DOCUMENT RESUME

ED 347 189 TM 018 665

AUTHOR Freund, David S.; Rock, Donald A.
TITLE A Preliminary Investigation of Pattern-Marking in 1990 NAEP Data.
PUB DATE Apr 92
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Age Differences; Black Students; Cohort Analysis; *Educational Assessment; Grade 8; Grade 12; Hispanic Americans; *National Surveys; *Response Style (Tests); Secondary Education; *Secondary School Students; Sex Differences; Standardized Tests; Student Motivation; *Testing Problems; White Students
IDENTIFIERS *National Assessment of Educational Progress; *Pattern Marking (Data)

ABSTRACT

The National Assessment of Educational Progress (NAEP) focuses on providing information on what demographic subgroups of America's students know and can do. Because the NAEP does not report for individual students, it could be classified by some students as a low-risk test. Consequently, some students may lack proper motivation for giving their best efforts. This paper describes a method to identify students who may have randomly marked or systematically marked (pattern marked) their responses. The magnitude of the difference between responses (in absolute value) is used to compute a variance for each student's response string. The variances for students who take a particular block of items are compared; students who have a very small variance may have lacked motivation and are classified as potential pattern markers. Approximate sample sizes for each block (for both mathematics and science) were 3,700 students in the age 13 years/grade 8 cohort and 3,600 students in the age 17 years/grade 12 cohort of the 1990 NAEP. Subgroup analysis indicates that males are identified as potential pattern-markers more frequently than are females. Black and Hispanic American students are classified as potential pattern markers more often than are White students. Pattern marking tends to occur more frequently when the item block is near the end of the test. Seven tables, five figures, and three references are included. (Author/SLD)

A Preliminary Investigation of Pattern-marking in 1990 NAEP Data

David S. Freund and Donald A. Rock

Educational Testing Service, Princeton, New Jersey

Presented at AERA, San Francisco, CA
April 24, 1992

Abstract

The primary goal of the National Assessment of Educational Progress (NAEP) is to provide information on what demographic subgroups of America's students know and can do. Because NAEP does not report results for individual students, NAEP could be classified by some students as a low-risk test. Consequently, some students may lack proper motivation for giving their best effort. This paper describes a method to identify students who may have randomly marked or systematically marked, i.e., pattern-marked, their responses. The magnitude of the difference between responses (in absolute value) is used to compute a variance for each student's response string. The variances for students who take a particular block of items are compared; students who have a very small variance may have lacked motivation and are classified as potential pattern-markers. Subgroup analysis results indicate that males are identified as potential pattern-marks more frequently than females; black and hispanic students tend to be classified as potential pattern-markers more often than white students. There is a tendency for pattern-marking to occur more frequently when the item block is near the end of the test.

Introduction

The primary purpose of the National Assessment of Educational Progress (NAEP) is to provide information to educators and policymakers on what America's students know and can do in various subject areas such as reading, mathematics, and science. The National Assessment reports student performance for the entire nation and for important subpopulations, including gender, race/ethnicity, parental education, region of the country, and type of community. (Since 1990, NAEP has also included a state-level component called the Trial State Assessment Program.)

NAEP has always maintained the confidentiality of individual students, schools, school districts, and counties. Students are told in advance of the assessment session that their responses will remain confidential, and that NAEP is not a test--no grades will be assigned. The administrator of each NAEP session reads aloud to the students from a script that includes the following:

> "The National Assessment of Educational Progress is sponsored by the U.S. Government. Its goal is to find out what students your age know and can do in different school subjects. ... When the study is over, your answers will be combined with information from other students all over the country. The results will help government leaders, school administrators, and teachers to determine what students are learning. Because the study will have an impact on schools and students all over the country, we hope that you will do the best that you can!" (Westat, 1992)

In response to requests by some schools, NAEP will provide the school with a parental information letter that explains the assessment. An excerpt from that letter includes:

> "To ensure confidentiality, National Assessment results are not reported to, or about, individual students, school, or districts. Students' names are not recorded on any of the assessment materials taken from the school or reported in any way. Participation in the program will <u>not</u> affect your child's grades or progress in school." (Westat, 1992)

When students enter a classroom, cafeteria or library to participate in the National Assessment, they are aware that their responses will remain confidential. Consequently, NAEP

could be considered a low-risk test to students, a test in which some students may lack motivation to put forth their best effort. There may be some individuals who do not take the test seriously, including some who may randomly mark responses, or give responses in a systematic pattern to multiple-choice questions without concern for identifying the correct response. (Open-ended items do not provide a way for students to randomly mark or pattern-mark responses.) The goal of this paper is to identify a subset of students who may lack motivation, and as a consequence may randomly grid or systematically pattern-mark some, or all, of their responses. Given that a subset of students who exhibit this behavior can be identified, are there systematic differences in the magnitude of this behavior by:

- gender subgroups or race/ethnicity (White, Black, and Hispanic) subgroups?
- age/grade cohort (age 13/grade 8 versus age 17/grade 12)?
- block sequence in the test? (A NAEP test typically is composed of three separately timed blocks of cognitive items.) Is lack of motivation more evident for students in the third block in the test than in the first and/or second block?

Methods Used

In order to attempt to identify students who are potential random-markers or pattern-markers, the variance of the response pattern will be computed for separately timed blocks of items. Students respond to questions by gridding a response of A, B, C, D, or E; these responses are mapped to numeric values of 1, 2, 3, 4, and 5, respectively. Two different response pattern variance models will be used. In both models, the difference (in absolute value) between pairs of responses for a given student will be used in the computation of the variance of the response pattern.

- The *adjacent-pairs model* will compute the variance of adjacent pairs of nonblank responses (this model focuses on the magnitude in change between adjacent responses.) If, say, there are N items in a given block, the adjacent-pairs model will examine N-1 pairs of consecutive responses. The adjacent-pairs variance is defined as:

$$\frac{\sum_{i=1}^{N-1}(R_i - R_{i+1})^2 - \frac{(\sum_{i=1}^{N-1}|(R_i - R_{i+1})|)^2}{N-1}}{N-1}$$

where R_i = response for item i
N = number of items in the block

For example, suppose you have a response string of **ACDCB**. This string (of length five) is translated to the numerical values of **13432**. The absolute value of the differen between the adjacent responses results in the values **2, 1, 1, 1**, producing an adjacent-pairs variance of 0.19 or

$$\frac{(4+1+1+1) - \frac{(2+1+1+1)^2}{4}}{4} .$$

- The *all-pairs model* will examine all possible combinations of pairs of responses and compute the variance of these combinations. If there are N items in the block, then there are (N * (N-1)) / 2 pairs of responses. The all-pairs response pattern variance is defined as:

$$\frac{\sum_{i=1}^{N-1}\sum_{j=i}^{N-1}(R_j-R_{j+1})^2 - \frac{(\sum_{i=1}^{N-1}\sum_{j=i}^{N-1}|(R_j-R_{j+1})|)^2}{N\,(N-1)}}{N\,(N-1)}$$

where R_j = response for item j
N = number of items in the block

The response pattern of **ACDCB** results (after mapping to numeric) in the creation of responses pairs -- **1-3, 1-4, 1-3, 1-2, 3-4, 3-3, 3-2, 4-3, 4-2, 3-2.** The differences (in absolute value) of these pairs is **2, 3, 2, 1, 1, 0, 1, 1, 2, 1,** resulting in an all-pairs variance of 0.64.

Data Source

The 1990 NAEP provides data for examining student motivation. The subject areas of reading, mathematics, and science were assessed in the cross-sectional 1990 NAEP for three age/grade cohorts. The age 13/grade 8 and age 17/grade 12 cohorts will be analyzed in this paper. Students at age 9/grade 4 will not be examined, since these younger students have had less exposure to testing and thus are likely to extend their best effort. Analyses will be conducted using data from the mathematics and science assessments; the reading assessment will be disregarded in these analyses, to avoid the complications that may be associated with the extensive passages that are an essential part of the reading assessment.

The 1990 NAEP mathematics and science cross-sectional samples for age 13/grade 8 and age 17/grade 12 are based on a focused balanced incomplete block (BIB) design (see Kline, 1992). Each subject area contains seven blocks of cognitive items. Each cognitive block appears in three different booklets--in the first, second, and third position in the booklet. Each of these booklets is administered to a randomly equivalent subsample of eighth graders/13-year-olds or

twelfth graders/17-year-olds by "spiralling" the booklet forms within each assessment site. As a result of this design, seven different booklets are required for each subject area. Students respond to science items by gridding their responses on a separate answer sheet. Students taking math booklets respond directly in the test booklet. Therefore, differences in pattern-marking may occur for the science and mathematics booklets. Sample sizes for each block (for both math and science) are approximately 3,700 for age 13/grade 8 and 3,600 for age 17/grade 12. Since each block of cognitive items is presented in the first, second, and third position in the booklet, there are between 1,200 and 1,250 students per block for each position in the booklet.

The focus here will be only on multiple-choice items, ignoring the complexities associated with open-ended (free-response) item constructs. In determining which blocks are most appropriate for analysis, there are some important points to consider:

- Blocks selected should contain all, or nearly all, multiple-choice items; avoid blocks that are composed of many open-ended items (these items sometimes are more difficult, typically take longer to respond to, and do not provide a means for the student to guess, or randomly-mark/pattern-mark a response).
- Avoid blocks that require the use of rulers, protractors, or calculators.
- Select blocks that have considerable overlap in items between the age 13/grade 8 and age 17/grade 12 cohorts.

Two blocks that met the analysis requirements were identified for the science assessment and the math assessment for age 13/grade 8 and age 17/grade 12. These blocks contain all, or nearly all, multiple-choice items with either four or five choices per item. The science blocks are identified as SD and SG; math blocks are identified as MD and MG. (These blocks are sometimes referred to as S4, S7, M4, and M7 in other NAEP reports/documents.) Table 1 provides the characteristics of these blocks. The overlap items between age 13/grade 8 and age 17/grade 12 (26,

20, 20, and 16 items for blocks SD, SG, MD, and MG, respectively) will be the focus of these analyses.

Table 1
Item Representation within Blocks

Block	Age 13 / Grade 8		Age 17 / Grade 12		Overlap Multiple-choice items
	Multiple-choice items	Open-ended items	Multiple-choice items	Open-ended items	
SD	26	0	29	0	26
SG	20	3	23	3	20
MD	21	0	22	0	20
MG	17	1	18	3	16

Results

Table 2 summarizes the characteristics of the items to be analyzed including the response pattern variance of the correct (keyed) responses for both variance models, and the number of four- and five-choice items in each block. The response pattern variance of the correct responses is an important baseline for evaluating response pattern variances for individual students. The variances of the correct response string are considerably higher for block MG than the other blocks. Block MG includes five items that have a key of 'E' or '5', resulting in increased variances. No other block has more than two items where 'E' is the correct response.

Table 2
Profile of Blocks

Block	Number of Items	Response Pattern Variance of Correct (Keyed) Responses		Number of 4-Choice Items	Number of 5-Choice Items	Answer Mode
		Adjacent-Pairs	All-Pairs			
SD	26	0.56	0.86	25	1	Answer Sheet
SG	20	0.78	1.11	16	4	Answer Sheet
MD	20	0.78	0.97	14	6	In Book
MG	16	1.57	1.96	2	14	In Book

Response pattern variances are computed using both models for each student who took block SD, SG, MD, or MG. Distributions for each of the blocks are shown in the Appendix (Figures 1-4). Each graph includes distributions for both age/grade cohorts for both response-pattern variance models. The shapes of the distributions differ considerably for each block. However, each distribution suggests that the response pattern variances are approximately normally distributed. Students whose response pattern variance is in the lower tail of the distribution will be the focus of these analyses. In particular, any student whose response pattern variance is 1.5 standard deviations or more below the mean response pattern variance for the age 13/grade 8 distribution for either the adjacent-pairs model or the all-pairs models is considered a potential pattern-marker. (In a normal distribution, 6.7% of the population is 1.5 standard deviations or more below the mean.) The choice of 1.5 standard deviations may seem to be arbitrary; however, it was chosen for the following reasons:

- Inspection of the actual response vectors 'around' 1.5 standard deviations below the mean response pattern variance indicated that these students appeared to be students

who might have pattern-marked.

- Sufficient sample sizes are required to analyze subgroup differences. If a cut-off point larger than 1.5 standard deviation units below the mean response-pattern variance is established, sample sizes are too small, so race/ethnicity comparisons are not plausible.

The 13-year-old/eighth-grade distribution was chosen as the "standardization" distribution because it was felt that of the two cohorts, the younger students would be more likely to put forth their best effort when compared with the 17-year-olds/twelfth graders. The critical values for ascertaining if a student is a potential pattern-marker are presented in Table 3.

Table 3
Critical Values for Determining Potential Pattern-markers

Block	Adjacent-pairs	All-pairs
SD	0.41	0.67
SG	0.40	0.69
MD	0.50	0.71
MG	0.53	0.75

A sample of response patterns from students who were identified as potential pattern-markers and their associated response pattern variances are presented in Table 4. Patterns 1-6 are also included in the Appendix (Figure 5) with the students' responses gridded on the answer sheet. Undoubtedly, these students are pattern-markers. Generally, both variance models coincide in identification of response patterns as pattern-markers. However, examples #3, #5, and #6 are response patterns that were identified only using the adjacent-pairs model; the all-pairs model did not recognize these patterns as pattern-marking. On the other hand, the all-pairs model may at times (examples #8 and #9) evaluate response patterns as potential pattern-marking while the adjacent-pairs model does not.

Table 4
Sample Response Patterns
(* indicates pattern is in lower tail of distribution, i.e., potential pattern-marker)

#	Block	Responses	Response Pattern Variance	
			Adjacent-Pairs	All-Pairs
1	SG	11111111111111111111	0.0 *	0.0 *
2	SG	12121212121212121212	0.0 *	0.25 *
3	SG	54321234321234321234	0.0 *	0.89
4	SG	55544444444444444454	0.13 *	0.23 *
5	SG	55543212343432123454	0.09 *	1.12
6	SG	12343212341234321234	0.20 *	0.77
7	MG	2444444444444442	0.46 *	0.72 *
8	SD	23142313232313232323132332	0.44	0.56 *
9	MG	3222113223121311	0.60	0.49 *
10	MD	31121212213212211131	0.47 *	0.48 *

Tables 5 includes the percent of students identified as potential pattern-markers for both age/grade cohorts for the four blocks of items in this study. Also included in Table 5 is the mean number correct, i.e., the mean number of items that the students answered correctly (and the corresponding standard deviation), both for the group of students classified as potential pattern-markers and for the total group of students who took that block. The results are presented separately for each position in the booklet and for all three positions combined. Some observations of these results include:

- The percent of students identified as pattern-markers within a block ranges from 4.9% to 10.4% with an average of 7.1%.

- Students identified as pattern-markers perform poorly in terms of mean number of items correct in the block. The difference in mean number correct between potential pattern-markers and the total sample is generally close to one standard deviation unit.
- There is a tendency for pattern-marking to occur more frequently when the block is in the third (last) position in the book. Chi-square tests were performed for the number of students identified as pattern-markers versus the number not identified as pattern-markers by position in the block for each age/grade. For both age/grade cohorts, block SG produced a significant chi-square test result (at alpha level = .05), suggesting that there may be a block position effect (the other six chi-square tests produced values ranging from 0.4 to 4.4).
- The difference in performance (mean number correct in the block) between potential pattern-markers and the total sample tends to be larger for age 17/grade 12 than for age 13/grade 8.

Table 5
Percent Pattern-markers and Mean Number Right by Position in Booklet

Block	Age/ Grade	Position in book	Potential Pattern-markers Percent	Mean (St Dev)	Total Mean
SD	13/08	1	7.5%	10.9(4.5)	13.8(4.7)
		2	8.8%	10.3(4.9)	13.6(5.0)
		3	8.9%	10.1(4.9)	13.5(5.1)
		Total	8.4%	10.4(4.8)	13.6(4.9)
	17/12	1	6.3%	13.0(6.1)	16.6(5.0)
		2	6.9%	12.2(6.1)	16.9(5.1)
		3	6.4%	12.3(5.9)	16.6(5.1)
		Total	6.6%	12.5(6.1)	16.7(5.1)
SG	13/08	1	7.3%	5.7(2.1)	8.2(3.6)
		2	7.0%	5.4(2.2)	8.2(3.6)
		3	10.4%	5.2(2.5)	7.6(3.4)
		Total	8.2%	5.4(2.3)	8.0(3.5)
	17/12	1	4.9%	5.9(3.0)	10.7(4.2)
		2	5.9%	5.8(2.9)	10.6(4.2)
		3	7.4%	5.7(2.9)	10.1(4.4)
		Total	6.1%	5.8(2.9)	10.5(4.3)
MD	13/08	1	6.8%	9.4(3.8)	10.9(3.6)
		2	6.3%	8.5(3.4)	10.9(3.8)
		3	8.4%	6.7(4.2)	10.6(3.8)
		Total	7.2%	8.1(4.0)	10.8(3.7)
	17/12	1	5.5%	11.0(3.8)	13.4(3.7)
		2	6.2%	10.4(4.8)	13.5(3.7)
		3	7.5%	9.5(4.6)	13.1(3.8)
		Total	6.4%	10.2(4.5)	13.4(3.8)
MG	13/08	1	8.6%	5.0(2.5)	6.7(3.2)
		2	8.6%	5.0(2.5)	6.7(3.2)
		3	9.3%	4.0(2.9)	6.3(3.3)
		Total	8.8%	4.6(2.7)	6.6(3.3)
	17/12	1	5.8%	5.7(2.8)	8.9(3.8)
		2	4.9%	5.6(3.2)	9.0(3.7)
		3	5.2%	4.8(3.1)	9.0(3.8)
		Total	5.3%	5.4(3.1)	9.0(3.8)

Subgroup Results

The next step is to examine if there are any gender or race/ethnicity differences in the number of students identified as potential pattern-markers. Table 6 includes these results, presented in terms of the percentage of the subgroup identified as potential pattern-markers, enumerated by block and age/grade. All three block positions are combined to provide sufficient sample sizes for subgroup analysis. In general, male students are classified as being possible pattern-makers more frequently than are female students. Black and Hispanic students are categorized as pattern-markers more often than are White students.

Table 6
Percent of Students by Subgroup Identified as Potential Pattern-Markers

Block	Age/Grade	Total	Male	Female	White	Black	Hispanic
SD	13/08	8.4	9.6	7.2	6.4	12.6	11.7
	17/12	6.6	8.1	4.9	5.9	9.4	8.4
SG	13/08	8.2	9.5	6.9	6.4	11.8	12.6
	17/12	6.1	7.6	4.4	4.5	10.9	10.9
MD	13/08	7.2	8.4	5.9	6.1	7.8	11.3
	17/12	6.4	7.3	5.5	5.9	6.9	8.8
MG	13/08	8.8	9.0	8.7	8.1	10.4	10.6
	17/12	5.3	6.1	4.5	5.2	6.6	4.4

To evaluate whether there is a significant difference by gender or race/ethnicity for the number of students identified as potential pattern-markers, a chi-square test is employed. The test is performed by block for both age 13/grade 8 and age 17/grade 12. These results (presented in

Table 7) indicate that there is a difference in the number of males identified as potential pattern-markers versus the number of females for all but one of the eight chi-square tests. Only block MG at age 13/grade 8 failed to produce a significant difference. Chi-square tests were also performed for race/ethnicity subgroups; significant differences were found for block SD, SG, and MD for each age/grade cohort. Block MG neglected to produce a significant difference. This may be caused by the large variances associated with the correct response pattern for block MG; perhaps a different model for identifying potential pattern-markers might be more successful for this block.

Table 7
Chi-Square Test Results by Gender and Race/Ethnicity

	Gender		Race/Ethnicity	
Block	Age 13/Grade 8	Age 17/Grade 12	Age 13/Grade 8	Age 17/Grade 12
SD	7.0 *	14.7 *	33.8 *	10.0 *
SG	7.9 *	15.4 *	34.6 *	49.0 *
MD	8.5 *	5.0 *	18.3 *	5.2 *
MG	0.1	4.3 *	5.4	2.5

* Indicates significant chi-square at alpha = 0.05. Degrees of freedom = 1 for gender and 2 for race/ethnicity.

Conclusions

Results of the investigation of pattern-marking in the 1990 NAEP data suggest that:

- There is a gender difference in the potential for pattern-marking. Male students appear more likely to pattern-mark than female students by the present criteria.
- A possible race/ethnicity effect exists in the potential for pattern-marking. Generally, a lower percentage of White students were identified as potential pattern-markers than were Black or Hispanic students.
- There is some suggestion of a block positioning effect in pattern-marking. Students appear to pattern-mark more frequently towards the end of the test.

If students can be clearly identified as pattern-markers, perhaps they should be excluded from item parameter estimation (IRT) (see Mislevy, 1992). These students probably add very little to the accuracy of the item parameters and, in fact, may add unnecessary variance to the item parameters.

As the emphasis on national testing, in general, and NAEP, in particular, continues to grow, students are being exposed to an increasing number of "low-risk" tests. This increased exposure may result in pattern-marking becoming more prevalent in the future. If test development proceeds in the direction of more examinee-constructed responses, pattern-markers may be more difficult to identify, since a pattern-marker may choose simply to not respond to constructed-response items. The net effect may be that response patterns of nonmotivated students, who may have pattern-marked, may closely resemble response patterns of low-ability students.

Future Research

In recognition of the potential problems associated with student motivation in NAEP, the 1992 NAEP Assessment includes a question that asks the student "How important was it to you to do well on this test?" Students' responses to this question will provide additional data to investigate student motivation and may assist in identifying pattern-markers.

References

Westat, Inc. (1992). *Manual for assessment administrators - Fourth grade* (Appendix B-1, p. 41). Rockville, MD: Westat, Inc.

Kline, Debra L. (1992). Assessment instrumentation. In E.G. Johnson and N.L. Allen, *The NAEP 1990 technical report.* (No. 21-TR-20). Washington, DC: National Center for Education Statistics.

Mislevy, R. J. (1992). Scaling procedures. In E.G. Johnson, and N.L. Allen, *The NAEP 1990 technical report.* (No. 21-TR-20). Washington, DC: National Center for Education Statistics.

Appendix

- Response-pattern variance distributions for block SD, SG, MD, and MG (figures 1, 2, 3, and 4, respectively) for age 13/grade 8 and age 17/grade 12 for the adjacent-pairs and the all-pairs response pattern variance models.

- Examples of pattern-marking (figure 5) for block SG.

Figure 1

Distribution of Response Pattern Variance
Block SD

(Shaded Area is 1.5 Std. Dev. Below Mean)

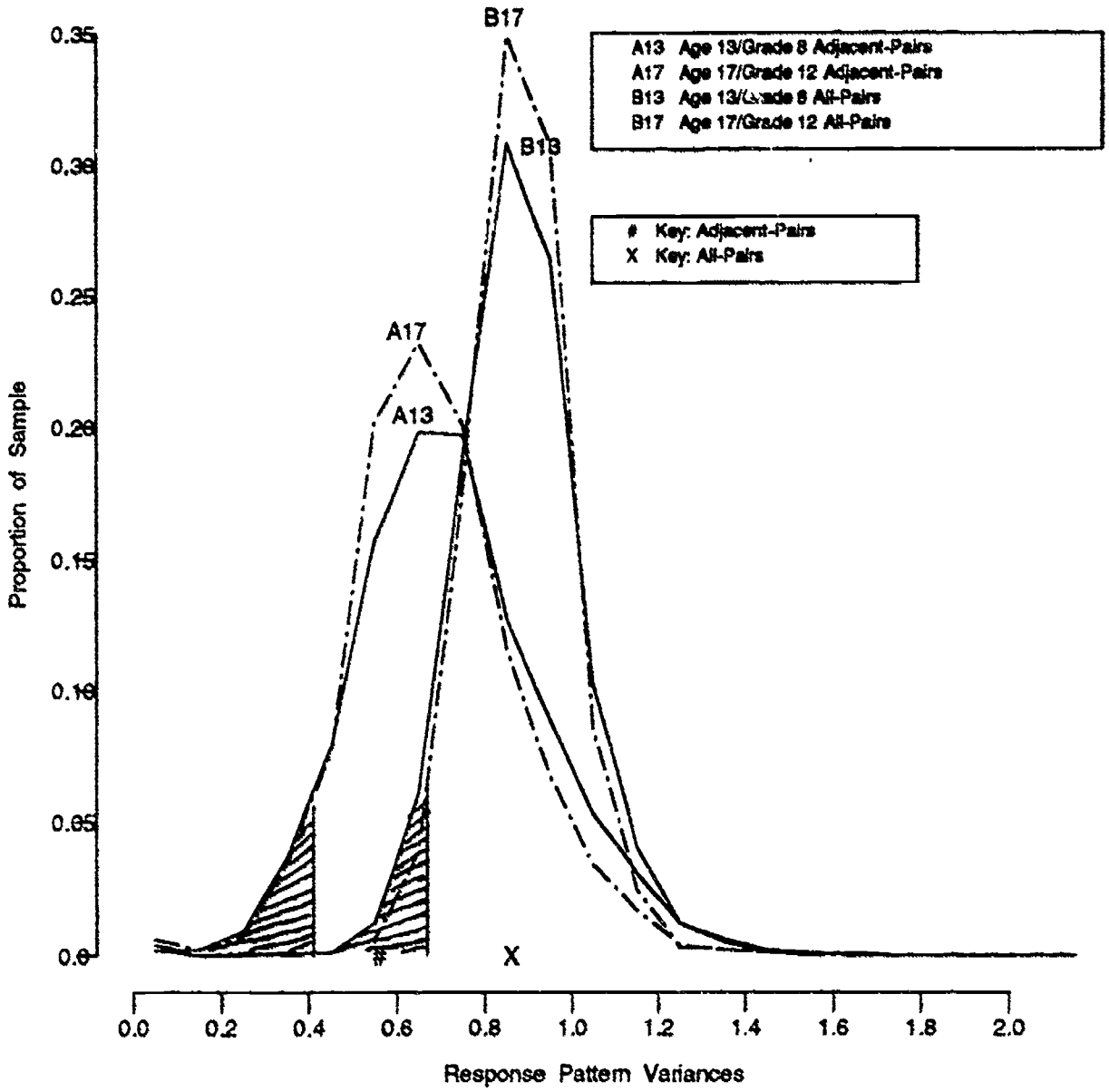


Figure 2

Distribution of Response Pattern Variance

Block SG

(Shaded Area is 1.5 Std. Dev. Below Mean)

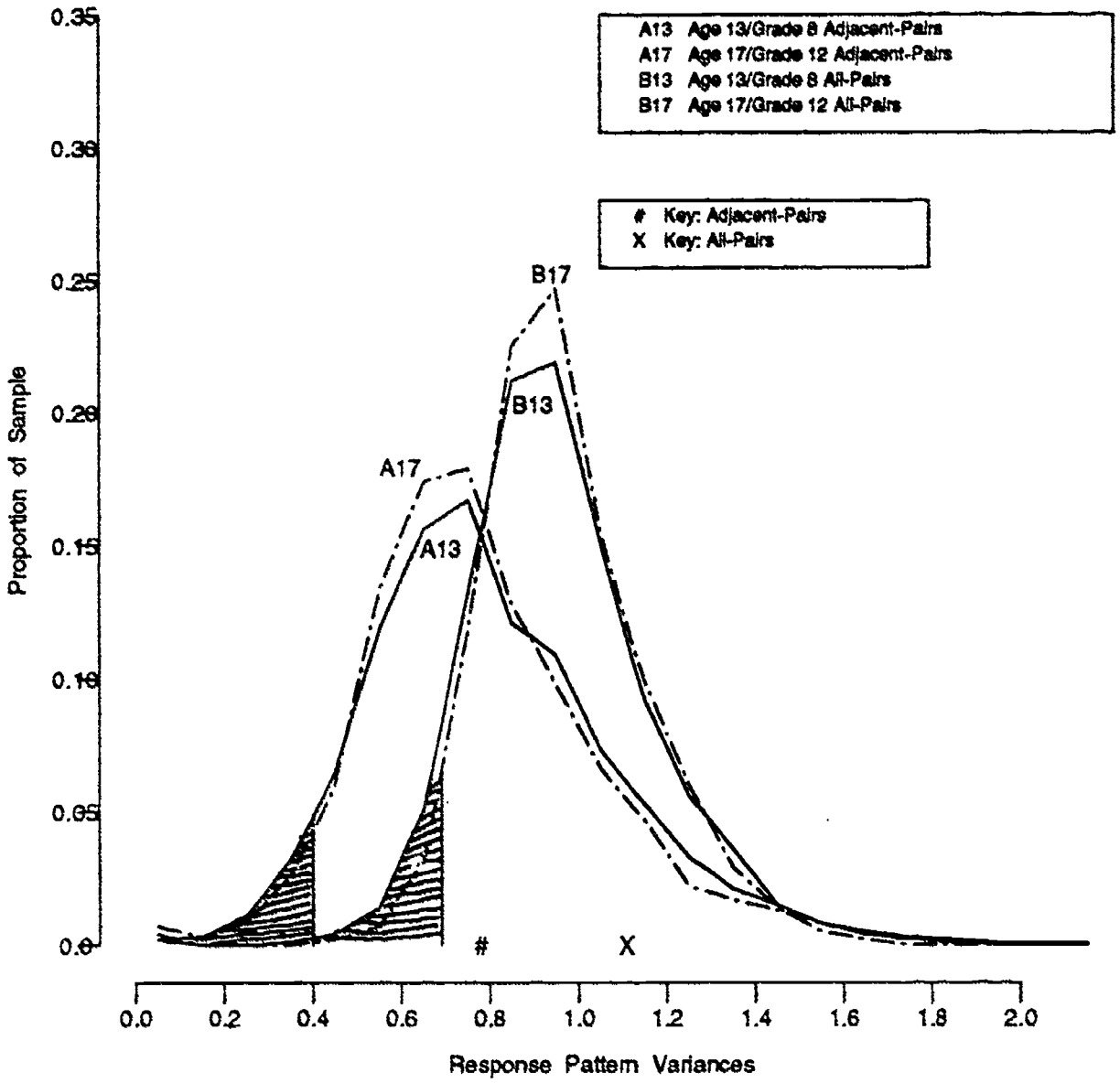


Figure 3
Distribution of Response Pattern Variance
Block MD
(Shaded Area is 1.5 Std. Dev. Below Mean)

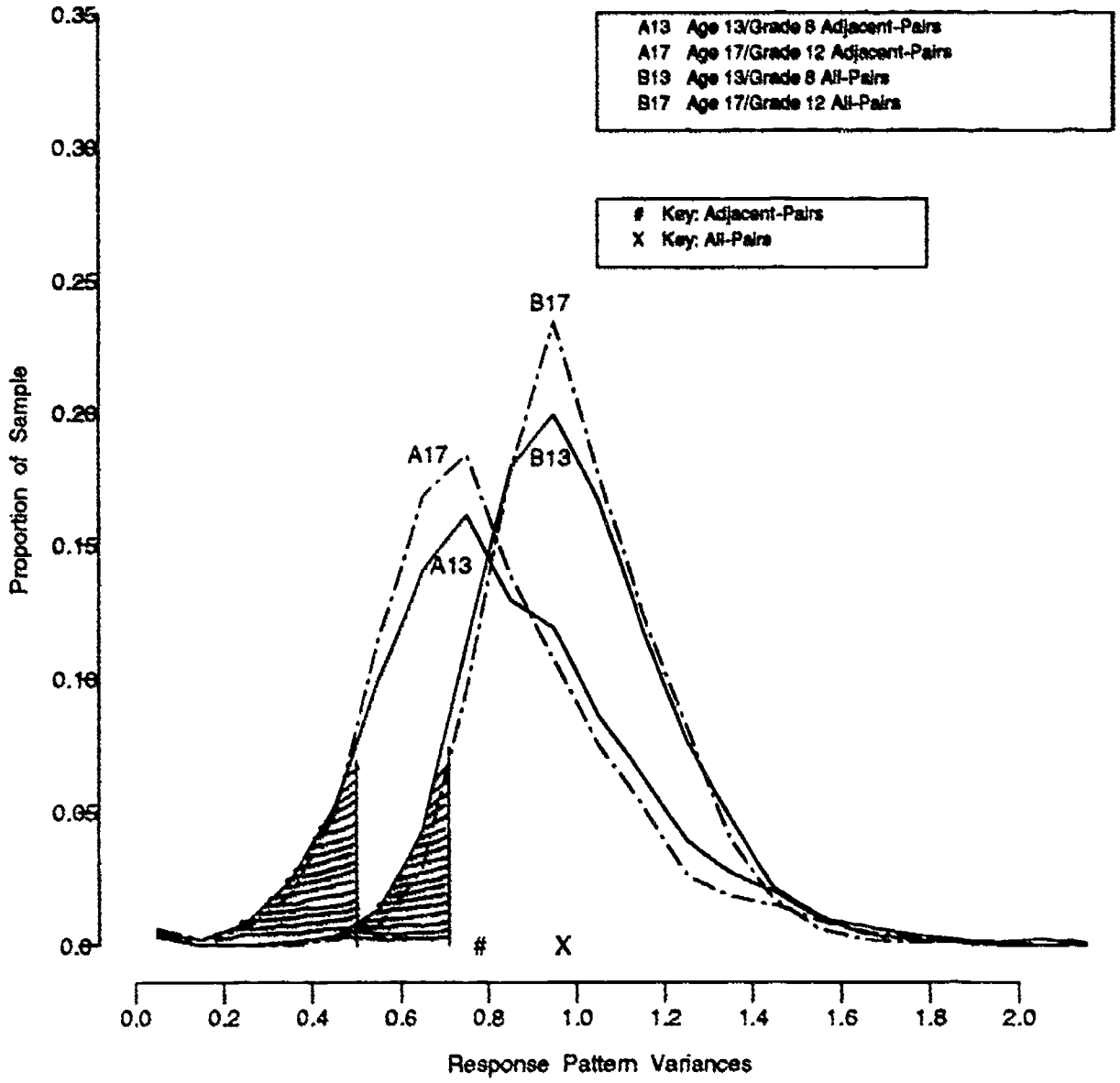


Figure 4
Distribution of Response Pattern Variance
Block MG
(Shaded Area is 1.5 Std. Dev. Below Mean)

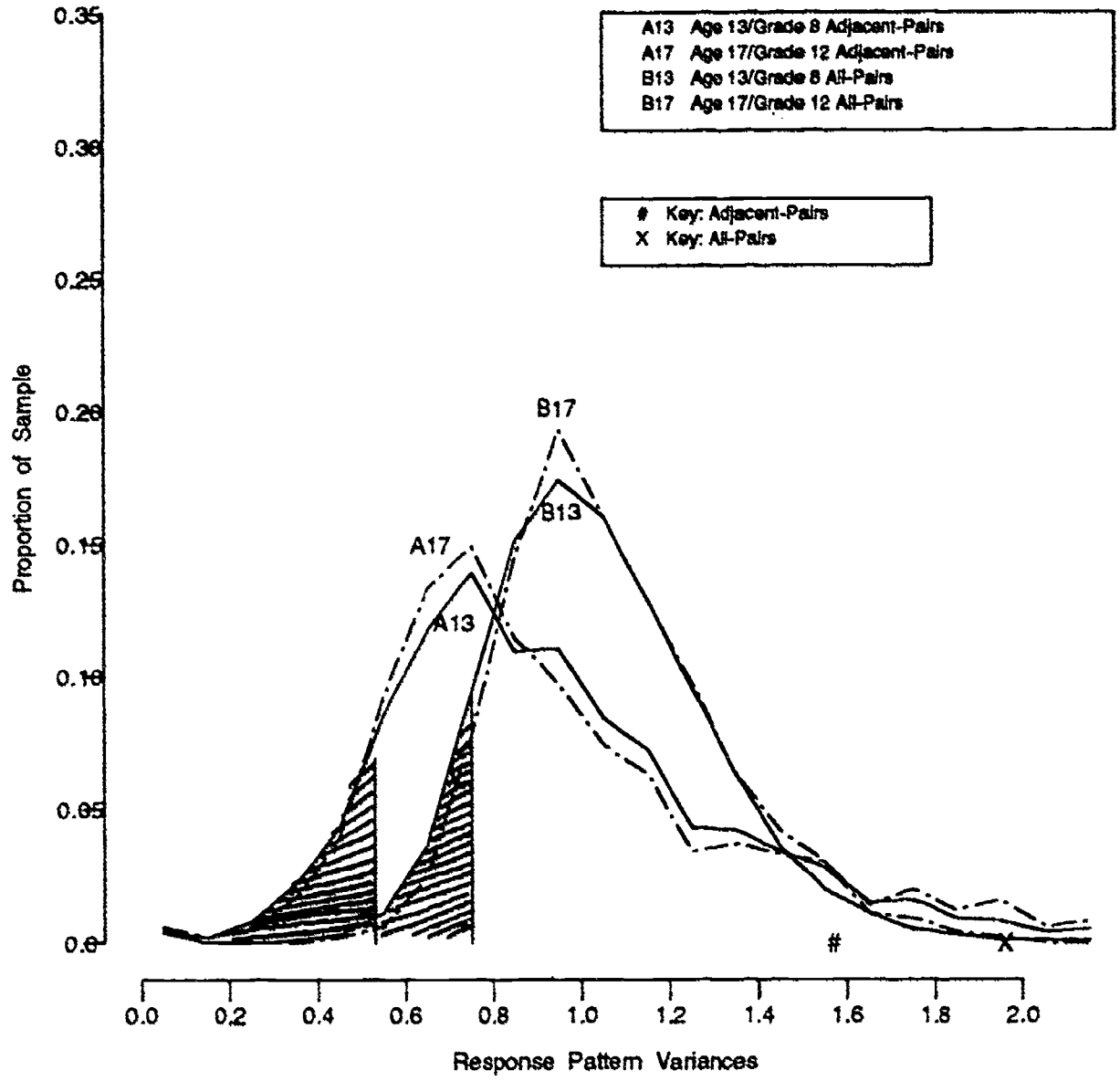
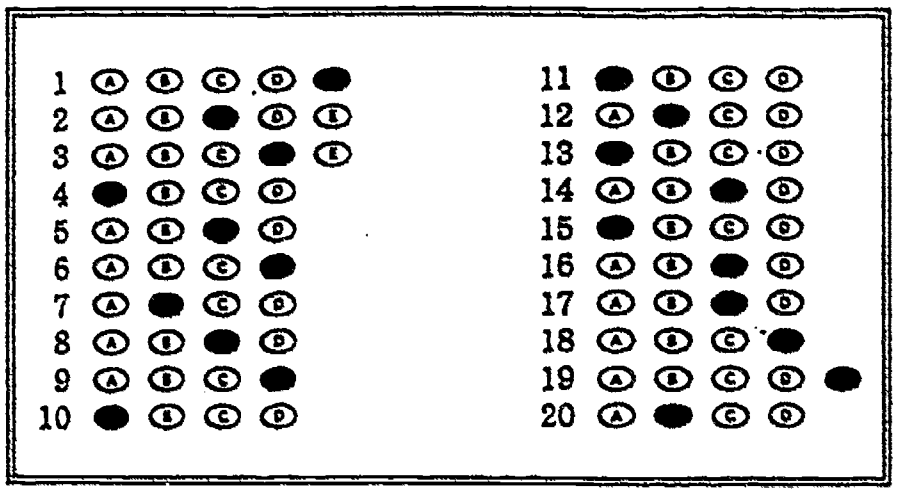
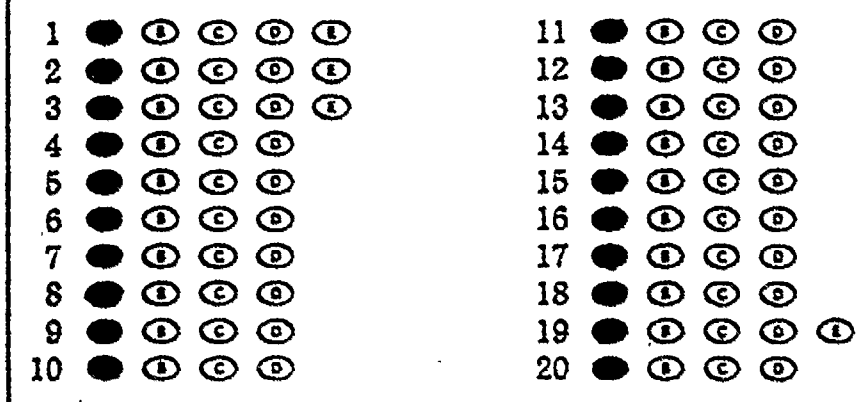


Figure 5
Examples of Pattern-marking for Block SG

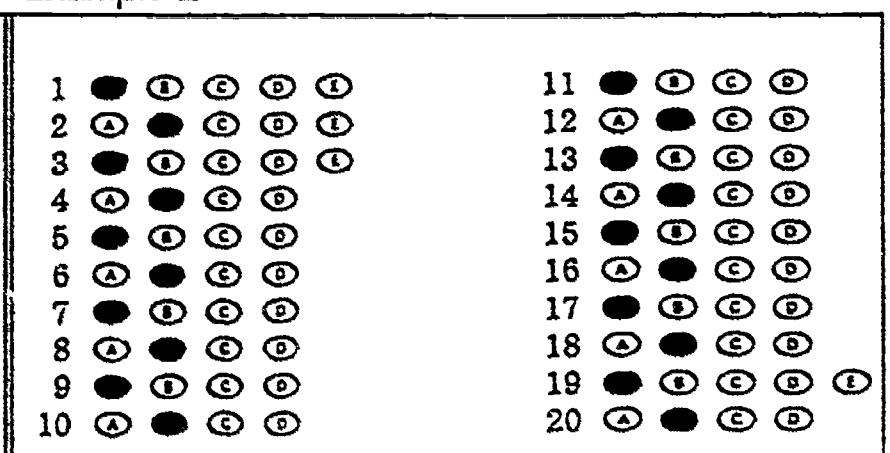
KEY



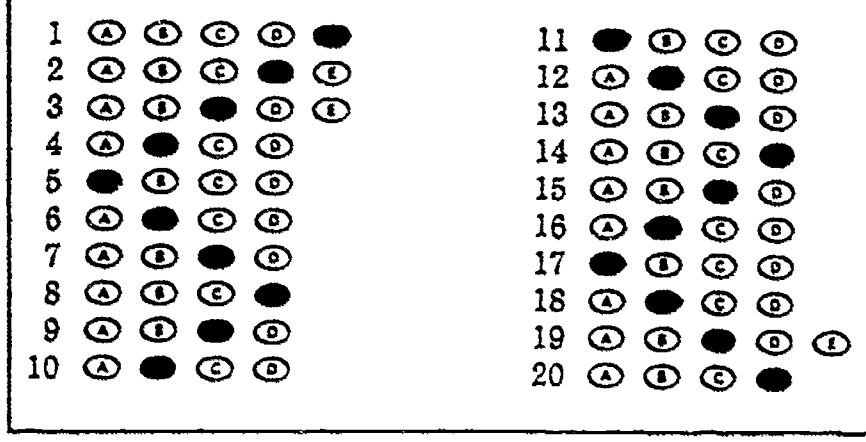
Example 1



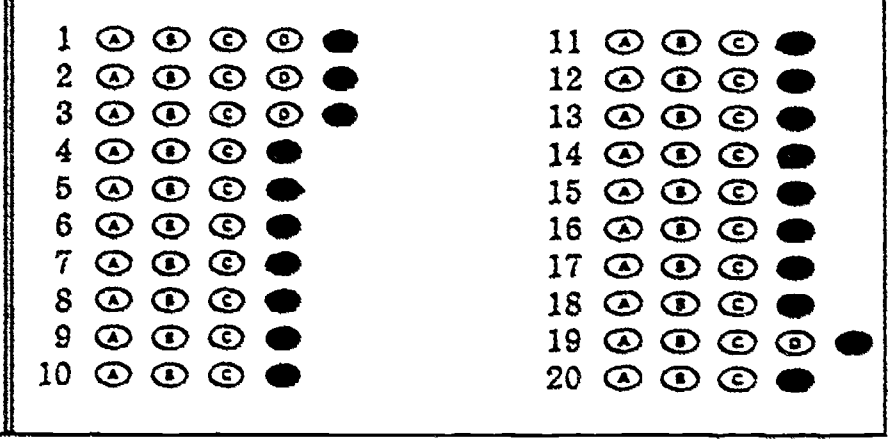
Example 2



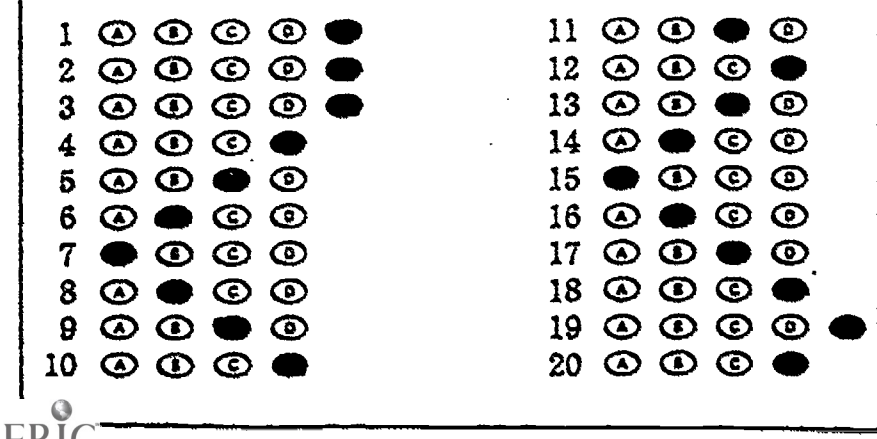
Example 3



Example 4



Example 5



Example 6