

DOCUMENT RESUME

ED 465 784

TM 034 169

AUTHOR Bastick, Tony
TITLE Critique of the Scoring Construct Validity of Dichotomously Scored Instruments.
PUB DATE 2002-04-00
NOTE 8p.; Paper presented at the Annual Meeting of the Western Psychological Association (Irvine, CA, April 14-17, 2002).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Adolescents; Foreign Countries; *High School Students; High Schools; *Reliability; Scores; *Scoring; *Validity
IDENTIFIERS *Dichotomous Scoring; Jamaica

ABSTRACT

This paper makes two criticisms of dichotomously scored instruments. One is that dichotomous scoring restrains the scores to ipsative measures that should not be compared, and the other is that dichotomous scoring ignores the strength with which a subject endorses a response so that the resulting count may imply a different construct measure from that indicated by the sum of rating responses. These criticisms were explored by comparing the reliability and validity of dichotomously scored responses and rating response scores from the Impulsivity Scale (L. Hirschfield, 1965). The concurrent validity of both scoring methods was established by comparison with scores on the Children's Perceived Self-Control Scale (P. Humphrey, 1982). Adolescent ninth graders (n=687) from 14 Jamaican schools were tested and retested with these instruments 2 weeks apart. Results show that rating scores were more reliable in stability and consistency and had greater validity than dichotomous scores. These results give empirical support to the criticisms raised in the paper. (Author/SLD)

ED 465 784

Critique of the Scoring Construct Validity of Dichotomously Scored Instruments

Tony Bastick
University of the West Indies

**Paper presented at the Western Psychological Association
WPA 2002 Convention
Irvine, CA
USA**

Publication Date: April, 2002

TM034169

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL HAS
BEEN GRANTED BY

T. Bastick

BEST COPY AVAILABLE

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

Critique of the Scoring Construct Validity of Dichotomously Scored Instruments

TM

Tony Bastick
University of the West Indies

Abstract

This paper makes two criticisms of dichotomously scored instruments, (i) that dichotomous scoring restrains the scores to ipsative measures that should not be compared, and (ii) that dichotomous scoring ignores the strength with which a subject endorses a response and so the resulting count may imply a different construct measure from that indicated by the sum of rating responses.

These criticisms were explored by comparing the reliability and validity of dichotomously scored responses and rating response scores from the Impulsivity Scale (IS). The concurrent validity of both scoring methods was established by comparison with scores on the Children's Perceived Self-Control (CPSC) Scale. Adolescent 9th grade Jamaican students (n=687) from 14 schools were tested and re-tested with these instruments two weeks apart.

Results showed that rating scores were more reliable in stability and consistency, and had greater validity than dichotomous scores. These results give empirical support to the criticisms raised in the paper.

Introduction

Much research in psychology is based on analyses of dichotomously scored instruments. These are instruments whose scoring categorises subjects' responses into one of two categories. Each category represents a psychological construct. The constructs are measured for each subject by summing the responses of the subject that fell into each category. These constructs are then analysed by comparing the measures derived from the subjects. Common examples are paper and pencil instruments using Agree/Disagree, Like me/Unlike me, or Yes/No response options – such as Rotter's Locus of Control (LOC) Scale or Hirschfield, Sutton-Smith, and Rosenberg's Impulsivity Scale (IS).

This paper makes two negative criticisms of these instruments. (i) The scoring forces the sum of the subject's construct measures to equal the number of responses allowed by the instrument, which is often a fixed number, such as the number of questions on a paper and pencil instrument. Yet, there is often no corresponding theoretical 'zero-sum' restriction requiring all subjects to be equal in the total 'strengths' of the constructs. For example, in LOC theory, one subject may be 'stronger' in both External LOC and Internal LOC than another subject, but the scoring system disallows this possibility. (ii) The scores are ipsative measures, proportions relative only to each subject, whose absolute size is not known. Yet, analyses of the constructs commonly treat these scores as absolute numbers by adding, averaging, or correlating them. For example, in the case of LOC, the responses of two subjects may be equally categorised. Their responses could be half Internal and half External. What we may not know is that perhaps Subject A endorsed both the Internal and External questions more strongly than Subject B, because A had a stronger Internal and External LOC than B.

For example, consider six questions each of which may be rated from -9 to +9, and let the first 5 questions be rated -1 and the sixth question be rated +9. Using a dichotomous scoring scheme that counts negative response as -1 and positive responses as +1, the final count would be -4. However, simply totalling the ratings yields a score of +4. Similarly, subjects with the same count for their dichotomous score can have very different totals for their ratings. Practical examples of these anomalies are given in this paper.

This research explored these criticisms by comparing the reliability and validity of dichotomous with continuous scoring of the responses of 687 adolescent Jamaican students to the Impulsivity Scale (IS). Concurrent validity was analysed using correlations with the subjects' responses to the Children's Perceived Self-Control (CPSC) Scale.

Method

Subjects (n=687) were tested with the two instruments and, in order to collect data for stability analysis, they were retested with the same instruments. The mean time between test and retest was approximately 2 weeks (14.76 days).

Subjects

Fourteen schools were chosen at random from secondary schools in Jamaica. Principals and teachers from these schools agreed to the administration of the research instruments to 18 classes of adolescent students. The number of students in each class ranged from 35 to 49 with a median of 40. In all, 309 male and 378 female black grade 9 students participated in the study. The inter-quartile range of their ages was 14-15 years ages with a media age of 14 years 7 months.

Instruments

The two instruments used were the Impulsivity Scale (IS) (Hirschfield, 1965) and the Children's Perceived Self-Control (CPSC) Scale (Humphrey, 1982). The IS is the main focus of analysis in this paper with the CPSC being used to test the concurrent validity of the IS. Subjects' dichotomous responses and the strengths of their endorsements on a 10-point scale were both recorded for each question on both instruments.

The IS is reported to have good criterion-referenced concurrent validity, correlating significantly with teacher ratings and with behavioural observations (Hirschfield, 1965). The original IS consists of the 19 questions and response instructions shown in Figure 1.

Figure 1: Questions and response instructions for the original Impulsivity Scale (IS)

Decide whether each statement is true as applied to you or false as applied to you.

If a statement is True or Mostly True as applied to you, circle T. If a statement is False or Mostly False as applied to you, circle F.

- | | |
|---|--|
| T F 1. I like to keep moving around. | T F 11. I get into tricks at Halloween. |
| T F 2. I make friends quickly. | T F 12. I like being "it" when we play games of that sort. |
| T F 3. I like to wrestle and to horse around. | T F 13. It's fun to push people off the edge into the pool. |
| T F 4. I like to shoot with bows and arrows. | T F 14. I play hooky sometimes. |
| T F 5. I must admit I 'm a pretty good talker. | T F 15. I like to go with lots of other kids, not just one. |
| T F 6. Whenever there's a fire engine going someplace, I like to follow it. | T F 16. I like throwing stones at targets. |
| T F 7. My home life is not always happy. | T F 17. It's hard to stick to the rules if you're losing the game. |
| T F 8. When things get quiet, I like to stir up a little fuss. | T F 18. I like to dare kids to do things. |
| T F 9. I am restless. | T F 19. I'm not known as a hard and steady worker |
| T F 10. I don't think I'm as happy as other people. | |

Ten local teachers considered that questions 4, 6, 7, 11, 13, 14, and 16 should be rephrased to make them more culturally appropriate. Question 10 was reversed. The response instructions were also amended to allow for both dichotomous scoring (Agree/Disagree) and continuous scoring (-9 to +9). The final IS instrument used is shown with its response instructions in Figure 2. However, the question numbers were excluded from the instrument because it was thought they might have influenced the ratings.

Figure 2: Culturally amended Impulsivity Scale (IS) with alternative response instructions

The following sentences are about you. For each sentence (i) If you Agree you are like it says then circle the A. If you Disagree it is like you then circle the D. (ii) Then write a number on the line, from 0 to 9, to show how strongly you agree or disagree.

0-1 means 'slightly agree/disagree' 2-3 means 'agree/disagree a little'
 4-5 means 'mostly agree/disagree' 6-7 means 'strongly agree/disagree'
 8-9 means 'very strongly agree/disagree'

- 1 A D ___ I like to keep moving around.
- 2 A D ___ I make friends quickly.
- 3 A D ___ I like to wrestle and to horse around.
- 4 A D ___ I like to shoot with a slingshot or catapult.
- 5 A D ___ I must admit I 'm a pretty good talker.
- 6 A D ___ Whenever there's a float going someplace, I like to follow it.
- 7 A D ___ My home life is not always happy.
- 8 A D ___ When things get quiet, I like to stir up a little fuss.
- 9 A D ___ I am restless.
- 10 A D ___ I'm just as happy as other people.
- 11 A D ___ I get into tricks on April 1st, Tom fool day.
- 12 A D ___ I like being "it" when we play games of that sort.
- 13 A D ___ It's fun to push people off the edge into the river.
- 14 A D ___ I scull sometimes.
- 15 A D ___ I like to go with lots of other kids, not just one.
- 16 A D ___ I like throwing stones at dogs.
- 17 A D ___ It's hard to stick to the rules if you're losing the game.
- 18 A D ___ I like to dare kids to do things.
- 19 A D ___ I'm known as a hard and steady worker.

The CPSC is an 11-item instrument that measures self-control from a cognitive-behavioral perspective. Hence, it is measuring substantially the same construct, in reverse, as the IS and is thus a suitable instrument with which to test the validity of the IS. The CPSC was developed on a sample of suburban, middle class, fourth and fifth graders (372 boys and 391 girls) for which it has a reported test-retest reliability of 0.71 (Humphrey, 1982).

Analysis and Results

Each IS response was dichotomously scored +1 if 'A' was circled and -1 if 'D' was circled. The positive scores were then counted to give the subjects' Total Dichotomous IS scores (TDIS). This scoring is commensurate with the original scoring of the instrument except that, for greater accuracy, zero rated responses were excluded because zero agreement may be considered no different from zero disagreement. In addition, each rating, zeros included, was coded positively if 'A' was circled and negatively if 'D' was circled. The total of these coded ratings gave the subjects' Total Continuous IS Scores (TCIS). This may be considered as an alternative continuous score. A few subjects did use non-integer ratings.

The mean TDIS was 5.27 with a standard deviation of 2.05. The mean of the TCIS was -1.98 with a standard deviation of 43.09. The TDIS test-retest correlation was 0.595** ($p < 0.0005$).

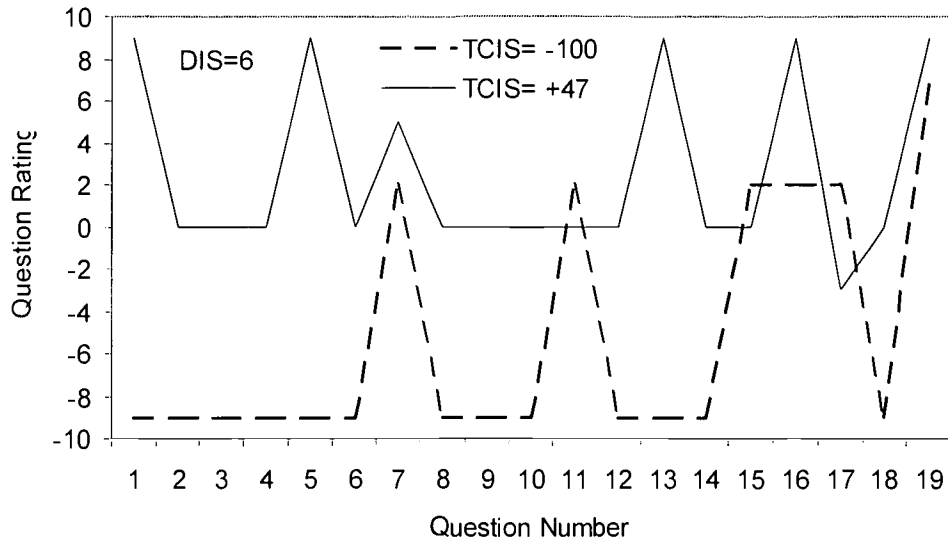
Scoring construct validity

The original dichotomous scoring ignores the varying strengths of individual's endorsements of their Agree/Disagree responses. So that it is possible for two subjects to have the same count of dichotomous scores but very different total rating scores. Figure 3 illustrates two such cases from the data.

Figure 3: Contrasting profiles of subjects with the same dichotomous scores (DIS), who have very different total rating scores (TCIS).

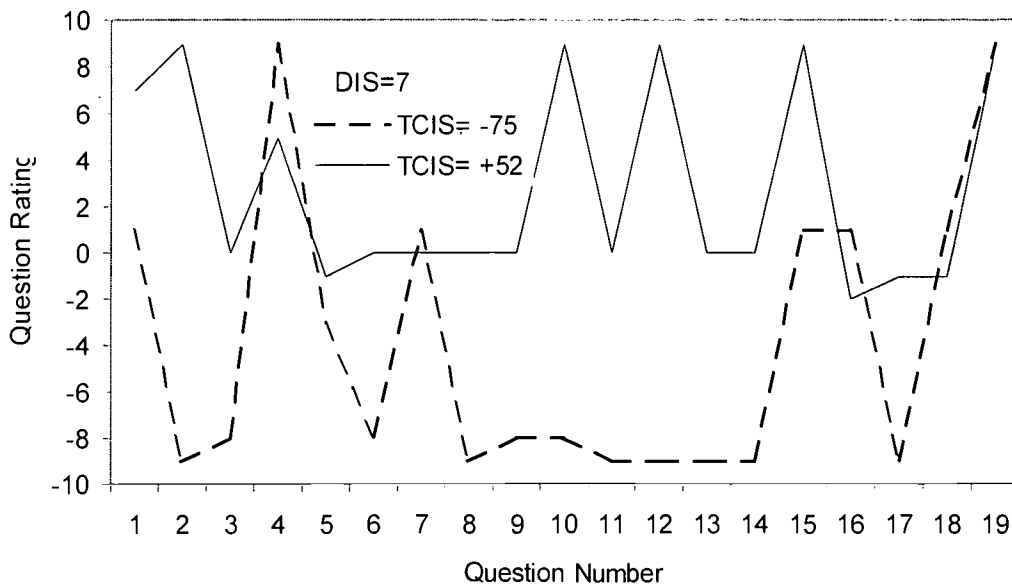
Ratings for subjects m=376 and m=184 who both have a Dichotomous score of 6 (DIS=6)

m	Qu. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
376	TCIS= -100	-9	-9	-9	-9	-9	-9	2	-9	-9	-9	2	-9	-9	-9	2	2	2	-9	7
184	TCIS= +47	9	0	0	0	9	0	5	0	0	0	0	0	9	0	0	9	-3	0	9



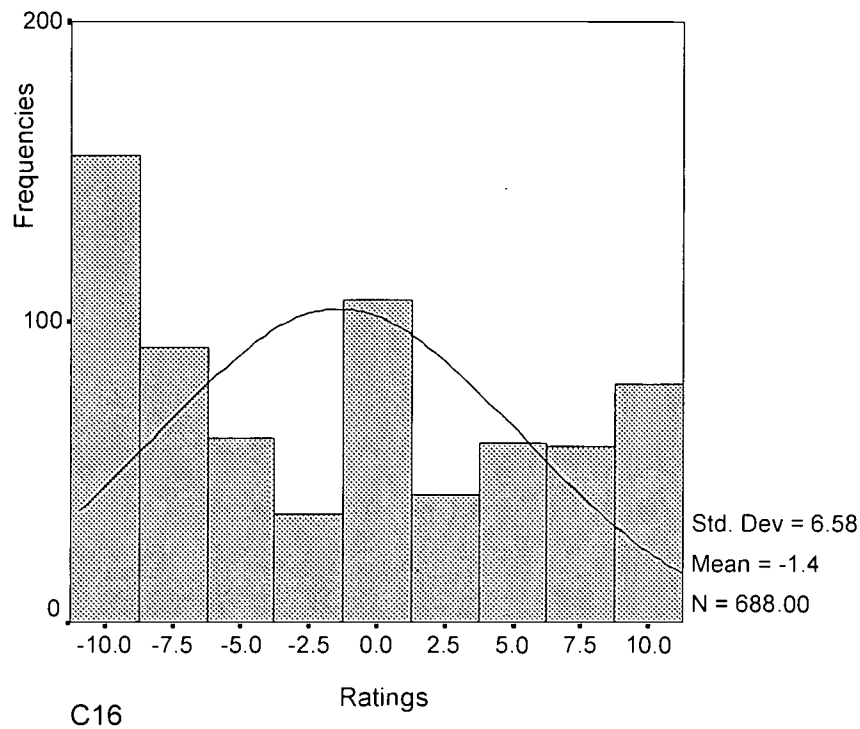
Ratings for subjects m=533 and m=177 who both have a Dichotomous score of 7 (DIS=7)

m	Qu. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
533	TCIS= -75	1	-9	-8	9	-3	-8	1	-9	-8	-8	-9	-9	-9	-9	1	1	-9	1	9
177	TCIS= +52	7	9	0	5	-1	0	0	0	0	9	0	9	0	0	9	-2	-1	-1	9



The spread of continuous ratings illustrates that subjects do indeed vary in the strengths of their endorsements. This is illustrated by the histogram of responses for question 16 shown in Figure 4.

Figure 4: Variations in the strengths of endorsements for question 16 'I like throwing stones at targets'.



A t-test indicated that, as one might expect, this spread was due to a significantly higher male endorsement of this question. The mean rating of question 16 by males was -0.16 compared to only -2.45 by females ($p < 0.0005$). These mean ratings indicate that males were 37% more impulsive than females. However, of the 254 subjects who rated Q16 positively, 138 were male and 116 female. On dichotomous scoring this indicates that males were only 19% more impulsive than females.

The reliability statistics were as follows: The correlation between the TDIS and the TCIS was 0.778** ($p < 0.0005$, $n = 690$). The test retest correlations for TDC and TCIS were $r = 0.595$ ($p < 0.0005$, $n = 547$) and $r = 0.708$ ($p < 0.0005$, $n = 547$) respectively. The internal consistency of the TDIS (C-alpha=0.6328 $n = 690$), as measured by Cronbach Alpha, was lower than the internal consistency of the TCIS (C-alpha=0.6835 $n = 673$).

The concurrent validities of the TDIS and the TCIS were compared by correlating with subjects' scores on the CPSC. The correlation between the CPSC and the TDIS was $r = 0.353$ ($p < 0.0005$, $n = 689$). The correlation between the CPSC and the TCIS was $r = 0.429$ ($p < 0.0005$, $n = 689$).

Discussion

The maximum Total Dichotomous Impulsivity Scale (TDIS) score is 19. The mean TDIS of 5.27 ($n = 689$, 9th graders) found in this study is lower than 8.24 reported by Hirschfield (1995) ($n = 127$, 5th/6th graders). This lower mean is likely to be due to the 0-9 rating scale as zeros were excluded from the dichotomous scores which would lower the TDIS mean. The test-retest stability in this study of 0.595 was lower than $r = 0.85$ reported by Hirschfield (1995). This may be because Hirschfield's subjects could have given more consideration to their responses.

The original dichotomous scoring ignores the subjects' strengths of endorsements that were captured by the continuous scoring. For example, the rating of question 16 'I like throwing stones at targets' had a standard deviation of 6.58 ($n=688$, $\text{mean}=1.4$) which illustrated a significant difference in the mean strength of males' and females' endorsements for this question; females disagreeing more strongly than males. However, dichotomous scoring showed that males were only 19% more impulsive than females, because 19% more males agreed with the statement; whereas, the difference in males' and females' mean endorsements of this question indicated that males were 37% more impulsive than females.

The test-retest correlations showed that the Total Continuous Impulsivity Scale (TCIS) was more stable over the two weeks than was the TDIS ($r=0.708>0.595$), a 19% improvement. Similarly, the TCIS had a higher internal consistency than the TDIS ($C\text{-alpha}=0.6835>0.6328$), an 8% improvement.

The checks on the concurrent validity of the two scoring methods showed that the TCIS scoring was more valid than the TDIS scoring ($r=0.429>0.353$), a 22% improvement.

We may conclude that although the responses of the Jamaican students may not have been so reliable as those on which the authors of the Impulsivity Scale, Hirschfield, Sutton-Smith, and Rosenberg, developed the instrument, the continuous scoring is more consistent, stable and valid than the dichotomous scoring. This empirically supports the criticisms of dichotomous scoring schemes given in the paper. However, continuous scoring requires considerably more effort from both the subjects and the researcher. Hence, one must decide whether the increased validity and reliability provided by the extra effort of continuous scoring is warranted by the requirements of one's applications of the IS scores.

References

- Hirschfield, P. P. (1965). Response set in impulsive children. *Journal of Genetic Psychology*, 107, 117—126.
- Humphrey, L. L. (1982). Children's and teachers' perspectives on children's self-control: The development of two rating scales. *Journal of Consulting and Clinical Psychology*, 50, 624—633. ■



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



TM034169

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Critique of the Scoring Construct Validity of Dichotomously Scored Instruments	
Author(s): Bastick, Tony	
Corporate Source: Paper presented at the Western Psychological Association, WPA 2002 Convention, Irvine, CA. USA.	Publication Date: 2002, April

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 MICROFORM, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

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

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: Tony Bastick, Research Coordinator, Dr.	
Organization/Address: University of the West Indies, Department of Educational Studies, Mona Campus, Kingston 7, Jamaica	Telephone: (876)927-2130	FAX: (876)977-0482
	E-Mail Address: tbastick@uwimona.edu.jm	Date: 4th April 2002



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:

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

4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200

Toll Free: 800-799-3742

FAX: 301-552-4700

e-mail: ericfac@inet.ed.gov

WWW: <http://ericfac.piccard.csc.com>