DOCUMENT RESUME

ED 322 181 TM 015 334

AUTHOR Gafni, Naomi; Estela, Melamed
TITLE Differential Tendencies To Guess as a Function of Gender and Lingual-Cultural Reference Group.
PUB DATE 88
NOTE 16p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Bound Students; *College Entrance Examinations; Comparative Testing; *Cultural Differences; Foreign Countries; *Guessing (Tests); Higher Education; High Schools; High School Students; Languages; Predictor Variables; *Response Style (Tests); *Sex Differences; Test Format; Testing Problems; *Test Wiseness
IDENTIFIERS Israel

ABSTRACT

The objective of this study was to investigate differential tendencies to avoid guessing as a function of three variables: (1) lingual-cultural-group; (2) gender; and (3) examination year. The Psychometric Entrance Test (PET) for universities in Israel was used, which is administered in Hebrew, Arabic, English, French, Spanish, and Russian. The PET is a battery of five subtests, and encompasses about 200 test items. Three of the five subtests were used in this study: figural reasoning, mathematical reasoning, and English. Data for 12,440 male and 10,532 female examinees were analyzed. The tendency to avoid guessing was measured by the proportion of two types of unanswered items: unreached items, and omitted items. A factor analysis using VARIMAX rotation indicated a strong two-factor structure, in which all indices based on omitted items loaded on the first factor and all indices based on unreached items loaded on he second factor. An analysis of covariance with the corrected-for-guessing scores as a covariate indicated that all three effects and paired interactions were significant. The examination year appeared to have the strongest effect on the proportion of omitted items, while the language-version seemed to affect the proportion of unreached items most strongly. The gender effect was not found to be as strong as were the other two main effects. Three data tables are included. (RLC)

Differential Tendencies to Guess as a Function of Gender and Lingual-Cultural Reference Group

Gafni Naomi, National Institute for Testing & Evaluation, Israel

Melamed Estela, National Institute for Testing & Evaluation, Israel[1]

Mailing Address: Naomi Gafni

National Institute for Testing & Evaluation

4 Rav Ashi St.

Jerusalem, 93593

Israel

Index keywords: guessing, unreached items, omitted items

[1] Authors' names appear in alphabetical order

Abstract

The objective of this study was to investigate differential tendencies to avoid guessing as a function of three variables: lingual-cultural-group, gender, and examination year. The Psychometric Entrance Test (including five subtests) to the universities in Israel, administered in five different languages, was used as the instrument. The tendency to avoid guessing was measured by the proportion of two types of unanswered items: unreached items and omitted items.

A factor analysis with VARIMAX rotation indicated a strong two-factor structure, where all indices based on omitted items loaded on the first factor and all indices based on unreached items loaded on the second factor. ANCOVA with the corrected-for-guessing scores as a covariate indicated that all three effects and paired interactions were significant. Exam year appeared to have the strongest effect on the proportion of omitted items, while language-version seemed to affect the proportion of unreached items most strongly. The gender effect was not found to be as strong as the other two main effects.

The National Institute for Testing and Evaluation (NITE) tests approximately 40,000 applicants for the universities and colleges in Israel every year, by means of the Psychometric Entrance Test (PET). The test is constructed in Hebrew but approximately twelve percent of the applicants are not native Hebrew speakers; as a result, NITE faces the unique challenge of translating the test into several languages and then equating the scores of the translated versions to the original Hebrew form. The largest non-Hebrew-speaking minority taking a translated version of PET is the Arabic-speaking population. Other languages into which PET is translated are: English, French, Spanish, and Russian.

No correction for guessing is used in scoring the exam, and examinees are encouraged to guess when they do not know the correct answer. Yet only 75% to 93% of the examinees (depending on the specific subtest in the PET battery) respond to all the items in the test. Since PET scores are used for admissions decisions, differences in scores that result from differential tendencies to guess, however small or rare, are perceived to be important.

The literature on guessing behavior has chiefly dealt with three issues: (1) the relative advantages and disadvantages of different kinds of test instructions (e.g., Angoff and Schrader, 1981; Cronbach, 1970; Swineford and Miller, 1953); (2) individual differences in guessing as related to personality traits such as "gambling tendency" (Swineford, 1938), cautiousness (Gulliksen, 1950) and risk taking perception (Cohen, 1960); and (3) group differences (e.g., gender differences; Sini, 1984).

Different indices were used to measure guessing (Angoff &

Schrader, 1981; Slakter, 1967; Swineford, 1938; Ziller, 1957) in some of which the proportion of unanswered items was considered an indication of the tendency to avoid guessing. In this study two types of unanswered items were considered: unreached items and omitted items. Unreached items are identified by working from the end of an examinee's response string toward the beginning, taking unanswered items as unreached until an answer is encountered. Unanswered items preceding this last answered item are considered intentionally omitted items (Mislevy, 1988).

The objective of this study was to investigate differential tendencies to avoid guessing (as defined above) as a function of the following factors:

a) Language-group. Since the different language-groups taking PET come from different countries and cultures, it was postulated that they might manifest different guessing behaviors. For example, it was expected that the English-speaking group would be more familiar with multiple-choice tests and, therefore, would be more likely to closely follow the test instructions. On the other hand, the Russian-speaking group, less acquainted with this type of test, might be less inclined to guess.

b) Gender. Previous findings supported the existence of differences in guessing behavior between the two genders (Sini, 1985; Slakter, Koehler, Hampton & Grennel, 1971; Swineford, 1941) and related them to differences between the two groups in personality traits such as tendency for risk taking (Keinan, Meir & Gome-Nemirovsky, 1984; Slakter, 1967).

c) Examination date. Since the first year in which NITE administered

its exams (1984), various evidence had accumulated, suggesting that the population of examinees had become more familiar with the test and was better prepared for it. This increasing familiarity with the test format was considered likely to reduce the effect of personality tendencies, so that current examinees would be more inclined to profit from the test directions and omit a smaller number of items than past examinees.

Method

Instrument

PET is a battery of five subtests, encompassing about 200 test items. Three of the five subtests, which were identical for the different language-versions, were used for this study. There were slight differences in the number of items included in each of the PET forms used in the study. The three subtests were:

1. Figural Reasoning (FO), which contained between 25 to 27 items.
2. Mathematical Reasoning (MA), which contained between 30 to 35 items.
3. English (EN), which contained between 48 to 50 items, was a test of English as a foreign language and included reading comprehension, sentence completion and restatement items.

Each language-version was essentially a translation of the Hebrew version. The FO and EN subtests were identical for all language-versions. The MA subtest was translated and reviewed by bilingual experts. Scores on the different language-versions were equated to those on the Hebrew version using FO and MA as anchor tests. Each sub-

test was scored separately on a scale of mean = 100 and standard deviation = 20. The final PET score was a simple mean of the five subtests, and was reported on a scale with mean = 500 and standard deviation = 100.

Subjects

Table 1 presents the distribution of examinees by language-version and gender, for Form 1 which was administered in all languages in 1984, Form 17 which was administered in Arabic and Hebrew in April 1987, and Form 18 which was administered in five languages (except for Arabic) in April 1987.

[Insert Table 1 about here]

Procedure

Two indices were used: the proportion of unreached items and the proportion of omitted items as defined above. In order to examine the question of whether the two types of unanswered items indicated two separate tendencies, a factor analysis was run on each of the within group correlation matrices of the ten variables (two indices by five subtests, including a General Knowledge - GK and an Analytical Reasoning - RE - subtests, which were not included in the other analyses because they were not identical for all languages).

For each of the six dependent variables (two indices x three subtests) a covariance analysis (using the SAS procedure GLM) was performed with language group, gender, and exam date as independent variables, and with the formula score as a covariate (adjusted for the slightly different number of items in the different test forms). This

score was preferred over the number right score because it was hoped that it would moderate the confounding of number-right score with the proportion of unanswered items (a similar analysis was performed using the scaled number-right score as the covariate, yielded highly similar results).

Results and Discussion

Factor Analysis

The factor analysis (rotated by VARIMAX method) indicated a strong two-factor structure, where all indices based on omitted items were found loaded on the first factor and those indices based on unreached items loaded on the second factor. This structure was evident for all language-versions by gender groups. Table 2 presents the rotated factor pattern matrix for the Hebrew-speaking males examined in Form 17; similar matrices were obtained for all groups. This structure is an indication of two distinct tendencies: one, a tendency to omit items when being uncertain of the correct answer; the other, a disposition to randomly guess at the end of a subtest when unable to try all items, due to a time limit.

[Insert Table 2 about here]

Language-Group Effect

Table 3 presents the means and standard deviations of the proportions of omitted and unreached items on the three subtests, for the six language groups and the two genders, examined in 1984 and 1987.

A language-group effect was found for both types of unanswered

items, and in particular, for the proportion of unreached items. Russian-, Arabic- and French-speaking examinees tended to omit more items than Hebrew-, English- and Spanish-speaking examinees in 1984; in 1987 Russian-speaking examinees tended to omit more items than all other groups. The Russian-speaking group tended not to answer (i.e., did not reach) more items at the end of subtests than the other groups, in both 1984 and 1987. It is worth noting that while for the Hebrew- and Russian-speaking examinees the proportion of unreached items was larger than the proportion of omitted items, for the Arabic-speaking examinees the proportion of omitted items was larger. For the other groups, no such difference was found.

[Insert Table 3 about here]

Gender Effect

A gender effect and an interaction effect of gender with language-version and with examination date were found for all variables, indicating a smaller difference between the two gender groups in 1987. A gender effect was found for Form 17, mainly due to the gender differences found for the Arabic-speaking group, for both types of unanswered items. An interaction effect of language-version and gender was found on this form for the unreached items indices only, where the difference between the two gender groups was larger for the Arabic version. Although significant, this effect was not found to be as strong as the other two main effects (language group and examination date) or as the interaction effects.

Examination Date Effect

The proportion of both types of unanswered items significantly dropped from 1984 to 1987. These results were attributed to an intensive educational operation being implemented among the candidates with respect to test preparation. An interaction effect was found for exam date with language group. While the Arabic-speaking examinees tended to omit and not-reach more items than the Hebrew-speaking examinees in 1984, they tended to answer more items (guess more) than their Hebrew-speaking counterparts in 1987 (the scores of the Arabic-speaking examinees were about one standard deviation lower than those of the Hebrew-speaking counterparts in both years).

Summary

The difference between the tendencies to omit and to "unreach" items has been already recognized: for example, in the context of item parameter estimation (e.g., Mislevy, 1988; Mislevy & Bock, 1986). There is also some preliminary evidence suggesting that the two different types of "missing values" affect the assessment of test dimensionality (Ben-Simon & Cohen, personal communication). The results obtained in this study support the idea that omitting items and not-reaching items are two separate phenomena. They also suggest that guessing behavior can be taught but not entirely eliminated. Exam date appeared to have the strongest effect on the proportion of omitted items, while language-version seemed to affect the proportion of unreached items most strongly. In addition, exam date had a greater effect on the proportion of omitted items, than on the proportion of unreached items, indicating that it is probably easier to train people not to leave an item unanswered if it has already been tried, than to train them to

randomly guess untried items at the end of the test.

People with differing cultural backgrounds, as well as the two genders, differ in their tendency to guess. It is probable that some of the lower scores of certain groups on multiple choice tests can be partially explained by their tendency to avoid guessing, as are some of the differences in performance among the language groups. It is recommended to emphasize the importance of test directions, in particular, among members of groups known to avoid guessing.

References

Angoff, W. A., & Schrader, W. B. (1981). A study of alternative methods for equating right scores to formula scores (ETS RR-81-8). Princeton NJ: Educational Testing Service.

Cohen, J. (1960). Chance, Skill and Luck. Baltimore, MD: Penguin.

Cronbach, L. J. (1970). Essentials of Psychological Testing (3rd ed.). New-York: Harper & Row.

Gulliksen, H. (1950). Theory of Mental Tests. New York: John Wiley & sons, Inc.

Keinan, G., Meir, E. & Gome-Nemirovsky, T. (1984). Measurement of risk-takers personality. Psychological Reports, 55, 163-167.

Mislevy, R. J. (1988). Inferring examinee ability when some item responses are missing (ETS RR-88-48). Princeton NJ: Educational Testing Service.

Mislevy, R. J. & Bock, R. D. (1986). PC-BILOG - Item Analysis and Test Scoring with Binary Logistic Models. Mooresville IN: Scientific Software, Inc.

Sini, J. (1985). Gender differences in guessing tendencies in multiple choice tests (submitted as partial fulfillment of requirements for a degree of Master of Arts - in Hebrew). Jerusalem: Hebrew University.

Slakter, M. J. (1967). Risk taking on objective examinations. American Educational Research Journal, 4, 31-43.

Slakter, M. J., Koehler, R. A., Hampton, S. H., and Grennell, R. L. (1971). Sex, grade level and risk taking on objective examinations. Journal of Experimental Education, 39, 65-68.

Swineford, F. (1938). The measurement of a personality trait. Journal of Educational Psychology, 29, 295-300.

Swineford, F. (1941). Analysis of a personality trait. Journal of Educational Psychology, 32, 438-444.

Swineford, F. & Miller, P. M. (1953). Effects of directions regarding guessing on item statistics of a multiple choice vocabulary test. Journal of Educational Psychology, 44, 129-140.

Ziller, R. C. (1957). A measure of the gambling response-set in objective tests. Psychometrika, 22, 289-292.

Table 1

Frequencies of Examinees by Version and Gender

	Test Form					
	Form 1		Form 17		Form 18	
Language	Males	Females	Males	Females	Males	Females
Hebrew	520	467	3786	3815	2810	2716
Arabic	1876	839	2383	1460		
English	121	225			154	183
French	82	132			94	137
Spanish	123	146			367	304
Russian	63	52			61	56

Table 2

Rotated Factor Pattern Matrix for Proportion of Omitted (denoted by "A" suffix) and Unreached Items (denoted by "R" suffix) for Five Subtests in Form 17 for Male Hebrew Examinees

Subtest	Factor 1	Factor 2
RE-R	.754	.190
MA-R	.707	.103
FO-R	.643	.209
EN-R	.625	.090
GK-R	.426	.063
MA-A	.095	.764
FO-A	.094	.602
RE-A	.080	.589
EN-A	.117	.471
GK-A	.171	.451

Table 3

Means and Standard Deviations of the Proportion of Omitted and Unreached Items (in percentages) for the Three Subtests Separately and Together for 1984 Examinees (Form 1) and for 1987 Examinees (Forms 17 and 18)

		Language Version																	
		Hebrew				Arabic				English				French				Spanish				Russian			
		Males		Females		Males		Females		Males		Females		Males		Females		Males		Females		Males		Females	
Variable	Form	X	SD	X	SD	X	SD	X	SD	X	SD	X	SD	X	SD	X	SD	X	SD	X	SD	X	SD	X	SD
FO-A*	1	0.9	(3.2)	1.6	(4.3)	1.8	(4.9)	3.7	(7.6)	1.0	(3.6)	1.1	(3.5)	1.6	(4.1)	1.7	(4.0)	0.7	(1.9)	2.1	(4.1)	0.8	(2.4)	1.1	(2.9)
	18	0.7	(2.5)	1.1	(3.7)					0.6	(2.4)	0.9	(3.0)	0.3	(1.1)	0.7	(2.5)	0.5	(1.7)	0.7	(2.6)	0.7	(2.0)	1.0	(3.3)
	17	0.7	(2.5)	1.1	(3.5)	0.7	(2.6)	0.9	(3.7)																
MA-A	1	1.1	(4.4)	3.1	(7.5)	2.6	(6.3)	5.8	(10.2)	1.8	(5.2)	2.4	(5.8)	2.9	(7.6)	4.4	(8.7)	1.2	(2.8)	4.0	(7.1)	2.3	(5.9)	4.6	(9.6)
	18	0.9	(3.0)	1.8	(5.2)					0.5	(1.9)	1.1	(4.4)	0.6	(2.1)	0.9	(4.9)	0.5	(1.6)	1.7	(5.0)	1.1	(2.7)	1.3	(4.6)
	17	0.9	(3.0)	1.8	(5.0)	0.8	(2.9)	1.4	(4.3)																
EN-A	1	0.7	(2.8)	1.8	(5.4)	1.7	(5.1)	3.9	(8.1)	0.3	(1.3)	0.2	(1.2)	1.3	(3.4)	0.6	(1.7)	0.8	(2.5)	0.9	(1.9)	2.9	(8.3)	2.3	(6.1)
	18	0.7	(2.9)	1.4	(4.8)					0.1	(0.3)	0.1	(0.8)	0.9	(4.1)	0.8	(3.6)	0.5	(2.8)	0.8	(2.7)	1.7	(4.9)	1.9	(6.2)
	17	0.7	(3.0)	1.3	(4.4)	0.7	(2.5)	1.4	(5.0)																
Total-A	1	0.9	(2.8)	2.1	(4.8)	2.0	(4.4)	4.4	(7.4)	0.9	(2.2)	1.0	(2.5)	1.8	(3.9)	1.9	(3.3)	0.9	(1.8)	2.1	(3.1)	2.2	(5.7)	2.7	(5.3)
	18	0.8	(2.1)	1.5	(3.7)					0.3	(1.1)	0.6	(1.9)	0.7	(2.3)	0.8	(3.3)	0.5	(1.6)	1.0	(2.7)	1.3	(2.8)	1.5	(4.5)
	17	0.8	(2.2)	1.4	(3.4)	0.7	(2.0)	1.3	(3.7)																
FO-R**	1	2.2	(6.4)	3.8	(8.7)	0.9	(4.0)	1.2	(4.5)	2.5	(6.9)	2.5	(7.1)	1.9	(6.1)	3.4	(8.4)	1.0	(5.0)	1.8	(5.6)	3.7	(10.7)	7.1	(13.7)
	18	1.8	(6.5)	2.5	(7.5)					1.6	(6.2)	2.2	(7.0)	0.7	(4.7)	0.9	(4.8)	1.0	(5.1)	1.4	(5.4)	3.3	(9.7)	4.0	(9.8)
	17	1.6	(5.7)	3.1	(3.5)	0.5	(3.4)	0.6	(2.8)																
MA-R	1	1.9	(7.0)	4.1	(10.1)	1.1	(4.9)	1.8	(6.6)	1.7	(5.1)	2.4	(7.1)	1.3	(6.7)	4.9	(11.5)	0.3	(3.6)	2.1	(7.7)	3.2	(9.8)	7.8	(14.7)
	18	0.9	(4.4)	1.5	(5.9)					0.6	(3.7)	0.6	(3.0)	0.5	(3.1)	0.8	(3.7)	0.5	(3.4)	1.3	(4.9)	2.4	(7.8)	3.7	(9.8)
	17	0.9	(4.4)	2.3	(7.5)	0.5	(3.2)	0.8	(3.5)																
EN-R	1	2.2	(7.2)	5.0	(10.7)	1.6	(6.6)	4.3	(10.7)	0.9	(5.5)	0.2	(2.0)	2.2	(6.1)	2.6	(6.9)	1.0	(5.1)	1.6	(5.1)	4.8	(12.2)	9.5	(14.4)
	18	1.7	(6.8)	2.7	(8.5)					0.1	(1.5)	0.0	(0.1)	0.3	(2.5)	0.7	(4.4)	0.6	(5.3)	1.2	(5.7)	2.9	(8.6)	5.9	(14.7)
	17	1.2	(5.5)	2.7	(8.7)	0.6	(4.1)	1.1	(4.8)																
Total-R	1	2.1	(6.0)	4.5	(8.7)	1.3	(4.4)	2.8	(6.8)	1.5	(4.2)	1.4	(3.5)	1.9	(4.7)	3.1	(7.2)	0.8	(3.9)	1.8	(4.9)	4.1	(10.2)	8.4	(11.1)
	18	1.4	(4.8)	2.3	(6.1)					0.7	(2.5)	0.7	(2.1)	0.5	(3.1)	0.8	(3.1)	0.7	(3.3)	1.3	(4.2)	2.9	(5.9)	4.7	(9.3)
	17	1.2	(4.2)	2.7	(6.9)	0.5	(3.1)	0.9	(3.0)																

* "A" denotes the proportion of omitted items

** "R" denotes the proportion of unreached items