DOCUMENT RESUME

ED 399 285 TM 025 512

AUTHOR	Wolfe, Edward W.
TITLE	A Report on the Reliability of a Large-Scale Portfolio Assessment for Language Arts, Mathematics, and Science.
PUB DATE	Apr 96
NOTE	19p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New York, NY, April 9-11, 1996).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Decision Making; Generalizability Theory; Interrater Reliability; *Language Arts; *Mathematics Instruction; Pilot Projects; *Portfolio Assessment; Portfolios (Background Materials); Science Instruction; Scores; *Scoring; Secondary Education; Secondary School Students; *Test Reliability; Test Use
IDENTIFIERS	*Large Scale Assessment

ABSTRACT

Although portfolio assessment is becoming increasingly popular, it may not survive unless portfolio scoring can meet the demands of large-scale assessment standards. The results of studies of interrater reliability with large-scale portfolio assessments have been mixed. This paper reports the scoring results of a nationwide portfolio pilot in which over 2,000 secondary students submitted portfolios from language arts, mathematics, and science classes. For language arts, both interrater reliability and score reliability were at reasonable levels. For mathematics, the interrater reliability was adequate, but the score reliability was low. For the science portfolio, neither the interrater reliability nor the score reliability was adequate. Generalizability studies also suggest that adequate reliability for student level decisions can be achieved with scores derived from five portfolio entries, each scored by two raters. With changes to the scoring rubrics and student and teacher manuals, more reliable scores should result in the second year of the project. (Contains nine tables and nine references.) (Author/SLD)

Running head: Portfolio Reliability

A report on the reliability of a large-scale portfolio assessment for language arts, mathematics, and science

Edward W. Wolfe

American College Testing, Iowa City, Iowa

Paper presented at the Annual Meeting of the National Council for Measurement in Education in New York, NY (April, 1996).

Addendum to:

Wolfe, E.W. (1996, April). *A report on the reliability of a large-scale portfolio assessment for language arts, mathematics, and science.* Paper presented at the Annual meeting of the National Council for Measurement in Education, New York, NY.

The average generalizability (G) coefficients reported on page 15 of the manuscript (G_{LA} = .73, G_M = .33, and G_S = .31) were computed as a weighted average of the Fischer z transformation of the G coefficients reported in Tables 7, 8, and 9. According to Dr. Robert L. Brennan of the University of Iowa, a better method for estimating the reliability realizable under a specific scoring model can be obtained by computing a G coefficient based on the average of the variance components across the G studies.

The table below show the average variance component for each facet in the G study design and reports the G and ϕ coefficients for a scenario in which each student submits five portfolio entries, each scored by two raters. These values are slightly higher than those reported on page 15 of the manuscript.

Average Variance Components

Content Area	Variance Components	G Coefficient	ϕ Coefficient
Language Arts	p = .3383 i = .0931 r:i = .0120 pi = .2759 pr:i = .4060	.78	.75
Mathematics	p = .1203 i = .3990 r:i = .0075 pi = .5478 pr:i = .4249	.44	.34
Science	p = .0405 i = .4135 r:i = .0057 pi = .2134 pr:i = .2787	.36	.21

Abstract

Portfolio assessment is becoming increasingly popular as an assessment tool because portfolios allows teachers to determine how well students work on long-term projects, collaborate with others, develop a piece of work over time, and reflect on what they have learned. Although classroom-based portfolios facilitate good instruction, this form of assessment may not survive unless portfolio scoring can meet the demands of large-scale assessment standards (Freedman, 1993).

A number of researchers have investigated interrater reliability with large-scale portfolio assessments. The results of these studies have been mixed, producing interrater correlations ranging between .44 and .94 (Herman, Gearhart, & Baker, 1993; Koretz, Klein, McCaffrey, & Stecher, 1994; LeMahieu, Gitomer, & Eresh, 1995; and Nystrand, Cohen, & Dowling, 1993). This paper reports the scoring results of a nation-wide large-scale portfolio pilot in which over 2,000 secondary students submitted portfolios from language arts, mathematics, and science classes. Our analyses show that the interrater reliability from this pilot project matched and, in some cases, surpassed those found for state and regional portfolio assessments. Generalizability studies also suggest that adequate reliability for student level decisions can be achieved with scores derived from five portfolio entries, each scored by two raters.

A report on the reliability of a large-scale portfolio assessment for language arts, mathematics, and science

The use of portfolios, as an assessment tool, allows teachers to determine how students work on long-term projects, collaborate with others, develop a piece of work over time, and reflect on what they have learned. As a result, portfolios are being considered as an alternative to multiple-choice tests by educators who are interested in assessing broader, more complex educational outcomes in contexts that require authentic, every-day uses of those skills. Portfolio assessment is also becoming a popular format for assessing student outcomes because it provides a means of linking classroom instruction to large-scale testing. Because of the need to keep test blueprints and item pools secure, large-scale multiple choice tests are not well-suited to guiding classroom instruction. Portfolios, on the other hand, provide students and teachers with clearly-defined standards and show models of student work that demonstrate varying degrees of accomplishment within those standards.

However, the complexity and comprehensiveness of the student outcomes that can be assessed with portfolios comes at a price. Educators must rely on human judgements of a portfolio's quality if the assessment results are to be used as the basis for educational decisions. Because the use of human raters introduces sources of measurement error that are not associated with items that are scored in a more "objective" manner, it is important that test developers find ways to control this source of construct-irrelevant variance. Unless portfolio scoring can meet the demands of large-scale assessment standards, portfolio assessment may not remain a viable assessment format, regardless of the instructional benefits associated with it (Freedman, 1993).

The degree to which raters introduce measurement error into scores from performance assessments is not clearly agreed upon. Some proponents of generalizability theory suggest that the amount of error introduced by raters, relative to the amount introduced by other potential sources of error, is trivial (Shavelson, Baxter, & Gao, 1993). According to these researchers, very little, if any, of the overall measurement error in performance assessment scores is accounted for by the error variance associated with raters. On the other hand, the amount of variance contributed by person-by-task and person-by-task-by-occasion interactions is relatively large. Proponents of Rasch measurement theory, on the other hand, suggest that the amount of error contributed to performance assessment scores by raters is significant (Engelhard, 1994 and Lunz, Wright, & Linacre, 1990), regardless of its relative size. These researchers also show that the dependability of an individual student's score can be improved by taking this error into account. In fact, these researchers have proposed scaling methods that can be used to eliminate some of the error introduced by human raters (Linacre, 1994).

Regardless of the magnitude of the error contributed to portfolio assessment scores by raters, this issue has become the primary focus of psychometric research associated with performance assessments. A number of recent articles have focused the likelihood that test developers can control measurement error associated with raters to a degree that will allow scores from performance assessments to be valid for making decisions about individual students. The preliminary work has been encouraging. For example, Herman, Gearhart, and Baker (1993) report interrater correlations ranging from .76 to .94 for a school-wide pilot of an elementary level writing portfolio. Similarly, LeMahieu, Gitomer, and Eresh (1995) report interrater correlations ranging from .74 to .87 for a district-wide pilot of a middle school and secondary level writing portfolio.

Larger portfolio assessment projects have been similarly successful. Nystrand, Cohen, and Dowling (1993) report results from a follow-up scoring of university level writing portfolios. They achieved interrater correlations ranging from .44 to .86 and generalizability coefficients around .55. Koretz, Klein, McCaffrey, and Stecher (1994) report interrater reliabilities for the Vermont portfolio program. They found interrater correlations for composite mathematics portfolio scores ranged from .53 to .79 for elementary and middle school students. Slightly lower correlations were reported for writing portfolios, ranging from .49 to .63.

These researchers have shown that, with practice, reasonable levels of interrater agreement can be achieved for portfolio assessments. However, most of these efforts have been rather limited in scope--focusing only on assigning scores to students from a single school, district, or state and focusing primarily on portfolio assessment in the content area of writing. To be truly useful as a vehicle for promoting specific educational standards, portfolio assessments must be usable on a national level, and they must be usable in multiple content areas. This study extends prior research concerning the rater and score reliability of portfolio assessments by examining scores from the pilot of a large-scale portfolio project in the content areas of language arts, mathematics, and science.

Method

Results from the 1994/1995 pilot of the *ACT Portfolio System* are reported here. Portfolios from three content areas were scored ($N_{language\ arts} = 477$, $N_{mathematics} = 451$, and $N_{science} = 440$). All portfolios came from seven Design Partner schools that were selected for their geographic and demographic diversity.

Portfolios for each content area were scored by raters who had obtained a minimum of a bachelor's degree in that content area. Raters were trained for each Work Sample Description within their content area, scoring all of the student work for a particular Work Sample Description prior to being trained for the next Work Sample Description. Scores were assigned according to a six-point rubric for all Work Sample Descriptions in each content area. Once all Work Sample Descriptions had been scored, raters were trained to assign a holistic score to the portfolio based on all five entries. Scores were assigned according to a four-point rubric for holistic scoring in each content area.

At least 10% of the portfolios were scored by a randomly selected second rater. Because students and teachers were allowed to choose the Work Sample Descriptions for which they submitted entries and because some students submitted fewer than the five requested samples of work, the number of portfolio entries scored for each Work Sample Description varied greatly. As a result, reliability analyses are restricted to those Work Sample Descriptions for which second scores were assigned to a minimum of 25 student entries.

At the conclusion of the scoring project, indices of interrater agreement (percent of scores in *perfect*, *adjacent*, and *outside of adjacent* agreement and *interrater correlations*) were computed for each Work Sample Description satisfying the 25 double score minimum. *Perfect* agreement was achieved when both raters assigned the same score to the student's entry. *Adjacent* agreement was achieved when the two scores assigned to the student's entry were within one point of each other. *Outside of adjacent* agreement was achieved when the absolute difference between the two scores assigned to the student's was greater than one. In addition, generalizability coefficients were computed for pairs of Work Sample Descriptions for which there were a minimum of 25 doubles scores for students who submitted an entry for each Work Sample Description in that pair. A preliminary ANOVA revealed that pairs of raters were interchangeable, so the following design was used for the generalizability study:

s(r:i), where
s is the student facet
r is the rater facet (nested within items)
i is the item facet

Decision studies were also computed based on the anticipated design of the *ACT Portfolio*--two raters score each of five entries for each student.

Results

Interrater Agreement

Table 1 shows the interrater agreement for each language arts Work Sample Description. Table 2 shows the interrater correlations for the language arts Work Sample Descriptions.

Table 1: Interrater Agreement for Language Arts Work Sample Descriptions

Work Sample Description	N	Agreement	Percent
1	90	Perfect	48
		Adjacent	46
		Outside	7
2	56	Perfect	61
		Adjacent	29
		Outside	11
4	40	Perfect	55
		Adjacent	35
		Outside	10
5	158	Perfect	43
		Adjacent	47
		Outside	10
6	70	Perfect	34
		Adjacent	56
		Outside	10
7	67	Perfect	46
		Adjacent	34
		Outside	20
8	73	Perfect	40
		Adjacent	47
		Outside	14
9	26	Perfect	39
		Adjacent	46
		Outside	15
12	107	Perfect	40
		Adjacent	44
		Outside	16

Table 2: Interrater Correlations for Language Arts Work Sample Descriptions

Work Sample Description	N	r_p
1	90	.79
2	56	.72
4	40	.75
5	158	.60
6	70	.47
7	67	.66
8	73	.55
9	26	.50
12	107	.47

Table 3 shows the interrater agreement for each mathematics Work Sample Description. Table 4 shows the interrater correlations for the mathematics Work Sample Descriptions.

Table 3: Interrater Agreement for Mathematics Work Sample Descriptions

Work Sample Description	N	Agreement	Percent
1	121	Perfect	75
		Adjacent	17
		Outside	8
2	193	Perfect	78
		Adjacent	11
		Outside	12
3	174	Perfect	80
		Adjacent	11
		Outside	9
5	117	Perfect	48
		Adjacent	35
		Outside	17
7	32	Perfect	91
		Adjacent	9
		Outside	0
8	160	Perfect	69
		Adjacent	23
		Outside	8
9	80	Perfect	43
		Adjacent	38
		Outside	20

Table 4: Interrater Correlations for Mathematics Work Sample Descriptions

Work Sample Description	N	r_p
1	121	.73
2	193	.59
3	174	.60
5	117	.61
7	32	.96
8	160	.74
9	80	.46

Table 5 shows the interrater agreement for each science Work Sample Description. Table 6 shows the interrater correlations for the science Work Sample Descriptions.

Table 5: Interrater Agreement for Science Work Sample Descriptions

Work Sample Description	N	Agreement	Percent
1	156	Perfect	56
		Adjacent	41
		Outside	3
2	43	Perfect	54
		Adjacent	44
		Outside	2
4	59	Perfect	36
		Adjacent	55
		Outside	9
5	46	Perfect	64
		Adjacent	32
		Outside	3
9	82	Perfect	56
		Adjacent	42
		Outside	2
11	93	Perfect	33
		Adjacent	54
		Outside	13

Table 6: Interrater Correlations for Science Work Sample Descriptions

Work Sample Description	N	r_p
1	156	.55
2	43	.54
4	59	-.04
5	46	.51
9	82	.40
11	93	.44

Generalizability Analyses

The following tables show the results of generalizability and interrater reliability studies for the *ACT Portfolio*. Generalizability studies were run for all pairs of Work Sample Descriptions for which more than 25 students had double scores on both Work Sample Descriptions in the pair. The design used in this study contained students crossed with raters who are nested within items, *s(r:i)*. This design is not as desirable as a fully balanced and completely crossed design involving all items taken by all students, but such a design was not economically or logistically feasible for our pilot study. Variance components were obtained from this G study and were used to estimate generalizability coefficients and phi coefficients (D studies) for the case of five items per student, each item rated by two raters.

Table 7 shows the generalizability estimates for the language arts portfolio.

Table 7: Generalizability for Language Arts Work Sample Descriptions

WSDs	Variance Components	G Coefficient	ϕ Coefficient
1 and 5	p = .8543 i = .0000 r:i = .0238 pi = .3092 pr:i = .3177	.90	.90
1 and 6	p = .4323 i = .0000 r:i = .0362 pi = .1600 pr:i = .4444	.85	.84
1 and 7	p = .5108 i = .0000 r:i = .0139 pi = .4592 pr:i = .3419	.80	.80
1 and 8	p = .3102 i = .3433 r:i = .0000 pi = .3558 pr:i = .3426	.75	.64
5 and 6	p = .2809 i = .0253 r:i = .0132 pi = .0405 pr:i = .4013	.85	.84
5 and 7	p = .2404 i = .0395 r:i = .0053 pi = .2763 pr:i = .3565	.73	.71
5 and 8	p = .2234 i = .2030 r:i = .0000 pi = .1671 pr:i = .3919	.75	.66
6 and 7	p = .1777 i = .0000 r:i = .0154 pi = .0462 pr:i = .5615	.73	.73
7 and 8	p = .0146 i = .2269 r:i = .0000 pi = .6685 pr:i = .4964	.07	.06

Table 8 shows the generalizability estimates for the mathematics portfolio.

Table 8: Generalizability Analyses for Mathematics Work Sample Descriptions

WSDs	Variance Components	G Coefficient	ϕ Coefficient
1 and 2	p = .0118 i = .9283 r:i = .0081 pi = .4155 pr:i = .2914	.10	.04
1 and 3	p = .1549 i = .5460 r:i = .0050 pi = .4326 pr:i = .2450	.58	.41
1 and 5	p = .1752 i = .0000 r:i = .0203 pi = .4421 pr:i = .5778	.55	.54
1 and 8	p = .2977 i = .5953 r:i = .0123 pi = .3971 pr:i = .3499	.72	.56
2 and 3	p = .1061 i = .0672 r:i = .0000 pi = .2470 pr:i = .2475	.59	.55
2 and 5	p = .2393 i = 1.0352 r:i = .0114 pi = .6080 pr:i = .4774	.59	.39
2 and 8	p = .0021 i = .0106 r:i = .0017 pi = .5802 pr:i = .2806	.01	.01
2 and 9	p = .0084 i = .4028 r:i = .0000 pi = .4492 pr:i = .5197	.06	.04
3 and 5	p = .0000 i = .7863 r:i = .0171 pi = .8372 pr:i = .4739	.00	.00
3 and 8	p = .1403 i = .0000 r:i = .0056 pi = .5762 pr:i = .2891	.49	.49
3 and 9	p = .0000 i = .1839 r:i = .0000 pi = .6796 pr:i = .5064	.00	.00
5 and 8	p = .3317 i = .6521 r:i = .0153 pi = .5281 pr:i = .5806	.67	.53

Table 8: Generalizability Analyses for Mathematics Work Sample Descriptions--continued

WSDs	Variance Components	G Coefficient	ϕ Coefficient
5 and 9	p = .0000 i = .1558 r:i = .0154 pi = .7478 pr:i = .8132	.00	.00
7 and 8	p = .3365 i = .4796 r:i = .0000 pi = .4804 pr:i = .0992	.76	.63
8 and 9	p = .0000 i = .1420 r:i = .0000 pi = .7967 pr:i = .6225	.00	.00

Table 9 shows the generalizability estimates for the science portfolio.

Table 9: Generalizability Analyses for Science Work Sample Descriptions

WSDs	Variance Components	G Coefficient	φ Coefficient
1 and 2	p = .0462 i = .0439 r:i = .0139 pi = .0912 pr:i = .2767	.50	.45
1 and 5	p = .0465 i = .2569 r:i = .0023 pi = .0409 pr:i = .2530	.58	.35
1 and 9	p = .0000 i = .1230 r:i = .0001 pi = .2439 pr:i = .2539	.00	.00
1 and 11	p = .0751 i = .1890 r:i = .0082 pi = .2816 pr:i = .3594	.45	.36
2 and 5	p = .0000 i = .0519 r:i = .0000 pi = .1258 pr:i = .2072	.00	.00
2 and 9	p = .0762 i = .0081 r:i = .0000 pi = .1746 pr:i = .1923	.58	.58
2 and 11	p = .0667 i = .7304 r:i = .0000 pi = .4429 pr:i = .3064	.36	.20
5 and 11	p = .0542 i = 1.337 r:i = .0227 pi = .2537 pr:i = .2784	.41	.13
9 and 11	p = .0000 i = .9810 r:i = .0034 pi = .2691 pr:i = .3812	.00	.00

Discussion

These results show that our efforts to implement a nationally-based portfolio project in multiple content areas has been as successful as prior attempts to implement smaller-scale projects. The average interrater reliability for mathematics and language arts were satisfactory (.66 and .62, respectively). The average interrater reliability for science, however, was lower than would be desirable (.44). Although these interrater correlations are not as high as would be necessary if scores were to be used to make decisions about individual students, they are reasonable for a first year pilot. Generalizability coefficients, on the other hand were quite high for language arts portfolios (averaging .73), but quite low for mathematics and science portfolios (.33 and .31, respectively).

Overall, we found both interrater reliability and score reliability to be at reasonable levels for the language arts portfolios. On the other hand, the lower reliabilities observed with the mathematics and science scores raise some concerns. For the mathematics portfolios, the interrater reliability was adequate, but the score reliability was low. This may indicate that, although the scoring criteria is easy to understand and apply, the Work Sample Descriptions for mathematics are tapping multiple constructs (i.e., they are multidimensional). For the Science portfolio, however, neither the interrater reliability nor the score reliability were adequate. This may be an indication that our scoring standards need revision.

As a result of the pilot of the *ACT Portfolio System*, numerous revisions have been implemented with hopes of increasing score reliability in the future. For all three content areas, the distributions of student scores were considerably less variable than was expected. For most Work Sample Descriptions, the distributions were heavily positively skewed so that only a small portion of the students obtained scores of five and six on the six-point scale. In some cases, no examples of student work were assigned to these score categories.

Three problems were identified that may have led to these results. First, it seems that the original standards that we set for students were too high. After consulting with teachers in the project, it became apparent that although our expectations might be reasonable for some students, the majority of students were not able to perform above the first and second levels of our scoring rubrics. As a result, our scoring rubrics are being revised so that the expectations of students are more reasonable and attainable.

Second, most of the students and teachers involved in the project had a difficult time producing five pieces of their best work in the amount of time they were given to complete the portfolios. Because of time constraints, most teachers did not introduce the portfolio package to their students until well after the school year began. Our scoring rubrics, on the other hand, were designed based on the assumption that students would have an entire academic year to complete the portfolio. As a result, the quality of the work students included in their portfolios was not as good as it could be. During the second year of our pilot, training took place early in the school year to remedy this problem.

A third problem arose because of ambiguity in our training materials. Because of the complexity of the concepts contained in the menu of Work Sample Descriptions and because of the short amount of time teachers and students had to work with these materials, many of the samples of student work submitted for each Work Sample Description were poor fits for that category. As a result, student work was evaluated more negatively than it would have been if the work had been evaluated in a better fitting Work Sample Description. Again, steps have been taken to remedy this problem during the second year of our project. We have revised the teacher and student guides to include more descriptive language about the scoring rubrics as well as examples of classroom activities that would likely elicit appropriate samples of student work.

We expect that the changes we have made to the scoring rubrics and to the student and teacher manuals will result in more variability between students, and thus more reliable scores for the second year of our project.

References

Engelhard, G.J. (1994). Examining rater errors in the assessment of written composition with a many-faceted Rasch model. *Journal of Educational Measurement, 31*(2), 93-112.

Freedman, S.W. (1993). Linking large-scale testing and classroom portfolio assessments of student writing. *Educational Assessment, 1*(1), 27-52.

Herman, J.L., Gearhart, M., & Baker, E.L. (1993). Assessing writing portfolios: Issues in the validity and meaning of scores. *Educational Assessment, 1*(3), 201-224.

Koretz, D., Klein, S., McCaffrey, D., & Stecher, B. (1994). The Vermont Portfolio Assessment Program: Findings and implications. *Educational measurement: Issues and practice, 13*(3), 5-16.

LeMahieu, P.G., Gitomer, D.H., & Eresh, J.T. (1995). *Portfolios beyond the classroom: Data quality and qualities.* Princeton, NJ: Educational Testing Service.

Linacre, J.M. (1993). *Many-facet: Rasch measurement.* Chicago, IL: MESA Press.

Lunz, M.E., Wright, B.D., & Linacre, J.M. (1990). Measuring the impact of judge severity on examination scores. *Applied Measurement in Education, 3*(4), 331-345.

Nystrand, M., Cohen, A.S., & Dowling, N.M. (1993). Addressing reliability problems in the portfolio assessment of college writing. *Educational Assessment, 1*(1), 53-70.

Shavelson, R.J., Baxter, G.P., & Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement, 30*(3), 215-232.

TM 025512

AERA April 8-12, 1996

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

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: A Report of the Reliability of a large-scale portfolio assessment for language arts, mathematics & Science

Author(s): Edward W Wolfe

Corporate Source:

Publication Date:

II. REPRODUCTION RELEASE:

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

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

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

Sample sticker to be affixed to document

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

[ ] **or here**
Permitting reproduction in other than paper copy.

Sign Here, Please

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

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

Signature: Edward W Wolfe	Position:
Printed Name: Edward W Wolfe	Organization: ACT
Address: ACT-41 POB 168 Iowa City, IA 52240	Telephone Number: (319) 337-1548
	Date: 4/8/96

CUA

THE CATHOLIC UNIVERSITY OF AMERICA
Department of Education, O'Boyle Hall
Washington, DC 20064
202 319-5120

February 27, 1996

Dear AERA Presenter,

Congratulations on being a presenter at AERA[1]. The ERIC Clearinghouse on Assessment and Evaluation invites you to contribute to the ERIC database by providing us with a written copy of your presentation.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed and electronic versions of *RIE*. The paper will be available through the microfiche collections that are housed at libraries around the world and through the ERIC Document Reproduction Service.

We are gathering all the papers from the AERA Conference. We will route your paper to the appropriate clearinghouse. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality.

Please sign the Reproduction Release Form on the back of this letter and include it with **two** copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can drop off the copies of your paper and Reproduction Release Form at the **ERIC booth (23)** or mail to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to:

AERA 1996/ERIC Acquisitions
The Catholic University of America
O'Boyle Hall, Room 210
Washington, DC 20064

This year ERIC/AE is making a **Searchable Conference Program** available on the AERA web page (http://tikkun.ed.asu.edu/aera/). Check it out!

Sincerely,

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

[1]If you are an AERA chair or discussant, please save this form for future use.