DOCUMENT RESUME

ED 451 241 TM 032 469

AUTHOR	Fraas, John W.; Newman, Isadore
TITLE	Testing Non-Nil Null Hypotheses Using Resampling Stats Add-In for Microsoft Excel.
PUB DATE	2001-02-00
NOTE	23p.; Paper presented at the Annual Meeting of the Eastern Educational Research Association (Hilton Head, SC, February 14-17, 2001).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Computer Software; *Hypothesis Testing; *Statistical Significance
IDENTIFIERS	Microsoft Excel; *Nonnull Hypothesis; *Randomization

ABSTRACT

J. Frass and I. Newman (2000) proposed a hypothesis testing procedure that incorporated the following three key elements: (1) the establishment of a practical significance value; (2) the construction of a non-nil null hypothesis that incorporated the practical significance value; and (3) statistical testing of the non-nil null hypothesis with a randomization test. One of the difficulties researchers may encounter with this testing procedure is the implementation of the randomization test. This paper describes, through use of an example, how researchers can conduct a randomization test with relative ease with the use of the computer software Resampling Stats Add-in for Excel. In the final section of the paper, the randomization test results for the example data are compared to independent samples t-test results. The outcome of this comparison suggests that future investigation of the relative results of the two types of statistical tests may be beneficial to researchers. An appendix contains the computer program for the randomization test. (Contains 44 references.) (Author/SLD)

Running head: TESTING NON-NIL NULL HYPOTHESES

Testing Non-Nil Null Hypotheses Using Resampling Stats Add-In for Microsoft Excel

John W. Fraas
Ashland University

Isadore Newman
The University of Akron

Paper presented at the annual meeting of the
Eastern Educational Research Association
Hilton Head, South Carolina
February 14-17, 2001

Abstract

Fraas and Newman (2000) proposed a hypothesis testing procedure that incorporated the following three key elements: (a) the establishment of a practical significance value; (b) the construction of a non-nil null hypothesis that incorporated the practical significance value; and (c) statistical testing of the non-nil null hypothesis with a randomization test. One of the difficulties researchers may encounter with this testing procedure is the implementation of the randomization test. This paper describes, through the use of an example, how researchers can conduct a randomization test with relative ease with the use of the computer software Resampling Stats Add-In for Excel. In the final section of this paper, the randomization test results for the example data were compared to independent-samples t test results. The outcome of this comparison suggests that future investigation of the relative results of the two types of statistical tests may be beneficial to researchers.

Testing Non-Nil Null Hypotheses Using
Resampling Stats Add-In for Microsoft Excel

We have proposed (Fraas & Newman, 2000) that current research practices can be strengthened if researchers incorporate into their work the use of nil-null hypotheses that are based on effect sizes deemed important by researchers and practitioners. The testing procedure we proposed incorporated three key elements. First, researchers and practitioners must establish a practical significance value. Second, a non-nil null hypothesis that incorporates the established practical significance value is formulated. Third, the non-nil null hypothesis is statistically tested with a randomization test.

In our original paper on this testing procedure (Fraas & Newman, 2000), we statistically tested a non-nil null hypothesis with a randomization test by employing the computer software Resampling Stats (Simon, Weidenfeld, Bruce, & Puig, 1999). This software required the construction of a set of commands in order to statistically test the non-nil null hypothesis (see Appendix A for a copy of the commands). We believe researchers will be more inclined to employ our testing procedure if the testing of a non-nil null hypotheses can be conducted in a more user-friendly computer environment. Thus, we described with the use of an example the relative simplicity of testing a non-nil null hypothesis when the randomization test is executed by the Resampling Stats Add-In for Excel (Blank, Seiter, & Bruce, 1999) computer software.

In addition to demonstrating how researchers can use the Resampling Stats Add-In software (Blank, Seiter, & Bruce, 1999) in conjunction with our suggested hypothesis testing procedure, we compared the randomization test results obtained from our example to the independent-samples t test results produced by the SPSS version 10.0 (SPSS Inc., 1999) computer software. The purpose of this comparison was to determine, at least for our example, if similar results would be obtained by the two types of statistical tests.

Suggested Modifications to the Null Hypothesis Statistical Testing Procedure

Our proposal of an analytical statistical testing procedure that uses a randomization test of a non-nil null hypothesis (Fraas & Newman, 2000) was prompted by the concerns expressed by researchers over the years regarding the statistical testing of a nil-null hypothesis. Kirk (1996) stated that nil null hypothesis significance testing has been an integral part of the research process for almost 70 years. He also noted that nil null hypothesis significance testing has been surrounded by controversy for most of that time. As early as 1938, Berkson (1938) challenged the use of nil null hypothesis statistical testing. Over time, Berkson's challenges have been supported by numerous authors (Carver, 1978, 1993; Cohen, 1990, 1994; Falk, 1986; Falk & Greenbaum, 1995; Huberty, 1987, 1993; Meehl, 1967; Rozeboom, 1960; Shaver, 1980, 1993; Thompson, 1989a, 1989b, 1996, 1997, 1998, 1999a, 1999b, 1999c). It should be noted, however, that other authors have defended the use of nil null hypothesis testing (Cortina & Dunlap, 1997; Frick, 1996, 1999; Hagen, 1997; Levin, 1993, 1996, 1998; Levin and Robinson, 2000; Robinson & Levin, 1997).

As noted by Thompson (1999a) "a few scholars have called for the banning of statistical significance tests. However, the fact that many psychologists misinterpret statistical significance tests is not a reasonable warrant for banning these tests. Consequently, attention has now turned toward ways to improve practice" (p.169). Much of this attention has been directed towards the reporting of practical significance levels, e.g., effect sizes, and the testing of non-nil null hypotheses.

The Importance of Practical Significance

The importance of supplementing a statistical test of a null hypothesis is not a new idea. Fisher (1925) proposed that researchers report eta values, which measure the strength of association between the independent and dependent variables, along with the statistical tests contained in analysis of variance results. Later, Cohen (1969) introduced the concept of expressing the size of the population treatment effect in units of the common population standard deviation, which was labeled d. Cohen (1988) even provided the following guidelines for interpreting the magnitude of d: (a) a d value of .5 is a median effect; (b) a d value of .2 is a small effect; and (c) a d value of .8 is a large effect.

Much of the recent debate on the use of statistical tests and the importance of practical significance has centered on the question: Should researchers consider both concepts, and if so, how should it be done? Thompson (1996) expressed the view that formal statistical hypothesis testing might be an optional companion to the reporting of practical significance levels, i.e., effect sizes. Robinson and Levin (1997) took issue with Thompson's position. They believe that declarations of statistical significance should regularly precede deliberations of substantive significance. In light of this position, Levin and Robinson proposed a two-step data analysis process (Levin & Robinson, 2000; Robinson & Levin, 1997). In their two-step procedure the researchers would first determine whether the observed effect was statistically significant. Only if the observed effect was statistically significant would the researchers implement the second step in which they would assess the practical significance of the observed effect.

The Use and Testing of Non-Nil Null Hypotheses

Cohen (1994) expressed the view that "even null hypothesis testing complete with power analysis can be useful if we abandon the rejection of the point nil hypotheses [nil null hypotheses]" (p. 1002). Thompson (1999a) stated that researchers continue to use nil null hypotheses, however, for two reasons. First, most computer packages assume the researchers are testing nil null hypotheses. Thus, they are not equipped to invoke the necessary changes in calculations. As noted by Selin and Lapsley (1985, 1993), such changes include the use of critical values obtained from noncentralized t and F distributions. Second, Thompson noted that some of the complexities of using non-nil null hypotheses are not yet readily applicable in many designs.

In spite of these two roadblocks, Edgington (1995) recommends a testing technique to researchers who believe it is important to test non-nil hypotheses. He suggests that researchers can readily employ non-nil null hypotheses if they utilize randomization testing techniques. Edgington expressed the view that: "A randomization test null hypothesis need not be simply one of no differential treatment effect [a nil null hypothesis] . . . but can . . . [reflect] response magnitudes [a non-nil null hypotheses]" (pp. 319-320).

Suggested Testing Procedure

In light of the debate and views expressed by researchers regarding practical significance versus statistical significance and their opinions related to the use of non-nil null hypotheses, we suggested that researchers use an analytic technique which incorporates three major elements (Fraas & Newman, 2000). These three elements are as follows: (a) the establishment of a practically significant value; (b) the incorporation of the practical significance value into a non-nil null hypothesis and its alternative; and (c) the use of a randomization test to statistically test the non-nil null hypothesis.

In initial presentation of our suggested testing procedure (Fraas & Newman, 2000), we noted two key features of the procedure. First, we recommend that the non-nil null hypothesis should be statistically tested with a randomization test. We suggested this type of test because a randomization test will generate the distribution needed by the researcher to determine if the test statistic is statistically significant. Thus, the researcher would not need to incorporate special critical values as required by the use of the t and F values generated by most standardized statistical programs. In addition, a random sample is not required when a randomization test is conducted. Edgington (1995) stated the position that: "A randomization test is valid for any kind of sample, regardless of how the sample is selected. This is an extremely important property because the use of nonrandom samples is common in experimentation" (p.6).

Second, this procedure reflects our philosophical position regarding statistical hypothesis testing. That is, we believe the concepts of statistical and practical significance are both essential components of the evaluation process. And the level of change in a variable or the difference between group means that is defined to be practically significant must directly and thoughtfully be determined by the researchers and practitioners.

A Concern With the Implementation of the Testing Procedure

We believe researchers will be more inclined to utilize our suggested hypothesis testing procedure if they find it to be a relatively easy procedure to implement. One concern that we have with our testing procedure is the difficulty researchers may encounter in attempting to conduct a randomization test of a given non-nil null hypothesis. In our initial presentation of this testing process (Fraas & Newman, 2000), we utilized the Resampling Stats (Simon et al., 1999) computer software to conduct the randomization test of a non-nil null hypothesis. We believe that due to its generally non-user friendly computer environment, researchers may be discourage from using it, and thus, non-nil null hypotheses. The following section of this paper illustrates how researchers can simplify the process of conducting a randomization test of a non-nil null hypothesis by utilizing the computer software Resampling Stats Add-In for Excel (Blank et al., 1999), which is used in conjunction with the Microsoft Excel software.

An Illustration of the Recommended Testing Procedure

In the initial presentation of our suggested testing procedure we demonstrated its application to a set of data contained in a study conducted by Piirto, Beach, Cassone, Rogers, and Fraas (2000). In the Piirto et al. study, the authors were interested in determining whether high-school aged gifted students had higher intellectual scores than high-school aged non-gifted students. The intellectual scores measured the students' levels of desire for knowledge and inquiry. This illustration included a total of 49 gifted students and 51 non-gifted students. Each intellectual score was multiplied by 100 to facilitate the presentation of this illustration.

For the Piirto et al. (2000) data, the practical significance level for the difference between the two group means was set at a difference of four points, with the gifted group mean expected to exceed the non-gifted group by at least that many points. Thus, the non-nil null hypothesis and its corresponding alternative hypothesis were as follows:

H_0: The mean of the gifted students does not exceed the mean of the non-gifted students by more than four points.

H_1: The mean of the gifted students does exceed the mean of the non-gifted students by more than four points.

Resampling Stats Computer Software

In our original study (Fraas & Newman, 2000), this non-nil null hypothesis was tested with a randomization test, which was generated by the Resampling Stats (Simon et al., 1999) computer software. The specific program used to conduct the randomization test required to statistically test our non-nil null hypothesis is listed in Appendix A. Before the students scores were subjected to the randomization test through this program, however, the value of four was subtracted from each gifted student's score. The scores for the non-gifted students were not modified in this manner. The data file consisted of two variables. The first variable, which was entitled "group," was a dummy variable consisting of the values zero and one. The zero and one values represented the gifted and non-gifted students, respectively. The second variable, which was entitled "ntanew" consisted of the gifted students' modified scores and the non-gifted students non-modified scores.

The mean of the gifted students in the sample was 23.65. The mean modified score of the gifted students was 19.65, which was four points lower than their mean non-modified score. The standard deviation of gifted students' modified scores was 16.04, which, matched the standard deviation of their non-modified scores. The mean and standard deviation values for the non-gifted students' non-modified scores were 15.55 and 14.75, respectively. Due to its importance in the randomization test, the difference of 4.1 points between the mean modified score for the gifted students and the mean non-modified score for the non-gifted students was calculated.

Once the scores of the gifted students and the non-gifted students were entered into the randomization test program, it generated a distribution of 10,000 differences between the mean of the students randomly assigned to the gifted group and the mean of the students randomly assigned to the non-gifted group. The program calculated the proportion of the 10,000 values in the distribution that exceeded the value of 4.1, which was the difference between the mean of the gifted students' modified scores ($\bar{x}$ = 19.65) and the mean of the non-gifted students' non-modified scores ($\bar{x}$ = 15.55). The generated proportion of .092 was compared to an established maximum proportion of values in the distribution that we were willing to obtain and still reject the non-nil null hypothesis. We established this maximum proportion to be .05. Since the calculated proportion of .092 exceeds the established maximum proportion of .05, we were not willing to reject the non-nil null hypothesis. Thus, we concluded that any difference between the means of the gifted students and the non-gifted students in excess of four points, is more likely to occur by chance at a level greater than we were willing to accept.

Resampling Stats Add-In for Excel Computer Software

We believe that some researchers may find the construction of the programs required by the Resampling Stats (Simon et al., 1999) computer software to be difficult if not intimidating. Thus, they may be hesitant to utilize non-nil null hypotheses in their studies. This problem may be avoided if researchers consider using the Resampling Stats Add-In for Excel (Blank et al., 1999) computer software to conduct the randomization tests of their non-nil null hypotheses.

The Structure of the Data File. When utilizing the Resampling Stats Add-In for Excel (Blank et al., 1999) computer software, the data file does not have the same structure as the file used with the Resampling Stats (Simon et al., 1999) software. As was the case for the data file used with the Resampling Stats software, the data file constructed for the Resampling Stats Add-In for Excel software consisted of two variables. The variables, however, were not the same. The first variable consisted of the 49 modified scores for the gifted group. The modified scores for the gifted students were placed in rows 1 through 49 under column A. In Microsoft Excel terminology the scores were located in A1:A49. The second variable contained the 51 non-modified scores for the non-gifted group. These scores were placed in B1:B51.

The Steps Used to Obtain the Randomization Test. Once the data were entered into the Microsoft Excel file, the randomization test of the previously stated non-nil null hypothesis was obtained by completing the follow steps:

1. The area A1 to B51, which contained the modified scores of the gifted group and the non-modified scores of the non-gifted group, was highlighted. This action identified the scores that were to be randomly assigned to the two groups.

2. The "S" on the resampling toolbar, which is provided by the Resampling Stats Add-In for Excel (Blank et al., 1999) software, was clicked. This action produced a dialog box entitled "Matrix Shuffle." The highlighted area was automatically placed into the row entitled "Input Range" of this dialog box. In addition, the cell location "E1" was typed into the row labeled "Top Left Cell of Output Range." Finally, the "OK" button located in this dialog box was clicked. These actions identified the position on the worksheet for the random placement of students into the two groups. That is, the E1 specification set the cell position for the first score

randomly assigned to the gifted group. The other 48 scores randomly assigned to the gifted group were placed in the cells E2 through E49. The scores randomly assigned to the non-gifted group were placed in the cells F1 through F51.

3. The commands required to calculate the averages of the scores randomly assigned to the gifted and non-gifted groups were placed into the E52 and F52 cells, respectively. Cell E52 contained the command [=average(E1:E49)] and cell F52 contained the command [=average(F1:F51)]. Note that all commands presented in this section are contained in brackets, which are not part of the command. The command required to calculate the difference between the averages of the scores for the students randomly assigned to the gifted and non-gifted groups was [=E52-F52]. This command was placed in cell H52.

4. The area containing the cells E1 through F51 was highlighted. This operation designated where the scores would be placed each time they were randomly assigned to the groups.

5. The "RS" on the resampling toolbar is clicked, which accesses the "Repeat and Score" program. This action produces a dialog box entitled "Multiple Scoring" in which the "OK" button was clicked. Next, the cell H52 was double clicked. This action, which changes the color of cell H52, specified that the difference between the group means would be calculated for each of the random trials conducted. Next, a blank cell was double clicked. This action, which specifies the end of the cell selection process, produced another dialog box entitled "Multiple Score Cells". Once its "OK" button was clicked another dialog box entitled "Repeat and Score" was produced. The number of random trials, which was set at 10,000 for this example, was entered into the row entitled "Number of Trials." The "OK" button in this dialog box was

clicked. This action initiated the execution of the 10,000 trials and the calculation of the 10,000 differences between the two group means.

6. Once the computer completed the randomization process and the calculation of the 10,000 differences between the group means, a dialog box entitled "Execution Time" was displayed. The "OK" button in this dialog box was clicked. This action produced a dialog box entitled "View Output Sheet". The "Yes" button in this dialog box was clicked. This action allowed us to view the 10,000 difference values that were placed in cells A1 through A10000 on the output page.

7. Cell A10001 was highlighted on the output page and the command [=countif(A1:A10000,">=4.1")/10000] was inserted. This action instructed the computer to calculate the proportion of the 10,000 differences between the group means that exceeded the difference between the sample groups, which was 4.1 points.

Results of the Randomizaton Test. The value of .094, which was contained in cell A10001, corresponds to the proportion (.092) produced by the program executed by the Resampling Stats (Simon et al., 1999) computer software. Thus, the conclusions regarding the non-nil null hypothesis produced by each software resulted in the same decision, i.e., it was not rejected. It should be noted that due to the nature of randomization tests, the proportions generated by either randomization program will vary slightly from one analysis to another for any given set of data.

A Comparison of Randomization Test and t-Test Results

One issue that researchers may raise regarding the use of our suggested testing procedure relates to the use of a randomization test rather than an independent-samples t test to statistically test the non-nil null hypothesis. As previously noted, Selin and Lapsley (1985, 1993) suggested that the use of non-nil null hypotheses may require the use of critical values obtained from noncentralized t and F distributions. In addition, they noted that some of the complexities of using non-nil null hypotheses are not yet readily applicable in many designs. The question is: Do these concerns appear to significantly influence the testing of the non-nil null hypothesis for the data contained in our example? That is, could researchers obtain the same results if they used an independent-samples t test rather than a randomization test?

As an initial examination of this issue, we conducted an independent-samples t test of the modified scores for the gifted group and the non-modified scores for the non-gifted group. We used the SPSS version 10.0 (SPSS, Inc., 1999) computer software. The analysis produced an independent-samples t value of 1.33, which generated a one-tailed probability of .093. This probability value (.093), which is very close to the randomization test proportion value (.094) produced by the Resampling Stats Add-In for Excel (Blank et al., 1999) software, results in the same type of conclusion being drawn by either statistical test. That is, we were not willing to conclude that the scores for the gifted group and the non-gifted group differ by at least four points, which is the practical significance level.

In spite of the similar results produced by the randomization test and the independent-samples t test for our example, an important question remains. That is, would the randomization test and the independent-samples t test produce similar test values under various conditions such as unequal sample sizes, unequal variances, various practical significance levels or some combination of these conditions? We believe that an investigation of this question, possibly through a Monte Carlo study, may provide important information for researchers when deciding whether one should use a randomization test or an independent-samples t test in connection with our suggested hypothesis testing procedure.

Summary

In an earlier paper (Fraas & Newman, 2000) we proposed non-nil null hypotheses, which incorporated practical significance levels, be tested with randomization tests. One of the concerns we had with our suggested testing procedure was the difficulty researchers may encounter when conducting randomization tests. In this paper we illustrated how the Resampling Stats Add-In for Excel (Blank et al., 1999) computer software provides a simple way of conducting a randomization test of a non-nil null hypothesis that incorporates a practical significance level. We believe that researchers who are even slightly familiar with the Microsoft Excel computer software, will find that conducting randomization tests with the Resampling Stats Add-In for Excel computer software is a rather straight forward and simple process. We hope exposure to the Resampling Stats Add-In for Excel software in conjunction with our proposed hypothesis testing procedure will encourage researchers to use non-nil null hypotheses that incorporate practical significance levels.

In addition to demonstrating the use of the Resampling Stats Add-In for Excel (Blank et al., 1999) computer software, we compared the results obtained from an application of a randomization test and an independent-samples t test to the data contained in our example. For this example, the two types of statistical tests produced nearly identical test values and the same conclusion regarding the disposition of the non-nil null hypothesis. We believe that further investigation of the relative results produced by these two types of tests under various conditions may be important to researchers. If Monte Carlo studies indicate under what type of conditions

the results of the two tests are similar and under what type of conditions they are not, it may assist researchers in selecting an appropriate test.

References

Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. Journal of the American Statistical Association, 33, 526-542.

Blank, S., Seiter, C., & Bruce, P. (1999). Resampling Stats Add-In for Excel [Computer software]. Arlington, VA: Resampling Stats, Inc.

Carver, R. (1978). The case against statistical significance testing. Harvard Educational Review, 48, 378-399.

Carver, R. (1993). The case against statistical significance testing, revisited. Journal of Experimental Education, 61, 287-292.

Cohen, J. (1969). Statistical power analysis for the behavioral sciences. New York: Academic Press.

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

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

Cohen, J. (1990). Things I have learned (so far). American Psychologist, 45, 1304-1312.

Cortina, J. M. & Dunlap, W. P. (1997). Logic and purpose of significance testing. Psychological Methods, 2, 161-172.

Edgington, E. S. (1995). Randomization tests. New York: Marcel Dekker.

Falk, R. (1986). Misconceptions of statistical significance. Journal of Structural Learning, 9, 83-96.

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

Fisher, R. A. (1925). Statistical methods for research workers. London: Oliver & Boyd.

Fraas, J. W., & Newman, I. (2000, October). Testing for statistical and practical significance: A suggested technique using a randomization test. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL.

Frick, R. W. (1996). The appropriate use of null hypothesis testing. Psychological Methods, 1, 379-390.

Frick, R. W. (1999). Defending the statistical status quo. Theory & Psychology, 9, 183-189.

Hagen, R. L. (1997). In praise of the null hypothesis statistical test. American Psychologist, 52 (1), 15-24.

Huberty, C. J. (1987). On statistical testing. Educational Researcher 16 (8), 4-9.

Huberty, C. J. (1993). Historical origins of statistical testing practices: The treatment of Fisher versus Neyman-Pearson views in textbooks. The Journal of Experimental Education, 61 (4), 317-333.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. Educational and Psychological Measurement 56 (5), 746-759.

Levin, J. R. (1993). Statistical significance testing from three perspectives. The Journal of Experimental Education, 61 (4), 378-382.

Levin, J. R. (1996). Discussant of significance tests: Should they be banned in APA journals? Symposium presented at the annual meeting of The American Psychological Association, Toronto, Canada.

Levin, J. R. (1998). To test or not to test H_0? Educational and Psychological Measurement, 58 (2), 313-333.

Levin, J. R. & Robinson, D. H. (2000). Rejoinder: Statistical hypothesis testing, effect-size estimate, and the conclusion coherence of primary research studies. Educational Researcher, 29 (1), 34-36.

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. Philosophy of Science, 34, 103-115.

Piirto, J., Beach, L., Cassone, G., Rogers, R., & Fraas, J. (2000, August). Is overexcitability a differential personality attribute of high IQ youth? A comparison study of identified gifted and vocational high school teenagers using the overexcitability questionnaire (OEQ). Paper presented at the European Council for High Ability (ECHA) Conference, Debrecen, Hungary.

Resampling Stats [computer software] (1999). Arlington, VA: Resampling Stats.

Robinson, D. H. & Levin, J. R. (1997). Reflections on statistical and substantive significance, with a slice of replication. Educational Researcher, 26 (5), 21-26.

Rozeboom, W. W. (1960). The fallacy of the null-hypothesis significance test. Psychological Bulletin, 57 (5), 416-428.

Serlin, R. A. & Lapsley, D. K. (1993). Rational appraisal of psychological research and the good-enough principle. In G. Keren & C. Lewis (Eds.), A handbook for data analysis in the behavioral sciences: Methodological issues (pp. 199-228). Hillsdale, NJ: Erlbaum.

Serlin, R. C. & Lapsley, D. K. (1985). Rationality in psychological research: The good-enough principle. American Psychologist 40 (1), 73-83.

Shaver, J. P. (1980, April). Readdressing the role of statistical tests of significance. Paper presented at the annual meeting of The American Educational Research Association, Boston.

Shaver, J. P. (1993). What statistical significance testing is, and what it is not. Journal of Experimental Education, 61 (4), 293-316.

Simon, J., Weidenfeld, D., Bruce, P., & Puig, C. (1999). Resampling Stats (Version 4.2b2) [Computer software]. Arlington, VA: Resampling Stats, Inc.

SPSS, Inc. (1999). SPSS for Windows (Version 10.0) [Computer software]. Chicago, IL: SPSS, Inc.

Thompson, B. (1989a). Statistical significance, result importance, and generalizability: Three noteworthy but somewhat different issues. Measurement and Evaluation in Counseling and Development, 22, 2-5.

Thompson, B. (1989b). Asking "what if" questions about significance tests. Measurement and Evaluation in Counseling and Development, 22, 66-68.

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

Thompson, B. (1997). Editorial policies regarding statistical significance tests: Further comments. Educational Researcher, 26 (5), 29-32.

Thompson, B. (1998). Review of what if there were no significance tests? Educational and Psychological Measurement, 58 (2), 334-346.

Thompson, B. (1999a). If statistical significance tests are broken/misused, what practices should supplement or replace them? Theory and Psychology, 9 (2), 165-181.

Thompson, B. (1999b). Statistical significance tests, effect size reporting and the vain pursuit of pseudo-objectivity. Theory and Psychology, 9 (2), 191-196.

Thompson, B. (1999c). Journal editorial policies regarding statistical significance tests: Heat is to fire as p is to importance. Educational Psychology Review, 11 (2), 157-169.

Appendix A

Computer Program for the Randomization Test

```
add 10000 0 rep
maxsize default 10000
read file "data effect tab" group ntanew
count group=0 groupg 'Number of observations in Group0'
count group=1 groupv 'Number of observations in Group1'
print groupg groupv
add groupg+1 minv
add groupg groupv maxv
print minv maxv
tagsort group key
take ntanew key value$
sort group group$
take value$ 1,groupg g 'these numbers will depend on the number of observations in the gifted
group'
take value$ minv,maxv v 'these numbers will depend on the number of observations in the
vocational group'
mean g meang
mean v meanv
stdev g SDg
stdev v SDv
subtract meang meanv diff
print meang SDg meanv SDv diff
repeat rep
        shuffle ntanew all$
        take all$ 1,groupg gifted$
        take all$ minv,maxv voc$
        mean gifted$ meang$
        mean voc$ meanv$
        subtract meang$ meanv$ diff$
        score diff$ z
end
count z >= diff k
divide k rep propor
print  propor
histogram z
```

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE

(Specific Document)

TM032469

I. DOCUMENT IDENTIFICATION:

Title: Testing Non-Nil Null Hypotheses Using Resampling Stats Add-Ins for Microsoft Excel	
Author(s): John W. Fraas; Isadore Newman	
Corporate Source: Ashland University	Publication Date: 2/16/01

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] ← Sample sticker to be affixed to document

Check here
Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document → [ ]

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

or here
Permitting reproduction in other than paper copy.

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: John W Fraas	Position: Trustees' Professor
Printed Name: John W. Fraas	Organization: Ashland University
Address: 220 Andrews Hall, Ashland University, Ashland, Ohio 44805	Telephone Number: (419) 289-5930
	Date: 2/21/01

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC , *or*, if you wish ERIC to cite the availability of this document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents which cannot be made available through EDRS).

Publisher/Distributor:	
Address:	
Price Per Copy:	Quantity Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name and address of current copyright/reproduction rights holder:

Name:

Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

If you are making an unsolicited contribution to ERIC, you may return this form (and the document being contributed) to:

ERIC Facility
1301 Piccard Drive, Suite 300
Rockville, Maryland 20850-4305
Telephone: (301) 258-5500