ED 455 279 TM 033 082

AUTHOR	Fan, Xitao; Yin, Ping
TITLE	Sample Characteristics and Measurement Reliability: An Empirical Exploration.
PUB DATE	2001-04-00
NOTE	20p.; Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Error of Measurement; *Measurement Techniques; *Reliability; Sample Size; *Sampling
IDENTIFIERS	Heterogeneity of Variance; Iowa Tests of Basic Skills; Texas Assessment of Academic Skills

ABSTRACT

The literature on measurement reliability shows the consensus that group heterogeneity with regard to the trait being measured is a factor that affects the sample measurement reliability, but the degree of such effect is not entirely clear. Sample performance also has the potential to affect measurement reliability because of its effect on the relative magnitude of error score variance. This paper empirically examines the effects of these two sample characteristics on measurement reliability. Two large extant data sets are used in the investigation. One set of data was for 50,000 students from the criterion-referenced Texas Assessment of Academic Skills, and the other was for 10,000 students from the norm-referenced Iowa Tests of Basic Skills. The results suggest that both group variability and group performance level affect measurement reliability, and measurement error tends to be smaller for high-performance samples than for low performance samples. (Author/SLD)

Sample Characteristics and Measurement Reliability:

An Empirical Exploration

Xitao Fan
University of Virginia

Ping Yin
University of Iowa

Running Head: Sample Characteristics

Correspondence about this paper should be addressed to:

Xitao Fan, Ph.D.
Curry School of Education
University of Virginia
405 Emmet Street South
P. O. Box 400277
Charlottesville, VA 22904-4277

Phone (804)243-8906
Fax (804)924-1384
E-Mail xfan@virginia.edu
Web http://www.people.virginia.edu/~xf8d

The authors gratefully acknowledge the support of Iowa Testing Program for providing the ITBS data used in this study.

Paper presented at the 2001 Annual Meeting of the American Educational Research Association, April 10-14 (*Session #44.68*), Seattle, WA.

Abstract

The literature on measurement reliability shows the general consensus that group heterogeneity with regard to the trait being measured is a factor that affects the sample measurement reliability, but the degree of such effect is not entirely clear. Potentially, the sample performance may also affect measurement reliability, because of its effect on the relative magnitude of error score variance. This paper empirically examines the effects of these two sample characteristics on measurement reliability. Two large extant data sets (criterion- and norm-referenced, respectively) are used in the investigation. The results suggest that both group variability and group performance level affect measurement reliability, and measurement error tends to be smaller for high-performance samples than for the low-performance samples.

KEYWORDS: score reliability, sample heterogeneity, sample performance, measurement error

It is generally recognized that measurement reliability should be considered as the characteristic of test scores, rather than test itself (e.g., Crocker & Algina, 1986; Thompson & Vacha-Haase, 2000; Yin & Fan, 2000). As discussed by Crocker and Algina, "reliability is a property of the scores on a test for a particular group of examinees" (p. 144). In the reliability literature, group heterogeneity with regard to the trait being measured has been identified as one factor that affects sample measurement reliability estimate. In general, other things being equal, measurement reliability would be higher for a group that is heterogeneous with regard to the trait being measured than that of a more homogeneous group.

Although sample heterogeneity/homogeneity is generally recognized as one prominent factor that affects the measurement reliability, the degree of such effect is not entirely clear. In the discussion about the relationship between group heterogeneity and its effects on measurement reliability (e.g., Gulliksen, 1987, Chapter 10; Magnusson, 1967), it is typically assumed that the error score variances of the groups are equal. As discussed by Gulliksen, "the effect of group heterogeneity on test reliability has been derived on the assumption that the error variance is the same for the two groups, the entire difference in observed variance being attributed to a difference in true (score) variance of the two groups" (p. 124, emphasis original).

Under the assumption that the error score variance remain constant, the difference in the observed score variance is entirely attributed to the difference in true score variance. Let s_x and r_{xx} represent the standard deviation and score reliability for Group X, and let s_y and r_{yy} represent the standard deviation and score reliability for Group Y. The measurement error for Group X and Group Y are as follows:

$$\begin{aligned} s_{e_x} &= s_x\sqrt{1-r_{xx}} \\ s_{e_y} &= s_y\sqrt{1-r_{yy}} \end{aligned} \qquad (1)$$

Under the assumption of invariant error score variance for the two groups, the following relationship exists:

$$s_x\sqrt{1-r_{xx}} = s_y\sqrt{1-r_{yy}} \qquad (2)$$

Thus, theoretically, the amount of change in reliability to be expected from any given change in observed score variance, under the assumption that this change is due entirely to the difference in true score variance, can be derived from Equation 2 and represented as follows (Gulliksen, 1987):

$$r_{yy} = 1 - \frac{s_x^{\;2}}{s_y^{\;2}}(1 - r_{xx}) \qquad (3)$$

where s_x^2 is the variance of an existing group (X) on a given test, and r_{xx} is the measurement reliability for this group. s_y^2 is the variance of a new group (Y) on the same test, and r_{yy} is the expected reliability for the new group (Y). Little empirical research, however, has been conducted to evaluate the extent to which the assumption of invariant error variance, and the theoretical relationship represented above, are tenable in measurement practice.

In addition to group heterogeneity/homogeneity, another potential factor that may affect measurement reliability is the performance level of a group. As discussed previously, it is usually assumed that measurement error is invariant across groups that differ in their performance levels. And it is under this assumption that the formula (Equation 3) presented previously was developed. This assumption, however, is far from being certain in practice (Crocker & Algina, 1986). As early as in the 1920s, this assumption was seriously questioned (Holzinger, 1921). As discussed by Gulliksen (1987), ". . . the error of measurement for a given test may vary with the ability of the group" (p. 111). It is thus reasonable to hypothesize that the relative magnitude of error score variance may be related to the performance level of the group to a certain degree. The relative magnitude of error score variance, in turn, will affect the measurement reliability. This potential factor has received little attention in empirical research in the area of measurement reliability.

This paper intends to empirically examine the two issues discussed above. More specifically, we attempted to address the following two questions:

(1) To what extent the group heterogeneity/homogeneity with regard to the trait being measured may affect measurement reliability? And to what extent the theoretical relationship between group heterogeneity/homogeneity and measurement reliability under the assumption of invariant measurement error will hold empirically?

(2) How does sample performance level affect measurement reliability? Will measurement error remain invariant for groups with different performance levels?

Methods

Data Description

Two data sets, one from Texas Assessment of Academic Skills (TAAS), a primarily criterion-referenced testing program, and the other from Iowa Test of Basic Skills (ITBS), a primarily norm-referenced testing program, are used for the empirical investigation in this paper.

TAAS Data

One data source used in this study is from the Texas Assessment of Academic Skills (TAAS) tests administered in 1992 and taken by 11th Grade students at the time. Designed for assessing the mastery of school instructional objectives, TAAS was a state-mandated criterion-referenced test battery consisting of Reading, Math and Writing tests. Used in this study were the Reading (48 items) and the Math (60 items) tests that consisted of multiple-choice items scored dichotomously as either correct or incorrect. Un-attempted items were scored as incorrect responses. The examinee pool for the database has over 193,000 subjects. A random sample of 50,000 was used in this study. TAAS was designed to be a test battery for assessing minimum-competency of students in Texas public schools. As is typically the case for mastery tests in education, TAAS test items were primarily curriculum content-based, and test score distributions were somewhat negatively skewed, indicating some ceiling effect of the score distributions.

ITBS Data

A random sample (n=10,000) from the 1999-2000 administration of Iowa Test for Basic Skills (ITBS) taken by 6th Graders were obtained from the Iowa Testing Program. The ITBS data have 44 items for ITBS Reading, and 121 items for ITBS Math. The Math portion of ITBS has three sections: Math Concepts (48 items), Math Problems (32 items), and Math Computations (41 items). All items are in multiple-choice format and are dichotomously scored as either correct or incorrect. Un-attempted items are counted as incorrect responses. ITBS is designed primarily for use as norm-referenced measurement instrument, and the sample test score distributions indicate that they are more symmetric than the TAAS distributions, with the skewness being closer to zero. Table 1 presents the test score distribution characteristics for both TAAS and ITBS data used in this study.

Insert Table 1 about here

Examinee Sampling for Group Heterogeneity

For examining the effect of group heterogeneity on measurement reliability estimates, samples are drawn in such a way that the performance level is controlled but the sample variance is varied. More specifically, samples were repeatedly drawn from each of the following conditions:

a) Random samples were drawn with no restriction on group heterogeneity, and these samples represent the original test data population;

b) Random samples were drawn from the middle 90% of the original test data population, with those above 95th percentile and those below the 5th percentile being excluded from being sampled. This condition represents the lowest degree of range restriction implemented in this study in terms of score variance.

c) Random samples were drawn from the middle 80% of the original test data population, with those above 90th percentile and those below the 10th percentile being excluded from being sampled. This condition represents somewhat more severe range restriction implemented in this study in terms of score variance.

d) Random samples were drawn from the middle 50% of the original test data population, with those above 75th percentile and those below the 25th percentile being excluded from being sampled. This condition represents the most severe range restriction implemented in this study in terms of score variance.

Because the range of restriction was implemented symmetrically from the two tails of the score distributions, the samples drawn under each of the previous four conditions have approximately the same performance level as measured by group means, but the samples under different sampling conditions systematically differ in terms of the group score variance. Measurement reliability estimates were obtained from each sample for later analyses.

Examinee Sampling for Group Performance

For examining the effect of performance level on measurement reliability, samples are drawn in such a way that the sample performance level is systematically varied while we attempted to control the sample score variance. More specifically, the following sampling conditions were implemented for drawing random samples from the two test data sets:

i) a) Random samples were drawn from the upper 75% of the original test data population, and those below the 25th percentile on the test in question were excluded from such sampling;

b) Random samples were drawn from the lower 75% of the original test data population, and those above the 75th percentile on the test in question were excluded from such sampling;

ii) c) Random samples were drawn from the lower 50% of the original test data population, and those above the median on the test in question were excluded from such sampling;

d) Random samples were drawn from the upper 50% of the original test data population, and those below the median on the test in question were excluded from such sampling.

For a truly symmetric score distribution, each pair of a sampling condition above [e.g., a) and b) under i)] would allow systematic variation in performance level, while holding the sample score variability constant. Unfortunately, because the data distributions used in this study are skewed (see Table 1), it would not be possible to hold sample variability constant while performance level changes.

For TAAS data, TAAS Reading and TAAS Math were treated as separate tests, and sampling plans discussed above were implemented independently for each section. For ITBS data, ITBS Reading and ITBS Math were also treated as separate tests with sampling plans implemented independently for each. The three sub-tests under ITBS Math (Math Concepts, Math Problems, and Math Computations) were not considered as separate tests, and the distinction among them was ignored in this study.

For each of the sampling plans discussed above, random samples of n=500 were drawn from the TAAS data. Score reliability estimate in the form of Cronbach's coefficient α was computed for each sample for the test involved (TAAS Reading or TAAS Math), and the reliability estimate was saved for later analyses. For ITBS data, random samples of n=100 were drawn from the data, and reliability estimate in the form of Cronbach's coefficient α was computed for each sample for the test involved (ITBS Reading or ITBS Math), and the reliability estimate was saved for later analyses. At the same time, each sample's mean score and standard deviation were also obtained and saved together with the reliability estimate.

Results and Discussion

Table 2 presents the results for assessing the relationship between group variability and measurement reliability. In Table 2, for each test, four sample variability conditions were used: samples drawn from the full range of the original data "population" with no range restriction, samples drawn from the middle 90% (5th – 95th percentile), the middle 80% (10th – 90th percentile), and the middle 50% (25th-75th percentile), of the original data "population". As shown in Table 2, for each test, the performance levels of samples under different conditions are reasonably comparable. For example, for TAAS Reading Test data, the average $\overline{X}$ across samples under each of the four sampling conditions are 37.30, 38.20, 38.65, and 38.90, respectively. As expected, the difference in group variability is systematic and obvious (7.52, 5.96, 4.98, and 2.81, respectively). The average reliability estimates, under the column heading "Average α", for the four conditions are also obviously different. Corresponding to the decreasing group variability, the measurement reliability decreasing from samples drawn from the full range of the original data "population" ($\bar{\alpha}$=.8874) to samples drawn from the original middle 50% of the data "population" ($\bar{\alpha}$=.2396).

The observed relationship between group variability and measurement reliability is certainly expected. The question is, however, to what extent this empirical relationship conforms to the theoretical relationship between the two as specified by Equation 3 presented previously? To answer this question, theoretically-expected reliability coefficients were obtained, using the full-range sample as the reference group for each test, and these theoretically-expected reliability coefficients were presented under the column heading "Formula-Based α". For example, for TAAS Reading Test, the "full range" condition was used as the reference group, and the sample variability (sd=7.52) and the reliability coefficient (α=.8874) were empirically obtained. We then ask the question, if the sample variability changes to sd=5.96, 4.98, and 2.81, respectively, how would the change in group variability theoretically impact the measurement reliability? If the theoretically-expected reliability coefficients have the same magnitude as the empirically estimated reliability coefficients, it would indicate that the empirical relationship between group variability and measurement reliability conforms well to the theoretically expected relationship between the two.

As shown in Table 2 under the column heading "Formula-Based α", the theoretically-expected reliability coefficients for the given group variability are generally very close (with difference in the 3rd decimal place for most cases) to the empirically obtained reliability coefficients (under the column heading "Average α"), except in cases where the restriction of range is severe (samples drawn only from the middle 50% of the original data "population"). These results are consistent across the subject tests within the same test battery (Reading and Math), and are consistent across the two test batteries (TAAS and ITBS). This consistency across test batteries and across different content areas indicates that the results observed here are likely to be replicable in other measurement situations. So, for the first research question in this paper ("To what extent the group heterogeneity/homogeneity may affect measurement reliability, and the theoretical relationship between group heterogeneity/homogeneity and measurement reliability under the assumption of invariant measurement error will hold empirically?"), the results here suggest that, when performance levels are comparable, the assumption of invariant measurement error is empirically tenable. Furthermore, measurement reliability largely depends on group variability, as classical reliability theory predicts.

Table 3 presents the measurement reliability estimates for samples with different performance levels, and the results here were intended to answer the second research question, "How performance level of groups affect measurement reliability? Will measurement error remain invariant for groups with different performance levels?". Unfortunately, the situation here is complicated because of the confounding of both performance level difference and group variability difference at the same time. For example, for TAAS Reading, between Upper 75% and Lower 75% data "populations", the difference in performance is obvious (average of sample means of 41.03 vs. 34.88, respectively). But in addition to this performance difference, the two data "populations" also differ in terms of their variability (average of sample standard deviations of 3.90 vs. 6.85, respectively). This difference in variability is the result of the skewed distribution in TAAS Reading scores (skewness=-.93, see Table 1). The skewness makes the TAAS Reading score distribution non-symmetrical, with the long tail pointing to the lower end of the distribution. Consequently, the Upper 75% of the original distribution has much less variability than the Lower 75% of the original data distribution. Similar situation exists for other high-performance vs. low-performance comparisons for TAAS Math, and to a lesser degree, for

ITBS Reading, which is less negatively skewed than TAAS tests. ITBS Math, however, has almost symmetrical distribution (skewness=.03, see Table 1). As a result, for ITBS Math, the high-performance vs. low-performance comparisons (Upper 75% vs. Lower 75% "populations", Upper 50% vs. Lower 50% "populations") has little confounding of variability differences (very similar standard deviations).

The confounding of group performance and group variability in Table 3 makes it somewhat difficult to assess the "pure" impact of performance on measurement reliability, because the difference in reliability estimates between the high-performance and the low-performance groups may be due either to group variability difference, or to performance difference, or to certain combination of the two. To make it possible to assess the impact of performance level on measurement reliability, some adjustment is needed for the group variability difference. From previous analysis (Table 2) for the condition of comparable performance, the validity of the theoretical relationship between group variability and measurement reliability is largely confirmed empirically. If group performance level had no impact on measurement reliability, we would expect that, after adjusting for the group variability difference, the measurement reliability for the high-performance group would be very close to that of the low-performance group. For example, for the comparison of Upper 75% vs. Lower 75% "populations" on TAAS Reading, the observed alpha is .6768 for the high-performance group, substantially lower than the .8407 for the low-performance group. But the standard deviations for the two "populations" are also very different (3.90 vs. 6.85). We ask the question, "Given that the high-performance group has lower _sd_ of 3.90, and alpha of .6768, if this group had the same _sd_ as the low-performance group (6.85), what would the theoretically-expected measurement reliability be for this high-performance group?" The same question was asked for all other high- vs. low-performance group comparison, and the adjusted alphas for the high-performance groups in each comparison after adjusting for the variability difference (i.e., sd) are presented under the column heading of "SD-adjusted α".

After adjusting for the group variability difference, there is the tendency that, after taking into the consideration of group variability difference, the high-performance group has higher measurement reliability. For example, in Table 3, for the first high- vs. low-performance comparison (TAAS Reading, Upper 75% of the data "population" vs. Lower 75% of the data "population"), the sd-adjusted α for the high-performance group is .8952. This adjustment assumes that the high-performance group had the same sd of 6.85 as the low-performance group. If the performance level had no impact on measurement reliability, it would be expected that this adjusted α would be equivalent to the observed measurement reliability of the low-performance group. The fact that the sd-adjusted α for the high-performance group (.8952) turned out to be higher than the low-performance group's measurement reliability (.8407) suggests that, if the two groups had the same sample variability, the measurement reliability would be higher for the high-performance group than that of the low-performance group. In other words, for the two measurement data sets examined, there tends to be more measurement error for the low-performance group. Similar observation was made for all the other high- vs. low-performance comparisons in Table 3. Even for the last two comparisons involving ITBS Math where high- and low-performance groups have very similar group variability, thus little adjustment actually occurred, the high-performance group still shows higher measurement reliability than the low-performance group. The measurement reliability difference between the groups is much noticeable when the group performance difference is large (Upper 50% vs. Lower 50%) than when the group performance difference is smaller (Upper 75% vs. Lower 75%).

Conclusions

The preliminary findings from this exploratory empirical investigation suggest that, as expected in measurement theory, sample variability with regard to the trait being measured has obvious effect on measurement reliability, with measurement reliability being reduced by group variability restriction. Classical test theory usually assumes that observed group score variability difference is the result of true score variance difference only, and the measurement error variance remains invariant. When performance levels of the groups are comparable, this assumption appears to be tenable, because the theoretically predicted measurement reliability estimates are largely consistent with the empirically observed measurement reliability estimates.

Group performance level also appears to affect measurement reliability. For the data examined, after adjusting for the difference in group variability, measurement scores of the lower performing group tend to contain more measurement error, and consequently, their scores have lower measurement reliability. The larger the performance difference is, the more noticeable the difference in measurement reliability between the high- and low-performing groups. The preliminary findings here, however, are not as clear-cut, because the skewed score distributions made it difficult to control for group variability while examining the impact of performance. Although ad hoc adjustments were made to control for the difference in group variability while examining the impact of performance level on measurement reliability, the confounding between the two factors for the data warrants caution in drawing any conclusions in this regard. Despite the need for this caution, the preliminary findings here suggest that reliability generalization studies should consider sample characteristics as relevant variables, both in terms of sample group variability, and in terms of sample group performance level.

The most obvious limitation of this exploratory study is our lack of experimental control for examining the effects of different measurement score characteristics on score reliability. The lack of experimental control resulted in confounding between group variability and group performance, making it difficult to draw any definite conclusions about the effect of group performance level on measurement reliability. Given extant measurement data, such control is very difficult, if not impossible. Future research in this area may benefit from Monte Carlo approach by which measurement data with researcher-specified data characteristics are generated under an experimental design so that any potential confounding of the two factors can be avoided.

References

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Fort Worth, TX: Holt, Rinehart and Winston, Inc.

Gulliksen, H. (1987). Theory of mental tests. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Holzinger, K. J. (1921). On the assumption that errors of estimate are equal in narrow and wide ranges. Journal of Educational Research, 4, 237-239.

Magnusson, D. (1967). Test theory. Boston, MA: Addison-Wesley.

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

Yin, P., & Fan, X. (2000). Assessing the reliability of Beck Depression Inventory scores: Reliability generalization across studies. Educational and Psychological Measurement, 60, 201-223.

Table 1 Score Distribution Characteristics of Data Used

	TAAS		ITBS				
	Reading	Math	Reading	Math	(Concpt	Prob.	Comp) [a]
Mean	37.30	42.53	27.43	68.95	28.53	18.24	22.19
Sd	7.53	11.20	8.87	19.25	7.80	5.99	7.41
Skewness	-.94	-.85	-.28	.03	-.17	-.01	.09
Kurtosis	.66	-.47	-.85	-.72	-.59	-.77	-.64
100%ile (Max.)	48	60	44	118	48	32	41
75%ile (Q3)	43	52	35	83	34	23	28
50%ile (Mdn)	39	44	28	69	29	18	22
25%ile (Q1)	33	35	21	54	23	14	17
0%ile (Min.)	0	0	3	20	5	2	1

[a] These are the three sub-tests under ITBS Math: Math Concept, Math Problems, and Math Computation.

Table 2 Group Heterogeneity and Measurement Reliability Estimates

Test	Hetero- geneity	# of Samples	Average $\overline{X}$	Average sd.	Average α	Formula- Based α
TAAS	Full Range	100	37.30	7.52	.8874	--
Reading	Middle 90%	91	38.20	5.96	.8256	.8207
	Middle 80%	81	38.65	4.98	.7551	.7433
	Middle 50%	50	38.90	2.81	.2396	.1936
TAAS	Full Range	100	43.00	10.07	.8857	--
Math	Middle 90%	90	43.59	8.41	.8347	.8361
	Middle 80%	81	44.21	7.40	.7896	.7883
	Middle 50%	49	44.36	4.28	.4277	.3673
ITBS	Full Range	100	27.43	8.85	.8995	--
Reading	Middle 90%	89	27.81	7.56	.8582	.8623
	Middle 80%	80	28.38	6.64	.8149	.8215
	Middle 50%	50	28.82	3.99	.4668	.5056
ITBS	Full Range	100	68.95	19.21	.9384	--
Math	Middle 90%	90	69.25	16.21	.9121	.9135
	Middle 80%	80	69.09	13.93	.8800	.8829
	Middle 50%	50	69.06	8.20	.6495	.6619

Table 3 Performance Level and Measurement Reliability Estimates

Test	Performance	# of Samples	Average $\overline{X}$	Average sd.	Average α	sd-Adjusted α [a]
TAAS	Upper 75%	74	41.03	3.90	.6768	.8952
Reading	Lower 75%	76	34.88	6.85	.8407	.8407
	Upper 50%	51	43.04	2.63	.4500	.8961
	Lower 50%	49	31.10	6.05	.7550	.7550
TAAS	Upper 75%	74	47.88	6.28	.7430	.8493
Math	Lower 75%	74	38.60	8.20	.8104	.8104
	Upper 50%	49	51.52	4.12	.5464	.8364
	Lower 50%	51	34.90	6.86	.7183	.7183
ITBS	Upper 75%	75	31.39	5.99	.7962	.8549
Reading	Lower 75%	74	23.71	7.10	.8215	.8215
	Upper 50%	50	35.02	3.86	.5985	.8058
	Lower 50%	50	20.11	5.55	.6920	.6920
ITBS	Upper 75%	74	77.33	14.22	.8923	.8906
Math	Lower 75%	75	60.74	14.11	.8788	.8788
	Upper 50%	50	84.73	10.95	.8320	.8162
	Lower 50%	50	52.73	10.47	.7741	.7741

[a] These are measurement reliability estimates after adjusting for the difference in group variability (group standard deviation). The sd. of the lower-performing group was used as the reference sd. for both groups.

TM033082

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

ERIC

Reproduction Release

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Sample Characteristics and Measurement Reliability: An Empirical Exploration

Author(s): Xitao Fan, Ping Yin

Corporate Source: University of Virginia

Publication Date: April, 2001

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 in the indicated space following.

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)	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)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
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.

Signature:

Printed Name/Position/Title: Xitao Fan, Associate Professor

Organization/Address:
Curry School of Education
University of Virginia
PO Box 400277
Charlottesville, VA 22904-4277

Telephone: (804)243-8906

Fax: (804)924-1384

E-mail Address: xfan@virginia.edu

Date: May 29, 2001