DOCUMENT RESUME

ED 372 115 TM 021 951

AUTHOR	Lewis, J. C.
TITLE	The Effect of Context and Gender on Assessment of Estimation.
PUB DATE	Apr 94
NOTE	28p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 5-7, 1994).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Comparative Analysis; *Context Effect; Educational Assessment; Elementary Education; Elementary School Students; *Estimation (Mathematics); Females; Knowledge Level; Males; Numbers; Sex Bias; *Sex Differences; *Test Items
IDENTIFIERS	Cognitive Abilities Test; Iowa Tests of Basic Skills; *Pictures

ABSTRACT

Whether boys and girls perform differently on mathematics estimation items with a picture format (applied context [AC] items) compared with items with a numbers-only (NC) format was studied when effects of computational skill, conceptual knowledge, and quantitative ability were controlled. Subjects were approximately 80,000 students from grades 4 through 8 who participated in the 1992 joint national standardization of the Iowa Tests of Basic Skills, Form K, and the Cognitive Abilities Test, Form 5. Because of the way items were selected for the estimation subtest, it was not meaningful to compare performance on AC items versus NC items alone. However, the interaction of gender with item type as mediated by computational skill, conceptual knowledge, and quantitative ability was examined. In general, males performed slightly better than females on these items, but there did not seem to be a consistent pattern of differences favoring one item type over the other for either gender group. In addition, differences were so small that there seemed to be little need for concern about gender bias attributable to applied context versus the numbers-only context. Eight tables and seven figures present the analyses. (Contains 13 references.) (SLD)

The Effect of Context and Gender on Assessment of Estimation

J. C. Lewis

The University of Iowa

Paper to be presented at 1994 NCME Annual Conference in New Orleans.

The Effect of Context and Gender on Assessment of Estimation

The purpose of this study was to investigate whether boys and girls performed differently on mathematics estimation items with a picture format compared to items with a numbers-only format when the effects of computational skill, conceptual knowledge, and quantitative ability were controlled.

Educational Significance

The ability to estimate is closely related to the understanding of numbers and operations and is sometimes included under the construct "number sense" (Greeno, 1991). Both number sense and computational estimation are areas of instruction emphasized by the National Council of Teachers of Mathematics (1989) in their curriculum and evaluation standards for mathematics.

Estimation is a mathematical skill commonly used in everyday life. The widespread use of computers to perform computation has opened access to large quantities of information. For example, computers make possible aggregation and comparison of national educational data that was impractical thirty years ago. Estimation allows quick assessment of how reasonable the assertions about such data are. Students frequently use calculators to perform computation and need to decide quickly whether the answers they get are plausible. Consequently, students need to know when using an estimate is appropriate and desirable, several ways to make an estimate, and how to choose the best estimate.

Part of the process of implementing the new mathematics standards is assessing how well the students have learned the desired concepts and skills. Research and practice on how to best assess estimation skills, especially in a standardized format, are still in the early stages. Silver (1990) pointed out that children tend to seek exact answers and suggested that making the estimation items less like a familiar arithmetic task would increase the likelihood that students actually used estimation techniques. Silver also suggested that children were more able to apply their understanding of numbers when that knowledge was embedded in real-life mathematical activity. Greeno's (1991) discussion of number sense as "situated cognition" proposed that children do not necessarily have to understand the symbolic representations of ideas in order to approach difficult problems. Instead, Greeno hypothesized that children create mental models that represent the objects, properties and relationships in the situations they encounter. Thus, presenting estimation items with pictures or a short description might encourage children to make connections with prior experience and allow them to attempt more complex problems than they ordinarily could when the problem is presented in a symbolic form. Reys (1986) suggested that presenting estimation items in an applied context rather than a numbers-only context could improve children's performance. An applied context might include a picture, table, or story that approximates as much as possible a real-life situation, while a numbers-only context would look like a standard computation question that requests an estimate rather than an exact answer. Schoen (1990) showed that adding an applied context to estimation items made the items significantly easier for students at each level for grades five through eight. Contexts which were most familiar to the students such as money showed the most improvement due to context.

Since other types of mathematics items have shown differences in performance between boys and girls (Fennema & Carpenter, 1981), boys and girls may perform differently on estimation items as well. In particular, it is unclear whether the presence of visual materials that might require spatial skills would affect mathematics performance for girls more than for boys (Connor & Serbin, 1985; Tartre, 1990). Martin & Hoover (1987) showed that the pattern of gender differences changed at different ability levels and across grades. They found differences favoring males on some math subtests above the 70th percentile and in the upper grades, whereas at lower levels and in the lower grades, differences often favored females. Because estimation is one of the areas of mathematics which is proposed for increased attention by the 1989 NCTM curriculum standards, it is important to understand whether methods of assessing students' performance in estimation work equally well for boys and girls at each ability level.

Methods

Subjects. The subjects for this study were the approximately 80,000 students from grades 4 to 8 who participated in the 1992 joint national standardization of the Iowa Tests of Basic Skills, Form K (Hoover et al., 1993a) and the Cognitive Abilities Test, Form 5 (Thorndike & Hagen, 1993). The standardization sample (Hoover et al., 1993b) included students from public, Catholic, and non-Catholic private schools in all regions of the country. The schools were selected to provide a range of socio-economic status from low to high and enrollments from very small to very large. Students enrolled full-time in special education classrooms were not included in the sample.

Instruments Used. The Iowa Tests of Basic Skills, Form K (ITBS) include a section of estimation items in the Mathematics Concepts and Estimation subtest. The estimation test is different for each grade, although about half the items for adjacent grades are the same. The number of items ranges from 16 at grade four to 24 at grade eight, and the items are divided equally between numbers-only items (NC) and applied context items (AC) which include a picture, table or short description. See Figure 1 for examples of these two item types. The items are all multiple-choice using item formats based on the research by Schoen et al. (1990). The items are intended to assess the estimation strategies of standard rounding, front-end rounding, and other special methods such as compatible numbers and compensation.

The score from the concepts section of the ITBS Mathematics Concepts and Estimation subtest (CN), the score from the ITBS Computation subtest (CP), and the quantitative score (Q) from the Cognitive Abilities Test Form 5 (CoGat) are included as measures of mathematics skills and abilities that could affect estimation performance. The Q score is derived from a quantitative concepts subtest, an equation building subtest and a number series subtest.

The SAS System for OS/2 Version 6.08 was used for all analyses.

Design. The purpose of the study was to investigate gender and grade differences in performance between AC and NC estimation items with the effects of computational skill, conceptual knowledge and quantitative ability controlled. Because the estimation items used in this study were selected for an achievement test rather than a research project, the items with applied context (AC) were not simply numbers-only (NC) items with pictures, tables or stories added. They were completely different items. Therefore, it was not meaningful with these data to compare performance on AC items versus NC items. After the preliminary statistics were computed, the AC and NC scores were standardized to remove the main effect due to item type. Two different methods of controlling effects of CN, CP, and Q were used. First the effects of CN, CP, and Q were controlled by using them as covariates in an analysis of covariance. Secondly, CN, CP, and Q were categorized and used as the block variables in an analysis of variance.

Analysis of covariance assumes that the covariates do not interact with the categorical groups. However, the three mathematics covariates could have significant interactions with gender. Thus AC and NC first were predicted with the regression model, AC NC = Q | CN | CP, by gender for each grade and residual scores computed for AC and NC. These residual scores as well as the raw scores for AC and NC were standardized separately across gender within grade to have means of 50 and standard deviations of 10. The standardized residual scores were entered in analyses of covariance with gender (2 between) by item type (2 within) for each grade. This allowed analysis of covariance with the effects of the mathematics covariates completely removed. Analyses of covariance using the standardized raw scores were also computed to compare the results with the influence of Q, CN, and CP ignored.

In the second set of analyses Q, CN, and CP were trichotomized in each grade into low, medium and high groups using cut scores to divide the number in each group as close to 33% and 66% as possible. For each grade 3B x 3B x 3B x 2B x 2W mixed design analyses of variance were computed on Q, CN, CP, gender (G), and item type (T). Finally, selected means derived from these analyses were examined.

Results

The number of students per grade varied from approximately 18,000 in grade 5 to 12,000 in grade 8. There were approximately equal numbers of boys and girls in each grade. The average percent correct AC and NC scores for boys and girls in grades 4 to 8 are shown in Table 1 along with the means and standard deviations for Q, CN, and CP. Comparing boys to girls, there was very little difference except that girls had higher mean computation scores in all grades. The intercorrelations are shown in Table 2. The correlations ranged from 0.52 to 0.78 with the concepts and quantitative scores being the most highly correlated for both boys and girls at all grades. Because the quantitative test includes sections on concepts, equation building, and number series, the ITBS concepts test and the CoGat quantitative test do cover overlapping material and could be expected to be highly correlated.

Missing data in Q, CN, and CP caused about a ten percent loss of data. Discriminant analyses were performed at each grade to see if the missing data group could be identified relative to the non-missing data group. In none of the five discriminant analyses (one at each grade 4, 5, 6, 7, and 8) were any of the students in the missing data group correctly classified into this group. Thus, it was concluded that the students lost to the analyses due to lack of complete data were not unique compared to those remaining.

The results of the regression analyses testing significant interaction of gender with Q, CN, and CP for the AC and NC scores are shown in Table 3. Although the interaction of gender with the Q, CN, and CP scores was statistically significant in most cases, the increase in R^2 was at most 0.0016, which was less than 0.2% of the total variance to be explained.

Achieving a gain of less the 0.2% in R^2 at the cost of 15 degrees of freedom did not seem worthwhile. Thus the variance due to the interaction of gender and math ability was included in the e· or variance in the analyses of covariance shown in Table 4. In all grades there was a significant gender effect both using the standardized residual scores and using the standardized raw scores. Note that the effect of item type was purposely removed, but a significant gender by type interaction occurred in grades 4 and 8 for both analyses. Also, there was a significant interaction of gender and type at grade 7 in the analysis using residual scores but not in the analysis with raw scores. Similarly, the interaction of gender and type was significant for grade 5 using raw scores but not with residual scores.

The comparisons of the means and standard deviations of the total estimation score for boys and girls with the residual scores and the raw scores are shown in Table 5. Note that this table is based on the data in which the AC and NC scores were each standardized to a mean of 50 and a standard deviation of 10. Removing the effects of the computation, concepts, and quantitative scores increased the differences between the scores for boys and girls slightly in every grade. The differences between the mean scores for boys and girls were significant at the 0.05 level for both the residual and the raw scores in all grades. Even so, the difference between mean scores for boys and girls was at most 15% of a standard deviation, so its practical significance seems questionable.

The means and standard deviations of the AC and NC scores by gender and grade are shown in Table 6. Again, there was very little difference between the raw scores and the residual scores. Two patterns emerged. In grades 4 and 6 girls had a higher AC score than NC score while boys had a lower AC score than NC score. Grades 5, 7, and 8 showed the opposite pattern. As examples, Figures 2 and 3 show the two opposite patterns in grades 4 and 8. In all cases boys had higher mean scores than girls did. However, when the whole range of scores was considered, these differences were very small (Figures 4 and 5).

Looking at the data from the second perspective, Table 7 shows the within-subjects effects of the analyses of variance using the categorized math variables. Theses analyses are based on the standardized raw score data. The hierarchical (Type I) sums of squares are shown. The gender by type interaction was significant for all grades except grade 6. Grades 4, 6, and 7 had a significant three-way interaction among gender, item type and quantitative score. Grades 4 and 5 also had a significant four-way interaction among item type and the three math variables. The gender by type interaction (Table 6) has already been discussed. Table 8 shows comparisons between selected groups to illustrate part of the interaction between item type and math variables. Given the relatively high correlations between the Q, CN, and CP scores, it was not surprising that the three largest groups of children in each grade fell into the groups which were either low, medium or high in all three areas. The AC and NC scores were quite similar within each of these groups, and there seemed to be no consistent pattern favoring AC or NC in any of these groups. However, those students who were weak conceptually and quantitatively but had excellent computation skills seemed to do better on the NC items which were more similar to the kind of item in a computation test. Figures 6 and 7 show the patterns in grades 4 and 5, which were the two grades in which the interaction among item type and the three math variables was significant.

Summary

Because of the way the items were selected for the estimation subtest, it was not meaningful with these data to compare performance on AC items versus NC items alone. However, the interaction of gender with item type as mediated by computational skill, conceptual knowledge and quantitative ability was examined thoroughly. In general, boys performed slightly better than girls on these items. There did not seem to be a consistent pattern of differences favoring one type of item over the other for either gender group. In addition, the differences were so small that, from a practical standpoint, there seemed to little need for concern about gender bias due to an applied context versus a numbers-only context. Consequently context can be included in estimation items to make them both more interesting and easier without worrying about gender effects.

References

Commission on Standards for School Mathematics of the National Council of Teachers of Mathematics (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: National Council of Teachers of Mathematics.

Connor, J. M. and Serbin, L. A. (1985). Visual-spatial skill: Is it important for mathematics? Can it be taught? In S. F. Chipman, L. R. Brush, and D. M. Wilson (Eds.), Women and mathematics: Balancing the equation (pp. 151-174). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Fennema, E. and Carpenter, T. P. (1981) Sex-related differences in mathematics: Results from the national assessment. Mathematics teacher, 74 (7), 554-559.

Greeno, J. G. (1991). Number sense as situated knowing in a conceptual domain. Journal for research in mathematics education, 22 (3), 170-218.

Hoover, H. D., Hieronymus, A. N., Frisbie, D. A., and Dunbar, S. B. (1993). The Iowa tests of basic skills form K. Chicago: Riverside Publishing Co.

Hoover, H. D., Hieronymus, A. N., Frisbie, D. A., and Dunbar, S. B. (1993). The Iowa tests of basic skills form K: Norms and score conversions. Chicago: Riverside Publishing Co.

Martin, D. J. & Hoover, H. D. (1987). Sex differences in educational achievement: A longitudinal study. Journal of early adolescence, 7 (1), 65-83.

Reys, R. E. (1986). Evaluating computational estimation. In H. L. Schoen & M. J. Zweng (Eds.), Estimation and mental computation (pp. 225-238). Reston, VA: National Council of Teachers of Mathematics.

Shoen, H. L. (1990). Assessing computational estimation: Research and new directions. Paper presented at the United States/Japan Joint Seminar, Computation for the 21st century: Cross cultural perspectives, Honolulu, HI.

Schoen, H. L., Blume, G., and Hoover, H. D. (1990). Outcomes and processes on estimation test items in different formats. Journal for research in mathematics education, 21 (1), 61-73.

Silver, E. A. (1990). Treating estimation and mental computation as situated mathematical processes. Paper presented at the United States/Japan Joint Seminar, Computation for the 21st century: Cross cultural perspectives, Honolulu, HI.

Tartre, L. A. (1990). Spatial skills, gender, and mathematics. In E. Fennema and G. C. Leder (Eds.), Mathematics and gender (pp. 27-59). New York: Teachers College Press.

Thorndike, R. L. & Hagen, E. P. (1993). Cognitive abilities test, multi-level edition, form 5. Chicago: Riverside Publishing Co.

Numbers-Only Versus Applied Context

1. Applied Context

6  17 

18  26 

The closest estimate of the number of cars in the train is ________.

A) 40

B) 50

* C) 70

D) 80

2. Numbers-Only Context

The closest estimate of 6 + 18 + 17 + 26 is ______

A) 40

B) 50

* C) 70

D) 80

Fig. 1

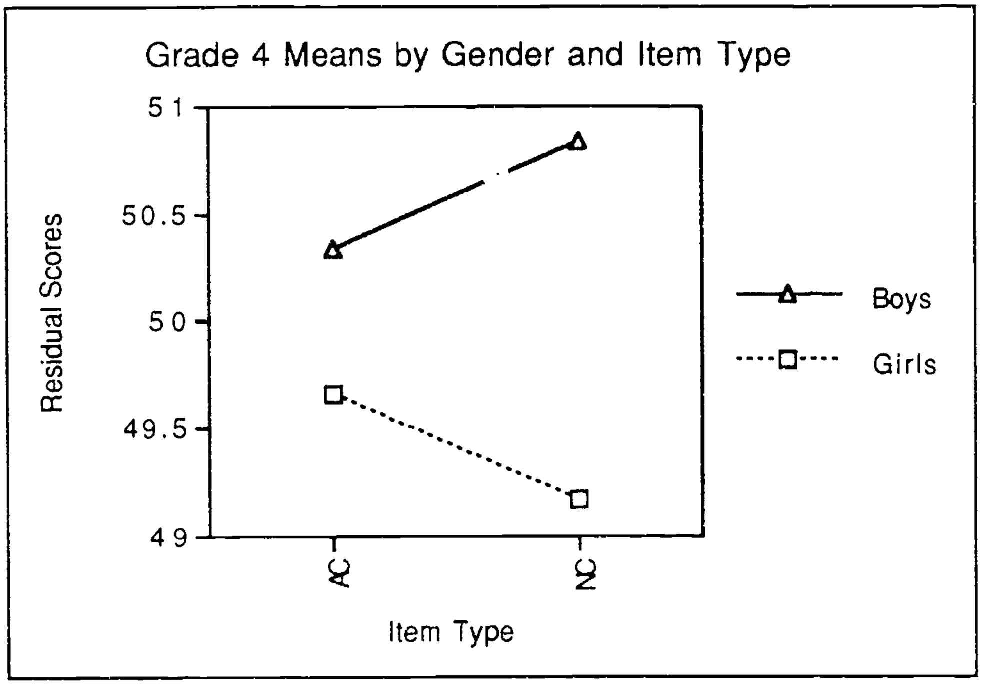
Grade 4 Means by Gender and Item Type
Residual Scores
51
50.5
50
49.5
49
AC
NC
Item Type
Boys
Girls

Fig. 2

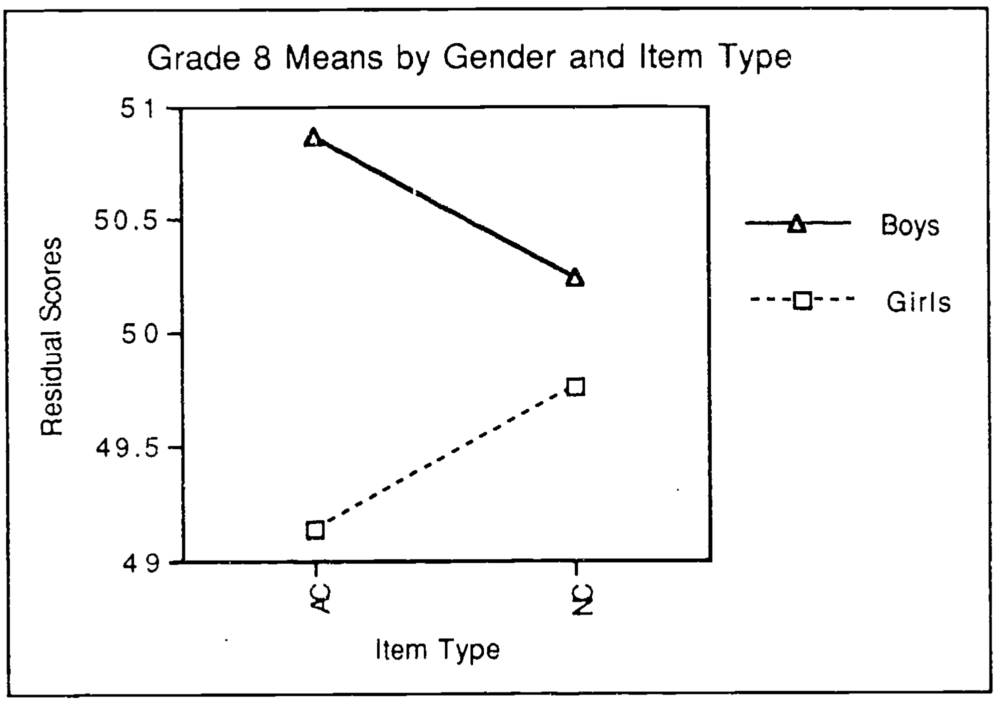
Grade 8 Means by Gender and Item Type
Residual Scores
51
50.5
50
49.5
49
AC
NC
Item Type
Boys
Girls

Fig. 3

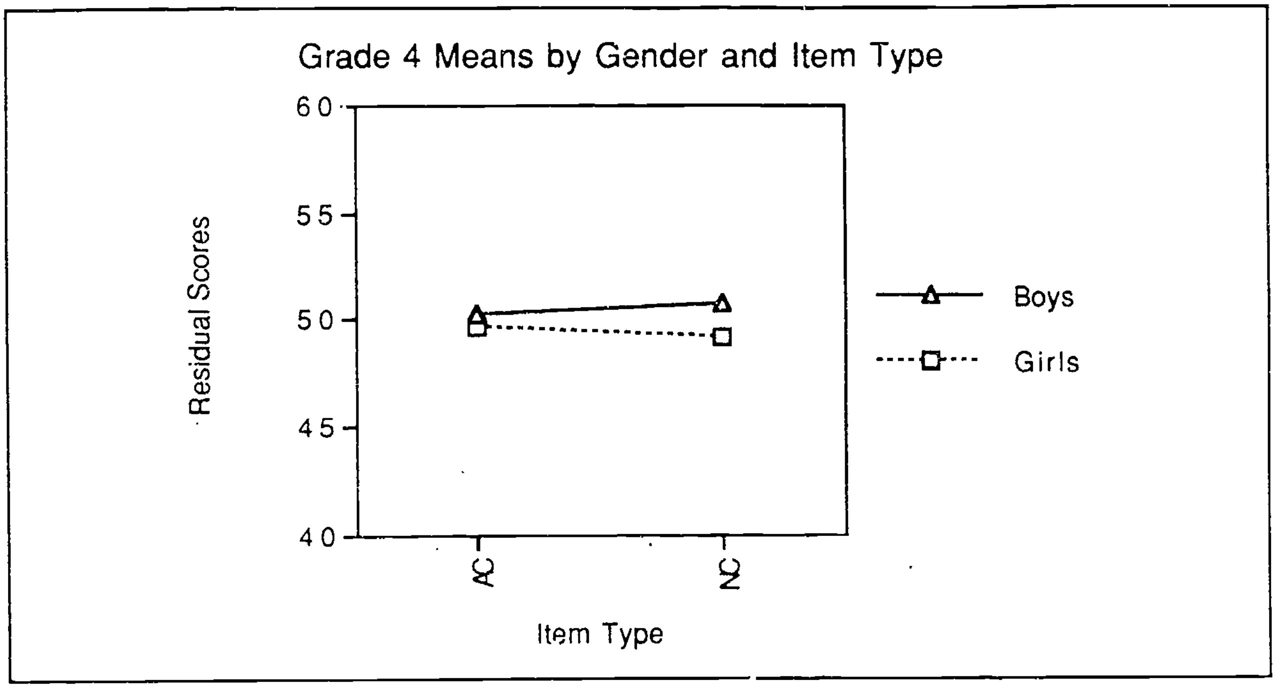
Grade 4 Means by Gender and Item Type
60
55
50
45
40
Residual Scores
AC
NC
Item Type
Boys
Girls

Fig. 4

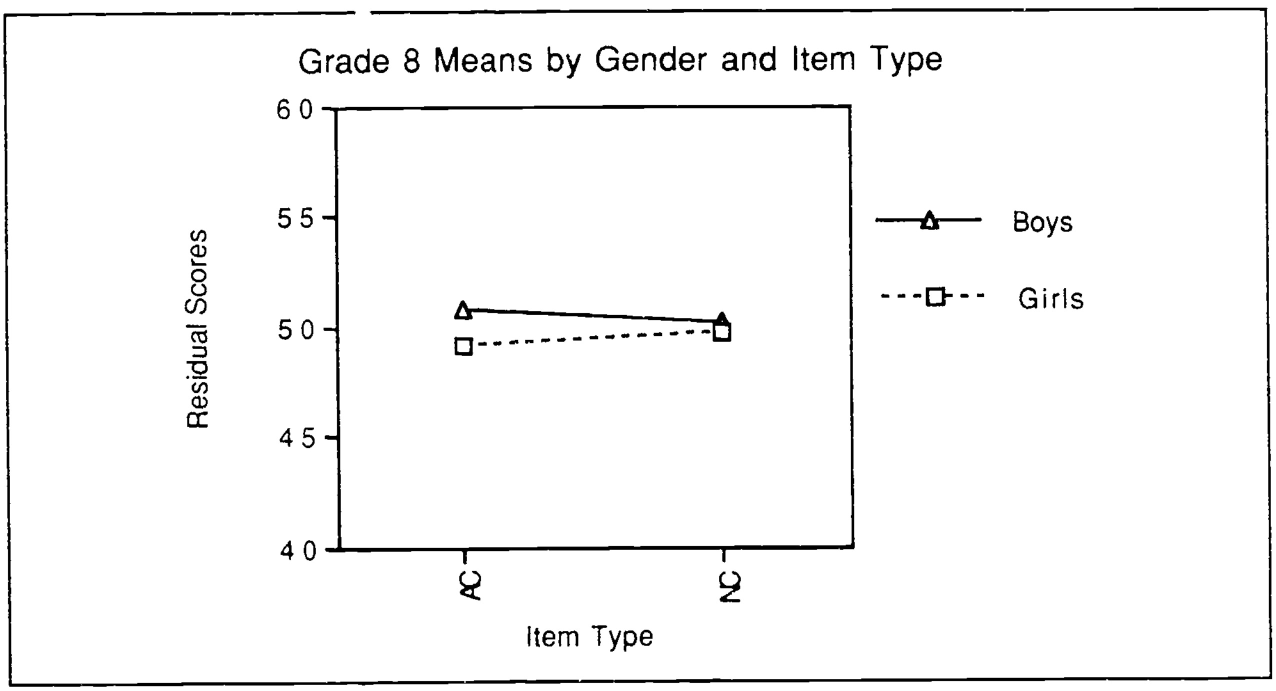
Grade 8 Means by Gender and Item Type
60
55
50
45
40
Residual Scores
AC
NC
Item Type
Boys
Girls

Fig. 5

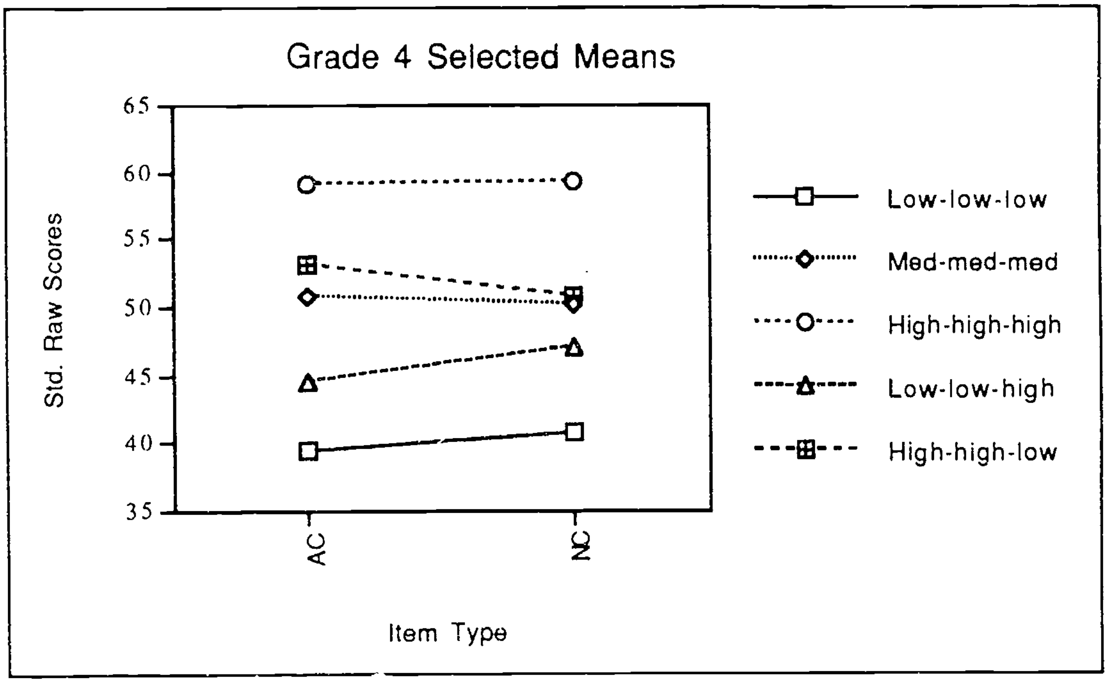
Grade 4 Selected Means
65
60
55
50
45
40
35
Std. Raw Scores
AC
NC
Item Type
Low-low-low
Med-med-med
High-high-high
Low-low-high
High-high-low

Fig. 6

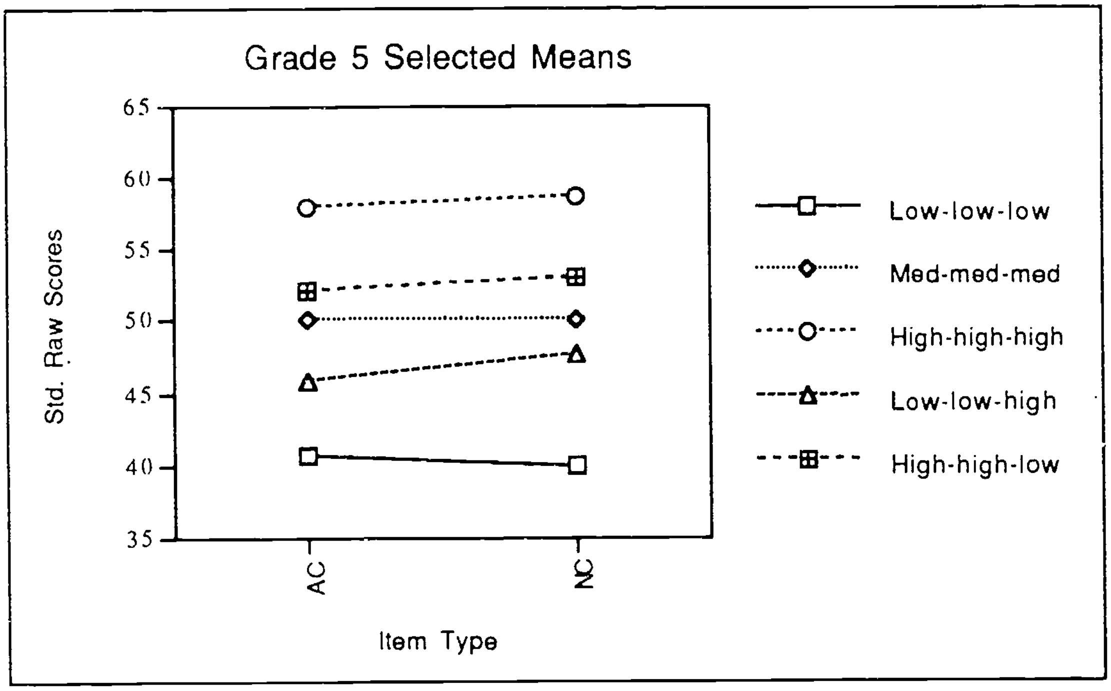
Grade 5 Selected Means
65
60
55
50
45
40
35
Std. Raw Scores
AC
NC
Item Type
Low-low-low
Med-med-med
High-high-high
Low-low-high
High-high-low

Fig. 7

Table 1

Estimation, Concepts, Computation and Quantitative Ability Scores by Grade and Gender

Number, Means, and Standard Deviations

	GIRLS		BOYS	
GRADE 4:				
Number	8217	50.2%	8145	49.8%
	Mean	**S D**	**Mean**	**S D**
A C (Percentage Correct)	0.56	0.24	0.58	0.25
N C (Percentage Correct)	0.51	0.21	0.54	0.22
CN	16.9	4.34	17.2	4.64
CP	25.6	7.25	24.5	7.85
Q	101.8	14.78	101.5	16.60
GRADE 5:				
Number	8715	50.6%	8511	49.4%
	Mean	**S D**	**Mean**	**S D**
A C (Percentage Correct)	0.56	0.20	0.57	0.21
N C (Percentage Correct)	0.56	0.22	0.56	0.24
CN	18.3	4.60	18.5	4.90
CP	27.2	7.70	25.3	8.49
Q	101.5	14.71	101.6	16.83
GRADE 6:				
Number	7830	50.4%	7711	49.6%
	Mean	**S D**	**Mean**	**S D**
A C (Percentage Correct)	0.55	0.21	0.57	0.22
N C (Percentage Correct)	0.52	0.20	0.54	0.21
CN	18.6	5.28	18.6	5.69
CP	27.4	7.72	25.1	8.66
Q	101.6	13.93	101.5	16.12
GRADE 7:				
Number	6397	50.5%	6278	49.5%
	Mean	**S D**	**Mean**	**S D**
A C (Percentage Correct)	0.61	0.22	0.62	0.23
N C (Percentage Correct)	0.50	0.20	0.52	0.23
CN	20.5	5.58	20.1	6.34
CP	26.1	8.18	23.3	8.70
Q	102.7	14.47	102.0	16.49
GRADE 8:				
Number	5535	50.4%	5442	49.6%
	Mean	**S D**	**Mean**	**S D**
A C (Percentage Correct)	0.53	0.20	0.56	0.21
N C (Percentage Correct)	0.51	0.19	0.51	0.21
CN	17.4	5.79	18.2	6.24
CP	24.8	8.73	22.5	9.04
Q	101.1	14.27	101.2	16.39

Table 2

Correlation Coefficients By Grade and Gender

(Girls in Lower Half, Boys in Upper Half of Table)

Grade 4 Girls \ Boys

	AC	NC	CN	CP	Q
AC	1.00	0.59	0.66	0.60	0.64
NC	0.53	1.00	0.60	0.58	0.60
CN	0.62	0.54	1.00	0.65	0.74
CP	0.54	0.52	0.59	1.00	0.69
Q	0.61	0.54	0.71	0.64	1.00

Grade 5 Girls \ Boys

	AC	NC	CN	CP	Q
AC	1.00	0.58	0.58	0.54	0.58
NC	0.54	1.00	0.61	0.61	0.63
CN	0.54	0.57	1.00	0.65	0.75
CP	0.49	0.56	0.63	1.00	0.71
Q	0.54	0.57	0.74	0.66	1.00

Grade 6 Girls \ Boys

	AC	NC	CN	CP	Q
AC	1.00	0.61	0.63	0.59	0.63
NC	0.53	1.00	0.62	0.59	0.62
CN	0.59	0.54	1.00	0.70	0.78
CP	0.55	0.52	0.68	1.00	0.71
Q	0.59	0.55	0.76	0.68	1.00

Grade 7 Girls \ Boys

	AC	NC	CN	CP	Q
AC	1.00	0.69	0.67	0.58	0.65
NC	0.61	1.00	0.68	0.63	0.68
CN	0.61	0.61	1.00	0.69	0.78
CP	0.50	0.57	0.68	1.00	0.72
Q	0.61	0.61	0.77	0.69	1.00

Grade 8 Girls \ Boys

	AC	NC	CN	CP	Q
AC	1.00	0.65	0.65	0.56	0.63
NC	0.56	1.00	0.66	0.60	0.64
CN	0.57	0.58	1.00	0.68	0.78
CP	0.49	0.53	0.65	1.00	0.70
Q	0.57	0.59	0.77	0.67	1.00

Table 3

Summary Tables for Forward Selection Regression By Grade

Dep. Var.	Variable Group	No. of Vars.	Model R^2	Partial R^2	F	Pr > F
Grade 4:						
AC	MAIN EFFECTS	8	0.4889		917.24	0.0001
	INTERACTIONS	15	0.4894	0.0005	2 34	0.0218
NC	MAIN EFFECTS	8	0.4172		639.15	0.0001
	INTERACTIONS	15	0.4176	0.0004	1.46	0.1776
Grade 5:						
AC	MAIN EFFECTS	8	0.3774		1304.00	0.0001
	INTERACTIONS	15	Did not enter the equation.			n.s.
NC	MAIN EFFECTS	8	0.4483		767.47	0.0001
	INTERACTIONS	15	0.4488	0.0004	2.03	0.0471
Grade 6:						
AC	MAIN EFFECTS	8	0.4460		706.10	0.0001
	INTERACTIONS	15	0.4463	0.0003	1.30	0.2437
NC	MAIN EFFECTS	8	0.4068		546.89	0.0001
	INTERACTIONS	15	0.4077	0.0009	3.24	0.0019
Grade 7:						
AC	MAIN EFFECTS	8	0.4618		587.91	0.0001
	INTERACTIONS	15	0.4633	0.0015	4.93	0.0001
NC	MAIN EFFECTS	8	0.4978		628.48	0.0001
	INTERACTIONS	15	0.4994	0.0016	5.95	0.0001
Grade 8:						
AC	MAIN EFFECTS	8	0.4281		429.55	0.0001
	INTERACTIONS	15	0.4292	0.0011	2.96	0.0042
NC	MAIN EFFECTS	8	0.4558		471.59	0.0001
	INTERACTIONS	15	0.4574	0.0016	4.67	0.0001

Table 4
Analysis of Covariance by Grade

Grade 4		Standardized Residual Scores			Standardized Raw Scores		
Source	df	ms	F	Pr > F	ms	F	Pr > F
Gender	1	11334	94.7	0.0001	5332	34.3	0.0001
Between Error	16360	120			155		
Item Type	1	0	0	1.0000	0	0	1.0000
Gender * Type	1	1969	24.8	0.0001	1254	28.4	0.0001
Within Error	16360	79			44		

Grade 5							
Source	df	ms	F	Pr > F	ms	F	Pr > F
Gender	1	7230	57.6	0.0001	1065	6.8	0.0001
Between Error	17224	125			156		
Item Type	1	0	0.0	1.0000	0	0.0	1.0000
Gender * Type	1	89	1.2	0.2740	338	7.7	0.0055
Within Error	17224	74			44		

Grade 6							
Source	df	ms	F	Pr > F	ms	F	Pr > F
Gender	1	16503	132.0	0.0001	3038	19.3	0.0001
Between Error	15539	125			157		
Item Type	1	0	0	1.0000	0	0	1.0000
Gender * Type	1	129	1.74	0.1871	63	1.5	0.2240
Within Error	15539	74			43		

Grade 7							
Source	df	ms	F	Pr > F	ms	F	Pr > F
Gender	1	19171	143.8	0.0001	2106	12.8	0.0004
Between Error	12673	133			165		
Item Type	1	0	0	1.0000	0	0	1.0000
Gender * Type	1	609	9.3	0.0022	83	2.4	0.1211
Within Error	12673	65			35		

Grade 8							
Source	df	ms	F	Pr > F	ms	F	Pr > F
Gender	1	6798	52.3	0.0001	4785	29.8	0.0001
Between Error	10975	130			160		
Item Type	1	0	0	1.0000	0	0	1.0000
Gender * Type	1	2108	30.6	0.0001	1990	51.0	0.0001
Within Error	10975	69			39		

Table 5
Comparison of Means and Standard Deviations of Standardized Residual and Raw Scores By Gender and Grade

		Residual Score		Raw Score	
	N	Mean	SD	Mean	SD
Grade 4					
Girls	16,434	49.4	10.07	49.6	9.83
Boys	16,290	50.6	9.90	50.4	10.15
Grade 5					
Girls	17,430	49.5	9.98	49.8	9.73
Boys	17,022	50.5	10.00	50.2	10.26
Grade 6					
Girls	15,660	49.3	9.97	49.7	9.68
Boys	15,422	50.7	9.97	50.3	10.30
Grade 7					
Girls	12,794	49.1	9.89	49.7	9.51
Boys	12,556	50.9	10.04	50.3	10.47
Grade 8					
Girls	11,070	49.4	9.99	49.5	9.59
Boys	10,884	50.6	9.98	50.5	10.38

Table 6
Comparison of Means and Standard Deviations
of Standardized Residual and Raw Scores
By Gender, Item Type and Grade

				Residual Scores		Raw Scores	
Grade	Gender	Item Type	N	Mean	S.D.	Mean	S.D.
4	Girls	AC	8217	49.66	10.06	49.79	9.87
	Girls	NC	8217	49.17	10.07	49.40	9.79
	Boys	AC	8145	50.34	9.93	50.21	10.13
	Boys	NC	8145	50.84	9.86	50.60	10.18
5	Girls	AC	8715	49.50	9.96	49.73	9.77
	Girls	NC	8715	49.60	10.00	49.92	9.70
	Boys	AC	8511	50.51	10.02	50.28	10.22
	Boys	NC	8511	50.41	9.99	50.08	10.30
6	Girls	AC	7830	49.34	9.96	49.73	9.75
	Girls	NC	7830	49.21	9.93	49.64	9.61
	Boys	AC	7711	50.67	9.99	50.27	10.24
	Boys	NC	7711	50.80	9.95	50.36	10.37
7	Girls	AC	6397	49.29	9.93	49.77	9.62
	Girls	NC	6397	48.98	9.84	49.66	9.40
	Boys	AC	6278	50.72	10.02	50.23	10.36
	Boys	NC	6278	51.03	10.06	50.35	10.56
8	Girls	AC	5535	49.14	10.05	49.24	9.66
	Girls	NC	5535	49.76	9.92	49.84	9.51
	Boys	AC	5442	50.87	9.87	50.77	10.27
	Boys	NC	5442	50.25	10.08	50.17	10.48

Table 7
Repeated Measures Analysis of Variance
Within Subjects Effects by Grade: Interactions with Item Type

Grade 4:

Source	df	ms	F Value	Pr > F
T	1	0	0.00	1.0000
T*Q	2	1903	43.48	0.0001
T*CN	2	325	7.42	0.0006
T*Q*CN	4	75	1.72	0.1435
T*CP	2	489	11.18	0.0001
T*Q*CP	4	59	1.35	0.2492
T*CN*CP	4	54	1.24	0.2934
T*Q*CN*CP	8	90	2.07	0.0353
T*G	1	1281	29.26	0.0001
T*G*Q	2	219	5.01	0.0067
T*G*CN	2	32	0.73	0.4829
T*G*Q*CN	4	82	1.86	0.1140
T*G*CP	2	33	0.75	0.4735
T*G*Q*CP	4	23	0.52	0.7222
T*G*CN*CP	4	37	0.84	0.5015
T*G*Q*CN*CP	8	69	1.57	0.1277
Error	16308	44		

Grade 5:

Source	df	ms	F Value	Pr > F
T	1	0	0.00	1.0000
T*Q	2	789	18.05	0.0001
T*CN	2	60	1.37	0.2551
T*Q*CN	4	36	0.83	0.5067
T*CP	2	1397	31.98	0.0001
T*Q*CP	4	60	1.36	0.2442
T*CN*CP	4	19	0.44	0.7787
T*Q*CN*CP	8	128	2.92	0.0029
T*G	1	338	7.75	0.0054
T*G*Q	2	124	2.84	0.0587
T*G*CN	2	47	1.07	0.3447
T*G*Q*CN	4	75	1.71	0.1444
T*G*CP	2	64	1.46	0.2317
T*G*Q*CP	4	19	0.44	0.7833
T*G*CN*CP	4	26	0.60	0.6637
T*G*Q*CN*CP	8	44	1.00	0.4320
Error	17172	44		

Table 7 Continued

Grade 6:

Source	df	ms	F Value	Pr > F
T	1	0	0.00	1.0000
T*Q	2	707	16.59	0.0001
T*CN	2	62	1.46	0.2329
T*Q*CN	4	35	0.82	0.5126
T*CP	2	21	0.49	0.6102
T*Q*CP	4	105	2.48	0.0420
T*CN*CP	4	39	0.92	0.4518
T*Q*CN*CP	8	11	0.26	0.9782
T*G	1	88	2.07	0.1506
T*G*Q	2	272	6.38	0.0017
T*G*CN	2	127	2.99	0.0504
T*G*Q*CN	4	32	0.75	0.5605
T*G*CP	2	23	0.54	0.5830
T*G*Q*CP	4	26	0.62	0.6466
T*G*CN*CP	4	10	0.24	0.9163
T*G*Q*CN*CP	8	108	2.54	0.0093
Error	16308	44		

Grade 7:

Source	df	ms	F Value	Pr > F
T	1	0	0.00	1.0000
T*Q	2	418	12.19	0.0001
T*CN	2	155	4.52	0.0109
T*Q*CN	4	25	0.72	0.5751
T*CP	2	999	29.10	0.0001
T*Q*CP	4	37	1.08	0.3670
T*CN*CP	4	6	0.19	0.9455
T*Q*CN*CP	8	35	1.02	0.4177
T*G	1	215	6.27	0.0123
T*G*Q	2	352	10.27	0.0001
T*G*CN	2	80	2.34	0.0963
T*G*Q*CN	4	73	2.13	0.0738
T*G*CP	2	83	2.42	0.0894
T*G*Q*CP	4	64	1.85	0.1155
T*G*CN*CP	4	22	0.66	0.6230
T*G*Q*CN*CP	8	22	0.65	0.7386
Error	12621	34		

Table 7 Continued

Grade 8:

Source	df	ms	F Value	Pr > F
T	1	0	0.00	1.0000
T*Q	2	158	4.07	0.0172
T*CN	2	303	7.79	0.0004
T*Q*CN	4	24	0.63	0.6431
T*CP	2	800	20.57	0.0001
T*Q*CP	4	32	0.82	0.5108
T*CN*CP	4	53	1.36	0.2439
T*Q*CN*CP	8	27	0.70	0.6877
T*G	1	1413	36.34	0.0001
T*G*Q	2	5	0.12	0.8893
T*G*CN	2	39	1.00	0.3679
T*G*Q*CN	4	47	1.20	0.3080
T*G*CP	2	39	1.01	0.3638
T*G*Q*CP	4	9	0.23	0.9207
T*G*CN*CP	4	46	1.19	0.3116
T*G*Q*CN*CP	8	22	0.56	0.8132
Error	10923	39		

Table 8

Comparison of Selected Ability Groups
(Based on Standardized Raw Scores)

GRADE	ITEM TYPE	N	Q	C N	C P	MEAN	S D
4	AC	2783	LOW	LOW	LOW	39.5	7.34
	NC	2783	LOW	LOW	LOW	40.8	7.65
	AC	1129	MED	MED	MED	50.8	8.04
	NC	1129	MED	MED	MED	50.1	8.28
	AC	2819	HIGH	HIGH	HIGH	59.2	5.36
	NC	2819	HIGH	HIGH	HIGH	59.3	7.01
	AC	195	LOW	LOW	HIGH	44.6	8.50
	NC	195	LOW	LOW	HIGH	47.0	8.56
	AC	146	HIGH	HIGH	LOW	53.1	8.33
	NC	146	HIGH	HIGH	LOW	50.7	8.03
5	AC	2888	LOW	LOW	LOW	40.6	8.77
	NC	2888	LOW	LOW	LOW	39.9	7.61
	AC	1104	MED	MED	MED	50.1	8.43
	NC	1104	MED	MED	MED	50.0	7.91
	AC	3385	HIGH	HIGH	HIGH	58.0	6.99
	NC	3385	HIGH	HIGH	HIGH	58.7	6.71
	AC	175	LOW	LOW	HIGH	45.8	9.31
	NC	175	LOW	LOW	HIGH	47.6	8.10
	AC	153	HIGH	HIGH	LOW	52.2	8.02
	NC	153	HIGH	HIGH	LOW	53.0	7.70
6	AC	2636	LOW	LOW	LOW	40.5	7.66
	NC	2636	LOW	LOW	LOW	40.7	7.90
	AC	1247	MED	MED	MED	50.0	7.98
	NC	1247	MED	MED	MED	49.7	8.04
	AC	3121	HIGH	HIGH	HIGH	59.1	6.81
	NC	3121	HIGH	HIGH	HIGH	58.5	7.58
	AC	114	LOW	LOW	HIGH	44.8	8.04
	NC	114	LOW	LOW	HIGH	47.0	8.64
	AC	145	HIGH	HIGH	LOW	52.7	8.26
	NC	145	HIGH	HIGH	LOW	52.3	8.98

Table 8 Continued

GRADE	ITEM TYPE	N	Q	C N	C P	MEAN	S D
7	AC	2201	LOW	LOW	LOW	40.3	8.24
	NC	2201	LOW	LOW	LOW	40.4	7.12
	AC	1014	MED	MED	MED	50.5	7.63
	NC	1014	MED	MED	MED	49.6	7.74
	AC	2504	HIGH	HIGH	HIGH	59.1	6.35
	NC	2504	HIGH	HIGH	HIGH	59.9	7.33
	AC	93	LOW	LOW	HIGH	43.9	8.63
	NC	93	LOW	LOW	HIGH	45.9	9.07
	AC	106	HIGH	HIGH	LOW	53.8	7.44
	NC	106	HIGH	HIGH	LOW	53.2	6.96
8	AC	1783	LOW	LOW	LOW	41.3	7.78
	NC	1783	LOW	LOW	LOW	41.1	7.36
	AC	859	MED	MED	MED	49.1	7.92
	NC	859	MED	MED	MED	49.2	7.62
	AC	2114	HIGH	HIGH	HIGH	59.5	7.69
	NC	2114	HIGH	HIGH	HIGH	59.9	7.44
	AC	103	LOW	LOW	HIGH	44.6	8.01
	NC	103	LOW	LOW	HIGH	46.0	8.47
	AC	92	HIGH	HIGH	LOW	54.8	7.53
	NC	92	HIGH	HIGH	LOW	54.1	7.99