A Study of Effect Sizes in Marketing Experiments

Roger Marshall, Nanyang University

Loi Soh Loi, Nanyang University

Na WoonBong, Kyunghee University

ABSTRACT - An investigation is undertaken to empirically document the reported effect sizes of experiments reported in four leading marketing journals from the period 1985 to 1995, following a previous study concerned with the prior decade (Peterson, Albaum & Beltramini, 1985). Using omega-squared (w2), Pearson's coefficient (r) and Cohen's index (d) as measures of effect size, the results of 269 experiments, yielding a total of 1,399 experimental effects, were analyzed for the proportion of variance explained. Average incidence of reported w2 values has increased from 14% to 25% of experiments since the prior decade, but are still somewhat low.

A Study of Effect Sizes in Marketing Experiments

Roger Marshall, Nanyang University
Loi Soh Loi, Nanyang University
Na WoonBong, Kyunghee University

ABSTRACT

An investigation is undertaken to empirically document the reported effect sizes of experiments reported in four leading marketing journals from the period 1985 to 1995, following a previous study concerned with the prior decade (Peterson, Albaum & Beltramini, 1985). Using omega-squared (ω^2), Pearson's coefficient (r) and Cohen's index (d) as measures of effect size, the results of 269 experiments, yielding a total of 1,399 experimental effects, were analyzed for the proportion of variance explained. Average incidence of reported ω^2 values has increased from 14% to 25% of experiments since the prior decade, but are still somewhat low.

Purpose of the study

In 1985 Peterson, Albaum and Beltramini conducted a meta-analysis of the effect sizes reported in a number of leading consumer behavior and marketing journals over the previous ten years. This paper reports a similar meta-analysis for the following decade, 1985-1995, allowing comparison of results of the two studies.

The major area of comparison is that of finding if there has been an increase in the proportion of experiments in the relevant journals that report effect size estimates, as measured by omega-squared (ω^2) values. There is also a comparison made between the mean ω^2 values in the two studies; that is, the average percentage of the variance in a response variable that was explained or accounted for by the statistically significant effect reported. In addition, the relationship between several different measures of effect size is also examined, as there is a literature to suggest a consistent difference between such measures should be expected (Oakes, 1986).

This paper, although not identical, is sufficiently similar to the Peterson *et al* study (hereafter called PAB) to allow some meaningful comparisons to be made. Less emphasis is placed, however, on the characteristics of the selected experiments (such as types of effects, experiments, subjects, and response variables), and more emphasis is placed on the comparison of the various measures of effect size, namely omega squared (ω^2), Pearson's coefficient (r) and Cohen's index (d).

RELEVANT LITERATURE

This brief review of the relevant literature touches first upon the importance of effect size to empirical research studies and how it is measured, then on the validity of meta-analyses. Finally, the PAB study –which provided the inspiration for the present work– is inspected in more detail.

Effect size and tests of significance

Effect size is defined as the strength of a relationship or the magnitude of a difference between variables (PAB). It is also referred to as the degree to which the phenomenon is present in the population or the degree to which the null hypothesis is false (Cohen, 1977). Effect sizes can be used both to facilitate the interpretation of research findings and to determine their practical implications (Cooper, 1981). The reporting of effect size is argued to be beneficial, especially so when reported as a supplement to significance tests.

The reason why it is beneficial to report effect size with tests of significance is because tests of significance alone do not accurately indicate of the magnitude of an effect. This is mainly because tests of significance are not independent of sample size. In fact, it would be a serious case of abuse to even attempt to derive the effect size of rejected null hypotheses purely by analysis of the significance level. The difference between test statistics from different manipulations is by no means an indicator of the extent of the variance between the effect sizes of these manipulations. Indeed, the question becomes not so much if there is a relationship between data in a set, as much as how big the sample must be before a significant relationship is found. Research in various fields has been criticised as being over-reliant on significance testing at the expense of effect sizes, possibly due to the difficulties in quantifying effect size (Irvine, Miles & Evans, 1979).

There are clearly situations when a large effect size is not important. In a practical sense even a tiny effect size may be thought important when multi-million dollar sales is the dependent variable. Even in some academic work it could be argued that it is enough to show the existence of an effect, not withstanding the size of it. This argument is hard to sustain, though, as even though a tiny effect might be considered "enough" by an author, surely the reader ought to have the prerogative of making that judgement themselves in the light of the sample statistics? The position taken here, then, supports the view of Peterson *et al*, that it is that it is generally better to include an effect size where the data is available as this provides information which should be revealed to the reader.

The need for a common measure of effect size

Various research models have been designed and various test statistics developed to measure effect size in experimental work, such as t, F or χ^2. In order to apply meta-analytic procedures to these different variables, it is necessary to convert all the various summary statistics into a common measure in order to allow meaningful comparison across the studies.

The most widely used index of effect size is Omega squared (ω^2), which was popularised by Hays (1963) and expounded on by Cohen (1977). Using ω^2, Cohen provided rough guidelines for categorising small (ω^2=0.01), medium (ω^2=0.06) and large (ω^2=0.15) effect sizes. This was in an attempt to propose conventional effect sizes to help the making of informed decisions about effect size in socio-behavioural research. Two of the most common effect size measures selected, for the purpose of conversion from the various test statistics, are Pearson's Correlation (r) and Cohen's effect size index d. Guidelines for converting some of the summary statistics to r and d (Wolf 1986) are detailed in the Appendix. Cohen (1977) and Glass, McGaw & Smith (1981) provide guidelines for transforming less common statistics.

It is important to note at this point that while there are various measures of effect size, only ω^2, r and d will be used in the study. It is also interesting to note that ω^2 is consistently 20-25% lower than r^2 (Oakes, 1986). The present study will, in passing, verify this observation.

Meta-analysis

Meta-analysis is an analysis of analyses (Glass 1981). It is a quantitative accumulation and analysis of various test statistics across studies, made without accessing the original study data. Rust, Lehmann and Farley (1990) challenged the assumption that, in meta-analysis, the sample of studies is a fair representation of all the work done in the relevant field. In their study, they proposed that

TABLE 1
Distribution of qualifying articles in various journals

Journal / Proceedings	(%)
Journal of Consumer Research	7.1
Journal of Marketing Research	23.3
Journal of Marketing	4.3
Journal of Advertising Research	4.3
Journal of Applied Psychology	7.8
Association for Consumer Research Proceedings	43.6
American Marketing Association Proceedings	9.6

Source: "A Meta-Analysis of Effect Sizes in Consumer Behavior Experiments". *Journal of Consumer Research*, 12(June 1985).

since journals tend to publish only statistically significant results, there could exist some publication bias, which they attempted to estimate. Similarly, Hedges and Olkin (1985) assume that since all statistically non-significant results have been omitted from the analysis, meta-analysis is not a justifiable method. However, this seems a rather unrealistic an assumption. Moreover, results of the 1990 study undertaken by Rust *et al* show that the estimated publication bias in meta-analyses is relatively small.

Previous research, the Peterson, Albaum and Beltramini (1985) study

The PAB study was conducted with the primary intention of empirically documenting the effect sizes in consumer behaviour research experiments reported in the marketing literature. In their study, articles published in seven major marketing journals or proceedings, from 1970 to 1982, were examined. Included in their sample were consumer behaviour experiments that employed at least one treatment manipulation. Furthermore, since ω^2 was selected as their measure of effect size, experiments selected also had to either report ω^2, or contain sufficient information to permit its computation. A total of 311 articles or papers were identified in the six journals or conference proceedings as indicated in Table 1.

Of the 311 articles or papers, 115 of them contained 118 experiments that either reported ω^2 or contained sufficient information to permit its computation (only 14.4% were reported). Their study involved 1,036 effects, of which 475 were statistically significant at 0.05 or less. The scope of their study was extended to include the analysis of other variables that are potentially related to effect size, such as the type of experiment, the number and type of subjects and factors.

The four characteristics that were found to be significantly ($p \leq 0.01$) correlated with ω^2 are type of effect (main or interaction), type of experiment (field or laboratory), type of subjects and type of response variable (behavioral or others). These four characteristics were subjected to an analysis of variance in order to examine their relationship to effect size. The results of the analysis showed that the type of experiment and type of effect can be expected to affect the effect size of the experiments. For example, field experiments can be expected to report larger effect sizes than laboratory experiments; main effects are associated with higher effect sizes than interaction effects. The type of the experimental subjects is also related to the type of response variable. For example, non-college students can be expected to have a higher probability of increased response homogeneity than college students.

RESEARCH METHOD

The same general approach used in PAB is also used here, to conduct a type of meta-analysis of effect sizes of consumer behavior experiments published in selected, leading marketing journals. The period of interest covers the years from 1985 to 1995.

The sample

Four leading marketing journals were chosen in this study; the *Journal of Marketing* (JM), the *Journal of Marketing Research* (JMR), the *Journal of Advertising Research* (JAR), and the *Journal of Consumer Research* (JCR). These were among the journals examined in the PAB study. It was determined not to follow the lead of PAB, however, and include a psychology journal or conference proceedings, but stick to the main-line marketing publications.

Articles were included in the study based on the following guidelines. First, the article must involve a consumer behaviour experiment. Second, the article must either report ω^2 values or, when ω^2 is not reported, have sufficient statistical information so that ω^2, r or d can be computed. Third, only results of main effects of experiments are to be considered in cases where ω^2 values have to be computed (this was due to resource limitation rather than to any academic reason). Finally, only test statistics from chi-squared-tests (χ^2), t-tests and F-tests were included.

185 suitable articles from the four marketing journals were identified using these criteria. These articles contained 269 consumer behaviour experiments altogether, and from these experiments a total of 1,399 unique sets of statistical values of experimental effects were gleaned. The *Journal of Marketing* provided 16 articles (8.7%) and 16 experiments (5.6%); the *Journal of Marketing Research* 44 articles (23.8%) and 59 experiments (22%); the *Journal of Advertising Research* 14 articles (7.6%) and 15 experiments (5.6%); and the *Journal of Consumer Research* 111 articles (60%) and 179 experiments (66.5%).

Methods to compute effect sizes

As the aim was to examine effect sizes as measured by ω^2, r or d, these values had to be computed, using statistical information obtained from the articles themselves, for experiments that did not report them. Three of the most common measures of effect size have been chosen as effect size estimates of the experimental results–ω^2, r and d values. Where sufficient information is provided, the chi-squared statistic (χ^2), the t-statistic or the F-statistic can be converted into the above-mentioned effect size measures via various formulae.

TABLE 2

Distribution between journals of articles and experiments that do and do not report ω^2

Journal	Articles Reporting ω^2		Articles not reporting ω^2		Experiments reporting ω^2		Experiments not reporting ω^2	
	No.	%	(No.)	(%)	(No.)	(%)	(No.)	(%)
JM	3	7.7	13	8.9	3	4.6	13	6.4
JMR	2	5.1	12	8.2	2	3.0	13	6.4
JAR	9	23.1	35	24.0	15	22.7	44	21.7
JCR	25	64.1	86	58.9	46	69.7	133	65.5
Total	39	100.0	146	100.0	66	100.0	203	100.0
% across	21.1		78.9		24.5		75.5	

TABLE 3

Distribution of experimental effects by reported levels of significance

Journal	Reported Significant Level of Effect					
	<0.0001	0.002-0.010	0.011-0.050	0.051-0.100	>0.1	Total
Journal of Marketing	31	33	29	17	2	112
Journal of Marketing Research	101	87	109	26	2	325
Journal of Advertising Research	36	18	76	0	0	130
Journal of Consumer Research	204	262	295	58	13	832
Total	372	400	509	101	17	1399

RESULTS

Distribution of the incidence and level of effect reporting between journals

Incidence of reported effects

Table 2 contains the distribution of reported effects, of both articles and experiments, between the journals in the sample. Thus, overall, 24.5% of the significance levels reported were accompanied by an effect size. Table 3 presents the distribution of the experimental effects in the various journals according to their reported levels of significance, whilst Table 4 shows the distribution of reported ω^2 values.

It can be seen from Table 4 that the majority (60.8%) of the reported ω^2 values fall between 0.01 and 0.09. A little more than 5% of the reported ω^2 values are less than 0.01, which, according to Cohen, renders them of no real effect. Large ω^2 values, of not less than 0.3, constitute a relatively small percentage. Following Cohen's suggested categorization of ω^2 values into 'small', 'medium' and 'large' effects, the distribution of the ω^2 values can be seen in Table 5.

It can also be seen that the 'small effects' category forms the majority (40.63%) of the ω^2 values reported, with 5% of these values having no experimental effect of importance. The average of these 352 ω^2 values is 0.102, with a standard deviation of 0.109.

Estimating ω^2 values from articles and experiments that do not report ω^2

There were 1,047 experimental effects for which ω^2 were not reported. Table 6 shows the distribution of these experimental effects among the three types of statistical tests. Of the 1,047 experimental effects which did not report ω^2, there was only sufficient information to compute397 values of ω^2. This amounts to only 37.92% of the reported experimental effects. The distribution of these estimated ω^2 values, computed from test statistics t and F between, first, the journals and then the test statistics, are shown in Tables 6 and 7. Note that Chi-squared tests were not included because no formula to convert the χ^2 statistic into ω^2 values could be found.

It is useful to consider the ω^2statistics for this group of effects according to Cohen's suggested categorization of effect size, as previously. This is shown in Table 9, which (as in the previous case) shows the majority of unreported effect size values fall into the category of 'small effects'. The mean of the estimated ω^2 value is 0.11 and the standard deviation is 0.13. The aggregate mean of all the effect sizes as measured by both reported and estimated ω^2 values is also 0.11.

Distribution of experimental effects measured by Pearson's r

The second measure of effect size used was Pearson's r. It was found that 3.79% of the effect size values based on Pearson's r are less than 0.1 and so can be considered as having no effect at all. 49.61% of the effect size values fall under the range of 'small effect,' while 13.81% belong to the 'large' effect category. Distributions between journals and by effect size are shown in Tables 10 and 11.

Distribution of experimental effects measured by Cohen's d

The third–and final–measure of effect size used was Cohen's d. The distribution of effect size values, as measured by d and shown

TABLE 4
Distribution of ω^2 values according to Cohen's categorization

Experimental Effect	Range of ω^2 values	Number of ω^2 values in range
None	$\omega^2 < 0.01$*	19 (5.40%)
Small	$0.01 < \omega^2 < 0.06$	143 (40.63%)
Medium	$0.06 < \omega^2 < 0.15$	111 (31.53%)
Large	$\omega^2 > 0.15$	79 (22.44%)
Total		352 (100.00%)

*The 'no effect' category is assumed by default.

TABLE 5
Distribution of reported ω^2 for experimental effects

	Reported ω^2 Values						
Journal	<0.009	0.01-0.09	0.10-0.19	0.20-0.29	0.30-0.39	>0.40	Total
Journal of Marketing	3	7	1	1	0	1	13
Journal of Marketing Research	0	55	13	7	1	0	76
Journal of Advertising Research	7	16	8	4	4	5	44
Journal of Consumer Research	9	136	42	16	12	4	219
Total (%)	19 (5.40)	214 (60.8)	64 (18.18)	28 (7.95)	17 (4.83)	10 (2.84)	352 (100)

TABLE 6
Distribution of experimental effects which did not report ω^2

	Types of Test			
Journal	By χ^2–test	By *t*-test	By *F*-test	Total
Journal of Marketing	2	41	56	99
Journal of Marketing Research	17	109	123	249
Journal of Advertising Research	11	15	60	86
Journal of Consumer Research	34	155	424	613
Total	64	320	663	1047

TABLE 7
Distribution of estimated ω^2 between journals

	Estimated ω^2 Values						
Journal	0.009 or less	0.01-0.09	0.10-0.19	0.20-0.29	0.30-0.39	0.40+	Total
Journal of Marketing	10	40	1	0	0	1	52
Journal of Marketing Research	7	62	33	12	7	2	123
Journal of Advertising Research	5	15	5	2	1	0	28
Journal of Consumer Research	6	106	49	19	10	11	201
Total (%)	28 (6.93)	223 (55.2)	88 (21.78)	33 (8.17)	18 (4.46)	14 (3.47)	404 (100)

TABLE 8
Distribution of estimated ω^2 among test statistics

Test Statistics	Estimated ω^2 Values: 0.009 or less	0.01-0.09	0.10-0.19	0.20-0.29	0.30-0.39	0.40 or more	Total
F-test	9	44	9	2	2	4	70
t-test	19	179	79	31	16	10	334
Total (%)	28 (6.93)	223 (55.2)	88 (21.78)	33 (8.17)	18 (4.46)	14 (3.47)	404 (100)

TABLE 9
Distribution of estimated ω^2 values according to Cohen's categorization

Effect	Range of ω^2	Number of ω^2
None*	$\omega^2<0.01^*$	28 (6.93%)
Small	$0.01<\omega^2<0.06$	154 (38.12%)
Medium	$0.06<\omega^2<0.15$	124 (30.69%)
Large	$\omega^2>0.15$	98 (24.27%)
Total		404 (100.00%)

*This particular category is assumed by default

TABLE 10
Distribution of estimated *r* between journals

Journal	Estimated *r* Values: 0.10 or less	0.10-0.29 (Small)	0.20-0.49 (Medium)	0.50 or more (Large)	Total
Journal of Marketing	14	66	13	1	94
Journal of Marketing Research	7	109	89	44	249
Journal of Advertising Research	12	40	22	4	78
Journal of Consumer Research	6	295	213	93	607
Total (%)	39 (3.79)	510 (49.61)	337 (32.78)	142 (13.81)	1028 (100.00)

TABLE 11
Distribution of estimated *r* between test statistics

Test Statistic	Estimated *r* Values: 0.10 or less (None)	0.10-0.29 (Small)	0.20-0.49 (Medium)	0.50 or more (Large)	Total
χ^2-test	2	36	21	5	64
F-test	22	314	201	93	630
t-test	15	160	115	44	334
Total (%)	39 (3.79)	510 (49.61)	337 (32.78)	142 (13.81)	1028 (100.00)

TABLE 12
Distribution of estimated Cohen's *d* between journals

Journal	Estimated *d* Values: 0.20 or less	0.20-0.49 Small	0.50-0.79 Medium	0.80 or more Large	Total
Journal of Marketing	14	48	23	7	92
Journal of Marketing Research	5	68	68	91	232
Journal of Advertising Research	23	34	11	11	79
Journal of Consumer Research	6	183	181	203	573
Total (%)	48 (4.92)	333 (34.12)	283 (29.00)	312 (31.97)	976 (100.00)

TABLE 13
Distribution of estimated *d* between test statistics

Test Statistic	Estimated *d* Values: 0.20 or less (None)	0.20-0.49 (Small)	0.50-0.79 (Medium)	0.80 or more Large	Total
t-test	15	103	109	107	334
F-test	33	230	174	205	642
Total	48 (4.92)	333 (34.12)	283 (29.00)	312 (31.97)	976 (100.00)

*Chi-squared test was not included because there was no equation found to convert chi-squared values into *d* values.

in Tables 12 and 13, is not as varied as was the case for ω^2 and *r*; the majority of the *d* values still fall within the category of 'small effect', however. There is, once again, a small number of effect size estimates that are so small as to be negligible (4.92%).

Relationship between ω^2 and r^2

Corresponding values of both ω^2 and r^2 from the same experiments were used to examine this relationship between these two measures of effect size. Regression analysis yields information about a close relationship; r^2=0.94 (F=20920.90, p<0.001). Furthermore, the calculated effect size, based on ω^2, is very large (ω^2=0.98).

DISCUSSION

General level of effect size reporting

About one quarter (24.54%) of the experiments studied reported ω^2. There are several possible reasons for not reporting effect size. Clearly, some test statistics employed by researchers may not allow for the use of ω^2 as a measure of effect size. There are also other measures of effect sizes besides ω^2 which might have been used. Nevertheless, it is tempting to speculate that ω^2 values may go unreported because of their small values. Researchers may be unwilling to disclose these small effect sizes as they will show highly significant results to be significant only because of the large sample size of the studies. In fact, this supposition is not born out here, as the incidence of no, small, medium and large effect sizes is very similar between those experiments reporting effect sizes and those not (see Tables 4 and 9). The fairly low rate at which ω^2 values are being reported could also be indicative of ignorance, a general lack of interest in discovering the extent to which an effect is significant or a genuine belief that only the existence of an experimental effect is important, no matter what it's size.

For all this, the present study has at least found an increase in the number of experiments that report ω^2 since the 1985 study conducted by Peterson *et al*. Only 14.4% of the total experiments studied in the period between 1970 and 1982 reported ω^2 values, while 24.63% of the experiments we studied reported ω^2. This suggests an increasing awareness of the importance of effect sizes as a complement to the traditional statistical tests of significance and the possibility that the PAB paper had an impact.

Cohen's Categorization of effect size

An analysis of the distribution of ω^2 values among the 'small', 'medium' and 'large' effect categories provides an insight into the nature of consumer behavior experiments. The results reported above are consistent with the opinion that effect sizes reported in consumer behavior experiments are usually quite small. The majority (60.80%) of reported ω^2 values documented here fall between the range of 0.01 to 0.09, while less than 10% of the experimental effects have values greater than 0.3. This could be attributed to the fact that many of these studies are subject to unidentified and uncontrollable effects or, more simply, that the social science view that explaining even a very small part of a human's behavior is very difficult (but, nonetheless, worthy). The mean value of ω^2 in the work considered is 0.102, with a standard deviation of 0.109. This means that, on the average, the value of ω^2 for a given experimental effect in the consumer behavior literature is about 10%. This confirms what was earlier mentioned about effect sizes being generally small for behavioral experiments. 5.4% of reported ω^2 values reported here were 0.009 or less. This small number of studies poses a problem because it could be that the reporting of these insignificant effects would not serve to enhance the quality of the report but rather would reduce the credibility of the study.

Articles and experiments that do not report ω^2

The majority (55.2%) of the estimated ω^2 values fall between the range of 0.01 and 0.09 as expected. The calculated mean value of ω^2 is 0.11 with a standard deviation of 0.13. This is very close to the mean value of significant effects (0.107) calculated in the study conducted by Peterson *et al* (1985).

All three measures of effect size consistently reveal a small percentage of the experimental effects to be insignificant. Although this percentage is quite small (ranging from 3.79% to 6.13%), it should be noted that this percentage is taken out of experimental effects that do report statistically significant results from various statistical tests (e.g. F-tests, t-tests or χ^2 tests). There is a clear message here, though, for the reviewers of marketing journals.

Relationship Between ω^2 and r^2

The results here do show that ω^2 and r^2 indeed have a strong linear relationship. However, it was found that ω^2 was consistently less than r^2 by 6%, rather than the 20-25% reported by Oakes (1986). This is rather important, as the consistency of the relationship would otherwise make it possible to estimate values of ω^2 given the values of ω^2 and r^2, and vice versa.

A possible reason for the discrepancy between this study and Oakes (1986) may lie in the fact that while r^2 can measure the proportion of the total variance explained by an experimental effect only if there is a linear relationship between the independent and dependent variables, ω^2 is not subject to the same limitation. Suppose there is no evidence of linearity between the independent and dependent variable, r^2 will be less accurate (compared to ω^2) in measuring the proportion of the total variance that is explained by the experimental effect. Therefore, the similarity between ω^2 and r^2 may be affected by the presence (or even the strength) of linearity between the dependent variables and the independent variables in the experiments.

Further research

There are several questions left begging by this study, which leaves the way for others to continue to expand the research. One such avenue follows immediately from the comments above. An investigation of how the strength and presence of linearity between the independent and dependent variables of an experiment affect the presence (as well as strength) of linearity between corresponding values of ω^2 and r^2 would clearly provide an interesting research study.

A second challenge lies with the sample bias problem. Of the 1,399 experimental effects considered here, more than half (59.47%) were taken from the Journal of Consumer Research. Whilst it is true that the nature and focus of this journal centres on consumer behaviour research, there still seems some justification for a claim of sample bias. As is so often the case with longitudinal studies of this type, a problem arises because precise replication of an original study may ignore other, emerging journals. Further research, then, to cope with publication bias more rigorously and to expand the work to other journals, would be of value.

REFERENCES

Cohen, Jacob (1977). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.

Cooper, Harris M. (1981). "On the Significance of Effects and the Effects of Significance". *Journal of Personality and Social Psychology*, 41(November), 1013–1018.

Glass, Gene V., Barry McGaw and Mary L. Smith (1981). *Meta-Analysis in Social Research*. Beverly Hills, CA: Sage.

Hays, William L. (1963). *Statistics for Psychologists*. New York: Holt, Rinehart and Winston, Inc.

Hays, William L. (1988). *Statistics, 4th edition*. Holt, Rinehart and Winston, Inc.

Hedges, L.V. and Ingram Olkin (1985). *Statistical Methods for Meta-Analysis*. Academic Press, Inc. (London) Ltd.

Irvine, J., I. Miles and J. Evans (1979). *Demystifying Social Statistics*. Pluto Press

Oakes, Michael (1986). *Statistical Inference: A Commentary for the Social and Behavioral Sciences*. John Wiley & Sons.

Peterson, Robert A., Gerald Albaum and Richard F. Beltramini (1985). "A Meta-Analysis of Effect Sizes in Consumer Behavior Experiments". *Journal of Consumer Research*, 12(June), 97–103.

Rust, Roland T., Donald R. Lehmann and John U. Farley (1990). "Estimating Publication Bias in Meta-Analysis". *Journal of Marketing Research*, 27(May), 220-226.

Wolf, Fredric M. (1986). Meta-Analysis: Quantitative Methods for Research Synthesis, in *Quantitative Applications in the Social Sciences Series*. Beverly Hills, CA: Sage.