THE EFFECT OF ITEM POSITION ON STATE MATHEMATICS ASSESSMENT

Pui Chi Chiu

University of Kansas

AERA 2012

Purpose

In this study, the goal is to investigate the effect of item position on students' performance on Mathematics assessments. This study focuses on grade 7 students who took Mathematics assessments in Spring 2009. The Mathematics assessment is part of a statewide assessment system in a Midwestern state. Items were arranged in two different ways to create two different test forms with the exact same items. Item difficulty (*p*-value) for each item will be calculated and examined. In addition, items were grouped by states content standards, and the average *p*-value of the grouped items was calculated and compared across test forms. Furthermore, the differences of item parameters between test forms were investigated using the Mantel-Haenszel method to assess the presence of differential item function (DIF).

Perspective

In testing situations, the use of alternate test forms constructed with the same items presented in different order is one of the strategies for deterring copying and enhancing test security in test administrations. Scrambling, the rearrangement of the same set of items within a test form, is often used to discourage examinee copying (Harris, 1991). However, the psychometric literature has shown that varied item and section orders can affect item and section characteristics (such as difficulty) and as a result have unintended effects on test scores (Pommerich & Harris, 2003; Zwick, 1991). These effects can make claims of test form interchangeability questionable, possibly violating testing industry standards (Moses, Yang, & Wilson, 2007).

Newman, Kundert, Lane, and Bull (1988) found that students (enrolled in an undergraduate educational psychology class) who received the forms with items in an increasing cognitive order scored higher on hard items, no matter what order of statistical difficulty; while

students who received items in an decreasing cognitive order and statistical difficulty orders scored the highest on medium difficulty level items. Hambleton and Traub (1974) studied 11th graders' performance on an Algebra II Mathematics Test. They discovered the average number of correct answers for test questions arranged from easy-to-difficult was significantly higher than the test questions arranged from difficult-to-easy.

Newman, Kundert, Lane, and Bull (1988) also found that students (enrolled in an undergraduate educational psychology class) who received the tests with items ordered in an increasing cognitive order (knowledge, comprehension, application, Bloom's (1956) taxonomy) received higher scores on hard items, no matter what order of statistical difficulty (easy, medium, hard or reversed order); while student who received items in an decreasing cognitive order and statistically difficulty orders scored the highest on medium different items.

Furthermore, Plake, Ansorge, Parker and Lowry (1982) suggested that gender interacts with item arrangement in mathematics achievement test; males obtained highest scores with an easy-hard ordering. Ryan and Chiu (2001) also found that both male and female test takers benefitted when the order of the algebra operations items were altered (easy to difficult within content area).

Data Sources and Methods

Participants and Assessments

This study focuses on grade 7 students who took the Mathematics assessments in a Midwestern state in Spring 2009. In each grade-level, two base forms and two corresponding scrambled forms were administered. However, one base and its scrambled forms (two parallel forms) were selected and analyzed in this study. Each test form was comprised of three parts, and each part consisted of about 27 to 29 items. Each parallel test form consisted of exactly the same

items but they were ordered in different ways within part. There were a total of 84 items on 7th grade assessments, which focused on operations and algebraic thinking, number and operations, measurement and data, and geometry. Table 1illustrates how items were ordered in each of the test forms. For instance, the same item appears as item 1 on the base form, as item 9 on the scrambled form on part 1. Table 2 illustrates the number of items and students in each test form.

The item format for these assessments is multiple-choice, with one correct answer to be selected from four response options. All forms were administered via computer-based testing (CBT), random selection and inclusion on a test form was possible for each student taking a set of items. Thus, test forms were randomly distributed to students ensuring that each test form was administered to a random group of students' representative of the student population subgroups in the Midwestern state. In addition, each test form is a part-based; each part of the test was administered separately to the students. Students do not have to take the entire test (three parts) at the same time; they can take each part of the test in multiple days while the test window is open. Special educational students (except gifted students) and students who were provided the read aloud accommodation were removed from the study.

Analyses

The effect of item ordering on students' performance was examined by looking at average percent correct scores of three test forms, items' proportion corrects (*p*-values), and item characteristic functions. The average percent correct scores were calculated for each test form, and they were compared across forms using a univariate analysis of variance (ANOVA). *P*-value for each item was calculated and compared across test forms. Items were grouped by states content standards, and the average *p*-value of grouped items was calculated and compared across test forms.

Item characteristic functions were compared across test forms using the Mantel-Haenszel method to assess the presence of differential item functioning (DIF). Differential Item Functioning Analysis System (DIFAS) (Penfield, 2005) was used to calculate the following: Mantel-Haenszel chi-square (MH CHI) (Mantel & Haenszel, 1959), Mantel-Haenszel common log-odds ratio (MH LOR), standard error of the Mantel-Haenszel common log-odds ratio (LOR SE), Breslow-Day chi-square (BD) (Breslow & Day, 1980), and the Educational Testing Service (ETS) categorization scheme (Zieky, 1993).

Results

A total of 12,247 students took 7th grade Mathematics assessment in Spring 2009. Of those, 6,119 students took base form, and 6,128 students took scrambled form. 7th grade assessment consisted of 84 items; 28 items in part 1, 29 items in part 2, and 27 items in part 3. Across the two forms, the average percent correct score obtained on the assessment was 73.4 (*SD* = 15.1), which is about 62 items answered correctly. Summary statistics for two test forms are given in Table 3.

A one-way ANOVA was conducted to explore differences in students' percent correct scores across test forms. The results indicate there were no statistically significant differences in students' percent correct scores across test forms. However, when looking at the *p*-value for each item and the average *p*-value for the grouped items, the results suggest the items had greater *p*-value when they were placed earlier on the test. Table 4 illustrates the *p*-value for each item and the average *p*-values of grouped items.

The Mantel-Haenszel method was used to examine item parameters' differences across test forms. Base form was used as the reference group in assessing the presence of DIF. The results of the Mantel-Haenszel chi-square statistics indicate that 34 items were identified as

uniform DIF. The results of the Breslow-Day chi-square indicate that four items were identified as non-uniform DIF (two of which were not identified by the Mantel-Haenszel chi-square statistic). Table 5 summarizes the results obtained from DIFAS.

Scientific significance

Over the years testing, cheating, and test security has increased as has the use of high stakes testing. The use of alternate test forms constructed with the same items presented in different order is one of the strategies for deterring copying and enhancing test security in test administrations. However, when scrambled forms of a test are used, the question of equity arises when scrambled versions of a test form are administered at the same time the base form is administered (Harris, 1991). Thus, caution would be used when scrambled forms are being administered in the state assessments, if item placement has statistically significant effect on the item parameters.

References

Bloom, B. (Ed.). (1956). *Taxonomy of Educational Objectives. Handbook 1: Cognitive Domain*. New York: McKay.

Breslow, N. E., & Day, N. E. (1980). *Statistical methods in cancer research: Volume 1 – The analysis of case-control studies*. Lyon: International Agency for Research on Cancer.

Hambleton, R. K. & Traub, R. E. (1974). The effects of item order on test performance and stress. *The Journal of Experimental Education, 43*(1), 40-46.

Harris, D. J. (1991). Effects of Passage and Item Scrambling on Equating Relationships. *Applied Psychological Measurement, 15*(3), 247-256.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719-748.

Moses, T., Yang, W., & Wilson, C. (2007). Using Kernel Equating to Assess Item Order Effects on Test Scores. *Journal of Educational Measurement, 44*(2), 157-178.

Newman, D. L., Kundert, D. K., Lane, D. S., & Bull, K. S. (1988). Effects of varying item order on multiple-choice test scores: Importance of statistical and cognitive difficulty. *Applied Measurement in Education, 1*(1), 89-97.

Penfield, R. D. (2005). DIFAS: Differential item functioning analysis system. *Applied Psychological Measurement, 29,* 150–151.

Plake, B. S., Ansorge, C. J., Parker, C. S., & Lowry, S. R. (1982). Effects of Item Arrangement, Knowledge of Arrangement Test Anxiety, and Sex on Test Performance. *Journal of Educational Measurement, 19*(1), 49–57.

Pommerich, M., & Harris, D. J. (2003). *Context Effects in Pretesting: Impact on Item Statistics and Examinee Scores*. Paper presented at the Annual Meeting of the American Educational

Research Association, Chicago.

Ryan, K. E., & Chiu, S. (2001). An Examination of Item Context Effects, DIF, and Gender DIF. *Applied Measurement in Education, 14*(1), 73-90.

Zieky, M. (1993). Practical questions use of DIF statistics it item development. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 337-364). Hillsdale, NJ: Lawrence Erlbaum.

Zwick, R. (1991). Effects of item order and context on estimation of NAEP reading proficiency. *Educational Measurement: Issues and Practice, 10*, 10–16.

Table 1

Item Orders for Each of the Test Forms.

Part	Standard	Base Form Question Number	Scrambled Form Question Number
1	Algebra	1	9
1	Algebra	2	10
1	Algebra	3	11
1	Algebra	4	12
1	Algebra	5	13
1	Algebra	6	20
1	Algebra	7	21
1	Algebra	8	22
1	Algebra	9	23
1	Geometry	10	1
1	Geometry	11	2
1	Geometry	12	3
1	Geometry	13	4
1	Geometry	14	24
1	Geometry	15	25
1	Geometry	16	26
1	Geometry	17	27
1	Geometry	18	28
1	Geometry	19	5
1	Geometry	20	6
1	Geometry	21	7
1	Geometry	22	8
1	Number & Computation	23	14
1	Number & Computation	24	15
1	Number & Computation	25	16
1	Number & Computation	26	17
1	Number & Computation	27	18
1	Number & Computation	28	19
2	Algebra	1	16
2	Algebra	2	17
2	Algebra	3	18
2	Algebra	4	19
2	Algebra	5	20
2	Algebra	6	21
2	Algebra	7	22
2	Algebra	8	23
2	Number & Computation	9	1
2	Number & Computation	10	2
2	Number & Computation	11	3
2	Number & Computation	12	4
2	Number & Computation	13	5
2	Algebra	14	24
2	Algebra	15	25
2	Algebra	16	26
2	Algebra	17	27
2	Algebra	18	28
2	Algebra	19	29
2	Geometry	20	11
2	Geometry	21	12
2	Geometry	22	13
2	Geometry	23	14
2	Geometry	24	15
2	Algebra	25	6
2	Algebra	26	7
2	Algebra	27	8
2	Algebra	28	9
2	Algebra	29	10

Part	Standard	Base Form Question Number	Scrambled Form Question Number
3	Number & Computation	1	8
3	Number & Computation	2	9
3	Number & Computation	3	10
3	Number & Computation	4	11
3	Number & Computation	5	12
3	Number & Computation	6	13
3	Number & Computation	7	14
3	Number & Computation	8	15
3	Geometry	9	21
3	Geometry	10	22
3	Geometry	11	23
3	Geometry	12	24
3	Geometry	13	25
3	Geometry	14	26
3	Geometry	15	27
3	Data	16	1
3	Data	17	2
3	Data	18	3
3	Data	19	4
3	Data	20	5
3	Data	21	6
3	Data	22	7
3	Data	23	16
3	Data	24	17
3	Data	25	18
3	Data	26	19
3	Data	27	20

Table 2

Numbers of Items and Numbers of Students for Each Form.

Test Form	Number of Students	Number of Items			
		Total	Part 1	Part 2	Part 3
Base Form	6,493	84	28	29	27
Scrambled Form	6,463				

Table 3

Summary Statistics for Each Test Form.

	Base Form	Scrambled Form
Number of Items	84	84
Percent Correct Score	73.39	73.36
Standard Deviation of Percent Correct Score	15.30	14.99
Average Correct Responses	61.93	61.92
Standard Deviation of Average Correct Responses	13.83	13.56

Table 4

Summary for Items' p-value and Average p-values of Grouped Items.

Part	Standard	Benchmark	Base Form			Scrambled Form		
			Question Number	P-value	Average (P-value)	Question Number	P-value	Average (P-value)
1	Algebra	Variables, Equations, & Inequalities	1	94.1	85.3	9	93.2	84.9
1	Algebra	Variables, Equations, & Inequalities	2	82.4		10	81.8	
1	Algebra	Variables, Equations, & Inequalities	3	86.3		11	84.5	
1	Algebra	Variables, Equations, & Inequalities	4	89.8		12	91.0	
1	Algebra	Variables, Equations, & Inequalities	5	73.9		13	74.1	
1	Algebra	Patterns	6	89.1	74.2	20	87.4	72.8
1	Algebra	Patterns	7	79.8		21	79.1	
1	Algebra	Patterns	8	42.0		22	40.9	
1	Algebra	Patterns	9	85.7		23	83.6	
1	Geometry	Transformational Geometry	10	98.6	81.7	1	98.4	81.7
1	Geometry	Transformational Geometry	11	81.4		2	81.3	
1	Geometry	Transformational Geometry	12	77.3		3	78.6	
1	Geometry	Transformational Geometry	13	69.3		4	68.5	
1	Geometry	Measurement & Estimation	14	77.0	59.5	24	76.3	57.1
1	Geometry	Measurement & Estimation	15	60.0		25	56.8	
1	Geometry	Measurement & Estimation	16	66.4		26	62.0	
1	Geometry	Measurement & Estimation	17	51.1		27	47.7	
1	Geometry	Measurement & Estimation	18	43.2		28	42.6	
1	Geometry	Measurement & Estimation	19	95.0	80.5	5	96.6	81.7
1	Geometry	Measurement & Estimation	20	66.8		6	68.8	
1	Geometry	Measurement & Estimation	21	86.7		7	87.6	
1	Geometry	Measurement & Estimation	22	73.3		8	73.9	
1	Number & Computation	Number Sense	23	86.1	70.8	14	87.4	74.8
1	Number & Computation	Number Sense	24	85.3		15	88.1	
1	Number & Computation	Number Sense	25	71.1		16	77.6	
1	Number & Computation	Number Sense	26	71.2		17	78.1	
1	Number & Computation	Number Sense	27	69.6		18	71.8	
1	Number & Computation	Number Sense	28	41.6		19	45.8	
2	Algebra	Variables, Equations, & Inequalities	1	97.0	76.5	16	95.9	75.7
2	Algebra	Variables, Equations, & Inequalities	2	86.8		17	88.2	
2	Algebra	Variables, Equations, & Inequalities	3	62.3		18	61.5	
2	Algebra	Variables, Equations, & Inequalities	4	77.7		19	77.8	
2	Algebra	Variables, Equations, & Inequalities	5	85.7		20	84.0	
2	Algebra	Variables, Equations, & Inequalities	6	64.5		21	63.5	
2	Algebra	Variables, Equations, & Inequalities	7	86.3		22	84.7	
2	Algebra	Variables, Equations, & Inequalities	8	51.8		23	50.1	
2	Number & Computation	Computation	9	69.0	68.5	1	48.6	69.5
2	Number & Computation	Computation	10	73.8		2	72.4	
2	Number & Computation	Computation	11	50.1		3	74.9	
2	Number & Computation	Computation	12	72.4		4	73.4	
2	Number & Computation	Computation	13	77.1		5	78.1	
2	Algebra	Variables, Equations, & Inequalities	14	86.6	73.4	24	84.5	70.1
2	Algebra	Variables, Equations, & Inequalities	15	86.2		25	83.4	
2	Algebra	Variables, Equations, & Inequalities	16	63.5		26	57.6	
2	Algebra	Variables, Equations, & Inequalities	17	68.9		27	67.2	
2	Algebra	Variables, Equations, & Inequalities	18	78.6		28	76.3	
2	Algebra	Variables, Equations, & Inequalities	19	56.3		29	51.7	
2	Geometry	Measurement & Estimation	20	79.6	71.8	11	79.9	72.2
2	Geometry	Measurement & Estimation	21	79.2		12	80.5	
2	Geometry	Measurement & Estimation	22	86.0		13	87.1	
2	Geometry	Measurement & Estimation	23	51.1		14	49.8	
2	Geometry	Measurement & Estimation	24	62.9		15	63.9	
2	Algebra	Patterns	25	92.1	82.5	6	94.0	84.5
2	Algebra	Patterns	26	85.1		7	86.0	
2	Algebra	Patterns	27	84.0		8	85.7	
2	Algebra	Patterns	28	75.6		9	78.1	
2	Algebra	Patterns	29	75.7		10	78.6	

Part	Standard	Benchmark	Base Form			Scrambled Form		
			Question Number	P-value	Average (P-value)	Question Number	P-value	Average (P-value)
3	Number & Computation	Computation	1	87.4	73.9	8	87.2	73.2
3	Number & Computation	Computation	2	77.8		9	77.1	
3	Number & Computation	Computation	3	81.0		10	82.0	
3	Number & Computation	Computation	4	93.4		11	92.7	
3	Number & Computation	Computation	5	77.7		12	77.7	
3	Number & Computation	Computation	6	54.5		13	54.0	
3	Number & Computation	Computation	7	48.9		14	47.1	
3	Number & Computation	Computation	8	70.1		15	68.0	
3	Geometry	Geometric Figures & Their Properties	9	82.3	71.6	21	80.6	70.0
3	Geometry	Geometric Figures & Their Properties	10	86.9		22	86.1	
3	Geometry	Geometric Figures & Their Properties	11	43.7		23	42.6	
3	Geometry	Geometric Figures & Their Properties	12	70.7		24	69.0	
3	Geometry	Geometric Figures & Their Properties	13	65.4		25	63.7	
3	Geometry	Geometric Figures & Their Properties	14	74.3		26	71.2	
3	Geometry	Geometric Figures & Their Properties	15	77.8		27	76.5	
3	Data	Statistics	16	97.2	75.7	1	97.5	76.8
3	Data	Statistics	17	88.1		2	88.1	
3	Data	Statistics	18	59.4		3	59.6	
3	Data	Statistics	19	46.8		4	48.7	
3	Data	Statistics	20	83.4		5	86.2	
3	Data	Statistics	21	73.3		6	74.7	
3	Data	Statistics	22	81.7		7	82.5	
3	Data	Statistics	23	78.1	65.1	16	78.7	66.0
3	Data	Statistics	24	76.0		17	77.7	
3	Data	Statistics	25	50.7		18	51.9	
3	Data	Statistics	26	61.4		19	63.0	
3	Data	Statistics	27	59.3		20	58.6	

Table 5

Base Form vs. Scrambled Form: DIF Statistics.

Part	Question Number	MH CHI	MH LOR	LOR SE	BD	ETS
1	1	7.877*	0.256	0.090	1.518	A
1	2	1.426	0.065	0.053	0.429	A
1	3	12.763*	0.211	0.059	0.507	A
1	4	3.834	-0.135	0.068	2.063	A
1	5	0.085	-0.015	0.048	0.037	A
1	6	8.415*	0.173	0.059	0.011	A
1	7	1.520	0.068	0.054	0.376	A
1	8	0.542	0.030	0.040	2.534	A
1	9	16.732*	0.243	0.059	0.526	A
1	10	1.186	0.181	0.158	0.340	A
1	11	0.217	0.026	0.052	2.149	A
1	12	3.448	-0.088	0.047	0.002	A
1	13	0.818	0.039	0.042	0.155	A
1	14	0.584	0.035	0.044	2.024	A
1	15	16.826*	0.179	0.044	0.010	A
1	16	30.644*	0.233	0.042	0.011	A
1	17	17.092*	0.183	0.044	2.604	A
1	18	0.127	0.015	0.041	15.913	A
1	19	17.144*	-0.401	0.097	0.350	A
1	20	7.496*	-0.135	0.049	1.076	A
1	21	2.184	-0.094	0.062	0.103	A
1	22	0.603	-0.039	0.048	0.121	A
1	23	4.144	-0.116	0.056	5.959	A
1	24	22.254*	-0.282	0.060	0.287	A
1	25	81.769*	-0.407	0.045	6.879	A
1	26	103.137*	-0.496	0.049	0.962	B
1	27	8.947*	-0.132	0.044	8.220	A
1	28	34.093*	-0.246	0.042	0.000	A
2	1	15.321*	0.413	0.104	3.073	A
2	2	6.059*	-0.142	0.057	0.508	A
2	3	0.701	0.037	0.043	3.688	A
2	4	0.005	-0.004	0.048	1.408	A
2	5	10.031*	0.180	0.056	0.694	A
2	6	1.297	0.046	0.040	0.648	A
2	7	8.242*	0.166	0.057	4.845	A
2	8	2.676	0.069	0.042	0.000	A
2	9	19.342*	-0.189	0.043	0.975	A
2	10	2.341	-0.072	0.047	0.145	A
2	11	2.473	0.066	0.041	1.801	A
2	12	2.225	-0.068	0.045	0.849	A
2	13	2.172	-0.073	0.048	0.760	A
2	14	13.365*	0.204	0.055	1.690	A
2	15	29.004*	0.316	0.058	0.878	A
2	16	53.435*	0.302	0.041	0.534	A
2	17	4.473	0.095	0.045	2.936	A
2	18	13.503*	0.179	0.048	0.785	A
2	19	30.066*	0.223	0.041	2.602	A
2	20	0.465	-0.033	0.047	0.832	A
2	21	3.821	-0.093	0.047	4.431	A
2	22	2.973	-0.099	0.056	0.877	A
2	23	1.712	0.059	0.044	8.389	A
2	24	2.153	-0.065	0.044	5.195	A
2	25	16.673*	-0.305	0.074	0.008	A
2	26	2.419	-0.090	0.057	0.319	A
2	27	6.710*	-0.143	0.055	0.621	A
2	28	11.467*	-0.165	0.048	0.069	A
2	29	17.375*	-0.210	0.050	0.137	A

*Significant at 0.01 alpha level and the corresponding critical value is 6.63
Note: Base Form as the reference group

Part	Question Number	MH CHI	MH LOR	LOR SE	BD	ETS
3	1	0.434	0.039	0.057	0.164	A
3	2	0.793	0.043	0.047	0.173	A
3	3	2.460	-0.080	0.050	1.064	A
3	4	2.337	0.119	0.076	0.813	A
3	5	0.051	0.013	0.050	0.066	A
3	6	0.048	0.010	0.044	0.118	A
3	7	3.559	0.077	0.041	1.821	A
3	8	6.502*	0.110	0.043	0.015	A
3	9	10.001*	0.162	0.051	0.327	A
3	10	3.487	0.113	0.060	0.585	A
3	11	3.104	-0.071	0.040	0.420	A
3	12	4.248	0.088	0.042	0.019	A
3	13	3.388	0.075	0.040	1.866	A
3	14	19.285*	0.200	0.045	0.339	A
3	15	3.201	0.082	0.045	1.605	A
3	16	0.926	-0.121	0.118	4.213	A
3	17	0.010	-0.008	0.058	3.212	A
3	18	0.215	-0.020	0.041	4.065	A
3	19	5.790*	-0.095	0.039	0.002	A
3	20	20.214*	-0.238	0.053	0.865	A
3	21	3.310	-0.083	0.045	1.355	A
3	22	1.772	-0.069	0.051	0.789	A
3	23	0.386	-0.033	0.052	2.426	A
3	24	5.133*	-0.111	0.048	0.541	A
3	25	2.981	-0.071	0.040	0.015	A
3	26	4.052	-0.081	0.040	0.065	A
3	27	0.288	0.022	0.040	0.026	A

*Significant at 0.01 alpha level and the corresponding critical value is 6.63
Note: Base Form as the reference group