駒澤大学心理学論集，2013，第15号，7-13
Komazawa Annual Reports of Psychology, 2013, 15, 7-13

原　著

APA 論文作成マニュアルにおける論文執筆の要旨
―効果量と信頼区間―

深堀　友覚

A summary of writing an article based on Publication Manual of the American Psychological Association
―Effect size and confidence interval―

Tomonari Fukabori (*Department of Psychology, Komazawa University, Japan*)

ABSTRACT

Reports on effect size and confidence interval are required by Publication Manual of the American Psychological Association, 6th edition. But in Japan, they aren't required. Because of a lack of the guidance, this study reported effect size and confidence interval with the data of Fukabori (2010). It also referred of wrong use of effect size and confidence interval to use them properly.

KEY WORDS: APA style, effect size, confidence interval, null hypothesis statistical significance testing (NHST)

アメリカ心理学会（American psychological Association；以下 APA と略記する）は 2010 年に『APA 論文作成マニュアル第 6 版（Publication Manual of the American Psychological Association, 6th edition）』を発表し，日本でも前田・江藤・田中（2011）によって邦訳された。APA 論文作成マニュアルは，第 3 版が 1973 年に出版されて以降，おおむね 10 年間隔で改訂がなされている。APA（2010）は，第 6 版への改訂の要点を 2 つ示している。1 つは，インターネットをはじめとする様々な技術革新に学術研究がどのように影響されてきたかという点，つまり，テクノロジーの進歩による研究の構想や実態，そして研究の成果の利用方法の変化をどのようにガイドライン化するのかという点である。実態として，さまざまな学術領域にて専用のデータベースが生まれ，利用者はオンラインで論文の閲覧やデータの共有が行えるようになっている。このような情報社会の進展が今日に至るまでに急速になされている。APA（2010）は，研究で用いた素データのような，紙媒体での表現に多大なコストを要する資料は原則として最低 5 年間は投稿者が保存すること，そして他の研究者がデータの開示を求めた際には速やかに提出するよう規定している（p.12）。現代心理学の 1 つの潮流として研究者間に認識されつつあるメタ分析（meta-analysis）において，先行研究の素データは情報源として必須である。APA（2010）は心理学的研究のさらなる進展のためにメタ分析を推奨しており，付録として 247 ページから 253 ページにかけて APA Publication and Communication Board Working Group on Journal Article Reporting Standards（2008）で示された表を引用し，メタ分析の進め方を説明している。データベースの利用や素データの開示や利用をオンラインで行う研究者は今後増えていくであろうし，利用されるデータベースの幅もますます広がっていくことが予想されるので，オンライン上でのデータの共有に関して明確なガイドラインを設ける必要があったのである。

マニュアル改訂におけるもう 1 つの重要な変更点は，心理統計学における結果の表記の方法が従来の指針から大幅に変更されたことである。Wilkinson & APA Task Force on Statistical Inference（1999）は APA 論文作成マニュアルの改編に先んじて，データの分析には，(1)必要最小限の分析を選ぶこと，(2)区間推定値，つまり信頼区間（confidence interval; *CI*）を明記すること，(3) *p* 値を具体的数値で示すこと，さらに効果量（effect size）を明記することを示した。この影響を受けて出版された APA 論文作成マニュアル第 5 版（2001）においては，効果量と信頼区間を「報告がなされることが望ましい」という推奨要件と

して明記した。つまり，効果量や信頼区間を表記することを望ましいとしながらも，その表記方法を明確にせず，研究者自身でその表現を工夫するよう示したのである。こうした指針のあり方に対して，Sohn（2000）はこの改訂を新たな統計学的手法のガイドラインとしては明確性に欠くと批判し，さらにFidler, Thomason, Cumming, Finch & Leeman（2004）は医学界では1980年代より取り入れられている区間推定や効果量の報告が心理学の領域で不徹底であることについて，数量的な情報が正しく報告されなければ研究の真に重要な要素を伝えることはできないとして効果量や信頼区間の報告を強く求めた。こうした批判に応える形で，APA論文作成マニュアル第6版（2010）では，“効果量，信頼区間，そして必要最低限の記述といった報告をもって，結果の意味するところをはじめて完全に表現できる”（p.33）とし，さらに，“帰無仮説有意差検定（null hypothesis statistical significance testing；以下NHSTと略記する）を行う際には，検定対象となるすべての仮説はもとより，適切な効果量と信頼区間の完全なる報告がAPAの投稿規定においては最低限度の基準である”（p.33）と明確に義務づけ，特に信頼区間の表記には各括弧を用いることなど，具体的な説明を付記した。

我が国における論文の執筆基準は，日本心理学会（2005）における『執筆・投稿の手引き』であるが，効果量や信頼区間についての報告については一切示されていない。我が国におけるNHSTでは検定統計量の値と自由度，そして危険率および棄却率しか報告されていない場合がほとんどである（鈴川・豊田，2012）ことからも，効果量や信頼区間についてはできる限り論文の中で報告されることが望ましいといえる。APA（2010）はそれ以前の統計学的情報が論文に不足していたことをふまえ，“統計学的手法は，仮説を正確に説明するための頑健性（robustness）が包含されることで分析の主旨が支持され，また，データに対してはあいまいではなく明確な洞察を統計学的手法によって提供しなければならない”（p.33）としている。つまり，効果量や信頼区間はデータの分析においてはその頑健性を増すための重要な要素であるということを示している。頑健性が増補されることにより，有意であるか否かの判断を行うNHSTにおいて，悉無率的な枠にとらわれることのない解釈を展開することが期待できる。解釈の可能性を広げることは，データを正しく考える上で非常に重要であり，また，NHSTにて有意差が認められないという結果においても効果量や信頼区間が報告されることによって新たな見解がもたらされる可能性を研究者がに考えられる。

そこで本論文では，日本ではなじみのないNHSTにおける効果量と信頼区間について，それぞれの意義，計算法，そして表記の仕方を実際の素データを元に簡易的に説明する。なお，本稿で説明されない他の効果量と信頼区間に関する詳しい説明や計算方法については大久保・岡田（2012）に詳しい。

方　　法

深堀（2010）における系列予言法（暗記学習）と調整法（ミューラー・リエル錯視）の素データをそのまま使用した。実験器具，実験参加者，手続きについては深堀（2010）を参照して頂きたい。

効果量

効果量とは効果の大きさのことであり，実験的操作の効果や変数間の関係の強さを表す指標である（Field & Hole, 2003）。芝・南風原（1990）は，効果量を測定単位に依存しない指標とし，単位の異なる変数を用いた研究の間でも実験効果の大きさを互いに比較することができるものと説明している。例えば回帰分析では相関係数の二乗（R^2）が決定係数（coefficient of determination）として用いられるが，この決定係数は回帰式においての説明力を示したものなので効果量とみなすことができる。平均値差の比較においては，決定係数のような広く認識されている効果量は存在しないので，平均値差の効果量について簡単に説明を行う。

t検定や分散分析など2群以上の平均値を比較する際には，効果量を求めるにあたり分散の等質性（homogeneity of variance）が満たされていることを前提とし，

$$\delta = \frac{\mu_1 - \mu_2}{\sigma}$$

と表す。ただし，σは母集団における標準偏差なので計算を行う際には標本統計量から母標準偏差を推定して用いることになる。一方，μは母平均値を表しているので標本平均値をそのまま用いる。

独立2群の効果量を表した代表的なものとして

はCohen's d（Cohen, 1988）があげられる。Cohen's d は，標本の平均値差を，標本からプールした標準偏差で除した値である。つまり，

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

となる（水本・竹内，2010）。この計算から得られる値は群ごとの平均値の差を標準化したものである。Cohen's d における判断の目安として，$d=.20$を効果量小，$d=.50$ を効果量中，$d=.80$ を効果量大とする。対応のある2群の効果量はCohen's d の計算式をそのまま用いるか，あるいは2群の相関係数を利用した計算式を用いる。相関係数を分母で用い，Cohen's d を分子とすると計算式は，

$$d_{Diff}=\frac{\text{対応なしの場合の } d}{\sqrt{2(1-\text{対応のあるデータの相関係数})}}$$

となる（水本・竹内，2010）。

分散分析における効果量は，要因平方和と全体平方和における比と考えるとわかりやすい。分散分析における効果量を $\hat{\eta}^2$ とすると計算式は，

$$\hat{\eta}^2=\frac{SS_A}{SS_T}$$

となる。この式において SS_Tは全平方和，SS_Aは要因平方和をそれぞれ示している。この式の意味するところは，要因平方和が，誤差平方和（SS_E）を含む全平方和においてどの程度の効果をもっているかを数値化したものといえる。ただし，この計算式は単純な一元配置分散分析でのみ適用可能であり，反復測定や二元配置以上の分散分析では若干モデルが複雑になる。繰り返しのある一元配置分散分析を例にすると，求められる全平方和は，要因平方和，誤差平方和，そして被験者内平方和（SS_S）の3つとなる。ここで要因平方和と全体平方和における比を想定して効果量を計算すると，研究の効果を説明する上で不要な平方和，つまり被験者内平方和が計算式に加味されてしまい，不要な値が効果量に反映されるという問題が生じる。したがって，反復測定分散分析においては，被験者内平方和を除外した計算式に準じて効果量を求めることになる。繰り返しのある一元配置分散分析における最終的な効果量の計算式は，

$$\hat{\omega}_p^2=\frac{df_A(MS_A-MS_E)}{df_A MS_A+(n-df_A)MS_E}$$

となる（大久保・岡田，2012）。この式において，MS_Aは要因平均平方，MS_Eは誤差平均平方，n はサンプル数を示す。式中の添え字にある p(partial)は，全体の要因から不要な要因を除いたことを意味している。なお大久保・岡田（2012）は，統計的指標の1つとして重要な意味をもつ効果量を他の研究者が確認できるようにするために，通常の論文では表記されることのない平方和や平均平方に関する情報，つまり分散分析表を示すことを推奨している。

信頼区間

母集団の推定値として通常求められる平均値は点推定（point estimation）によるものである。点推定とは，母数を1点の推定値で表現する方法である。サンプルサイズがある程度の大きさであれば点推定による平均値の誤差に大きな変動はないが，サンプルサイズが小さい場合，点推定による推定値では誤差変動が大きくなってしまう。点推定とは異なり，ある程度の区間を設けて母数の推定値を一定の範囲によって表す方法を区間推定（interval estimation）という。区間推定では，あらかじめ定められた確率で母数を含む区間を設定しなければならないが，一般的には95%がよく用いられる。よって，この区間推定値は95%信頼区間(95% CI)とよばれる。信頼区間は推定値の報告に付記することで極めて有効な情報をもつ（大久保，2012）。信頼区間に95%という数値が使われるのは，5%という慣例的な水準がそのまま踏襲されたものにすぎないので厳密さが要求されるような研究では99%信頼区間（99% CI）も用いられるが，区間の幅は95%よりも狭くなるため，その区間に母数の値が含まれていない可能性が高まるので扱いには注意が必要である。

信頼区間は，t 分布表の t の臨界値と標準誤差（standard error）を乗じたものに平均値を加減してその上限値と下限値を算出する。計算式は，

$$CI=M\pm\frac{s}{\sqrt{N}}\times t$$

となる。この式において，M は平均値，s は標準偏差，N はサンプルサイズをそれぞれ示している。例えば信頼区間を95%に据える場合には95%の t の臨界値（両側2.5%）を乗じることで求めることができる。なお，信頼区間を論文中に表記する際は角括弧を用い，グラフには信頼区間を示すエラーバーを併記することがAPA (2010) にて定められている。

結　　果

Table 1は，系列予言法の素データにおける平均値，標準偏差，標準誤差を示したものである。集中試行（massed practice）群における信頼区間は95% *CI* [9.33 16.23]，分散試行（distributed practice）群における信頼区間は95% *CI* [13.38 21.95] であった（Figure 1）。両群の分布の正規性が認められたことを確認し，両群の平均値差を検討するために t 検定を施行したところ $t=2.01$（$p=.06$），効果量は $d=.80$ であった。

Table 1　The number of times to complete memories with each method.

	massed practice	distributed practice
	16	12
	13	12
	19	14
	9	25
	11	21
	10	10
	20	21
	10	24
	7	20
mean	12.78	17.67
SD	4.58	5.68
SE	1.53	1.89

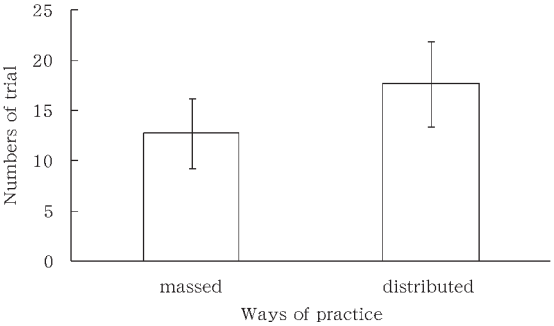

Figure 1. Means and confidence intervals on each method about the serial effect.

Table 2は，調整法の素データにおける各図形の平均値，標準偏差，標準誤差を示したものである。それぞれの図形における信頼区間は，15 mm 30°では95% *CI* [13.57 22.80]，30 mm 30°では95% *CI* [16.83 30.47]，35 mm 30°では95% *CI* [30.17 21.01]，45 mm 30°では95% *CI* [16.36 28.92]，15° 30 mmでは95% *CI* [21.28 31.09]，60°30 mmでは95% *CI* [14.34 22.19] であった（Figure 2）。6つの図形の錯視量の違いを検討するために反復測定による一元配置分散分析を施行したところ，30°に角度を固定した斜線の長さの違いにおける主効果は $F(3, 21)=7.36$，$p=.001$，$\hat{\omega}_p^2=0.70$ であった。Bonferroni法による多重比較の結果，15 mm 30°－30 mm 30°間，15 mm 30°－35 mm 30°間，15 mm 30°－45 mm 30°間，30 mm 30°－35 mm 30°間，30 mm 30°－30 mm 45°間，35 mm 30°－45 mm 30°間の平均値差と効果量はそれぞれ，5.46（$p=.114$，$d=0.76$，$d_{Diff}=1.95$），7.40（$p=.004$，$d=1.31$，$d_{Diff}=3.26$），4.45（$p=.224$，$d=0.66$，$d_{Diff}=1.44$），1.94（$p=1.00$，$d=0.27$，$d_{Diff}=0.54$），1.01（$p=1.00$，$d=0.13$，$d_{Diff}=0.34$），2.95（$p=.182$，$d=0.44$，$d_{Diff}=3.65$）であった。一方，30 mmに斜線の長さを固定した斜線の角度の違いにおける主効果は $F(2, 14)=9.15$，$p=.003$，$\hat{\omega}_p^2=0.67$ であった。Bonferroni法による多重比較の結果，15° 30 mm－30° 30 mm間，15° 30 mm－60° 30 mm間，30° 30 mm－60° 30 mm間の平均値差と効果量はそれぞれ，2.54（$p=.90$，$d=0.35$，$d_{Diff}=0.49$），7.93（$p=.004$，$d=1.45$，$d_{Diff}=2.38$），5.39（$p=.059$，$d=0.79$，$d_{Diff}=2.45$）であった。

Table 2　The quantities of optical illusion in each stimulus.

	30°				30 mm	
	15 mm	30 mm	35 mm	45 mm	15°	60°
	20.3	28.0	29.3	32.0	31.5	22.1
	15.6	15.6	22.4	17.9	26.9	14.4
	12.5	15.1	15.9	9.5	17.5	14.5
	13.3	11.4	23.6	18.0	21.9	12.4
	14.9	26.3	22.0	18.9	22.1	14.0
	22.3	29.1	33.1	28.1	33.8	23.5
	29.5	34.3	30.6	29.4	32.9	21.9
	17.1	29.4	27.8	27.3	22.9	23.3
mean	18.19	23.65	25.59	22.64	26.19	18.26
SD	5.66	8.37	5.62	7.70	6.01	4.82
SE	2.00	2.96	1.97	2.72	2.13	1.70

Table 3 An ANOVA table about the optical illusions fixed in the degree (30°).

	SS	*df*	*MS*	*F*	*p*
Factor A	235.77	3	78.59	7.36	0.001
Subject	1126.81	7			
Error	224.11	21	10.67		
total	1586.69	31			

Table 4 An ANOVA table about the optical illusions fixed in the length (30mm).

	SS	*df*	*MS*	*F*	*p*
Factor A	262.05	2	131.03	9.15	0.003
Subject	705.03	7			
Error	200.56	14	14.33		
total	1167.64	23			

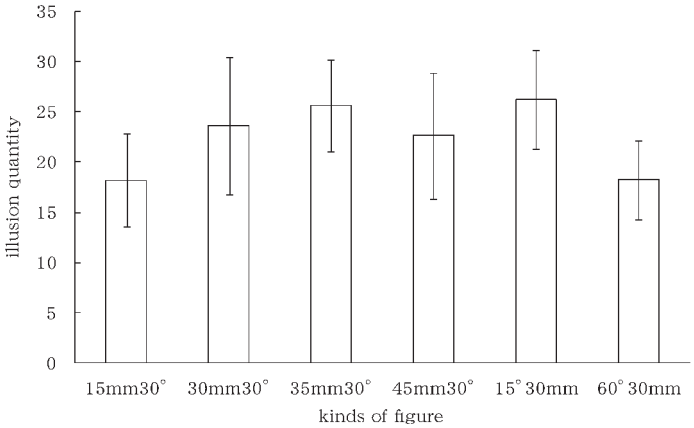

Figure 2. Means and confidence intervals on each method about the optical illusions.

考　　察

系列予言法における両群の平均値差に有意差は認められなかった($p<.05$)ものの，効果量が0.80と非常に高いことから，両群における学習の進度については集中試行群の方が速かったものと推察される。集中学習と分散学習では分散学習の方が一般的に学習の質は高い（野村，1974）が，深堀（2010）の通り，この実験で行った分散学習ではリハーサルを妨害するための計算課題が導入されたために学習達成に時間を多く費やしたものと解釈できる。有意差こそ認められないが，高い効果量が得られたことから両群における学習の質が異なること，つまり，この実験で用いられた項目の数，性質，さらに経過時間においては集中試行の方がその学習効果が高かったといえる。調整法では，角度あるいは長さの固定された錯視図形のいずれの錯視量において主効果が有意であり，効果量も 0.67 以上と高い値であった。多重比較では，有意差の認められた図形間においても高い効果量が得られた。特に斜線角度の差異が大きくなるほど錯視量が有意に小さくなることが，高い効果量をもって示された。ただし，有意差と効果量はまったく別の統計学的指標としてとらえなければならないことは既述した。たとえば，この実験において 35 mm 30°－45 mm 30°間 で は $p=.182$，$d=0.44$で，両要因間の効果量は中程度でみなされるが，対応のある 2 群の効果量では $d_{Diff}=3.65$ と高い数値を示した。一見すると高い効果量を示しているために両要因間の錯視量に何らかの効果があるものと解釈されるかもしれないが，この大きな数値の変動は両要因間の相関係数が.94 という非常に高い値であったことが影響しているためで，実験の目的として対応（repetition）が検討されている場合，たとえば実験群と統制群による反復測定のような条件でなければこの数値はまったく意味をもたない。Table 5 は，本実験における図形間の平均値差，対応のない効果量，相関係数，対応のある効果量を示したものであるが，相関係数が高くなるほど対応のある効果量は大きくなっていることがわかる。豊田（2009）によれば，対応のない場合の d と対応のある場合の d の違いは，その比較対象が要因間にあるのか，あるいは個人内にあるのかに起因するという。本研究では，提示順の順序効果を相殺することを目的とし，個人内変動をみることを目的としていないので，効果量

に関する情報は d_{Diff} ではなく d をみるべきである。実験条件に対応がある場合，検討事項の内容によって d と d_{Diff} は使い分けた方がよい。効果量は，その使用方法に関する報告を怠ってしまうと読者に誤った解釈をもたらすことにもなるので，効果量を報告する際には文献の明示，できれば計算式も提示することが望ましい（水本・竹内，2010）。

統計的検定は，有意水準，サンプルサイズ，効果量，検定力（power）の4つを用いることで求めることができる（水本・竹内，2010）。このことは，3つの要素が決まれば残り1つの要素も必然的に決まることを同時に意味している。例えば，Faul, Erdfelder, Lang, & Bunchner（2007）によるフリーソフトのG*Power（www.psycho.uni-duesseldorf.de/aap/projects/gpower/）を用いれば，有意水準，効果量，検定力を入力することで適切なサンプルサイズを自動計算するよう作られている。サンプルサイズが不当に大きい場合には第1種の過誤の確率が下がる一方で第2種の過誤（Type II error）を犯す確率が上がり，逆にサンプルサイズが過小であれば第1種の過誤を犯す確率を上げてしまう。このようなリスクを回避するためにも，適切なサンプルサイズを用いた実験計画が自ずと必要になり，そのためにも，効果量や信頼区間に対する正しい知識や使用法を研究者が個々に理解する必要がある。

NHSTでは，得られた標本が母集団を代表するものと仮定することから始まる。標本間（あるいは内）に差がないとする帰無仮説の生起確率が低いことを，あらかじめ研究者が設定した有意水準を用いて証明することによって，帰無仮説と対立する実験仮説の生起頻度を判断するのがNHSTの基本的な進め方である。研究者が設定する有意水準は主として5％，1％，0.1％が用いられるが，危険率が低い値で棄却されたということと，その両者の間の差に大きな効果があるものと解釈することは同義ではない。そもそも有意確率は，第1種の過誤（Type I error）を犯す可能性が低いということを意味するだけであって，仮に0.1％水準にて危険率が棄却されたとしても，それは両者の間における差の効果が大きいことを意味するものではなく，本当は帰無仮説が正しいのに誤って棄却してしまう確率が0.1％に満たないということだけしか意味しないのである。Table 6はデータ数の違いによって同じ平均値，標準偏差，効果量においても有意差が認められるケースと認められないケースを示したものであるが，危険率はサンプルサイズの影響を強く受けることが分かる（水本・竹内，2010）。したがって，サンプルサイズに影響されない効果量が大きな役割を果たすことになる。効果量は両者の差異をよりよく表現するための1つの目安ととらえ，積極的に使用されることが望まれる。ただし，現在の我が国の投稿論文では，例えば『基礎心理学研究』の1982年から2008年までに掲載された計199本の論文において，誤差範囲を示した論文は全体の47.24％であったが信頼区間の報告はわずか1件のみである

Table 5　The difference between normal d and revised d using correlation coefficient.

	30°						30 mm		
	15 mm −30 mm	15 mm −35 mm	15 mm −45 mm	30 mm −35 mm	30 mm −45 mm	35 mm −45 mm	15° −30°	15° −60°	30° −60°
d	0.76	1.31	0.66	0.27	0.13	0.44	0.35	1.45	0.79
r	0.80	0.80	0.77	0.75	0.81	0.94	0.65	0.69	0.84
d_{Diff}	1.95	3.26	1.44	0.54	0.34	3.65	0.49	2.38	2.45

Table 6　The comparison of p-value between two different samples.

データセット	グループ	n	平均値	標準偏差	平均値差	t検定	効果量 d
A	グループ1	50	40	10	3	$p=.137$	$d=0.3$
	グループ2	50	43	10			
B	グループ1	100	40	10	3	$p=.015$	$d=0.3$
	グループ2	100	43	10			

Note. This table is cited in Mizumoto & Takeuchi (2010).

こと，そしてほぼすべての論文で p 値が報告されている一方で効果量の報告は年代に関わらず少数でその割合は10.58%であった（大久保，2008）。ここで用いられた効果量についても，相関係数，回帰分析など効果量自体が考察の中心的対象となる分析方法を用いており，分散分析の効果量 ω^2 の値がわずかに1件だけ報告されたにすぎないという。このことからも，我が国の研究者に効果量および信頼区間に対する認識や知識が不足していることは否めないと考えざるを得ない。今後研究者が効果量と信頼区間の重要性を認識することで，統計処理の結果は自然とその表記が詳しくなり報告件数も増えてくるであろう。事実，小野寺・菱村（2005）も，いずれ数年の内には効果量を無視した研究報告はできないとしている。今後，心理学に携わる研究者は，従来のNHSTでは十分な情報が伝えられていないということを理解しなければならない。心理統計学に対する意識が変わり研究成果の報告に効果量や信頼区間が利用されることで，本邦での効果量や信頼区間の利用可能性や改善点もより明確になるであろう。

引用文献

American Psychological Association (2001). *Publication manual of the American Psychological Association*. 5th ed. Washington, D.C.: American Psychological Association.

American Psychological Association (2010). *Publication manual of the American Psychological Association*. 6th ed. Washington, D.C.: American Psychological Association.（前田樹海・江藤裕之・田中建彦(訳)（2011）．APA論文作成マニュアル第2版　医学書院）

APA Publication and Communication Board Working Group on Journal Article Reporting Standards (2008). Reporting standards for research in Psychology: Why do we need them? What might they be? *American Psychologist*, **63**, 848-849.

Cohen, J. (1988). *Statistical power analysis for the behavior science*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Faul, F., Erdfelder, E., Lang, A. G., & Bunchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical science. *Behavior Research Methods*, **39**, 175-191.

Fidler, F., Thomason, N., Cumming, G., Finch, S., & Leeman, J. (2004). Editors can lead researchers to confidence intervals, but can't make them think: Statistical reform lessons from medicine. *Psychological Science*, **15**, 119-126.

Field, A., & Hole, G. J. (2003). *How to design and report experiments*. London: Sage Publications.

深堀友覚（2010）．確率化テストを用いた基礎的実験の新たな解釈　駒澤大学心理学論集，**12**，1-6．

水本篤・竹内理（2010）．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―　外国語教育メディア学会関西支部メソドロジー研究部会報告論集，pp.47-73．

日本心理学会（2005）．執筆・投稿の手引き　社団法人日本心理学会

野村幸正（1974）．短期記憶課題の集中学習と分散学習の研究―分散効果と干渉理論―　心理学研究，**45**，130-139．

大久保街亜（2008）．日本における統計改革―基礎心理学研究を資料として―　基礎心理学研究，**28**，88-93．

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

小野寺正義・菱村豊（2005）．文科系学生のための新統計学　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　東京大学出版会

Sohn, D. (2000). Significance testing and the science. *American Psychologist*, **55**, 964-965.

鈴川由美・豊田秀樹（2012）．“心理学研究”における効果量・検定力・必要標本数の展望的事例分析　心理学研究，**83**，51-63．

豊田秀樹（2009）．検定力分析入門―Rで学ぶ最新データ解析―　東京図書

Wilkinson, L., & APA Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, **54**, 594-604.