DOCUMENT RESUME

ED 440 121 TM 030 732

AUTHOR	Thompson, Bruce
TITLE	The APA Task Force on Statistical Inference (TFSI) Report as a Framework for Teaching and Evaluating Students' Understandings of Study Validity.
PUB DATE	2000-04-00
NOTE	22p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*College Students; Comprehension; Computer Assisted Instruction; Evaluation Methods; Higher Education; Models; Research Reports; Research Utilization; *Statistical Inference; Statistics; *Student Evaluation; *Validity; *World Wide Web
IDENTIFIERS	*American Psychological Association

ABSTRACT

Web-based statistical instruction, like all statistical instruction, ought to focus on teaching the essence of the research endeavor: the exercise of reflective judgment. Using the framework of the recent report of the American Psychological Association (APA) Task Force on Statistical Inference (Wilkinson and the APA Task Force on Statistical Inference, 1999), this paper explores background for and potential instructional design of Web-based instruction involving: (1) effect-size reporting and interpretation and (2) score reliability evaluation. (Contains 57 references.) (Author/SLD)

rival_hy.wp1 3/11/00

The APA Task Force on Statistical Inference (TFSI) Report as a Framework for Teaching and Evaluating Students' Understandings of Study Validity

Bruce Thompson

Texas A&M University
and
Baylor College of Medicine (Houston)

Paper presented at the annual meeting of the American Educational Research Association (session #18.64), New Orleans, April 25, 2000.

Abstract

Web-based statistical instruction, like all statistical instruction, ought to focus on teaching the essence of the research endeavor: the exercise of reflective judgment. Using the framework of the recent report of the APA Task Force on Statistical Inference (Wilkinson & The APA Task Force on Statistical Inference, 1999), the present paper explores background for and potential instructional design of Web-based instruction involving (a) effect size reporting and interpretation and (b) score reliability evaluation.

In 1993, Carl Kaestle, prior to his term as President of the National Academy of Education, published in the Educational Researcher an article titled, "The Awful Reputation of Education Research." It is noteworthy that the article took as a given the conclusion that educational research suffers an awful reputation, and rather than justifying this conclusion, Kaestle focused instead on exploring the etiology of this presumed reality. For example, Kaestle (1993) noted that the education R&D community is seemingly in perpetual disarray, and that there is a

> ...lack of consensus--lack of consensus on goals, lack of consensus on research results, and lack of a united front on funding priorities and procedures.... [T]he lack of consensus on goals is more than political; it is the result of a weak field that cannot make tough decisions to do some things and not others, so it does a little of everything... (p. 29)

Although Kaestle (1993) did not find it necessary to provide a warrant for his conclusion that educational research has an awful reputation, others have directly addressed this concern.

The National Academy of Science evaluated educational research generically, and found "methodologically weak research, trivial studies, an infatuation with jargon, and a tendency toward fads with a consequent fragmentation of effort" (Atkinson & Jackson, 1992, p. 20). Others also have argued that "too much of what we see in print is seriously flawed" as regards research methods, and that "much of the work in print ought not to be there" (Tuckman, 1990, p. 22). Gall, Borg and Gall (1996) concurred, noting that "the quality of published studies in education and related disciplines is, unfortunately, not high" (p. 151).

Indeed, empirical studies of published research involving methodology experts as judges corroborate these impressions. For example, Hall, Ward and Comer (1988) and Ward, Hall and Schramm (1975) found that over 40% and over 60%, respectively, of published research was seriously or completely flawed. Wandt (1967) and Vockell and Asher (1974) reported similar results from their empirical studies of the quality of published research. Dissertations, too, have been examined, and have been found methodologically wanting (cf. Thompson, 1988, 1994).

Purpose of the Present Paper

These troubling realizations have led to some self-scrutiny on the part of professors of educational research as regards the training we provide to our students. Certainly, in an environment where less and less space in curriculum is allocated to methodological teaching (cf. Aiken, West, Sechrest, Reno with Roediger, Scarr, Kazdin, & Sherman, 1990), not all these problems can be laid at the doors of methodology professors.

Still, there is clearly some room for improvement in what we do. The present paper offers one perspective on potential vehicles for improvement.

Today increasing numbers of faculty are utilizing Web-based instructional tools to facilitate research training. Some applications allow students, for example, to "drag" data points in histograms or scattergrams, and watch the associated incremental changes in statistical indices. Applications such as these provide a user-friendly environment in which students can readily ask "what-if" questions and explore statistical dynamics.

One important skill that students must master is recognizing the various rival hypotheses that may explain the results in the literature they review, or in their own research. One way to teach such skills is to present synopses or excerpts from actual studies in a Web environment, and then allow students to enter "chat rooms" to offer alternative explanations for detected effects. Given the frailties of the human reviewer system that guards the gates of the publication citadel (Peters & Ceci, 1982), students must learn early to evaluate critically all that they read, or students will invariably otherwise rely on published specious claims.

One potential source of study vignettes is the popular books offered by Huck and his colleagues (cf. 2000; Huck & Cormier, 1996). Particularly relevant to the current focus are the vignettes presented by Huck and Sandler (1979).

Huck and Sandler (1979) presented a series of short study synopses in which various rival hypotheses might be invoked to explain reported results. The reader is then challenged to formulate these possibilities, and the back portion of the book presents possible alternative study explanations.

The purpose of the vignettes was characterized as facilitating "logical thinking" (p. xvi), and assisting "people in discriminating *possible* rival hypotheses and *plausible* rival hypotheses" (p. xiv). In other words, the purpose of the problems and their proposed solutions is to teach students to **think and evaluate critically** the claims in published (or unpublished) research!

The present treatise takes as a given both the utility and the import of just such an instructional emphasis. My own teaching is similarly focused. However, my pedagogic bias is frankly toward Socratic instruction with an emphasis on heuristic techniques requiring discovery learning on the part of students, rather than toward Web-based instruction, except as a fairly peripheral (but powerful) instructional aid.

Vignettes such as the Huck-Sandler examples might be used in Web-based instruction as a tool to help students think reflectively. However, my purpose here is to argue that any such instruction should be grounded in the contemporary analytic principles embedded within the recent report of the APA Task Force on Statistical Inference (Wilkinson & The APA Task Force on Statistical Inference, 1999).

Here I will advocate emphasis on two of these principles. Along the way, in each arena I will also cite some related illustrative features of Web-based instruction that I would find useful.

Principle #1: Report and Interpret Effect Sizes

Background

Statistical significance has a long history (cf. Huberty, 1993; Huberty & Pike, 1999). Recently, overreliance on statistical tests has been bluntly criticized (cf. Cohen, 1994; Daniel, 1998; Schmidt, 1996; Thompson, 1996, 1999c). For example, Tryon (1998) recently lamented,

> [T]he fact that statistical experts and investigators publishing in the best journals cannot consistently interpret the results of these analyses is extremely disturbing. Seventy-two years of education have resulted in minuscule, if any, progress toward correcting this situation. It is difficult to estimate the handicap that widespread, incorrect, and intractable use of a primary data analytic method has on a scientific discipline, but the deleterious effects are doubtless substantial... (p. 796)

Indeed, several <u>empirical</u> studies have shown that many researchers do not fully understand the statistical tests that they employ (Mittag & Thompson, in press; Nelson, Rosenthal & Rosnow, 1986; Oakes, 1986; Rosenthal & Gaito, 1963; Zuckerman, Hodgins, Zuckerman & Rosenthal, 1993).

Of course, even many defenders of statistical tests (cf. Abelson, 1997; Cortina & Dunlap, 1997; Frick, 1996; Robinson & Levin, 1997; also see Harlow, Mulaik & Steiger, 1997, and reviews by Levin, 1998 and Thompson, 1998) agree that the tests have sometimes been abused or misinterpreted. One area of agreement across many scholars writing on these topics is that researchers ought to report and interpret effect sizes (cf. Kirk, 1996; Thompson, 1996). Snyder and Lawson (1993) explain what effect sizes are and summarize the many available choices (e.g., Cohen's _d_, eta^2, $omega^2$).

In 1996, the APA Board of Scientific Affairs appointed its Task Force on Statistical Inference to make recommendations regarding whether statistical significance tests should be banned from APA journals (Azar, 1997; Shea, 1996). In its recently published article, the Task Force emphasized, "_Always_ provide some effect-size estimate when reporting a p value" (Wilkinson & The APA Task Force on Statistical Inference, 1999, p. 599, emphasis added). Later the Task Force also wrote,

> _Always_ present effect sizes for primary outcomes.... It helps to add brief comments that place these effect sizes in a practical and theoretical context.... We must stress again that reporting and interpreting effect sizes in the context of previously reported effects is _essential_ to good research. (p. 599, emphasis added)

Of course, the 1994 APA publication manual, incorporated by reference into the editorial policies of hundreds if not thousands of behavioral science journals, did "encourage" (p. 18) effect size reporting. However, as summarized by Vacha-Haase, Nilsson, Reetz, Lance & Thompson, in press), 11 _empirical_ studies of 1 or 2 post-1994 volumes of 23 different journals confirm that this "encouragement" has been ineffectual (cf. Keselman et al., 1998).

Thompson (1999b) explained why the APA "encouragement" has been so ineffective. He noted that only "encouraging" effect size reporting

> presents a self-canceling mixed-message. To present an "encouragement" in the context of strict absolute standards regarding the esoterics of author note placement, pagination, and margins is to send the message, "these myriad requirements count, this encouragement doesn't." (p. 162)

Consequently, various journals now *require* effect size reporting (e.g., Heldref Foundation, 1997, pp. 95-96; Murphy, 1997). Such journals include:

Educational and Psychological Measurement;
Journal of Agricultural Education;
Journal of Applied Psychology;
Journal of Consulting and Clinical Psychology;
Journal of Early Intervention;
Journal of Experimental Education;
Journal of Learning Disabilities;
Language Learning; and
The Professional Educator.

Editors at these journals will soon ask their editorial boards to approve such a requirement:

Journal of Mental Health Counseling; and
Research in the Schools.

<u>Web Instruction on Effect Size-related Concepts</u>

A fundamental concept in evaluating effect sizes as against statistical significance is the concept that

> The calculated p values in a given study are a function of several study features, but are particularly influenced by the confounded, joint influence of study sample size and study effect sizes. Because p values are confounded indices, in theory 100 studies with varying sample sizes and 100 different effect sizes could each have the same single $p_{CALCULATED}$, and 100 studies with the same single effect size could each have 100 different values for $p_{CALCULATED}$. (Thompson, 1999c, pp. 169-170)

There are various Web applications that could be employed to teach insights related to this concept.

One vehicle for such instruction might sequentially present a series of different studies, each with a fixed roughly-identical single effect size, but different n's and consequently each with different p values. Table 12 in my 1999 AERA Invited Address (Thompson, 1999a) presents just such a series. Students might then be asked both (a) to interpret each study's individual results and (b) to interpret the *set of studies as a holistic series*, as an emerging cumulating literature might be interpreted.

Another alternative would be to present a series of studies in which effect sizes and sample sizes varied but that each yielded an essentially fixed $p_{CALCULATED}$ value. Table 13 from Thompson (1999a) presents such a series. Again, students might be presented with the same two interpretation challenges. These exercises would force students to have the necessary "ah ha" experience related to the influences of sample sizes on *p* values, problems with interpreting *p* values without consulting effect sizes, and the importance of effect sizes.

A related series of vignette presentations might present both "uncorrected" (e.g., eta^2) and "corrected" (e.g., $omega^2$) effect sizes. This particular series would help students to understand what sampling error variance is and what three factors cause sampling error variance (Thompson, 1999a).

Principle #2: Evaluate, Report and Interpret Score Reliability

Background

In addition to strongly emphasizing the importance of effect sizes, the APA Task Force on Statistical Inference also emphasized that

> It is important to remember that a test is not reliable or unreliable. Reliability is a property of the scores on a test for a particular population of examinees (Feldt & Brennan, 1989). Thus, authors should provide reliability coefficients of the scores for the data being analyzed even when the focus of their research is not psychometric. Interpreting the size of observed effects requires an assessment of the reliability of the scores. (Wilkinson & The APA Task Force on Statistical Inference, 1999, p. 596)

Thompson and Vacha-Haase (2000) present a thorough (i.e., protracted) elaboration of these issues.

Unfortunately, empirical studies indicate that most authors do not evaluate and report the reliability coefficients for their own data (cf. Meier & Davis, 1990; Snyder & Thompson, 1998; Thompson & Snyder, 1998; Vacha-Haase, Ness, Nilsson & Reetz, 1999; Willson, 1980). Nor do authors who only merely cite reliability coefficients from previous studies even explicitly compare (a) their own sample compositions and (b) their own sample score variabilities with those in the previous studies, to thus establish that the previous coefficients might be generalized (Vacha-Haase & Kogan, in press)!

These dismal patterns of practice may occur because many researchers may not really understand what score reliability is (Thompson & Vacha-Haase, 2000). Certainly such misperceptions ought be expected, given the short shrift afforded measurement training in doctoral programs through the United States (Aiken, West, Sechrest, Reno with Roediger, Scarr, Kazdin, & Sherman, 1990).

Web Instruction on Score Reliability-related Concepts

Reliability is not a property of a test *per se*, and rather inures to a particular set of scores (Thompson & Vacha-Haase, 2000). Reliability is driven by score variability, and the generalizability of score reliability is driven by the comparability of the composition of samples with the sample composition used in a referenced prior reliability (e.g., normative) study (Crocker & Algina, 1986, p. 144).

The importance of **sample variability** as regards score reliability might be taught by building an applet to generate pairs of scores in ascending order for a fixed sample size, with a random number generator adding or subtracting a small random additive adjustment to each score in each pair. The applet might request as input the desired SD of the scores. The score pairs modeling test-retest reliability, for example, would then be generated, and the resulting reliability coefficient would be reported.

Students would then grasp at a deeper level **why** score variability impacts score reliability. In classical measurement theory reliability deals with the consistency with which individuals are rank ordered by measurement across parallel test forms, repeated measurements, and so forth. The degree of the homogeneity of the scores (i.e., $\underline{SD}_X$) directly affects the consistency (e.g., stability) of the score orderings because, as Cunningham (1986) explained,

> [W]hen scores are bunched together, a small [random measurement error] change in raw score will lead to large changes in relative position. If scores are spread out (variability is high), it is more likely that the relative position in the group will remain stable across the two forms of the test and the correlation coefficient will be relatively large. (p. 114)

In other words, "greater differences between the scores of individuals reduce the possibility of shifting positions" (Linn & Gronlund, 1995, p. 101).

The influence of **sample composition** on score reliability might be taught by generating a population scattergram for a test-retest situation in which score reliability differed across males and females. A Web applet might then allow sampling of different numbers of males and females, and report resulting reliability coefficients for each sample. Students would see score reliability coefficients fluctuate across every variation in sample composition.

Summary

Good statistical instruction is instruction that teaches students to understand dynamics within statistics as different characterizations of data. Statistics mastery does not equate with the rote memorization of formulae. Rather, the essence of conducting research is the exercise of reflective judgment. As Huberty and Morris (1988, p. 573) noted: "As in all of statistical inference, subjective judgment cannot be avoided. Neither can reasonableness!"

Web-based statistical instruction, like all statistical instruction, ought to focus on teaching the essence of the research endeavor: the exercise of reflective judgment. Using the framework of the recent report of the APA Task Force on Statistical Inference (Wilkinson & The APA Task Force on Statistical Inference, 1999), the present paper explored background for and potential instructional design of Web-based instruction involving (a) effect size reporting and interpretation and (b) score reliability evaluation.

References

Abelson, R.P. (1997). A retrospective on the significance test ban of 1999 (If there were no significance tests, they would be invented). In L.L. Harlow, S.A. Mulaik & J.H. Steiger (Eds.), What if there were no significance tests? (pp. 117-141). Mahwah, NJ: Erlbaum.

Aiken, L.S., West, S.G., Sechrest, L., Reno, R.R., with Roediger, H.L., Scarr, S., Kazdin, A.E., & Sherman, S.J. (1990). The training in statistics, methodology, and measurement in psychology. American Psychologist, 45, 721-734.

Atkinson, R.C., & Jackson, G.B. (Eds.). (1992). Research and education reform: Roles for the Office of Educational Research and Improvement. Washington, DC: National Academy of Sciences. (ERIC Document Reproduction Service No. ED 343 961)

Azar, B. (1997). APA task force urges a harder look at data. The APA Monitor, 28(3), 26.

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

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

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston.

Cunningham, G.K. (1986). Educational and psychological measurement. New York: Macmillan.

Daniel, L.G. (1998). Statistical significance testing: A historical overview of misuse and misinterpretation with implications for the editorial policies of educational journals. Research in the Schools, 5(2), 23-32.

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

Gall, M.D., Borg, W.R., & Gall, J.P. (1996). Educational research: An introduction (6th ed.). White Plains, NY: Longman.

Hall, B.W., Ward, A.W., & Comer, C.B. (1988). Published educational research: An empirical study of its quality. Journal of Educational Research, 81, 182-189.

Harlow, L.L., Mulaik, S.A., & Steiger, J.H. (Eds.). (1997). What if there were no significance tests?. Mahwah, NJ: Erlbaum.

Heldref Foundation. (1997). Guidelines for contributors. Journal of Experimental Education, 65, 95-96.

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

Huberty, C.J, & Morris, J.D. (1988). A single contrast test procedure. Educational and Psychological Measurement, 48, 567-578.

Huberty, C.J, & Pike, C.J. (1999). On some history regarding statistical testing. In B. Thompson (Ed.), Advances in social science methodology (Vol. 5, pp. 1-22). Stamford, CT: JAI Press.

Huck, S.W. (2000). Reading statistics and research (3rd ed.). New York: Addison Wesley Longman.

Huck, S.W, & Cormier, W.G. (1996). Reading statistics and research

(2nd ed.). New York: Harper Collins.

Huck, S.W, & Sandler, H.M. (1979). Rival hypotheses: Alternative interpretations of data based conclusions. New York: Harper and Row.

Kaestle, C.F. (1993). The awful reputation of education research. Educational Researcher, 22(1), 23, 26-31.

Keselman, H.J., Huberty, C.J, Lix, L.M., Olejnik, S., Cribbie, R., Donahue, B., Kowalchuk, R.K., Lowman, L.L., Petoskey, M.D., Keselman, J.C., & Levin, J.R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. Review of Educational Research, 68, 350-386.

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

Levin, J.R. (1998). To test or not to test H_0? Educational and Psychological Measurement, 58, 311-331.

Linn, R.L., & Gronlund, N.E. (1995). Measurement and assessment in teaching (7th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Meier, S.T., & Davis, S.R. (1990). Trends in reporting psychometric properties of scales used in counseling psychology research. Journal of Counseling Psychology, 37, 113-115.

Mittag, K.C., & Thompson, B. (in press). A national survey of AERA members' perceptions of statistical significance tests and other statistical issues. Educational Researcher.

Murphy, K.R. (1997). Editorial. Journal of Applied Psychology, 82, 3-5.

Nelson, N., Rosenthal, R., & Rosnow, R.L. (1986). Interpretation of significance levels and effect sizes by psychological researchers. American Psychologist, 41, 1299-1301.

Oakes, M. (1986). Statistical inference: A commentary for the social and behavioral sciences. New York: Wiley.

Peters, D.C., & Ceci, S.J. (1982). Peer review practices of psychological journals: The fate of published articles, submitted again. The Behavioral and Brain Sciences, 5, 187-255.

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

Rosenthal, R., & Gaito, J. (1963). The interpretation of level of significance by psychological researchers. Journal of Psychology, 55, 33-38.

Schmidt, F.L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. Psychological Methods, 1, 115-129.

Shea, C. (1996). Psychologists debate accuracy of "significance test." Chronicle of Higher Education, 42(49), A12, A16.

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. Journal of Experimental Education, 61, 334-349.

Snyder, P.A., & Thompson, B. (1998). Use of tests of statistical significance and other analytic choices in a school psychology journal: Review of practices and suggested alternatives. School Psychology Quarterly, 13, 335-348.

Thompson, B. (1988, November). Common methodology mistakes in dissertations: Improving dissertation quality. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY. (ERIC Document Reproduction Service No. ED 301 595)

Thompson, B. (1994, April). Common methodology mistakes in dissertations, revisited. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 368 771)

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

Thompson, B. (1998). Review of *What if there were no significance tests?* by L. Harlow, S. Mulaik & J. Steiger (Eds.). Educational and Psychological Measurement, 58, 332-344.

Thompson, B. (1999a, April). Common methodology mistakes in educational research, revisited, along with a primer on both effect sizes and the bootstrap. **Invited address** presented at the annual meeting of the American Educational Research Association, Montreal. (ERIC Document Reproduction Service No. ED 429 110)

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

Thompson, B. (1999c). If statistical significance tests are broken/misused, what practices should supplement or replace them?. Theory & Psychology, 9(2), 167-183.

Thompson, B., & Snyder, P.A. (1998). Statistical significance and reliability analyses in recent JCD research articles. Journal of Counseling and Development, 76, 436-441.

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics *is* datametrics: The test is not reliable. Educational and Psychological Measurement, 60, 174-195.

Tryon, W.W. (1998). The inscrutable null hypothesis. American Psychologist, 53, 796.

Tuckman, B.W. (1990). A proposal for improving the quality of published educational research. Educational Researcher, 19(9), 22-24.

Vacha-Haase, T. & Kogan, L.R. (in press). Author explicit comparisons of sample compositions and variabilities in published studies with those reported in test manuals: Implications for score reliability inferences. Educational and Psychological Measurement.

Vacha-Haase, T., Ness, C., Nilsson, J., & Reetz, D. (1999). Practices regarding reporting of reliability coefficients: A review of three journals. Journal of Experimental Education, 67, 335-341.

Vacha-Haase, T., Nilsson, J.E., Reetz, D.R., Lance, T.S. & Thompson, B. (in press). Reporting practices and APA editorial policies regarding statistical significance and effect size. Theory & Psychology.

Vockell, E.L., & Asher, W. (1974). Perceptions of document quality and use by educational decision makers and researchers. American Educational Research Journal, 11, 249-258.

Wandt, E. (1967). An evaluation of educational research published in journals (Report of the Committee on Evaluation of Research). Washington, DC: American Educational Research Association.

Ward, A.W., Hall, B.W., & Schramm, C.E. (1975). Evaluation of published educational research: A national survey. American Educational Research Journal, 12, 109-128.

Wilkinson, L., & The APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. American Psychologist, 54, 594-604. [reprint available through the APA Home Page: **http://www.apa.org/journals/amp/amp548594.html**]

Willson, V.L. (1980). Research techniques in AERJ articles: 1969 to 1978. Educational Researcher, 9(6), 5-10.

Zuckerman, M., Hodgins, H.S., Zuckerman, A., & Rosenthal, R. (1993). Contemporary issues in the analysis of data: A survey of 551 psychologists. Psychological Science, 4, 49-53.

AERA

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM030732

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: The APA Task Force on Statistical Inference (TFSI) Report as a Framework for Teaching and Evaluating Students' Understandings of Study Validity	
Author(s): BRUCE THOMPSON	
Corporate Source:	Publication Date: 4/25/00

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 media, and sold through the ERIC Document Reproduction Service (EDRS). 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 and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
XXX		
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic 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.

Sign here,→ please

Signature:	Printed Name/Position/Title: BRUCE THOMPSON, Prof	
Organization/Address: TAMU DEPT EDUC PSYC College Station, TX 77843-4225	Telephone: 979/845-1335	FAX:
	E-Mail Address:	Date: 3/11/00

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 the 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 that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

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

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

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com