DOCUMENT RESUME

ED 429 120 TM 029 663

AUTHOR Myerberg, N. James
TITLE The Relationship between Scoring Quality and Assessment Reliability.
PUB DATE 1999-04-00
NOTE 6p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal Quebec, Canada, April 19-23, 1999).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Assessment; Elementary Education; Elementary School Students; Mathematics Tests; Reading Tests; *Scoring; *Teachers; *Test Reliability
IDENTIFIERS *Alpha Coefficient

ABSTRACT

Whether the internal consistency reliability of a test changes as the quality of the scoring of the test changes was studied with data from reading and mathematics short-answer and extended-response assessments administered in grades 3 to 8 in the Montgomery County (Maryland) Public Schools. There were about 9,000 students in each grade, with data from 18 assessments. Each assessment was scored by about 50 teachers, and about 30% of the papers were scored twice to provide data about the quality of scoring and to help in the training of scorers. For each of the assessments an inter-rater correlation coefficient and a coefficient alpha were computed for the best and worst groups of scorers, yielding a total of 36 pairs. A wide range was achieved for both inter-rater correlations and the alpha coefficients. The analysis of these findings indicates that the internal consistency of an assessment changes as the quality of the scoring of the assessment changes. Thus, for tests that are not multiple choice, any report on test quality should also include data related to scoring quality. (Contains one table.) (SLD)

The Relationship Between Scoring Quality and Assessment Reliability

N. James Myerberg
Montgomery County (MD) Public Schools

Paper presented at the annual meeting of
The American Educational Research Association in April 1999
In Montreal

The Relationship Between Scoring Quality and Assessment Reliability

One of the major indicators of the quality of assessments is their internal consistency reliability as expressed by Coefficient Alpha. As many assessment programs have changed to include non-multiple choice questions, scorer consistency, i.e., inter-rater correlation, has become another indicator of the quality of the program. This study looks at the relationship between these two quality indicators by trying to answer the question – does the internal consistency reliability of the test change as the quality of the scoring of the test changes.

Data Source

The data are from reading and mathematics short answer and extended response assessments administered in Grades 3 to 8 in spring 1998 in the Montgomery County (MD) Public Schools. Most of these tests were developed by the school district. There were about 9000 students in each grade. Data from 18 assessments, 9 for each subject, were used in this study. Grades 4, 6, and 7 had two assessments in each subject. Each assessment was scored by a group of about 50 teachers. Papers were randomly assigned to scorers. About 30 percent of the papers were scored twice to provide data about the quality of scoring and to help in the training of teachers for scoring. These double scored papers are used to look at the relationship between scorer and test quality.

Method

Scorers for each assessment were ranked according to the inter-rater correlation (Pearson Product-Moment Coefficient) for the papers that they scored with a random sample of other scorers. This ranking was used to form two analysis groups for each assessment. The groups consisted of the 20 best scorers (highest correlations) and the 20 worst scorers (lowest correlations) for that assessment. The two groups were used for each assessment to assure a range in the quality of scoring. Thus, for each of the 18 assessments an inter-rater correlation coefficient and a Coefficient Alpha were computed for the best and worst groups providing a total of 36 pairs.

The strength of the relationship between the coefficients was determined by computing the Rank-Order Correlation between test and scorer quality. This was done for the 36 pairs of coefficients and also for the 18 pairs within each subject.

Results

A wide range was achieved for both the inter-rater correlations and the Alpha Coefficients. The inter-rater correlations ranged from .9913 for the best scorers on one of the seventh grade mathematics tests to .5009 for the worst scorers on one of the seventh grade reading tests. All but 4 of the correlations were at least .8400. The Coefficient Alphas ranged from .9162 for the best scorers on one of the seventh grade mathematics tests to .5632 for the worst scorers on one of the fourth grade reading tests. All but 3 of the coefficients were at least .7100. Table 1 presents the inter-rater correlation and Coefficient Alpha for the best and worst scorers for each assessment.

A strong relationship was generally found between the inter-rater correlations and Coefficient Alphas. Across all 36 pairs of data the rank order correlation was .7441. Broken down by subject the correlation for mathematics was even stronger, .8101. For reading the correlation was less, only .4221.

A possible reason for the lower correlation in reading was because there were two types of assessments involved. All six grades took a short-answer reading assessment on which each of 10 items was scored separately. Three of the grades -- 4, 6, and 7 – also took an extended writing assessment that was scored holistically on three domains. These domain scores were then added together for the total score. Rank-order correlations computed separately for the two different types of assessments are somewhat higher. The correlation from the short answer assessments was .8601. The correlation form the extended writing assessments was .6571.

Discussion

The results indicate that the internal consistency of an assessment changes as the quality of the scoring of that assessment changes. Thus, for non-multiple choice tests, any report on test quality should also include data related to scoring quality. If a test seems to have inadequate internal consistency, it could be the result of poor scoring, not because it is a poor assessment.

The data and results reported here are from one set of tests in one school district. Similar analyses should be carried out on data from other assessment programs to verify the generalizability of these findings.

Table 1

Scorer Quality (Inter-Rater Correlation) and Assessment Quality (Coefficient Alpha)

Assessment		Number	Number of Papers		Inter-Rater Correlation		Coefficient Alpha	
Subject	Grade	of Points	Best	Worst	Best	Worst	Best	Worst
Math	3	30	734	822	.9819	.9074	.8732	.8487
	4	30	826	638	.9823	.9825	.8421	.7485
	5	30	532	640	.9883	.9626	.9089	.8665
	6	30	824	932	.9848	.9541	.8772	.8322
	7	30	332	702	.9913	.9625	.9162	.8873
	8	30	684	1042	.9875	.9628	.9156	.8673
	4	12	3162	1884	.9554	.8826	.7216	.6984
	6	18	738	986	.9822	.9533	.8833	.8371
	7	18	576	522	.9795	.9135	.8838	.8244
Reading	3	30	652	798	.9649	.9072	.8661	.7948
	4	30	624	620	.9516	.7209	.8405	.5632
	5	30	942	856	.9367	.8460	.7922	.7129
	6	30	862	836	.9558	.8745	.8569	.7828
	7	30	726	776	.9372	.8424	.8671	.7479
	8	30	732	916	.9511	.8584	.7897	.6829
	4	12	650	834	.8800	.5071	.8707	.7623
	6	12	762	1142	.7827	.5713	.8533	.8161
	7	12	1312	1176	.7651	.5009	.8198	.8307

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

ERIC

TM029663

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: The Relationship Between Scoring Quality and Assessment Reliability

Author(s): N. James Myerberg

Corporate Source:

Publication Date: 4/99

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 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: N. James Myerberg	Printed Name/Position/Title: N. James Myerberg
Organization/Address: 5606 Ogden Rd. Bethesda, MD 20816	Telephone: 301-279-3194 FAX:
	E-Mail Address: Date: 4/5/99

(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:

THE UNIVERSITY OF MARYLAND
ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
1129 SHRIVER LAB, CAMPUS DRIVE
COLLEGE PARK, MD 20742-5701
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