ED 466 775 TM 034 286

AUTHOR	Du, Yunfei
TITLE	Sampling Theory and Confidence Intervals for Effect Sizes: Using ESCI To Illustrate "Bouncing" Confidence Intervals.
PUB DATE	2002-02-00
NOTE	18p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Austin, TX, February 14-16, 2002). Figures may not reproduce well.
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Computer Software; *Effect Size; Error of Measurement; *Meta Analysis; Sample Size; *Sampling; *Theories
IDENTIFIERS	*Confidence Intervals (Statistics); *Resampling Techniques

ABSTRACT

This paper discusses the impact of sampling error on the construction of confidence intervals around effect sizes. Sampling error affects the location and precision of confidence intervals. Meta-analytic resampling demonstrates that confidence intervals can haphazardly "bounce" around the true population parameter. Special software with graphical output, Exploratory Software for Confidence Intervals (G. Cuming, 2001; ESCI), is used to make this discussion concrete using a small heuristic sample. (Contains 3 tables, 1 figure, and 15 references.) (Author/SLD)

Running head: CONFIDENCE INTERVALS

Sampling Theory and Confidence Intervals for Effect Sizes: using ESCI to Illustrate "Bouncing" Confidence Intervals

Yunfei Du
University of North Texas

Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX, February 14-16, 2002.

Abstracts

The present paper discusses the impact of sampling error on the construction of confidence intervals around effect sizes. Sampling error affects the location and precision of confidence intervals. Meta-analytic re-sampling demonstrates that confidence intervals can haphazardly "bounce" around the true population parameter. Special software with graphical output is used to make this discussion concrete using a small heuristic sample.

Sampling theory and confidence intervals for effect sizes: using ESCI to illustrate "bouncing" confidence intervals

Statistical significance testing is one of the most frequently used statistical techniques in educational and psychological research. Huberty and Pike (1999) provide a through history of statistical significance in the social sciences. Finch, Cumming, and Thompson (2001) review the use of statistical significance testing in the Journal of Applied Psychology across six decades. Without doubt, null hypothesis significance testing has dominated the social science literature as the primary means of evaluating the import of findings. However, it has been well-documented that, given a large enough sample, a statistically significant result may always be found even when there is very little association between the independent and dependent variables (Cohen, 1994; Craig, Eison, & Metze, 1976). Therefore, some statisticians have suggested reporting effect sizes and make a distinction between statistical and practical significance (Cohen, 1994; Henson & Smith, 2000; Thompson, 1988). The reporting and interpretation of effect sizes is critical to good statistical practice. As the APA Task Force on Statistical Inference noted (Wilkinson & APA Task Force on Statistical Inference, 1999):

> It is hard to imagine a situation in which a dichotomous accept-reject decision is better than reporting an actual *p*-value or, better still, a confidence interval… *Always* provide some effect-size

> estimate when reporting a *p*-value. (p. 599, emphasis added)

Influenced by the Task Force report and current trends in the field, the recent fifth edition of the *APA Publication Manual* (APA, 2001) called the "failure to report effect sizes" a "defect in the design and reporting of research" (p. 5). The *Publication Manual* later observed: "For the reader to fully understand the importance of your findings, it is almost always necessary to include some index of effect size or strength of relationship in your Results section" (p. 25).

The fifth edition *Publication Manual* went further and also emphasized the role of confidence intervals (CIs) in result interpretation:

> The reporting of confidence intervals can be an extremely effective way of reporting results. Because confidence intervals combine information on location precision and can often be directly used to infer significance levels, they are, in general, the best reporting strategy. The use of confidence intervals is therefore *strongly* recommended. (p. 22, emphasis added)

Current trends in social science research support the inclusion and interpretation of effect sizes. The APA Publication Manual (APA, 2001) supports confidence intervals as one of the best reporting strategies. Therefore, the use of confidence intervals around effect sizes has obvious benefits for understanding research results. A confidence around an effect size informs the reader of (a) the point estimate of the population effect and (b) the degree of precision based on some level of confidence. Cumming and Finch (2001) provide an accessible introduction to the application of confidence intervals for effect sizes (i.e., Cohen's d).

The purpose of the present paper is to discuss the use of confidence intervals around effect sizes. More specifically, the impact of sampling error on the degree of precision in confidence intervals is examined. Because samples vary in the amount of sampling error present, the width of confidence intervals can vary from sample to sample even at the same level of confidence (e.g., 95%). Further, the possibility that confidence intervals around effects can be examined meta-analytically across studies is explored.

Definition of Confidence Intervals

Many textbooks give common definitions of confidence intervals (Felder & Thompson, 2001). For example, Hinkle, Wiersma, and Jurs (1998) defined a confidence interval (CI) in one-sample case for the mean as "a range of values that we are confident (but not certain) contains the population parameter" (p. 220). Moore and McCabe (1993) provided a more general definition: "A level C confidence interval for a parameter θ is an interval computed from sample data by a method that has probability C of producing an interval containing the true value of θ" (p. 433). In reality, whether or not a confidence captures a population parameter is dichotomous decision - either the interval captures the parameter or it does not (Cumming & Finch, 2001; Thompson, 2001). The level of confidence for an interval, say 95% or 99%, actually refers to the percentage of intervals that would capture the parameter across many studies. The mata-analytic perspective many help some researchers keep in mind that a given interval actually may not contain the population parameter even if the interval does not capture the null hypothesis value.

Purpose for using Confidence Interval with Effect Sizes

CIs are usually used in hypothesis testing. However, as Thompson (1998) stated, "If we mindlessly interpret a CI with reference to whether the interval subsumes zero, we are doing little more than nil hypothesis statistical testing" (p. 800).

CIs are most useful in comparison with intervals from related prior studies, instead of comparing with zero as the assumed value of the nil null hypotheses. In this vein, CIs can be most useful in meta-analysis as researchers meta-analytically evaluate CIs across studies. In fact, confidence intervals themselves can be meta-analytically synthesized which allows researchers to get, based on a history of research, the best point estimate of a population effect and a best estimate of the confidence intervals that should be around that effect. Just as the meta-analysis of effects yields greater accuracy in estimating true population effects, meta-analysis of confidence intervals increase our precision level and generally yields a more narrow, meta-analytic interval.

Recently, Cumming and Finch (2001) presented new software that automates and illustrates the concepts of non-central distributions and CIs for effect sizes. This user-friendly software runs under Microsoft Excel and is called Exploratory Software for Confidence Intervals (ESCI, pronounced "esky"). Using this software, the present paper illustrates the concept of sampling theory in the calculating of effects (and subsequently CIs) across studies, including the influence of sampling error.

Sampling Error and Sampling Distributions

In inferential statistics, a hypothesis is tested based on the assumption that the researcher can logically establish a hypothesized value for a population parameter, or in other words, we hope to make inferences about the population on the basis of what we observe in the sample. Of course, although we want to make inference from the sample to the population, what we actually do in null hypothesis testing is to assume something to be true in the population (i.e., the null hypothesis) and then test the likelihood of the null given our sample data. Therefore, the inference is actually from the population to the sample; not what we hope to do, but what we do nonetheless (cf. Cohen, 1994).

The general formula for constructing confidence interval is CI=statistic ± (critical value) (standard error of the statistic). CIs can be constructed around a great number of statistics (e.g., mean, coefficient alpha, kurtosis, etc.), but commonly researchers evaluate means. For example, if we assume population mean (μ) equals the sample mean ($\overline{X}$) and the population variance (σ) is known, and the sampling distribution of the mean is normally distributed, the formula for CI becomes:

$$CI = \overline{X} + (Z_{cv})\ (\sigma x)$$

Where

$\overline{X}$ = Sample Mean

Z_{cv} = critical value using the normal distribution (Z_{cv} equals 1.96 when $\alpha = .05$)

σx=standard error of the mean (Hinkle et al., 1998, p. 219). Hinkle et al. also suggested when computing confidence intervals for which the population variance (σ^2) is unknown, we need to find critical value by using the t distribution rather than the normal distribution.

The above formula makes clear that as the standard error of a sampling distribution increases, so does the width of the CI. This is important because it speaks to the degree of sampling error present in our sample, with more sampling error theoretically yielding larger standard errors. If we correctly conceptualize standard error as nothing more or less than the standard deviation of the sampling distribution, then we can understand that larger samples will yield smaller standard errors. Furthermore, holding sample size constant, smaller sample standard deviations will generally yield smaller standard errors.

All of this informs us that as sample size gets bigger and sample standard deviation gets smaller, the width of the confidence interval will become more narrow and be more precise, theoretically due to less sampling error and a smaller standard error. These same concepts will apply to CIs around effect sizes, although the derivation of CIs around effects is somewhat different.

Central and Non-central Test Distributions

When we assume the null hypothesis is true in population, the appropriate test distribution is the central distribution (e.g., central t test or central F test). Most of the popular statistical packages such as SPSS and SAS calculate CIs based on normal or "central" t-test statistic distributions. In the central distribution scenario, the confidence interval is obtained by "inverting" the hypothesis test, which means, the test of the null hypothesis is performed by checking to see whether the interval $\mu_0 \pm t$ (Standard error). However, when we believe the "nil" null hypothesis is not true, noncentral test distributions are necessary to compute accurate CIs for certain effect statistics such as Cohen's d, R^2, λ^2) (Cohen, 1994; Cumming & Finch, 2001; Filder & Thompson, 2001; Smithson, 2001). The reader is referred to Cummings and Finch (2001) and Smithson (2001) for reviews of CIS from central and non-central distributions.

Heuristic Example

Cumming's Exploratory Software for Confidence Intervals (ESCI) has six functions: NonCentral t, Power, CIjumping, CIoriginal, CIdelta, and MAthinking (Cumming, 2001). Two functions are relative to the present paper: CIoriginal and CIjumping. CIoriginal calculates and displays CIs for data with three simple experimental designs: Case 1 for single group, Case 2 for a two independent groups, and Case 3 for paired data. CIjumping takes repeated samples from a population to illustrate basic concepts of CIs and sampling error. To simplify the demonstration, I will use Case 1 for a single group and sampling error to illustrate CIs for effect size in this paper.

One Sample Case with CIoriginal Function

To help to understand a confidence interval, heuristic data was developed to illustrate CIs with CIoriginal (see Figure 1). The data for one sample were: 22, 23, 26, 31, 32, 33, 33, 35, 36, 49 (n=10, M=32, SD=7.70, CI=32 ± 5.51, $\mu_0 = 26$, $t = 2.46$). This figure illustrates the sample and the corresponding CI.

INSERT FIGURE 1 ABOUT HERE

The ESCI output in Figure 1 clearly illustrates the CI of this sample. The CI is a range between 32 ± 5.5103, with 95% of confidence level. The interval does not capture the null value and is statistically significant (p = .036).

One Case Sampling with CIjumping Function

Another ESCI function, CIjumping, takes repeated, independent samples from a normal population. Here I set the population means as 50 and population standard deviation as 20 and let the software re-sample 15 times (n=16 each time). ESCI generated graphical output of sample mean, SD, and standard error seen in Table 1.

INSERT TABLE 1 ABOUT HERE

CIjumping also generated the sampling distribution of the mean as seen in Figure 2. Note that this distribution would become less variable as sampling error decreased, or as sample size gets bigger and sample standard deviation gets smaller.

INSERT FIGURE 2 ABOUT HERE

Figure 2 also illustrates the population parameters ($\mu = 50$, $\mu_0 = 50$).

INSERT FIGURE 3 ABOUT HERE

Figure 3 graphically demonstrates the means and confidence intervals for the 15 samples drawn as listed in Table 1. This Figure visually demonstrates that, because sampling error varies from sample to sample, the location and the width of the intervals also varies from sample to sample. This demonstration reveals the fact that our confidence intervals (e.g., 95% or 99%) that are used in constructing confidence intervals actually apply to the percentage of intervals that will capture the population parameter across many samples. If the demonstration had been continued infinitely, 95% of the intervals would have been captured the population mean of 50.

Due to sampling error, even confidence intervals can haphazardly "bounce" around the population parameter, just as point estimates (i.e., the sample mean) also vary around the population mean. However, the synthesis of these confidence intervals meta-analytically would yield a much more precise indication of what the true population parameter might be.

This graphical output illustrates that confidence interval can be understood as distribution of different sampling cases, and they may or may not being captured in population means, with 95% of alpha level (see Figure 3). However, the bigger the sample size, the more confident we are in knowing the population parameter.

Discussion

Thompson (2001) asserted, "…like any other statistical estimates made using sample data, confidence intervals for effect sizes are impacted by sampling error variance" (p. 11). CIJumping output from ESCI software (see figure 3) illustrates that each time we take samples from population, CIs are different in terms of location and precision (i.e., width of interval).

This reality raises the question of whether a single CI (for example, 95% CIs) in a given study can be described as being 95% likely to capture the population parameter (as is often the case). This view is problematic because on a binominal basis the interval either <u>does</u> or <u>does not</u> capture the population parameter (Thompson, 2001, p. 15). ESCI further demonstrates that the confidence levels (e.g., 95% or 99%) often associated with CIs that do not refer to a single point estimate around the population mean, or other statistic, but rather to the percentage of CIs across studies that capture the population mean. ESCI also makes clear the impact of sampling error on the location and precision of confidence intervals, such that confidence intervals across studies are likely to “bounce” around the true parameter.

References

American Psychological Association (2001). *Publication Manual of the American Psychological Association* (5^{th} ed.). Washington, DC: Author.

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

Craig, J. R., Eison, C. L., & Metze, L. P. (1976). Significance tests and their interpretation: An examination utilizing published research and omega squared. *Bulletin of the Psychonomic Society,* 7, 280-282.

Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement*, 61, 532-574.

Cumming, G. User notes for ESCI (2001). URL:
http://www.psy.latrobe.edu.au/esci/Docs/ESCI_user_notes.pdf
Access date: December 10, 2001.

Filder, F. & Thompson, B. (2001). Computing correct confidence intervals for ANOVA fixed-and random-effects effect sizes. *Educational and Psychological Measurement, 61*, 575-604.

Finch, S., Cumming, G., & Thompson, N. (2001). Reporting of statistical inference in the Journal of Applied Psychology: Little evidence of reform. *Educational and Psychological Measurement, 61*, 181-210.

Henson, R. K., & Smith, A. D. (2000). State of the art in statistical significance and effect size reporting: A review of the APA Task Force report and current trends.

Journal of Research and Development in Education, 33, 285-296.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1998). *Applied statistics for the behavior sciences* (4th ed.). Boston, MA: Houghton Mifflin.

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.

Moore, D. S., & McCabe, G.P. (1993). *Introduction to practice of statistics* (2nd ed.). New York: Freeman.

Smithson, M. (2001). Correct confidence intervals for various regression effect sizes and parameters: The importance of noncentral distributions in computing intervals. *Educational and Psychological Measurement, 61*, 605-632.

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. (2001). Using the descriptive bootstrap to investigate effect replicability (with a comment about confidence intervals for effect sizes). Paper presented at the annual meeting of the American Educational Research Association , Seattle, WA.

Wilkinson, L., & American Psychological Association (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]

Figure 1. CIoriginal output

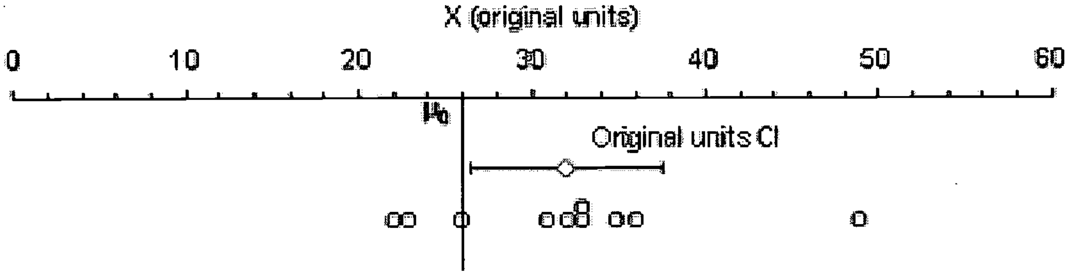


Table 1

CIJumping output

No.	Sampling mean	SD	SE	Bar Display
1	40.07333571	23.58907	5.897267	12.56974
2	46.27842016	23.154314	5.788578	12.33807
3	56.78747654	22.226357	5.556589	11.8436
4	55.54492772	15.531329	3.882832	8.276066
5	44.80125297	18.544907	4.636227	9.88189
6*	37.57639048	22.934678	5.73367	12.22103
7	53.71488795	19.363909	4.840977	10.3183
8	42.50622977	23.555014	5.888753	12.55159
9	41.7082917	19.921418	4.980355	10.61538
10	55.98832202	22.024905	5.506226	11.73625
11	56.76437253	25.842126	6.460531	13.77031
12	54.19252757	19.097086	4.774271	10.17612
13	51.14443851	5.797775	3.949444	8.418045
14	55.16810559	20.866201	5.21655	11.11882
15	49.32266377	20.875924	5.218981	11.124

Note: In case 6, the bar displayed didn't capture the confidence interval. (Sample size n=15)

Figure 2. CIJumping output of sampling distribution of the mean

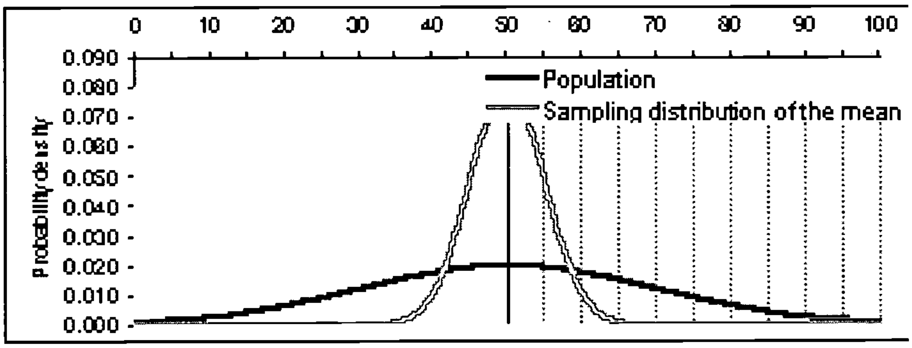
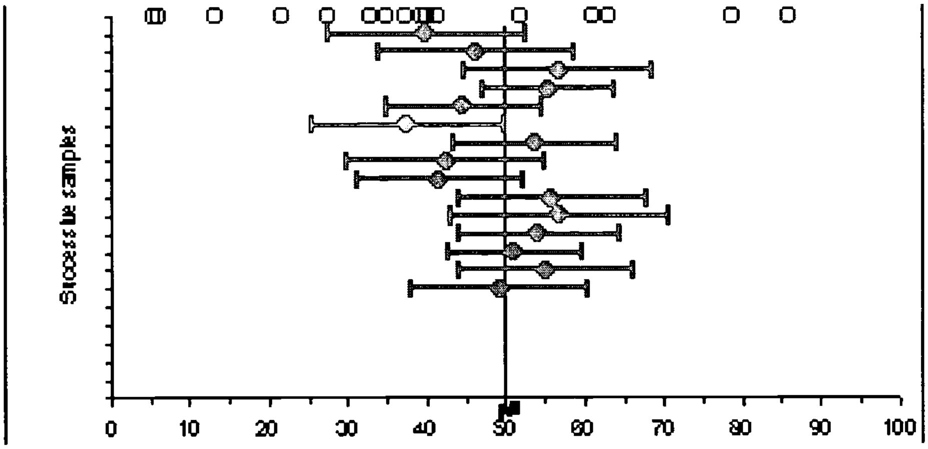


Figure 3. CIJumping output of CIs around effect sizes (sample size n=15, 14 cases captured, case 6 un-captured)



BEST COPY AVAILABLE

TM034286

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

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Sampling Theory and Confidence Intervals for Effect Sizes: Using ESCI to Illustrate "Bouncing" Confidence Intervals	
Author(s): YUNFEI DU	
Corporate Source: University of North Texas	Publication Date: Feb. 2002

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

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

↑

☒

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

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

Level 2A

↑

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

↑

☐

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 medie 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: Du, Yunfei	Printed Name/Position/Title: Yunfei Du / Res. Associate	
Organization/Address: School of Library & Information Sciences,	Telephone: (940)565-3569	FAX:
University of North Texas, Denton, TX 76201	E-Mail Address: Ydu@lis.admin.unt.edu	Date: 02/07/2002

(over)

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

-088 (Rev. 9/97)
PREVIOUS VERSIONS OF THIS FORM ARE OBSOLETE