DOCUMENT RESUME

ED 360 368 TM 020 272

AUTHOR Carlson, James E.
TITLE Dimensionality of NAEP Instruments That Incorporate Polytomously-Scored Items.
INSTITUTION Educational Testing Service, Princeton, N.J.
PUB DATE Apr 93
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Computer Simulation; Correlation; Elementary Secondary Education; Factor Analysis; *Mathematics Tests; Maximum Likelihood Statistics; Models; National Programs; *Reading Tests; Sample Size; *Scoring; Student Evaluation; Testing Programs; *Test Items
IDENTIFIERS Dichotomous Scoring; *Dimensionality (Tests); LISREL Computer Program; *National Assessment of Educational Progress; Polytomous Scoring; Residuals (Statistics)

ABSTRACT

In this article some results are presented relating to the dimensionality of instruments containing polytomously scored as well as dichotomously scored items, concentrating on the 1992 National Assessment of Educational Progress' (NAEP) mathematics and reading assessment data and several simulated datasets. The maximum likelihood factor analytic procedure of the LISREL 7 computer program was used. Results were evaluated through use of residuals from the fitted model. The square root of the mean squared residual was the statistic used. Overall sample sizes for mathematics were 1,125, 1,173, and 1,064 for grades 4, 8, and 12, respectively. For reading, the sizes were 1,169, 1,271, and 1,139 for each grade, respectively. Results suggest that the dimensionality of data structures in the NAEP assessment is generally not affected by the inclusion of polytomously scored items, but the data structures cannot be generalized to other situations. One reason is the size of the correlations among the scales of the NAEP, and another is the small number of conditions simulated in this study. In addition, the number of polytomously scored items was limited in the 1992 assessment. Eight tables present analysis results, and four figures illustrate the square roots of the mean squared residuals (Contains 32 references.) (SLD)

Dimensionality of NAEP Instruments That Incorporate Polytomously-scored Items

James E. Carlson

Educational Testing Service

Paper presented at the annual meeting of the American Educational Research Association, Atlanta, April, 1993. The work reported in this paper was supported under the National Assessment of Educational Progress as administered by the National Center for Education Statistics, Office of Educational Research and Improvement, U. S. Department of Education.

Dimensionality of NAEP Instruments That Incorporate Polytomously-scored Items

James E. Carlson

Educational Testing Service

Introduction

Carlson & Jirele (1992), Muthén (1991), Rock (1991), and Zwick (1986, 1987) studied the dimensionality of portions of the NAEP data using different techniques. All these studies, however, used data from instruments containing only dichotomously-scored items. In this paper some results are presented relating to dimensionality of instruments containing polytomously-scored as well as dichotomously-scored items. In particular the 1992 NAEP mathematics and reading assessment data are analyzed as well as several simulated datasets.

As pointed out by Carlson & Jirele (1992, p. 1)

> theoretical or empirical studies of dimensionality that involve statistical/psychometric techniques involve item-response data resulting from the examinee-item interaction and **not** the dimensionality of items as entities separate from examinees.

Thus in this paper, as in the 1992 paper, we will refer to the dimensionality of a set of item response data, with the understanding that such data result from the examinee-item interaction in a specific population.

Methods of Assessing Dimensionality

A number of different methods of assessing dimensionality underlying test items have been developed and studied by a many authors (e.g., Bock, Gibbons, & Muraki, 1988; Christoffersson, 1975; Hattie, 1984, 1985; Holland & Rosenbaum, 1986; Knol & Berger, 1991; McDonald, 1981, 1982a, 1982b, 1985; Mislevy, 1986; Muthén, 1978; Rosenbaum, 1984; Stout, 1983, 1987, 1990). These authors, however, have for the most part concentrated on dichotomously-scored items.

Of the computer programs available for assessing the dimensionality of test items, only LISREL 7 (Jöreskog & Sörbom, 1989) incorporates a procedure that has two facilities required in order to analyze polytomously-scored items administered using the Balanced Incomplete Block (BIB) spiral design of NAEP: an option to denote items as "not administered", and facility to use all of the information in polytomously-scored items (through computation of polychoric correlation coefficients). Hence only the maximum likelihood factor analytic procedure of that program was used in this study.

Previous dimensionality analyses of NAEP data

NAEP Reading assessment data collected during the 1983-84 academic year was studied for dimensionality by Zwick (1986, 1987) who also examined simulated data designed to mirror the NAEP reading item-response data but having known dimensionality. Principal components analysis (PCA) was applied to both phi and tetrachoric correlation matrices and full information item factor analysis

(Bock & Aitkin, 1981; Bock, Gibbons, & Muraki, 1988) implemented in the TESTFACT computer program (Wilson, Wood, & Gibbons, 1991) were applied to portions of the dataset, as was Rosenbaum's (1984, 1985) dimensionality testing procedures. Analysis of the simulated datasets allowed her to determine whether the BIB Spiraling design artificially increases dimensionality. Zwick found substantial agreement among the various statistical procedures, and that the results using BIB spiraling were similar to results for complete datasets. Overall she concluded that "it is not unreasonable to treat the data as unidimensional (1987, p. 306)."

The topic of Rock's (1991) investigation was "whether the presently reported subscale scores do span a multidimensional space defined by the content area subscales at each of the three grade levels in mathematics and science (p. 1)." He formed two parcels of items that are homogeneous with respect to content for each subtest of the NAEP mathematics and science tests from the 1990 assessment, and studied their dimensionality using confirmatory factor analysis. The resulting factor intercorrelations averaged across booklets ranged from .86 to .95 in mathematics, and from .94 to .96 in science. Rock's conclusion was that there was little evidence for discriminant validity except for the geometry subscale at the 8th grade level, and that "we are doing little damage in using a composite score in mathematics and science (p. 2)."

A second-order factor model was used by Muthén (1991) in a further analysis of Rock's mathematics data, to examine subgroup differences in dimensionality. Evidence of content-specific variation within subgroups was found but the average (across 7 booklets) percentages of such variation was very small, ranging from essentially zero to 22, and two-thirds of these percentages were smaller than 10.

Carlson & Jirele (1992) used full information item factor analysis (Bock, Gibbons, & Muraki, 1988) as implemented in the TESTFACT computer program (Wilson, Wood, & Gibbons, 1991), and normal harmonic factor analysis (McDonald, 1962, 1967, 1981) as implemented in the NOHARM program (Fraser, 1988) to examine 1990 NAEP mathematics data at three grade levels. Analyses of simulated one-dimensional data were also conducted, and the fit to these data, as measured by the Root Mean Square Residual (RMSR) and the Akaike Information Coefficient (AIC; Akaike, 1987), was slightly better than that to the real NAEP data. The simulated data were generated using a three-parameter logistic item response theory (IRT) model and a BIB spiralling design like that used in NAEP. Although there was some evidence suggesting more than one dimension in the NAEP data the strength of the first dimension led the authors to conclude that the data "are sufficiently unidimensional to support the use of a composite scale for describing the NAEP mathematics data, but that there is evidence that two dimensions would better fit the data than one (p. 31)."

Methods

As mentioned above, the nature of the NAEP datasets limits the applicability of some computer programs that are available for assessing dimensionality. Carlson and Jirele (1992) provided a description of the data as follows:

> NAEP test booklets are comprised of blocks of items. These blocks are paired and administered using a balanced incomplete block (BIB) spiraling design (Beaton, Johnson, & Ferris, 1987; Zwick, 1987). Hence no examinee is administered a complete set of all items in a subject area (or in any subscale of a subject area). The design, which is efficient for purposes of estimating group mean proficiency, precludes performing dimensionality analyses of the entire set of items. The incomplete nature of the entire dataset, with blocks of data missing by

design, would result in separate dimensions being identified within each block by any of the techniques used in this study (pp. 2-3).

Also, as mentioned above most programs are unable to handle all of the information in polytomously-scored items, the focus of this study. Hence the previously-referred-to LISREL computer program, using a maximum likelihood parameter estimation technique was used in this study to perform factor analyses. Results are evaluated, as suggested in McDonald's works referenced above, through use of residuals from the fitted model. Specifically the square root of the mean squared residual (RMSR) was the statistic used.

The data analyzed were the item-response data from selected 1992 NAEP main assessment mathematics and reading tests. Item response data from three booklets in each subject at each grade level from the BIB spiral designs were studied. The mathematics data contained four blocks of items (total of 40, 38, and 37 items at grades 4, 8, and 12, respectively) and the reading data three blocks (total of 30, 35, and 32 items at grades 4, 8, and 12, respectively). Books were selected so as to maximize the overlap of items, hence minimizing the amount of missing data. Tables 1 and 2 show examples (for 12th grade) of the block structure and sample sizes for each block. It can be seen that with approximately equal samples for each booklet, about one-third of the data is missing on each item in reading and about one-half of that in mathematics, except for items in block L for which about one-quarter of the data is missing. The overall sample sizes for mathematics were 1125, 1173, and 1064 for grades 4, 8, and 12, respectively. For reading these numbers were 1169, 1271, and 1139.

One- two- and three-dimensional solutions were fitted to matrices of polychoric correlation coefficients using LISREL. In deriving solutions for the mathematics data, target solutions using information about the 5 scales in the mathematics framework (content domain) to define the factor structure were used. For the reading data target factor structures were based on blocks of items. Each reading block that was used involves a single reading passage and is designed to measure one of three scales. In addition, target solutions separating items into polytomously- and dichotomously-scored subsets were fitted, as well as solutions separating items into multiple choice and open-ended subsets. Each item was specified to load on one factor and correlated factors were specified in the target solutions. Lower-dimensional solutions were specified by collapsing the two dimensions with the highest estimated correlations. If, for example, the highest interfactor correlation for a five-dimensional solution was that between the fourth and fifth dimensions, these dimensions were combined into one factor in specifying the target solution in four dimensions. Although four and five factor solutions were fitted as part of this procedure, only the one, two, and three factor solutions are reported in this paper because the higher dimensional solutions did not fit better than the that for three dimensions.

In addition to the actual NAEP data, simulated datasets were analyzed in order to compare analyses of actual NAEP data with similar data of known dimensionality. The simulated datasets were generated using both unidimensional (reading only) and multidimensional structures. Correlated latent dimensions were specified using correlations among the proficiency estimates of the scales in the actual NAEP data. These correlation coefficients are reported in Table 5 for mathematics and Table 8 for reading. Unidimensional data emulating the mathematics assessment were not studied because the correlations among the five mathematics scales are so high (.90 to .95) that the analyses of the multidimensional data appeared essentially unidimensional (as would be expected with such high correlations). Item parameter estimates based on the actual NAEP data were used as parameters for the generation technique which used the generalized partial credit IRT model (Muraki, 1992). This choice ensured that the simulated data structure would be as similar as possible to the actual data that were analyzed (the generating model is the model assumed in scaling NAEP data).

It should be noted that in the 1992 NAEP instruments used in this study there is only one polytomously-scored item in each block of the BIB spiral. Hence there were only 4 polytomously-scored items in each students' mathematics responses, and three in the case of reading. Additionally, at the twelfth grade there was one block of items that had no polytomously-scored item so these students were only administered two such items. In order to revisit one of the questions studied by Zwick (1986, 1987) complete datasets were simulated as well as datasets using the BIB design.

In most cases the matrices of polychoric correlation coefficients were not positive definite. In fact only the complete-data simulations resulted in positive definite matrices. In the non-positive definite cases the LISREL program employs a "ridge" technique of incrementing the diagonals of the matrix in order that a factor structure could be fitted. This procedure artificially increases the amount of error variance (uniqueness) in the matrix in order to stabilize the system.

Results

Table 3 presents the Root Mean Squared Residual (RMSR) statistics for the actual and simulated mathematics data for the three grade levels. Simulated data are presented only for the 12th grade. Within each grade level results of fitting one, two, and three factors with LISREL are shown. Also shown are the results of specifying a target factor structure in two dimensions with dichotomously-scored items loading on one factor and polytomously-scored items on the other ("Di vs Poly"), and a similar structure with multiple choice items loading on one factor and open ended items on the other ("MC vs OE", grade 12 only). It should be noted that the NAEP instruments include some open-ended items that are scored dichotomously so these two structures are different. The results displayed in Table 3 are also plotted in Figures 1 and 2. In some cases no proper solution was possible because of the high correlations. The LISREL program, in these cases, was trying to fit a structure with correlations of 1.0 or greater between factors, which resulted in the estimated correlation matrix becoming non positive definite.

As may be seen from the values of the RMSR statistics reported in Table 3, there is no obvious difference in the fit with one, two, or three factors at the twelfth grade level. At the lower grade levels there is some decrease in the RMSR when more factors are fitted but the increase is so minimal that the writer would consider the data to be essentially unidimensional. Types of items, one of the primary focuses of this research, do not appear to result in multidimensionality in the context of the types of structures in the NAEP mathematics data. That is, there are only minor differences between one-dimensional solutions and a two-dimensional solutions where the second dimension is defined by the polytomously-scored, or open-ended items, and the first by the dichotomously-scored or multiple choice items.

Table 4 presents the correlations among the factors in the various solutions and Table 5 contains the actual correlations among the five NAEP mathematics scales. The latter were used in the generation of simulated data. The large sizes of these correlations limits much possibility for multidimensionality in the data. One interesting value to note is the relatively low correlation (.83) between factors defined to contrast the dichotomously- and polytomously-scored items at the twelfth grade level. This might suggest some difference in structure according to the item types.

Tables 6, 7, and 8, and Figures 3 and 4 show similar results for the NAEP reading assessment. In the case of reading the lower correlations in the actual data suggested studying more than one simulated factor structure. Because of the specific blocks assembled into the NAEP reading instruments, the actual data used in this study never included items measuring more than one of the three NAEP reading scales. Each block, however, as pointed out above, consists of a reading passage and several items (9 to 13) about that passage. Hence the multidimensional simulated data were generated as if each passage defined a separate dimension. The correlations among the actual reading scales that were used in generating these multidimensional data, as may be seen in comparing Tables 5 and 8, are lower than those among the mathematics scales.

In the actual data, fitting more than one factor has more affect on the size of the RMSR statistics (Table 6) and interfactor correlations (Table 7) than was the case in mathematics, at least at the 8th and 12th grade levels. Again, however, there seems to be little or no effect associated with item type: dichotomously- versus polytomously-scored, or multiple choice versus open ended. In the case of simulating a complete data matrix of three dimensions at the 12th grade level the RMSR statistic does seem to indicate some lack of fit when 1 or 2 dimensions are fitted rather than the three that underlie the generation process. The trend in the actual 12th grade data shows less of an effect than in the simulated data suggesting less than three dimensions in the NAEP instruments.

Discussion

The present research, although suggesting that the dimensionality of data structures in the NAEP assessment is generally not affected by the inclusion of polytomously-scored items, cannot be generalized to other situations. One reason is the size of the correlations among the scales of the NAEP data, especially in mathematics. Another reason is the small number of conditions simulated in this study. Thirdly, the number of polytomously-scored items was limited in the 1992 NAEP assessment. The author is currently pursuing a larger simulation study designed to answer broader questions about the dimensionality of instruments containing various mixes of dichotomously- and polytomously-scored items.

The one case of a statistic suggesting some difference between dichotomously- and polytomously-scored items ("Di vs Poly" correlation of .83 at grade 12), although suggestive, is too little basis on which to reach any conclusions about such a difference.

The relative sizes of the RMSR statistics for the simulated as compared to actual data suggest that lack of fit may be more due to the BIB spiraling design of NAEP than the number of dimensions fitted. Consistent with findings by Zwick (1986, 1987), however, the incomplete design for data collection used in NAEP does not appear to be artificially inflating the dimensionality of the instruments. Note, as might be expected, that the sizes of the RMSR statistics for the Incomplete Simulation condition (a BIB design as in the actual NAEP assessment) are more like those of the real data than those of the case of simulation of a complete data matrix.

REFERENCES

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317-332.

Beaton, A. E., Johnson, E. G., & Ferris, J. J. (1987). The assignment of exercises to students. In A. E. Beaton, *Implementing the new design: The NAEP 1983-84 technical report* (pp. 97-118). Princeton, NJ: Educational Testing Service.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443-459.

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement, 12*, 261-280.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika, 35*, 283-319.

Carlson, J. E., & Jirele, T. (1992, April). *Dimensionality of 1990 NAEP mathematics data*. Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika, 40*, 5-32.

Fraser, C. (1988). *NOHARM A computer program for fitting both unidimensional and multidimensional normal ogive models of latent trait theory*. Armidale, Australia: The University of New England, Center for Behavioral Studies.

Hattie, J. (1984). An empirical study of various indices for determining unidimensionality. *Multivariate Behavioral Research, 19*, 49-78.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9*, 139-164.

Holland, P. W., & Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *Annals of Statistics, 134*, 1523-1543.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7 User's reference guide*. Mooresville, IN: Scientific Software, Inc.

Knol, D. L., & Berger, M. P. F. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research, 26*, 457-477.

McDonald, R. P. (1962). A general approach to nonlinear factor analysis. *Psychometrika*, *27*, 397-415.

McDonald, R. P. (1967). Nonlinear factor analysis. *Psychometric Monographs*, No. 15.

McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology, 34*, 100-117.

McDonald, R. P. (1982). Linear versus nonlinear models in item response theory. *Applied Psychological Measurement, 6*, 379-396. (a)

McDonald, R. P. (1982). Unidimensional and multidimensional models for item response theory. In D. J. Weiss (Ed.), *Proceedings of the 1982 Computerized Adaptive Testing Conference* (pp. 127-148). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program. (b)

McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Lawrence Erlbaum.

Mislevy, R. J. (1986). Recent developments in t[illegible] tor analysis of categorical variables. *Journal of Educational Statistics, 11*, 3-31.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*, 159-176.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika, 43*, 551-560.

Muthén, B. (1991, December). *Issues in using NAEP mathematics items to study achievement dimensionality, within-grade differences, and across-grade growth.* Paper presented at the meeting of the Design and Analysis Committee of the National Assessment of Educational Progress, Watsonville, CA.

Rock, D. A. (1991, November). *Subscale dimensionality*. Paper presented at the meeting of the Design and Analysis Committee of the National Assessment of Educational Progress, Washington, DC.

Rosenbaum, P. R. (1984). Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika, 49*, 425-436.

Rosenbaum, P. R. (1985). Comparing distributions of item responses for two groups. *British Journal of Mathematical and Statistical Psychology, 38*, 206-215.

Stout, W. (1983). A statistical test of unidimensionality of a parameter underlying bernoulli trails [sic] with applications. *Statistics & Probability Letters, 1*, 183-188.

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52*, 589-617.

Stout, W. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika, 55*, 293-325.

Wilson, D. T., Wood, R., & Gibbons, R. (1991). *TESTFACT: Test scoring, Item Statistics, and Item Factor Analysis 386/486 Version*. Mooresville, IN: Scientific Software.

Zwick, R. (1986). *Assessment of the dimensionality of NAEP year 15 reading data*. (ETS Research Report No. 86-4). Princeton: Educational Testing Service.

Zwick, R. (1987). Assessing the dimensionality of NAEP reading data. *Journal of Educational Measurement, 24*, 293-308.

Table 1
Booklet-Block Structure: Grade 12 Mathematics

Booklet	Blocks Used in Study				Blocks Not Used
M6	I		L		H
M7	I	J		M	
M10			L	M	C
M21		J	L		D
Total Sample Size	523	528	807	522	

Table 2
Booklet-Block Structure Grade 12 Reading

Booklet	Blocks		
R30	C	D	
R39		D	E
R40	C		E
Total Sample Size	757	737	748

Table 3
Mathematics: Root Mean Square Residuals

Grade	No. Factors	Actual Data	Incomplete Simulation	Complete Simulation
4	1	.122		
	2	.122		
	3	.120		
	Di vs Poly [a]	.122		
8	1	.103		
	2	.102		
	3	.102		
	Di vs Poly	.103		
12	1	.101	.109	.054
	2	.101	.108	.052
	3	.101	.108	.051
	Di vs Poly	.101	.108	.054
	MC vs OE	NS[c]	.108	.054

[a] Dichotomously- vs Polytomously-scored Items: 2 Factor Solution
[b] Multiple-choice vs Open-ended Items: 2 Factor Solution
[c] No Proper Solution Found

Table 4
Mathematics: Interfactor Correlations

Grade	No. Factors	Actual Data	Incomplete Simulation	Complete Simulation
4	2	.81		
	3	.87, .85, .66		
	Di vs Poly	> 1.0		
8	2	.90		
	3	.91, .89, .88		
	Di vs Poly	.99		
12	2	.96	.94	.89
	3	.97, .96, .95	.96, .94, .90	.91, .89, .79
	Di vs Poly	.83	.89	.97
	MC vs OE	> 1.0	> 1.0	> 1.0

Table 5
Correlations Among Mathematics Scales

1.00				
.93	1.00			
.91	.94	1.00		
.95	.90	.90	1.00	
.93	.92	.94	.92	1.00

Table 6
Reading: Root Mean Square Residuals

Grade	No. Factors	Actual Data	Incomplete Simulation		Complete Simulation	
			1 Dim.	3 Dim.	1 Dim.	3 Dim.
4	1	.077				
	2	.076				
	3	.076				
	Di vs Poly [a]	.077				
	MC vs OE [b]	.077				
8	1	.113				
	2	.110				
	3	.097				
	Di vs Poly	.112				
	MC vs OE	.113				
12	1	.083	.071	.074	.039	.055
	2	.081	.071	.066	.039	.048
	3	.078	NS	.065	.039	.044
	Di vs Poly	NS [c]	.071	NS	.039	.055
	MC vs OE	.082	.071	NS	.039	NS

[a] Dichotomously- vs Polytomously-scored Items: 2 Factor Solution
[b] Multiple-choice vs Open-ended Items: 2 Factor Solution
[c] No Proper Solution Found

Table 7
Reading: Interfactor Correlations

Grade	No. Factors	Actual Data	Incomplete Simulation 1 Dim.	Incomplete Simulation 3 Dim.	Complete Simulation 1 Dim.	Complete Simulation 3 Dim.
4	2	.92				
	3	.97, .95, .86				
	Di vs Poly	.93				
	MC vs OE	.99				
8	2	.83				
	3	.97, .90, .87				
	Di vs Poly	> 1.0				
	MC vs OE	> 1.0				
12	2	.85	.98	.78	.997	.83
	3	.83, .79, .78	NS	.95, .81, .71	all > 1.0	.83, .83, .76
	Di vs Poly	NS	.92	NS	.98	> 1.0
	MC vs OE	.87	> 1.0	NS	.99	NS

Table 8
Correlations Among Reading Scales

1.0		
.85	1.0	
.76	.84	1.0

14

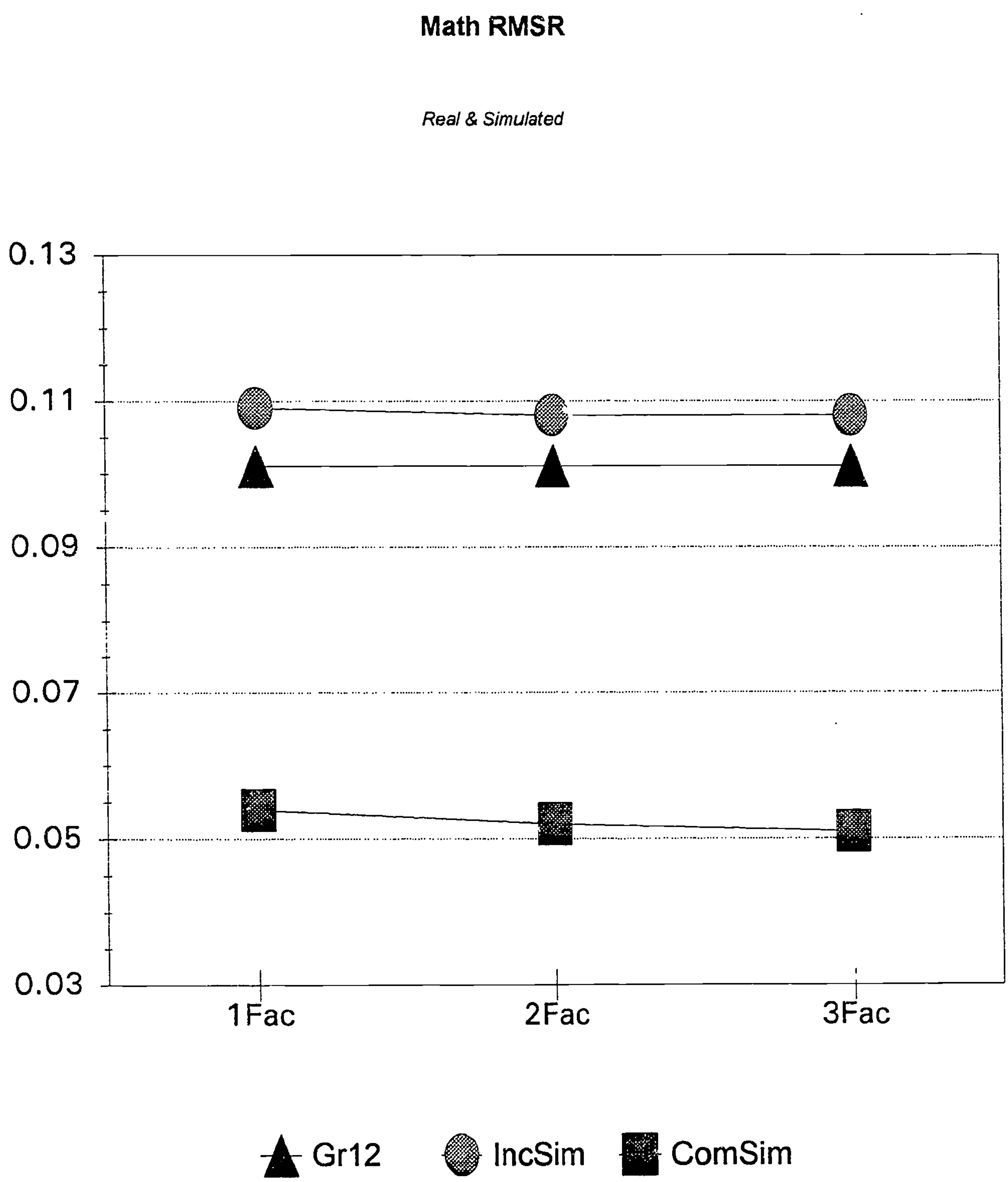
Math RMSR
Real & Simulated
0.13
0.11
0.09
0.07
0.05
0.03
1Fac
2Fac
3Fac
Gr12
IncSim
ComSim

Figure 1

Math RMSR

by grade

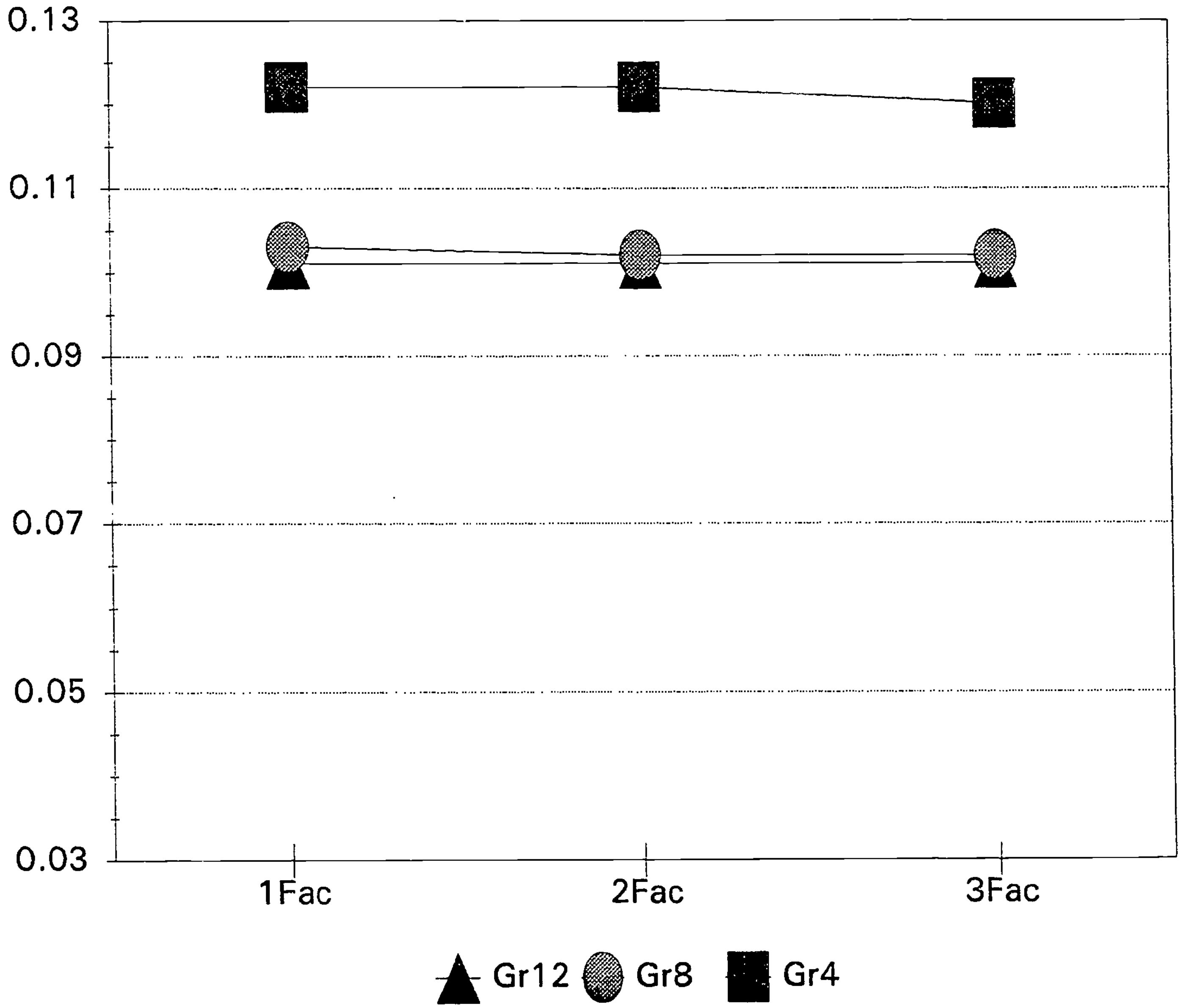

Figure 2

Reading RMSR

Real & Simulated

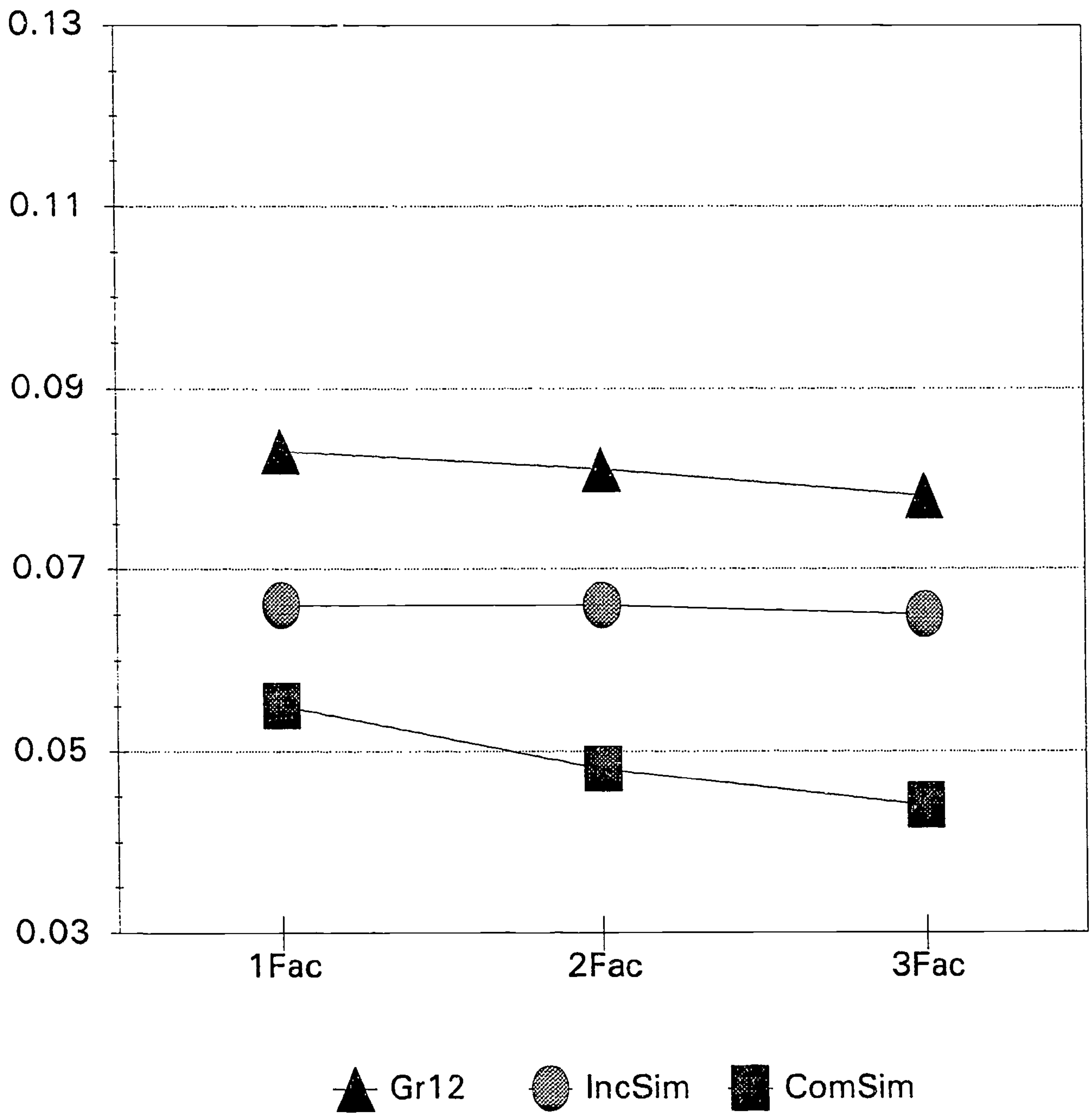

Figure 3

Readg RMSR

by grade

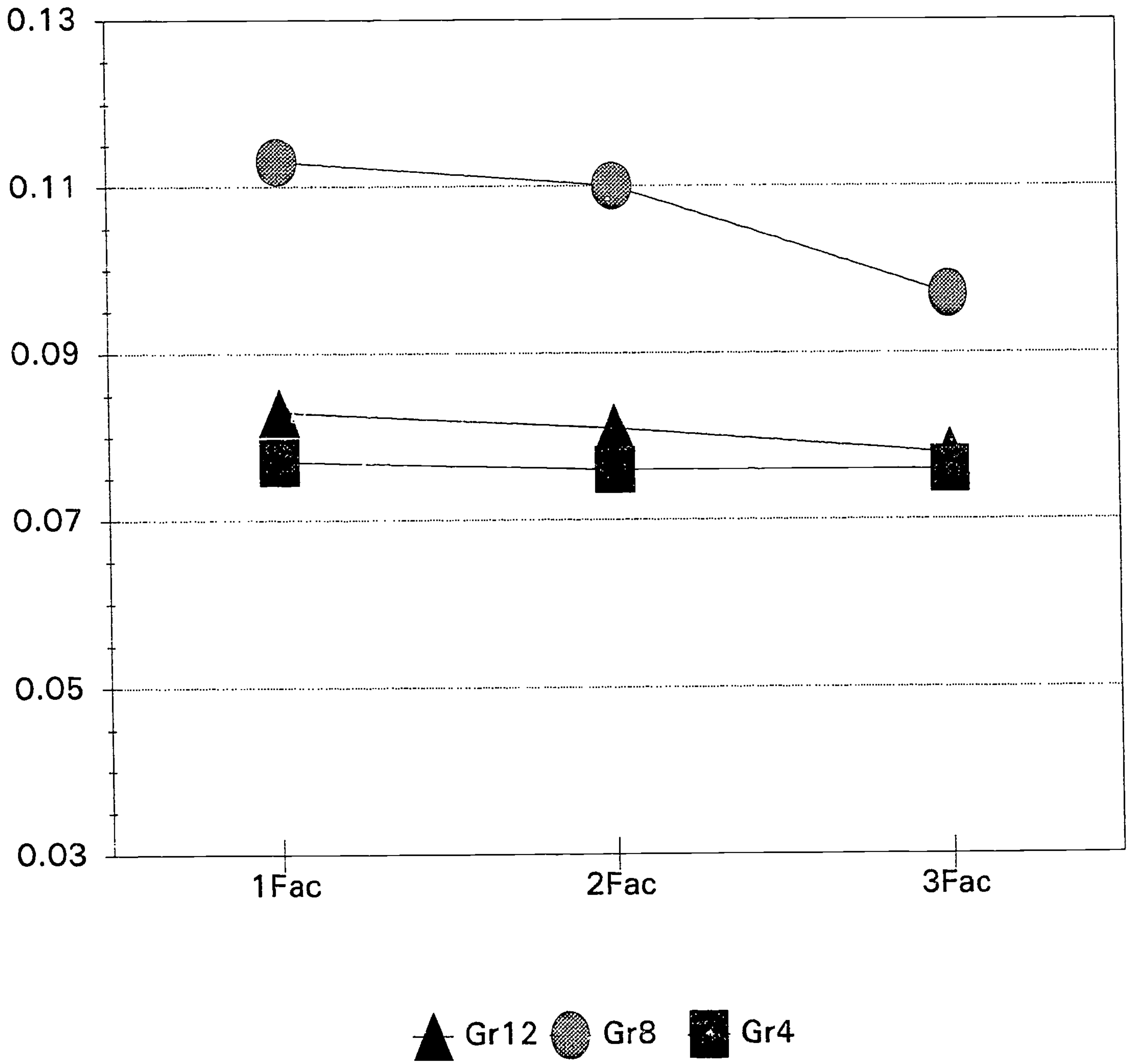

Figure 4