Practical Assessment, Research, and Evaluation

Volume 10 *Volume 10, 2005* Article 16

November 2019

Does It Matter If Non-Powerful Significance Tests Are Used in Dissertation Research?

Heping Deng

Recommended Citation
Deng, Heping (2019) "Does It Matter If Non-Powerful Significance Tests Are Used in Dissertation Research?," *Practical Assessment, Research, and Evaluation*: Vol. 10, Article 16.
DOI: https://doi.org/10.7275/qakz-t063
Available at: https://scholarworks.umass.edu/pare/vol10/iss1/16

Practical Assessment, Research & Evaluation
A peer-reviewed electronic journal.

Volume 10 Number 16, September 2005 ISSN 1531-7714

Does It Matter If Non-Powerful Significance Tests Are Used in Dissertation Research?

Heping Deng
Borough of Manhattan Community College

This study examines the statistical power levels presented in the dissertations completed in the field of educational leadership or educational administration. Eighty out of 221 reviewed dissertations were analyzed and overall statistical power levels were calculated for 2,629 significance tests. The statistical power levels demonstrated in the dissertations were satisfactory for detecting Cohen's large effect (d=0.80) and medium effect (d=0.50) but quite low for small effect (d=0.20). Therefore, the authors of analyzed dissertations had a very low probability of finding true significance when looking for Cohen's small effect.

Statistical significance is a function of several study features, including significance level (α), sample size, effect size, and statistical power level. The importance of understanding the interrelatedness between theses features has been repeatedly addressed in the literature (e.g., Carver, 1993, Thompson, 1989, Young, 1993). This allows a researcher to consider various issues, such as α level, effect size, sample size, and statistical power (Bergin & Garfield, 1994). In addition, when presented in the context of theses features, research results will be more meaningful.

In a significance test, the probability of rejecting the null hypothesis when it is false is defined as statistical power. Jacob Cohen has done more than anyone else to emphasize the importance of statistical power and to clarify the confusion and misunderstanding that surround the concept. Cohen (1962) suggested three levels of effect size (small, medium, and large) and identified the distinction between major and minor research hypotheses within each study and equated each non-parametric procedure with its analogous parametric test. Cohen (1965) offered possible solutions to the problems inherent in low statistical power and outlined step-by-step the process of a priori power analysis. Cohen (1988) provided a framework to determine the statistical power of a study subject to a specified significance level, effect size, and sample size. A researcher must specify (a) significance level, (b) minimum desired effect size, and (c) the desired power (Hill, 1990). The decisions have to be made at the design-planning stage of a research (Gall, Borg, and Gall, 2002). This process is known as statistical power analysis (Hallahan & Rosenthal, 1996). In social science research, especially with implications on policy decision-making, the importance of power analysis is clearly evident. Shavelson (1981) suggested that researchers should strive to design research with a high level of statistical power. Hinkle, Wiersma,

and Jurs (1994) indicated that a more powerful significance test can detect a smaller difference between two population means; hence, it is easier to reject the null hypothesis. In educational research, the consequences of accepting a false hypothesis is usually more serious than rejecting a true hypothesis. For example, when a new teaching method or new program that is in fact better than the conventional ones is not used (accepting a false hypothesis), this decision will result in the waste of resources applied to the design and development of the new teaching method or program. However, when a new method or program that is in fact not better than the conventional ones is used (rejecting a true hypothesis), this decision will not result in the waste of resources. Therefore, a powerful significance test can lead to better decision-making by educational leaders. Recent evidence indicates the growing interest on statistical power analysis (Nelson & Coorough, 1994). The importance of estimating the power of a significance test has received attention in fields such as applied statistics, education, psychology, and nursing (Gatti & Harwell, 1998). New statistical modules and software packages, which are relatively easy to use for calculating sample size and statistical power, are now widely available. An increasing number of researchers and graduate students are using these modules and packages to establish sample size and power estimates, *a priori*. Nevertheless, many behavioral science researchers fail to discover significant differences among sample means "even when differences among corresponding population means are substantial" (Rogers & Hopkins, 1988).

As no evidence can be found on power analytic surveys in educational leadership or educational administration research (See Appendix A), this paper attempts to examine the overall power level of significance tests conducted in this particular field. As dissertations reflect the current emphases in a research area (Nelson & Coorough, 1994) this study focuses on dissertations completed by doctoral students majoring in educational leadership or educational administration among doctoral higher education institutions in Tennessee.

METHODS

Data Collection

Data collection was limited to Tennessee simply for the ease of access although similarities could be found among other states. The target population for this study consisted of all quantitative dissertations successfully defended by doctoral students majoring in educational leadership or educational administration at five universities in Tennessee from January 1, 1996 through December 31, 1998. For inclusion in the power analysis, dissertations had to fit the following criteria:

1. They had to include statistical significance tests.
2. Sample sizes had to be reported.
3. Only those significance tests that were commonly used and associated with power tables in Cohen (1988) were included. They are: t-test, Pearson Product Moment Correlation Coefficient, Spearman Correlation Coefficient, Difference between Correlation Coefficients, Mann-Whitney, Chi-square test, F-test, Analysis of Variance and Covariance, z test, and multiple regression analysis.

A dissertation with an extremely large sample, which caused biased results, was excluded from analysis.

The following set of criteria was used in the power analysis:

1. When a preset α-level was provided by the researcher, this level was used for calculating statistical power.
2. When no preset α-level was provided, the α-level was assumed to be .05. In cases where multiple p-values were reported, an α-level of .05 was assumed if no other α-levels were reported.
3. When multiple comparisons were used, adjusted p-value was calculated using Sidak test formula: adjusted p-value=$1-(1-p_r)^K$, where p = probability of making Type I error, p_r = unadjusted p-value, K = number of

https://scholarworks.umass.edu/pare/vol10/iss1/16
DOI: https://doi.org/10.7275/qakz-t063

comparisons.

4. When an ANOVA was performed, adjusted p-value was calculated using the above formula but *K* was calculated using the following formula: $K = a(a-1)/2$, where a = the number of groups for comparison.
5. In a case where the researcher did not report whether a one-tailed or two-tailed alternative hypothesis was used, a two-tailed hypothesis was assumed.
6. Estimates of small, medium, and large effects were taken from Cohen (1988). These estimates were unique for each type of significance test.

Instruments

The main research tools used in this study are power tables found in *Statistical Power Analysis for the Behavioral Sciences* (Cohen, 1988). These power tables can be used for calculating statistical power from a given sample size and effect size at three different α levels of .01, .05, and .10 each. The power tables are available for each of the following statistical procedures: t-test for means, Pearson Product-Moment Correlation Coefficient, Difference Between Correlation Coefficients, the test that a proportion is .50 and sign test, differences between proportions, Chi-square, Multiple Regression and Correlation, set correlation, and multivariate methods (Cohen, 1988). McKean's (1990) recording instrument was used to record the data necessary to complete *a priori* power analysis and relevant background information from each dissertation.

The recorded sample size, α level, and directionality of each significance test were used to establish a post-hoc estimate of statistical power for detecting Cohen's (1988) small, medium, and large effects of that significance test but no actual effect size for each significance test was examined. The power levels were read directly from Cohen's (1988) power tables and then the estimated power levels for each selected dissertation were calculated. The average statistical power presented in all analyzed dissertations was obtained. While this meta-analysis was being conducted, close attention was also given to the findings of each analyzed dissertation to prevent misinterpretations of mean power levels. In addition to statistical power levels, the following descriptive data were analyzed and reported, too: percentage of analyzed dissertations over all reviewed dissertations by institution, percentage of all reviewed dissertations over all completed dissertations by institution, distribution of analyzed dissertations by institution, distribution of significance tests by institution, mean and median of each sample, distribution of mean sample size, mean sample size by institution, and sample size and optimal sample size by significance test as well as by institution.

Exploratory Data

Initially, 221 dissertations were reviewed, of which 80 met the criteria for inclusion in the study. The final list of dissertations includes five universities in Tennessee, which are referred to as U1, U2, U3, U4, and U5. These 80 selected dissertations used 2,629 statistical significance tests and the overall mean sample size was 187, median sample size was 74, and the overall or "typical" sample size was only 35. Many samples fell in the range of 1 to 49 (35%) and almost 60% of the reported significance tests were based on samples of fewer than 100. This suggests that while some doctoral students used large samples (more than 1,000), most used small samples. Given the direct relationship between power and sample size, the authors of these dissertations only had a reduction power level in their significance tests.

Statistical Power Analysis

The mean and median power levels were calculated across all significance tests for Cohen's (1988) small effect (d=0.20), medium effect (d=0.50), and large effect (d=0.80) (thereafter, indicated as small, medium, and large effects). The results are presented in Table 1.

Table 1. Mean, Median, and Standard Deviation of Statistical Power for All Significance Tests

	Small Effect	Medium Effect	Large Effect
Mean	.29	.75	.93
Median	.21	.85	1.00
Std dev.	.23	.26	.15

The significance tests examined, on average, had slightly higher than a one-fourth chance of detecting small effects, nearly a three-fourths chance of detecting medium effects, and a noticeably high chance of detecting large effects. Though these tests had realized better than conventional power level (.80) of detecting large effects, they had only average .29 power level of detecting the presence of small effects. Furthermore, examining other central tendencies, especially mode, we can see this average power level was even inflated by a few very large samples.

Of 2,629 significance tests, 91% had power levels lower than Cohen's recommended .80 for detecting small effects, 47% were below that level for medium effects, and almost 12% were below that level for large effects. Especially, more than 66% had power levels lower than .30 for detecting small effects. Compared to the mean and mode of the optimal sample size (73 and 64, respectively), the mean sample size used in these analyzed dissertations was much larger and the overall modal was much smaller. The average sample size is 2.5 times the average optimal sample size and the most frequently occurring sample size was nearly half of the model optimal sample size. The mean sample size for t-tests (M=80.6) was closest to its optimal sample size (M=63.5), while the mean sample size used for multiple regression (M=366) was the most different from its optimal size (M=82). The most frequently used significance test was t-test for means (n=948, 36%). The next commonly used significance test was Analysis of Variance or ANOVA (n=887, 34%). The third commonly used significance test was Chi-square (n=380, 15%). According to the recommendations of Cohen (1962), the Mann-Whitney U-tests were treated as if they were t-tests for means. Though Mann-Whitney U-tests and Difference between Correlation Coefficient Tests were included in this table, their distributions will not be included in thereafter illustrations because of their being only one of each type. Among the rest 7 types of tests, the overall mean statistical power of z-tests demonstrated the highest power levels for detecting small, medium, and large effects. While the mean power levels of both t and Chi-square tests were below the overall mean power level of these 7 tests for detecting small effects, the mean power level of ANOVA tests was below the overall mean power levels for detecting medium and large effects. Chi-square tests had the biggest mean sample size of 448 but they only had average power (.25) for detecting small effects due to their huge standard deviation (SD=1,152). Though 948 t-tests had the smallest mean sample size (80.6), they demonstrated low but not too bad mean power levels for detecting small, medium, and large effects due to their relatively small standard deviation (SD=80.6).

A little more than one third (37%) of the significance tests resulted in the rejection of null hypotheses at an alpha level of .05. The median power levels were .25 for detecting small effects, .90 for detecting medium effects, and .99 for detecting large effects. The median power-level for detecting small effects was much lower than Cohen's (1988) .80 power criterion but the median power-levels for detecting both medium and large effects exceeded this criterion.

Compared to the statistical power for detecting medium effects among the various types of tests, the lowest power was found in Analysis of Variance (ANOVA) at .71. The mean power level for

detecting medium effects did not reach Cohen's recommended .80 in the following tests: ANOVA, t-test, and Chi-square. The mean power level of the Chi-square tests for detecting medium effects was below .80 but still above the overall mean power of all significance tests (.76). The mean power levels of other tests were below the average level for detecting medium effects. This finding suggests that, particularly when using t-tests and ANOVA to compare group means, the number of subjects may be too small for detecting true effects.

Among five selected institutions, none met Cohen's (1988) .80 power criterion for detecting small effects. The overall mean power levels of the significance tests used in the dissertations completed at U1 were ranked highest for detecting small and large effects. The mean power for detecting medium effects of significance tests conducted at U1 and U5 were the only ones that met Cohen's recommended .80 power criterion for detecting medium effects. When large effects were considered, the mean power levels met the .80 power criterion at each of the five institutions. This finding indicates that if researchers were only interested in looking for large effects, the power levels exhibited in these 80 dissertations were sufficiently strong. However, if researchers were intending to detect small effects, the likelihood of doing so was much smaller.

A comparison between the statistical power estimates found in other fields shows that the mean statistical power level, when each significance test is considered the unit of analysis, demonstrated in this study for detecting a small effect (.29), is higher than the level found in a majority of the previous studies, and the mean power for detecting medium (.75) and large (.93) effects, were higher than the estimates found in all the other studies shown in Table 2.

Table 2. A Comparison of Mean Statistical Power Estimates Found in Various Fields

Discipline/Author	Statistical Power Estimates		
	Small Effect	Medium Effect	Large Effect
Abnormal Psychology:			
Cohen (1962)	.18	.48	.83
Sedlmeier & Gigerenzer (1989)	.14	.44	.90
Educational Research:			
Brewer (1972)	.14	.58	.78
Science Education:			
Penick & Brewer (1972)	.22	.71	.87
Health, Physical Education:			
Jones & Brewer (1972)	.15	.55	.81
Counselor Education:			
Haase (1974)	.10	.37	.74
Communication:			
Chase & Tucker (1975)	.18	.52	.79

Table 2. A Comparison of Mean Statistical Power Estimates Found in Various Fields (Continued)

Discipline/Author	Statistical Power Estimates		
	Small Effect	Medium Effect	Large Effect
Speech Pathology:			
Chase & Kroll (1975)	.16	.44	.73
Applied Psychology:			
Chase & Chase (1976)	.25	.67	.86
Occupational Therapy:			
Ottenbacher (1982)	.37	.63	.85
English Education:			
Daly & Hexamer (1983)	.22	.63	.86
Evaluation Research:			
Lipsey et al. (1985)	.28	.63	.81
Adult Education:			
West (1985)	.22	.66	.88
Educational Leadership:			
Deng (unpublished)	.29	.75	.93

DISCUSSION AND RECOMMENDATIONS

Neither optimal sample size nor formula for its calculation was used or mentioned in any of 80 analyzed dissertations. While the mean sample size used in 2,629 significance tests was 2.5 times greater than the mean optimal sample size, most significance tests still used samples that were much smaller than an optimal (or desired) size. Therefore, it appears that, little attention was given to the impact of sample size or statistical power and ultimately the quality of research findings. Perhaps a "rule of thumb" that calls for using whatever sample size the available resources allowed was being followed. However, if there are no other important concerns, using a suggested optimal sample size can help researchers determine a suitable sample size.

Neither Type II error nor statistical power was mentioned in analyzed dissertations. Therefore, it is hard to determine these dissertations were undertaken with consideration of type II error or the level of statistical power. Nevertheless the mean power levels demonstrated in these dissertations were higher than all reviewed studies in other fields for detecting medium and large effects.

The overall mean power levels, when each significance test was considered the unit of analysis, were .29 for detecting small effects, .75 for medium effects, and .93 for large effects. This study demonstrated the highest statistical power levels for detecting medium and large effects, with a relatively low level for detecting small effects. If a researcher was intending to detect a small effect, the likelihood of finding it was small.

With an overall mean power of .75 for detecting medium effects and .93 for large effects, doctoral students of analyzed dissertations had a three-fourths chance of rejecting the null hypotheses if they were seeking medium effects and a very high large chance (93%) of rejecting the null hypotheses if they were seeking large effects. These power levels are higher than all statistical power estimates found in other fields. Hence it appears that the statistical power levels demonstrated in all these dissertations are satisfactory. However, only 29 of 100 chances could these doctoral students correctly reject the null hypotheses if they were seeking small

effects. A question may be raised here as to whether they needed to detect or were expected to detect such a small effect. Obviously, research management involved in analyzed dissertations could be various and small effects would have not always been the targets for each test. But in non-experimental designs or new areas of research inquiry, effect size is likely to be small. In non-experimental research, the influence of uncontrollable extraneous variables can increase the amount of "instability" in the data, which makes it difficult to detect the existing differences or relations in a treatment group. When the more easily detectable effects have been partialled out, interaction effects are the targets of examination. Such interactions are often called for in the development and modeling stages of a growing field of inquiry as strategic management. Schendel and Hofer (1979) challenged researchers to pay "more attention to interactive effects of the various classes of variables contained in the (strategic management) model" (p. 530). Doctoral students majoring in educational leadership or educational administration, as potential educational researchers, need to be cautious about using non-experimental and exploratory research designs, and would better plan to detect relatively small effects in research for their dissertations. At planning stage, they need to find out the effect size on which they can set up appropriate α, statistical power level, and optimal sample size.

How can our doctoral students be made more aware of the importance of power analysis and made to conduct significance tests with sufficient statistical power in research for their dissertations? First, they should be required to use power analysis for determining sample size in the planning stage. Second, they may avoid using very low α levels in order to have sufficient power in their research. Third, they would be better required to estimate the power for each significance test to be conducted. "If the estimated power is too low, the paper should not be publicly published" (West, 1985).

The above discussion suggests the following:

1. The training doctoral students in educational leadership or educational administration receive in statistical data analysis may include a focus on statistical power or type II error and the interrelationship between effect size, sample size, and statistical power.
2. Doctoral students need to be taught how to decide upon a "best" sample size for their research. Cohen's (1988) optimal sample size tables and relevant formulas can be used in their studies.
3. They need to calculate and report effect size and statistical power for each performed significance test.

The following recommendations are made for further studies on statistical power analysis:

1. This study may be replicated in a 2-to-3 year cycle, using the same population to track the change of statistical power demonstrated in dissertations completed in most recent three years by doctoral students majoring in educational leadership or educational administration. The information obtained from these studies could be used to document how much emphasis is being put on this issue in the field of educational leadership or administration in Tennessee.
2. Similar studies may be conducted in other states, too.
3. A nationwide survey may be conducted to evaluate how much knowledge of statistical power the doctoral students majoring in educational leadership or educational administration have.

REFERENCES

Bergin, A.E.& Garfield, S. L. (1994). *Handbook of psychotherapy and behavior change*. New York: Wiley.

Brewer, J. K. (1972). On the power of statistical tests in the "American Educational Research

Journal." *American Educational Research Journal, 9,* 391-401.

Carter, D. C. (1997). The account taken of statistical power in research published in the 'British Journal of Psychology.' *British Journal of Psychology, 88*, 71-84.

Carver, R. P. (1993). The case against statistical significance testing, revisited. *The Journal of Experimental Education, 61*, 287-92.

Chase, L. J., & Baran, S. J. (1976). An assessment of quantitative research in mass communication. *Journalism Quarterly, 53*, 308-11.

Chase, L. J. & Chase, R. B. (1976). A statistical power analysis of applied psychological research. *Journal of Applied Psychology, 61,* 234-37.

Chase, L. J. & Kroll, R. M. (1975). Communication disorders: A power analytic assessment of recent research. *Journal of Communication Disorders, 8*, 237-47.

Chase, L. J. & Tucker, R. K. (1975). A power-analytic examination of contemporary communication research. *Speech Monographs, 42,* 29-41.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology, 65*, 145-53.

Cohen, J. (1965). Some statistical issues in psychological research. In Wolman, B.B. (Ed.), *Handbook of clinical psychology*. New York: McGraw Hill.

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

Cohen, J. (1994). The earth is round (p<.05). *American Psychologist, 49*, 997-1003.

Daly J.A. & Hexamer, A. (1983). Statistical power in research in English Education. *Research in the Teaching of English, 17*, 157-64.

Daniel, T. D. (1993). A statistical power analysis of the quantitative techniques used in the 'Journal of Research in Music Education', 1987 through 1991. Unpublished doctoral dissertation, Auburn University, Birmingham, Alabama.

Dawes, R. M. (1991). Probabilistic versus casual thinking. In D. Cicchetti & W. Grove (Eds.), *Thinking clearly about psychology: Vol. 1. Matters of public interest: Essays in honor of Paul Everett Meehl* (p.235-64). Minneapolis, MN: University of Minnesota Press.

Falk, R., & Greenbaum, C.W. (1995). Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory & Psychology, 5,* 75-98.

Gall, M. D., Borg, W. R., & Gall, J. P. (2002). *Educational research: An introduction*. New York: Longmon.

Gatti, G. G., & Harwell, M. (1998). Advantage of computer programs over power charts for the estimation of power. *Journal of Statistics Education*, *6*, 1-12. *Retrieved January 25, 1999 from the World Wide Web: http://www.amstat.org/publications/jse/*

Hallahan, M. & Rosenthal, R.(1996). Statistical power: Concepts, procedures, and applications. *Behavior Research and Therapy, 34*, 489-99.

Haase, R. F. (1974). Power analysis of research in counselor education. *Counselor Education and Supervision, 34,* 404-13.

Hill, O. W. (1990). Rethinking the "significance of the rejected null hypothesis." *American Psychologist,45*, 667-68.

Hinkle, D. E., Wiersma, W., & Jurs, S.G.(1994). *Applied statistics for the behavioral sciences (3rd ed.).* Boston: Houghton Mifflin.

Jones, B. J., & Brewer, J. K. (1971). An analysis of the power of statistical tests reported in the Research Quarterly. *Research Quarterly, 43(1)*, 23-30.

Katzer, J., & Sodt, J. (1973). An analysis of the use of statistical testing in communication research. *The Journal of Communication, 23*, 251-65.

Kosculek, J. F., & Szymanski, E. M. (1993). Statistical power analysis of rehabilitation counseling research. *Rehabilitation Counseling Bulletin, 36*, 212-19.

Lipsey, M. W., Crosse, S., Dunkle, J., Pollard, J., & Stobart, G. (1985). Evaluation: The state of the art and the sorry state of the science. *New Directions for Program Evaluation, 27,* 7-28.

McKean, K. E. (1990). *Statistical power analysis of doctoral dissertation research in educational psychology*. Unpublished doctoral dissertation. Oklahoma State University, Stillwater, Oklahoma.

Nelson, J. K., & Coorough, C. (1994). Content analysis of the PhD versus EdD dissertation. *Journal of Experimental Education, 62*, 158-68.

Osborne, J.W., Christensen, W.R., & Gunter, J. (2001). Educational psychology from a statistician's perspective: A review of the quantitative quality of our field. (a paper presented at the national meeting of the American Education Research Association, WA)

Ottenbacher, K. (1982). Statistical power and research in occupational therapy. *Occupational Therapy Journal of Research, 2*, 13-25.

Penick, J. E. & Brewer, J. K. (1972). The power of statistical tests in science teaching research, *Journal of Research in Science Teaching*, 9 (4), 377–81.

Rogers, W.T. & Hopkins, K.D. (1988). Power estimates in the presence of a covariate and measurement error. *Educational and Psychological Measurement, 48*, 647-56.

Rossi, J. S. (1990). Statistical power of psychological research: What have we gained in 20 years? *Journal of Consulting and Clinical Psychology, 58*, 646-56.

Sawyer, A. G., & Ball, A. D. (1981). Statistical power and effect size in marketing research. *Journal of Marketing Research, 18*, 275-290.

Schendel, D. E. and Hofer, C.W. (1979). *Strategic Management: A New View of Business Policy and Planning.* Boston: Little, Brown and Company.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power h ave an effect on the power of studies? *Psychological Bulletin, 105,* 309-16.

Shavelson, R. J. (1981). *Statistical reasoning for the behavioral sciences.* Boston: Allyn and Bacon.

Thompson, B. (1989). Statistical significance, result importance, and result generalizability: Three noteworthy but somewhat different issues. *Measurement and evaluation in Counseling and Development, 22*, 2-6.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher, 25(2)*, 26-30.

Young, M.A. (1993). Supplementing tests of statistical significance: Variation accounted for. *Journal of Speech and Hearing Search, 36*, 644-56.

West, R. F. (1985). A power analytic investigation of research in adult education:1970-1982. *Adult Education Quarterly, 35*, 131-41.

Appendix A:

Year	Analyzed Journal	Volume	# of Reviewed Articles	# of Analyzed Studies	Median Power			Results/Suggestions
					Small Effects	Medium Effects	Large Effects	
1972	American Educational Research Journal	1969 –1971	N/A		.14	.58	.78	
1972	Journal of Research in Science Teaching	1969 –1971	N/A		.22	.71	.87	
1972	The Research Quarterly	1969 –1971	N/A		.14	.52	.80	
1973	Journal of Communication	1971 - 1972	31		.23	.56	.79	
1974	Counselor Education and Supervision	1968 - 1971	234	60	.095	.365	.74	To increase power by increasing sample size and conduct a priori power analysis as a routine research planning
1975	The American Forensic Association Journal Central States Speech Journal Journal of Communication The Quarterly Journal of Speech Southern Speech Communication Journal Speech Monographs The Speech Teacher Today's Speech Western Speech	1973	N/A		.08 to .34 .18 (overall)	.26 to .76 .52 (overall)	.56 to .94 .79 (overall)	
1975	Two journals of speech pathology and audiology research		N/A		.16	.44	.73	
1976	Journalism Quarterly Journal of Broadcasting	1974	N/A		.34	.76	.91	Research in mass communication had been preformed with high power
1976	Journal of Applied Psychology	1974	N/A		.25	.67	.86	More studies with non-significant results should be published

Year	Analyzed Journal	Volume	# of Reviewed Articles	# of Analyzed Studies	Median Power			Results/Suggestions
					Small Effects	Medium Effects	Large Effects	
1981	Journal of Marketing Research	1979	60		.41*	.89*	.98*	Power can also be increased by other ways, such as increasing measurement and treat reliability, using covariance, and carefully selecting and manipulating independent variables.
1985	Adult Education	21 -32	N/A	65 with 1,666 tests reported	.22*	.15*		Researchers had less than a 16% chance of correctly rejecting the null hypothesis when using a small effect size. The lack of statistical power to detect small effects is a common problem throughout the social sciences. When medium effect sizes are considered, the estimates are not satisfying, either. However, for large effect sizes, the probability regarding a correct rejection of null hypothesis is sufficiently high.
1990	Journal of Abnormal Psychology Journal of Consulting and Clinical Psychology Journal of Personality and Social Psychology	1982	1,500	40,000 tests	.26*	.64*	.85*	
1993	Journal of Research in Music Education	1987 - 1991	109	78	.13	.64	.97	None reported any type of power analysis or any mention of an estimate of power or effect size.
1997	British Journal of Psychology	1993 – 1994	54	1,243 inferential statements	.80 or more*	.80 or more*	.80 or more*	
1990	Ph.D. – Level dissertations completed in 1989 in the field of educational psychology.		N/A		.169*	.541*	.795*	The levels of statistical power in the sampled dissertations were similar to whose reported in published literature reviews in the fields of education and behavioral sciences.

Year	Analyzed Journal	Volume	# of Reviewed Articles	# of Analyzed Studies	Median Power			Results/Suggestions
					Small Effects	Medium Effects	Large Effects	
1993	Rehabilitation Counseling Bulletin, V. 34, 1990-1991 Rehabilitation Psychology, Volume 35, 1990 Journal of applied Rehabilitation Counseling, V. 21, 1990 Journal of Rehabilitation, V. 56, 1990 Rehabilitation Education, V. 4, 1990		150	32	Note a	Note b	Note c	Only one study referred to power analysis, three authors discussed why a certain alpha or sample size was chosen, and none of them showed awareness of or concern about statistical power

Notes: * indicates mean power, rather than median power, to detect small, medium, and large effects.
a. None of these studies have a 50/50 chance of detecting small effect sizes
b. Only 12 of them have a 1in 2 chance of finding significant results assuming medium effects.
c. Nine percent of these studies showed less than a 50/50 chance of detecting large effects, and 3% of them showed less than 3 in 10 chances of detecting significant results assuming large effects.

Citation

Deng, Heping (2005). Does It Matter If Non-Powerful Significance Tests Are Used in Dissertation Research? *Practical Assessment Research & Evaluation*, 10(16). Available online: http://pareonline.net/getvn.asp?v=10&n=16

Author

Heping Deng, Higher Educational Assistant for the Office of Institutional Research at Borough of Manhattan Community College. He has Ed.D from East Tennessee State University with concentration on Institutional Research. His current research focuses on statistical power analysis, quantitative methodology, distance learning, instructional technology, assessment of student learning outcomes, etc. His email: cdeng@bmcc.cuny.edu. Phone: 212 220-8332. Fax: 212 220-8319.