramer, D. (2003). *An introduction to statistics in psychology* (Rev. 2nd ed.). Harlow: Prentice-Hall.
- 印南 洋・小泉利恵. (2007). 『統計分析ワークショップ：みんなでメタ分析を体験しよう』. 第 4 回言語テストインク・第二言語習得合同勉強会発表資料.
- 磯田貴道. (2004). 「テストの結果を比べる：3 クラス以上の場合」. 前田啓朗・山森光陽 (編) 磯田貴道・廣森友人 (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』 42-52. 東京：大修館書店.
- Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.
- 小池生男. (他 編). (2003). 『応用言語学事典』. 東京：研究社.
- Koizumi, R., & Katagiri, K. (2007). Changes in speaking performance of Japanese high school students: The case of an English course at a SELHi. *ARELE*, *18*, 81-90.
- Levine, R. T., & Hullett, R. C. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research*, *28*, 612-625.
- 前田啓朗. (2004). 「テストの得点間の関係の検討」. 前田啓朗・山森光陽 (編) 磯田貴道・廣森友人 (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』 64-72. 東京：大修館書店.
- Masgoret, A.-M., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: meta-analyses of studies by Gardner and associates. In Dörnyei, Z. (Ed.), *Attitudes, orientations and motivations in language learning* (pp. 167-210), (Refereed selection for The Best of Language Learning Series). Oxford: Blackwell.
- 増井健一. (2003). 『ここからはじめるメタ・アナリシス：Excel を使って簡単に』. 東京：真興交易(株)医書出版部.
- 村井潤一郎. (2006). 「サンプルサイズに関する一考察」 吉田寿夫 (編) 『心理学研究法の新しいかたち』 114-141. 東京：誠信書房.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417-528.
- 小野寺孝義・菱村 豊. (2005). 『文科系学生のための新統計学』. 京都：ナカニシヤ出版.
- 芝祐 順・南風原朝和. (1990). 『行動科学における統計解析法』. 東京：東京大学出版.
- Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics* (5th international ed.). Boston, MA: Pearson/Allyn & Bacon.
- 吉田寿夫. (1998). 『本当にわかりやすいすぐく大切なことが書いてあるごく初歩の統計の本』. 京都：北大路書房.
- Zhao, Y. (2003). Recent developments in technology and language learning: A literature review and meta-analysis. *CALICO Journal*, *21*, 7-27.