DOCUMENT RESUME

ED 479 331 TM 035 135

AUTHOR	Perkhounkova, Yelena; Koehly, Laura M.; Dunbar, Stephen B.
TITLE	Modeling the Dimensions of Language Achievement with Categorical Data Methods.
PUB DATE	2000-04-00
NOTE	18 p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 2000).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Goodness of Fit; *Junior High School Students; Junior High Schools; Language Proficiency; *Language Tests; Structural Equation Models
IDENTIFIERS	*Categorical Data; *Dimensionality (Tests); Iowa Tests of Basic Skills; Latent Variables

ABSTRACT

This study was conducted to investigate the dimensionality of language tests by means of latent variable models for categorical data. It differs from previous studies by conducting the analysis at the level of the original items using the structural modeling approach of B. Muthen (1984) for dichotomous and ordered polytomous variables. In this approach, a multivariate regression model describes the relationship between a set of outcome variables, whether continuous, dichotomous, or ordered categorical, and a set of latent predictor variables. Data were test scores of the national sample of seventh and eighth graders who participated in the joint administration of Form 1 of "Thinking about Language," a constructed response supplement to the Iowa Tests of Basic Skills (ITBS, and Form M of the ITBS and test scores of the national sample of seventh and eighth graders who participated in the ITBS fall 1992 national standardization of two ITBS forms. Fitting latent variable models to categorical data provides a direct means of assessing the extent to which conditional dependencies might exist among items with particular characteristics. The slightly better fit of the five latent variable model with one higher-order latent variable with paths to each first-order latent variable to all language tests in this study indicates the existence of some such dependencies unless latent variable models with content-specific dimensions are considered. This study advances the understanding of the dimensionality structures of different types of language tests and provides insights into using latent variable models for categorical data in the assessment of dimensionality. (Contains 4 tables and 15 references.) (SLD)

Modeling the Dimensions of Language Achievement with Categorical Data Methods

Yelena Perkhounkova

Laura M. Koehly

and

Stephen B. Dunbar

The University of Iowa

Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Introduction

A wealth of research has been devoted to understanding similarities and dissimilarities of results from tests with different response formats. The comparisons have involved tests composed of various kinds of multiple-choice (MC) items, constructed-response (CR) items, as well as combinations of items of different formats. Some researchers investigated whether the scores obtained on such tests could be considered as indicators of one construct or of different constructs (Ackerman & Smith, 1988; Breland & Gaynor, 1979; Bridgeman, 1992; Hoover & Bray, 1995; Ward, 1982). Others presumed that the changes in test format actually altered the measured construct (Frederiksen, 1984) and sought to reveal differences between constructs measured by the tests of different formats. Results vary greatly across content areas, format types, and purposes of assessment.

The whole body of research in this area can be viewed within the framework of test validity, that is, validation of the proposed interpretations and uses of the scores from the tests of different formats. Construct validity, as delineated by Messick (1989), "is based on an integration of any evidence that bears on the interpretation or meaning of the test scores" (p. 17). Messick distinguished between six sources of construct validity evidence: content, substantive, structural, external, generalizability, and consequential.

The structural aspect of construct validity, which was the focus of this study, includes investigation of the dimensionality of the test. A construct that is perceived as having a particular pattern of dimensionality would generate expectations of specific interrelationships among parts of the test. "The nature and dimensionality of the interitem structure should reflect the nature and dimensionality of the construct domain, and every effort should be made to capture this structure at the level of test scoring and interpretation" (Messick, 1989, p. 44).

This study's objective was to investigate the dimensionality of language tests by means of latent variable models for categorical data. Researchers have routinely employed related techniques for exploring the dimensionality of achievement tests, and format effects in particular. This study differs from previous approaches by conducting the analysis at the level of the original items using Muthén's (1984) structural modeling techniques for dichotomous and ordered polytomous variables. In this approach, a multivariate regression model describes the relationship between a set of outcome variables, that can be continuous, dichotomous or ordered categorical, and a set of latent predictor variables. Because of their assumption of multivariate normality, standard latent variable models are, strictly speaking, inappropriate for categorical item data.

Muthén assumes that item responses result from categorization of underlying normal variables and suggests using tetrachoric and polychoric correlations instead of Pearson's product moment correlations to measure interitem association because the latter would attenuate the actual relationship among the underlying variables and produce bias in chi-square tests of fit, parameter estimates, and standard errors (West, Finch, & Curran, 1995). The use of tetrachoric and polychoric correlations in a weighted least squares estimation procedure leads to unbiased, consistent, and efficient parameter estimates. Simulation studies (Muthén & Kaplan, 1985; Schoenberg & Arminger, 1989) suggest that Muthén's model is appropriate for situations in which the item response formats, like in the current study, allow for few categories and the distributions of the responses are sometimes highly and differentially skewed.

Previous investigation of dimensionality of language tests with different response formats by means of the Poly-DIMTEST procedure (Li & Stout, 1995) found an interaction of item content and item format (Perkhounkova & Dunbar, 1999). With regard to content, the MC tests were judged essentially unidimensional, whereas the analysis provided strong evidence that the CR tests approximately conformed to simple structure corresponding to the content specifications of the measures. Furthermore, content-related heterogeneity of the CR tests remained evident in the analysis combining the CR items and the MC items. This study expanded the investigation of the apparent interaction of item content and item format by including two additional MC language tests in the analysis, and by using techniques better suited to isolating sources of variation among items.

Method

Subjects

The following data sources were used in this study: (1) Test scores of the national sample of 7th and 8th graders who participated in the joint administration of Form 1 of *Thinking about Language: Constructed-Response Supplement to The Iowa Tests* and Form M of the *Iowa Tests of Basic Skills (ITBS)* during the winter of 1997 and (2) Test scores of the national sample of 7th and 8th graders who participated in the *ITBS* fall 1992 national standardization of Forms K and L for purposes of equating parallel forms.

Instruments

The *ITBS* is a battery of MC achievement tests in several subject areas (Hoover, Hieronymus, Frisbie, & Dunbar, 1993). The following *ITBS* tests were of interest for this research: the *Integrated Writing Skills Test* (*IWST*) (55 items at grade 7 and 57 items at grade 8), the four-part language test in the Complete Battery (separately timed tests in Spelling, Capitalization, Punctuation, and Usage and Expression) with a total of 138 items at grade 7 and 144 items at grade 8, and the Survey Battery test in Language (separately timed subsets of items in the five skill areas) with 64 items at grade 7 and 68 items at grade 8.

The test format of the *ITBS* Complete Battery Spelling, Capitalization, Punctuation, and Usage and Expression tests and the Survey Battery Language test differs from that of *IWST*. The *IWST* consists of several texts—stories, reports, and letters—that contain errors in spelling, punctuation, capitalization, written expression, and language usage incorporated throughout the test. In contrast, the Spelling, Capitalization, Punctuation, and Usage and Expression tests measure each language skill in isolation from the others in a set of separately timed administrations. The survey language test is a shortened version of the four-part language test in which the MC item response format is the same, but the administration takes place in a single session. In addition, the MC item response format of the *IWST* was different from that of the four-part language tests and the survey language tests.

The *Constructed-Response Supplement (CRS)* to the *ITBS* (Hoover, Hieronymus, Frisbie, & Dunbar, 1998) in the area of language includes three parts: editing, revising, and generating ideas. The language test includes 26 items (52 total score points) at grade 7 and 30 items (60 total score points) at grade 8. Depending on the complexity of the items, responses are scored, on a 0–1, a 0–1–2, or a 0–1–2–3 scale. Items within parts conform to the same general content specifications used in the MC tests (spelling, capitalization, punctuation, usage, and written expression). Similar to the *IWST*, the CR tests include several texts that contain multiple types of errors integrated throughout the test.

In particular, the sets of scores from the following *ITBS* tests were examined in separate analyses:

(1) *Thinking about Language*, Constructed-Response Supplement to The Iowa Tests (953 and 882 records in grades 7 and 8, respectively),
(2) *IWST*, Form M (953 and 882 records at grade 7 and 8, respectively),
(3) Survey Battery Language Test, Form K (1550 and 1622 records in grades 7 and 8, respectively),
(4) Complete Battery Language Tests in Spelling, Capitalization, Punctuation, and Usage and Expression, Form K (1566 and 1595 records in grades 7 and 8, respectively).
(5) Composite Language Test of *Constructed-Response Supplement* and *IWST* (952 and 882 records in grades 7 and 8, respectively).

Procedures

The dimensionality structures of the language tests described in the previous section were explored in this study by means of latent variable models for categorical data. The fit of various models suggested by previous research was examined.

The analysis was based on the content similarity of the tests under investigation. At both grades, the language tests included items that were designed to measure spelling, punctuation, capitalization, language usage, and written expression skills. All of the items in these tests could be classified into one of the five content categories. The composition of the tests allowed comparisons of the effects of the item content on the tests' dimensionality across test formats as well as grades.

The analysis included fitting models with a single latent variable, with five latent variables (5 LVs for the separate skill areas and 1 higher-order LV with paths to each first-order LV). These models were fit to the data from the CR tests, the four-part language tests, and the survey tests. Separate analyses were conducted for grades 7 and 8 for cross-validation purposes.

Parameters of these models were estimated with Mplus (Muthén & Muthén, 1998). Using the tetrachoric (for dichotomous data) or polychoric (for ordered categorical variables) correlations, the weighted least-squares parameter estimates with robust standard errors and mean-adjusted chi-square test statistic (Muthén, 1984) were estimated. Asymptotic theory for the estimator is discussed in Muthén (1984) and Muthén and Satorra (1995).

Although a variety of tests were included in this study, comparisons to evaluate the strength of latent variables defined by content and format should be based on tests that are comparable in terms of their administration times. Thus, comparison of results from the survey language test (30 minutes), *IWST* (40 minutes), and constructed-response test (35 minutes) as well as comparison of results from the four-part language test (60 minutes) and *IWST* combined with CR test (75 minutes) are emphasized to the extent that they correspond to patterns in goodness-of-fit.

Results

Tables 1 and 2 contain basic information about the samples and instruments used in this study. The number of items per skill area varies as a function of the overall length of the test. Note that although a separate entry appears for the combined CRS/IWST, this row of both Tables 1 and 2 describes the same student records, with the number of items aggregated.

Inspection of the distributions of scores from each test included in this study revealed raw-score distributions that were generally symmetrical. Individual items on each of these tests were examined for difficulty and discrimination prior to assembly of the final, published forms. Generally speaking, item difficulties for the multiple-choice tests ranged from about .30 to about .90 with a national mean difficulty in the neighborhood of .60. On the whole, the item data did not suggest that any serious artifacts would be introduced to the categorical modeling because of extreme distributions of item difficulty or poor correlations between item scores and total scores.

As described previously, latent variable models for these categorical data were estimated for each instrument. Models with one and five latent variables were fit to the item response data for the five sets of items. Fit of the data to the LV models was assessed by the mean-adjusted χ^2 goodness-of-fit measures produced by Mplus and root mean squared residuals (RMSR) computed subsequently. In addition, for each set of items the difference between the two latent-variable models (1 latent variable versus the hierarchical 5 latent variables) was assessed directly by computing residuals between the fitted matrices for each model. The results of all Mplus model fitting and model comparisons are given in Table 3.

The grade 7 and grade 8 samples were included in this study so that a replication condition could shed light on any consistency, or lack thereof, in model fit due to characteristics of the particular item set used at a given grade level or of the particular examinees included in the samples. Generally speaking, the fit statistics for the grade 7 and grade 8 samples are similar, regardless of the particular model or item set in question. The exception to this is the CRS, for which *slightly* better fit for the one-latent-variable (1LV) model was observed in grade 8 compared to grade 7. Other differences between residuals for the two grades appear to be essentially random.

To some degree, the similarities in fit across grades could be due to the composition of the item sets. Grade 7 and grade 8 forms of the *IWST* and the *ITBS* Complete Battery contain overlapping items as part of instrument design, so similarities in fit across grades are not surprising for those item sets. The *ITBS* Survey Battery was developed from Complete Battery items in such a way that no item overlap occurs, yet the fit in grades 7 and 8 was quite similar for each LV model. In contrast, the CRS item set, which contain no overlapping items across grade, showed small differences in degree of fit by grade level. Of further interest would be analyses of additional grade levels for both MC and CR item sets to determine if model fit for MC items shows greater consistency than model fit for CR items. The results presented here hint that this may be true.

The remainder of the results presented in Table 3 are indicative of substantial similarity in the fit of the 1LV and 5LVH models for nearly every item set included in the Mplus analyses. There is a consistent improvement in fit when additional latent variables are included for each grade and item set, and the differences between χ^2 likelihood ratio statistics, not surprisingly given the large samples, exceed critical values for any reasonable significance level. However, the increments in fit, as measured by residuals, are no greater in magnitude than the differences in fit between grade levels discussed previously.

Again, the exception to this rule appears to lie in the results for CRS, particularly in grade 7. The likelihood-ratio statistic for this case dropped by two-thirds between the 1LV and 5LVH models, and the drop in RMSR, while small, was markedly greater than for any other item set or grade level. The fact that the most noticeable effect of respecification of the 1LV model to a hierarchical model occurred for a CRS item set may be indicative of a format influence, possibly idiosyncratic with respect to the particular set of items in the test given to the grade 7 cohort.

Further examination of this possibility was conducted by means of a detailed inspection of residuals for the CRS 5LVH models in the grade 7 and 8 samples. This inspection revealed relatively large residual covariances related to six item pairs in the grade 8 CRS. Inspection of the test booklets revealed these pairs to correspond to either structural aspects of instrument design or of language skills. Three of the six item pairs involved the last three items in the test booklet, which were all based on the same writing task and which required the production of original ideas and sentences for a written report. Another item pair involved an editing situation in which two separately scored items appeared at the boundary of adjacent sentences. The dependency between these items involved the linguistic structure of the construction such that a particular type of change to one sentence would trigger a corresponding revision to the next sentence. The remaining item pairs were highly specific editing situations involving the use of commas in personal letters and in compound and complex sentences.

When the 5LVH model was respecified with free parameters corresponding to the covariances in these item pairs to account for the structural dependencies observed, the likelihood-ratio statistic dropped to about half of the value for the 1LV model (χ^2 = 760.12, df = 394). This improvement in fit was still not as large as that observed in the grade 7 CRS, but did reinforce the concern that in the case of the open-ended items on the CRS, that idiosyncratic effects could be the cause of the slightly more variable fit statistics.

Table 4 contains the path coefficients of the second-order LV on the first-order LVs. These weights indicate the second-order LV to be defined to a great extent by the first-order LVs in the areas of language usage and expression. The contribution of the LVs representing spelling and language mechanics was stronger for the MC item sets than it was for any item set that included items from the CRS, indicating that more unique variance was associated with the CR item format.

Discussion

Fitting latent variable models to categorical data provides a direct means of assessing the extent to which conditional dependencies might exist among items with particular characteristics. The slightly better fit of the 5LVH models to all of the language tests in this study suggests the existence of some such dependencies unless LV models with content-specific dimensions are considered. However, it should be recognized that all of the tests in this study measured a dominant dimension of general language skills accounting for the vast majority of covariation among test items.

The tests consisting of CR items seemed most vulnerable to conditional dependence among items related to some distinctive feature of skill content or format. Moreover, the limited evidence regarding variation in goodness-of-fit by grade level was observed for CR items. That an explanation for conditional dependencies among CR items could be based on highly specific features of item content—these features were also present in MC items in the *ITBS* Complete and Survey Batteries as well as in the *IWST*, though they didn't create local item dependencies—underscores the importance of careful item development procedures in the CR format. This finding is consistent with previous results described in Perkhounkova & Dunbar (1999).

Writing open-ended questions frees the test developer from many restrictions created by the controls in distractor choice necessary for good MC item characteristics. However, the downside of this freedom, from the standpoint of LV modeling, is that it may introduce complexities into the psychometric structure of the resulting instrument such that simple models for scaling, equating, or item selection may not apply uniformly across the levels of a multilevel achievement test battery. Finding local item dependencies that may exist is a bit like finding a needle in a haystack in the sense that the effects are small, difficult to distinguish from chance, yet no less pointed in their potential influence on model fit.

CR items and MC items are often combined in one test in an attempt to cover a broader range of assessed skills while maintaining acceptable level of reliability. Test developers face the need to aggregate the scores on such assessments so as to obtain a meaningful summary score for each examinee. Aggregating test scores often raises concerns about the dimensionality of the composite that should be addressed before using traditional IRT models for scaling tests (Wilson & Wang, 1995). This study advances our understanding of the dimensionality structures of different types of language tests and provides insights into using latent variable models for categorical data in the assessment of dimensionality.

References

Ackerman, T. A. & Smith, P. L. (1988). A comparison of the information provided by essay, multiple-choice, and free-response writing tests. *Applied Psychological Measurement, 12,* 117-128.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Breland, H. M. & Gaynor, J. L. (1979). A comparison of direct and indirect assessments of writing skill. *Journal of Educational Measurement, 16,* 119-128

Bridgeman, B. (1992). A comparison of quantitative questions in open-ended and multiple-choice formats. *Journal of Educational Measurement, 29,* 253-271.

Frederiksen, N. (1984). The real test bias: influence of testing on teaching and learning. *American Psychologist, 39,* 193-202.

Hoover, H. D. & Bray, G. B. (1995, April). *The research and development phase: Can a performance assessment be cost effective?* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Hoover, H.D., Hieronymus, A. N., Frisbie, D. A., & Dunbar, S. B. (1993). *Iowa Tests of Basic Skills, Forms K/L, Interpretive Guide*. Chicago: The Riverside Publishing Company.

Hoover, H.D., Hieronymus, A. N., Frisbie, D. A., & Dunbar, S. B. (1998). *Manual for Scoring and Interpretation: Constructed Response Supplement for the Iowa Test of Basic Skills (Levels 9-14).* Chicago: The Riverside Publishing Company.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.). New York: Macmillan.

Muthén, B. (1984). A general structural equation modeling with dichotomous, ordered, categorical, and continuous latent variable indicators. *Psychometrika, 49,* 115-132.

Muthén, L. K. & Muthén B. (1998). *Mplus User's Guide*. Los Angeles: Author.

Muthén B. O. & Satorra, A. (1995). Technical aspects of Muthén's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika, 60,* 489-503.

Perkhounkova, Y. & Dunbar, S. B. (1999). *Influences of Item Content and Format on the Dimensionality of Tests Combining Multiple-Choice and Open-Response Items: An application of the Poly-DIMTEST Procedure.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Ward, W. C. (1982). A comparison of free-response and multiple-choice forms of verbal aptitude tests. *Applied Psychological Measurement, 6,* 1-11.

Wilson, M. & Wang, W. (1995). Complex composites: Issues that arise in combining different modes of assessment. *Applied Psychological Measurement, 19,* 51-71.

Table 1
Data Set Summary

	Number of items		Number of examinees	
Item Set	Grade 7	Grade 8	Grade 7	Grade 8
CRS	26	30	953	882
IWST	55	57	953	882
ITBS Survey	64	68	1550	1622
ITBS Complete	138	144	1566	1595
CRS/IWST	81	87	953	882

Table 2
Content Breakdown by Item Set

	Spelling		Capitalization		Punctuation		Usage		Expression	
	Grade									
Item Sets	7	8	7	8	7	8	7	8	7	8
CRS	4	4	5	5	5	7	5	7	7	7
IWST	5	4	9	9	9	8	8	8	24	28
Survey	16	17	13	14	13	14	11	12	11	11
Complete	39	41	29	30	29	30	21	22	20	21
CRS/IWST	9	8	14	14	14	15	13	15	31	35

Table 3

Summary of Mplus Results

	Grade 7				Grade 8			
Item Set	χ^2	df	χ^2/df	RMSR	χ^2	df	χ^2/df	RMSR
CRS								
1LV	3021.58	299	10.11	.117	1579.24	405	3.90	.091
5LVH	1042.94	294	3.55	.074	1166.73	400	2.92	.083
1LV-5LVH				.092				.036
IWST								
1LV	2235.81	1430	1.56	.062	2201.69	1539	1.43	.064
5LV	2113.40	1425	1.48	.060	2131.54	1534	1.39	.063
1LV -5LVH				.015				.011
ITBS Survey								
1LV	4316.08	1952	2.21	.053	4887.61	2210	2.21	.052
5LVH	3598.43	1947	1.85	.049	4101.14	2205	1.86	.048
1F-5FH				.021				.020
ITBS Complete[a]								
1F	--	--	--	.058	--	--	--	.055
5FH	--	--	--	.051	--	--	--	.050
1LV-5LVH				.028				.024
CRS/IWST								
1LV	10077.78	3159	3.19	.076	6563.96	3654	1.80	.070
5LVH	7595.11	3154	2.41	.071	6188.02	3649	1.70	.068
1LV-5LVH				.037				.016

Notes. Dashes indicate the results were not obtainable because RAM requirements exceeded the capacity of a Pentium III processor with 392MBs. RMSR = root mean square residual. 1LV = one latent variable model; 5LVH = hierarchical model including 5 first-order latent variables based on content (spelling, capitalization, punctuation, usage, expression); 1LV-5LVH = RMSR of direct comparison of 1LV and 5LVH.

[a]RMSR obtained with ULS estimates.

Table 4

First-Order Factor Loadings

	Spelling		Capitalization		Punctuation		Usage		Expression	
	Grade									
Item Set	7	8	7	8	7	8	7	8	7	8
CRS	.66	.76	.60	.77	.85	.86	.93	.97	.84	.79
IWST	.97	.92	.91	.94	.99	.88	.99	.97	.91	.95
Survey	.84	.81	.91	.94	.97	.98	.94	.96	.91	.88
Complete[a]	.82	.82	.89	.90	.95	94	.88	.90	.86	.88
CRS/IWST	.82	.79	.71	.90	.92	.90	.98	96	.92	.94

Notes. [a]Estimates obtained with ULS.

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

REPRODUCTION RELEASE

(Specific Document)

TM035135

I. DOCUMENT IDENTIFICATION:

Title: Modeling the Dimensions of Language Achievement with Categorical Data Methods	
Author(s): Yelena Perkhounkova, Laura M. Koehly, Stephen B. Dunbar	
Corporate Source: The University of Iowa	Publication Date: 04/24/2000

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 MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	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.

Signature: Perkhounkova	Printed Name/Position/Title: Yelena Perkhounkova, Research Associate	
Organization/Address: American College Testing Davidsen Bulding, 500 ACT Drive, Iowa City, IA 52243-0168	Telephone: (319) 337-1065	FAX:
	E-Mail Address: yelena.perkhounkova@act.org	Date: 07/23/2003

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:

ERIC Clearinghouse on Assessment and Evaluation
University of Maryland, College Park
1129 Shriver Lab
College Park, MD 20742

EFF-088 (Rev. 4/2003)-TM-04-03-2003