Sex Roles, Vol. 32, Nos. 7/8, 1995

Cognitive Gender Differences Among Israeli Children[1]

Sorel Cahan[2] and Yael Ganor
The Hebrew University of Jerusalem

The study investigated gender differences among 11,000 Israeli children in Grades 4–6 with respect to verbal, spatial, and mathematical ability, as measured by 12 intelligence tests. Consistent differences in score variance were found across grades for 11 of the 12 tests. In each of these tests the variance for boys exceeded that for girls by 10%–20%. With respect to mean achievement, consistent cross-grade differences were found only for mathematical ability, where boys had the edge (about 0.20 SD). These findings diverge from those of recent American studies, which found no gender differences in any of these realms. Furthermore, they differ from the results of earlier Israeli studies in that the gender gap is limited to mathematical ability, and its size is much smaller. The revealed gender gap can be partially attributed to differences in response strategy: girls were found to be more likely to skip items for which they lack an answer (i.e., to take fewer risks in guessing). This implies that the performance of girls on intelligence tests will improve if they are encouraged to dare to guess.

Recent meta-analyses clearly indicate a continuous and considerable reduction in gender differences with respect to all cognitive areas among non-selective populations. Gender differences have virtually disappeared in the verbal realm (Feingold, 1988, 1992; Hyde & Linn, 1988; Linn & Hyde, 1989), and they are slight for most spatial tasks, although boys continue to

[1]This research was supported by the Henrietta Szold Research Institute. We thank Nora Cohen for providing the data base; Lavee Artman, Kalman Binyamini, Nora Cohen, Eyal Gamliel, and two anonymous reviewers for their helpful comments on previous drafts; and Helene Hogri for the editorial revisions.
[2]To whom correspondence should be addressed at School of Education, The Hebrew University, Jerusalem, 91905, Israel.

hold a clear advantage with respect to mental rotation (Linn & Hyde, 1989; Linn & Petersen, 1985). Even mathematical ability, long a province of males, has come to be relatively evenly distributed between boys and girls in non-selective elementary-school populations (Friedman, 1989; Hyde, Fennema & Lamon, 1990). Differences (in favor of boys) appear only in secondary school with respect to mathematical problem solving, and they are largest (about 0.5 SD) in the quantitative component of the Scholastic Aptitude Test (SAT). It should be kept in mind that SAT results are based on the selective population of college aspirants (Cole, 1990; Halpern, 1989; Hyde, Fennema & Lamon, 1990).

The above conclusions are not necessarily generalizable. First, they are based on weighted averages that may not reflect typical findings. For example, the 254 effect sizes included in Hyde, Fennema and Lamon's (1990) meta-analysis of gender differences in mathematics performance ranged between −0.89 and 0.88, with a mean of −0.05. Fifty-one percent of the effect sizes were positive and 43% were negative. It should be stressed that this considerable variation is not entirely attributable to random error. The homogeneity analyses performed by Hyde and colleagues (1990) indicated that the set of 254 effect sizes was heterogeneous in terms of true effect size. Therefore, its representation by means of a single average value is unwarranted and may be misleading.

Second, most if not all the studies included in the meta-analyses were based on North American samples. Yet gender differences in intelligence have been found to vary in both direction and size from one population to another (Born, Bleichrodt & Van Der Flier, 1987; Feingold, 1992; Linn & Hyde, 1989; Maccoby & Jacklin, 1974; Zeidner, 1986). There are, therefore, good grounds to examine the question in other populations and cultures (Feingold, 1992). This is particularly true with respect to gender differences in variability, which are inconsistent across countries: the well established U.S. findings of greater male variability in mathematical and spatial abilities do not always hold true in other cultures and nations (Feingold, 1994).

The present study investigates gender differences among elementary school children in Israel. Existing research on gender differences in Israeli society is scarce and focuses only on mean cognitive ability. Moreover, most of these studies were carried out on voluntary subpopulations, mainly university candidates. For instance, Nevo (1986) and Zeidner (1986) both found a stable advantage for males on the results of university entrance examinations. However, the selective nature of these groups precludes generalization to the Israeli population at large (see also Cole, 1990; Linn & Hyde, 1989). Indeed, Nevo's and Zeidner's findings may be attributed to the considerably higher proportion of female college applicants (the database of Zeidner, 1986, includes 690 men and 1,088 women).

A notable exception is Leiblich, Ninio & Ben Shachar-Segev's (1976) study of cognitive gender differences among 1100 children aged 6-16 (100 per age bracket), included in the 1973 standardization sample of the Hebrew version of the Wechsler Intelligence Scale for Children—Revised (WISC-R). According to this study, boys begin to have an advantage over girls at about age 9. Significant differences appear initially at age 11 in the area of verbal intelligence. By age 13, significant differences appear in almost all the subtests and all three IQ scores, and by the age of 16, the gender gap reaches 0.75 SD (Lieblich, 1983; Gafni, 1978).

There are several grounds for reexamining the Israeli situation. First, the Leiblich et al. study is rather outdated, and there is reason to believe that the gender gap may have changed over time. Second, the sample of that study is rather small; a broader one would allow for a more reliable quantitative estimation of gender differences. Finally, the earlier research did not address the issue of gender differences in variability.

The present study examines cognitive gender differences in Israel, taking advantage of a large existing database of scores on 12 intelligence tests administered to 11,000 fourth, fifth and sixth graders. An attempt was also made to examine specific explanations for gender differences stemming from differential test-taking strategies. Recent studies in both the U.S. (Doolittle & Cleary, 1987; Olson & Fennesy, 1990) and Israel (Ben-Shakhar & Sinai, 1991; Gafni & Melamed, 1990; Zohar, 1990) have shown that, on the average, females tend to guess less than males.

METHOD

This study was based on an existing database (the product of research which did not examine gender differences). Below we concisely outline the sample, tests and procedures used in the construction of this database (for additional information, see Cahan & Cohen, 1989; Cohen, in preparation).

Subjects

The sample included all fourth, fifth and sixth graders (aged 9-12) in Jerusalem's Hebrew-language (i.e., Jewish) secular and religious State schools during the 1986/87 school year, with a few exceptions: one school did not participate; children who immigrated to Israel after 1984 (about 150 children) were excluded, since the tests require a working knowledge of Hebrew; and about 1000 children were absent from school on the day their classes were tested. In all, 11,000 children from 61 schools (about 90% of the total target population) actually participated. At testing (the end of the school year) the students' mean ages were 10, 11 and 12 years for the 4th, 5th, and 6th grade, respectively.

It should be noted that grades 4-6 are noted for a particularly high rate of regular attendance and few selective processes (such as dropping out or movement to alternative frameworks). As a result, there is a maximal overlap between the *pupil* population and the population of *children* born in the same year.

Tests

Twelve tests were administered, covering a wide range of item content (e.g., classification, analogies, series) and varying in item modality (verbal, numerical, figural). Most of these were adapted from several widely used test batteries, namely, Thorndike & Hagen's (1971) Cognitive Abilities Test; Raven, Court & Raven's (1975) Standard Progressive Matrices; and Cattell & Cattell's (1965) Culture Fair Intelligence Test. In addition, some tests were taken from Ortar & Shachor's (1980) Hebrew-language intelligence test (Milta), widely used in Israel. All tests consisted of multiple-choice items only. There was a time limit for each test, ranging between three and five minutes. (For more details, see Table I in the Results section and the Appendix.)

Procedure

The tests were administered May 8-June 8, 1987. A few days prior to testing, subjects were given general explanations about the nature and purpose of the research. They practiced answering a few sample test items and filling out the answer sheets. The actual tests were administered at the beginning of the school day and lasted about 3 hours, with a break of about half an hour. Two examiners were present in each classroom, as was the homeroom teacher, who helped maintain discipline. The examiners were students at the Hebrew University of Jerusalem who had been trained in administration of the tests. No instructions concerning guessing were given.

Scoring

Two scores were calculated for each examinee on each of the twelve tests:

1 *Total score*: the number of correct answers.

Table I. Cognitive Gender Differences (Boys-Girls) by Test and Grade (in Pooled Within-Gender Standard Deviation Units)

Test	Grade 4	Grade 5	Grade 6	Cross-grade Mean
Verbal ability				
Classification[a] (16)[e]	–.05	–.07	–.03	–.05
Analogies[a] (21)	–.05	–.10	–.03	–.06
Vocabulary[b] (40)	.05	.02	.08	.05
Oddities[b] (19)	.17	.13	.19	.16
Sentence Completion[b] (28)	–.09	–.05	–.03	–.06
Spatial ability				
Classification[a] (18)	–.07	–.08	–.02	–.06
Analogies[a] (20)	–.03	–.08	–.02	–.04
Matrices[c] (16)	.06	.10	.09	.08
Series[d] (12)	–.04	–.05	–.07	–.05
Oddities[d] (14)	–.08	–.02	.04	[f]
Mathematical ability				
Series[b] (18)	.10	.15	.17	.14
Word Problems[b] (18)	.18	.25	.24	.23

[a]Cognitive Abilities Test, Levels A-H, Form 1 (Thorndike & Hagen, 1971).
[b]Milta, Elementary Level, Grades 4-6 (Ortar & Shachor, 1980).
[c]Standard Progressive Matrices (Raven, Court & Raven, 1975).
[d]Culture Fair Intelligence Test, Scale 2, Form A (Cattell & Cattell, 1965).
[e]Number of test items.
[f]Lack of cross-grade consistency in the direction of the difference.

2. *Omission score*: The percentage of questions skipped between the first question and the last one that was answered.

Gender differences were calculated for each score. To permit comparison between tests and between grades, differences were standardized using the pooled within-gender standard deviation of the scores in each grade.

Treatment of Random Error

Observed gender differences are not in themselves a basis for valid inferences about true gender differences, as they may reflect random within-grade differences. Hence, we did not consider an observed gender difference to be true unless the direction of the difference was consistent across the three grades. Only then did we estimate the size of the gap by the cross-grade mean difference. Moreover, we considered such a difference to be noticeable (i.e., of non-negligible size) only if it was greater than 1/10 SD. This procedure was performed separately for each of the 12 subtests. Underlying this approach was the conservative, yet plausible assumption that the direction of gender differences would hold constant over these three consecutive grade levels.

RESULTS

Mean gender differences in total scores per grade and subtest appear in Table I (see also Fig. 1). All findings are based on the 10,789 examinees who took at least 10 of the 12 tests: 3,393 fourth graders, 3,829 fifth graders and 3,567 six graders. In each of the tests, except for Figural Oddities, gender differences were consistent in direction across the three grades. Noticeable differences (greater than .10 SD) appear in favor of boys for both tests of mathematical ability and in the Verbal Oddities test. No noticeable

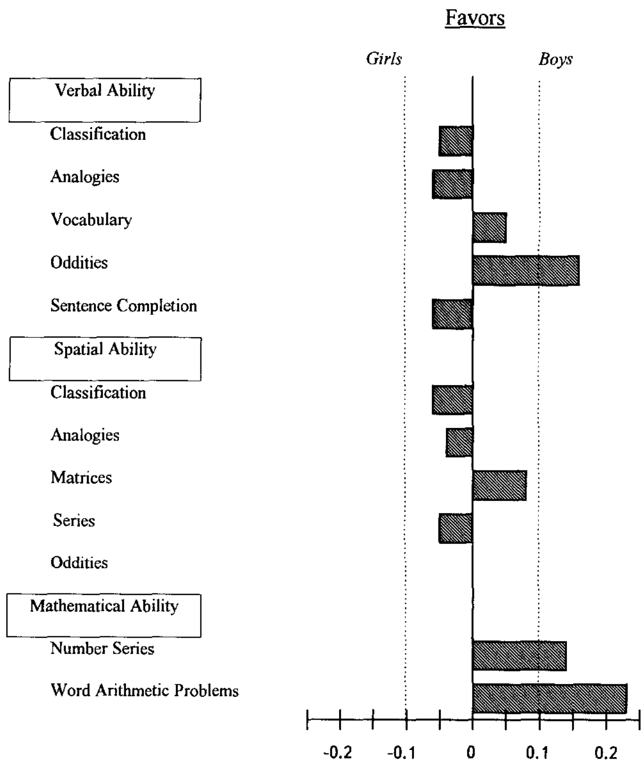


Fig. 1. Cognitive gender differences: Cross-grade means (in standard deviation units)

Table II. Boy-Girl Ratio at Top and Bottom 5% of Test-Score Distributions and Total Variance (Cross-Grade Means)

Test	Boy-girl variance ratio[a]	Ratio[b] Bottom 5%	Top 5%
Verbal ability			
Classification	1.2	1.5	1.2
Analogies	1.1	1.4	1.1
Vocabulary	1.1	1.2	1.4
Oddities	1.1	1.0	1.2
Sentence Completion	1.2	1.5	1.3
Spatial ability			
Classification	1.0	1.1	0.9
Analogies	1.1	1.3	1.3
Matrices	1.1	1.2	1.4
Series	1.2	1.5	1.1
Oddities	1.2	1.5	1.2
Mathematical ability			
Series	1.3	1.3	2.4
Word Problems	1.5	1.0	2.4

[a]Values greater than 1 denote greater variance among boys.
[b]Values greater than 1 denote greater proportions of boys.

differences were found for any of the spatial ability tests or for the large majority of verbal tests. It should be noted, moreover, that the difference found for the Verbal Oddities test derives mainly from the item on types of vehicles, a subject matter which apparently evokes more interest among boys. The omission of this item resulted in a 0.12 drop in the cross-grade gender difference on this subtest, from 0.16 to 0.04.

We also examined the ratio between variances of the total score for boys and girls (see the first column in Table II). Cross-grade consistency in the direction of gender differences was found for 11 tests. In each of these tests, the score variance for boys was larger or equal to that for girls. On the whole, it was greater by about 10%-20%.

The greater variance among boys is reflected in their over-representation at both the upper and lower extremes (top and bottom 5%) of the score distribution for almost all 12 subtests (see Table II). This holds true regardless of which gender has the higher mean score. The proportion of boys in the top 5% is especially high in the two mathematical tests, in which, it will be recalled, they had considerably higher means. At the bottom end of the mathematical score distribution, gender differences in means and in variance worked in opposite directions, and apparently the

Table III. Gender Differences (Boys-Girls) in Omission Scores: Cross-Grade Means (in Pooled Within-Gender Standard Deviation Units)

Test	Difference[a]
Verbal ability	
Classification	[b]
Analogies	-0.05
Vocabulary	-0.13
Oddities	-0.10
Sentence Completion	-0.05
Spatial ability	
Classification	-0.10
Analogies	-0.05
Matrices	-0.10
Series	-0.04
Oddities	-0.14
Mathematical ability	
Series	-0.04
Word Problems	-0.17

[a]Negative differences denote more skipped items for girls.
[b]Lack of cross-grade consistency in the direction of the difference.

latter had the stronger effect. That is, the greater variance among boys cancelled out the effect of their higher mean scores. Thus, it was they, rather than the girls, who dominated the lower extreme.

Gender differences with respect to omission scores are presented in Table III. A consistent cross-grade trend was revealed: in all tests but Verbal Classification, girls skipped more items than boys. Differences were noticeable (greater than 0.10 SD) in six of the tests—Vocabulary, Verbal Oddities, Figure Classification, Matrices, Figural Oddities and Word Arithmetic Problems. In all these tests (except for Figural Oddities, in which no consistent cross-grade differences were found), boys did better on the total score (see Fig. 1).

DISCUSSION

The main contribution of the present study is the reliable quantitative estimate of gender differences among Israeli children with respect to mean achievement and variability in various realms of cognitive ability. Regarding mean achievement, our study points to consistent gender differences in mathematical ability at a magnitude of about .20 SD. While a difference of this size might be considered small (Cohen, 1969), it can nevertheless hold considerable social importance (Hyde, 1981; Hyde, Fennema, & Lamon, 1990; Rosenthal & Rubin, 1982), particularly because it is often expressed by a disproportionate representation of the genders at the extremes of the ability distribution. Indeed, much recent research has focused on gender differences in these special subpopulations, especially the over-representation of boys at the upper extreme (Benbow & Stanley, 1980, 1982, 1983; Cole, 1990; Feingold, 1988; Hyde, Fennema & Lamon, 1990; Jacklin, 1989). The findings of the present research point in this direction as well. It is interesting to note that over-representation at the top was not accompanied by under-representation at the bottom, owing to the greater variance among boys.

Our finding of gender differences in mean mathematical ability differs from that reported by Hyde, Fennema and Lamon (1990), which is based on a meta-analysis of 21 studies of mathematical problem solving at ages 11-14: they found a very slight mean advantage for girls (.02 SD). Thus, it seems that, in contrast to the United States, where the gender gap in math seems to have disappeared, there is still a nonnegligible advantage for boys among the Jewish population in Israel.

The contrast between our findings and the American ones can be interpreted in two ways: as artifactual or as a reflection of true differences. In keeping with the former interpretation, disparate results may derive from a possible difference in the nature of the tests involved. Recent American tests (like the WISC III and K-ABC) have been constructed with the explicit aim of minimizing gender differences by deleting those items found to obtain the largest gap (see also Feingold, 1992). It is possible that the test items used to glean mathematical ability in the present study were not as gender-free. Following this explanation, then, the inconsistency between the Israeli and American findings stems, at least partially, from differences in the gender bias of the tests—either in favor of boys (Israeli tests) or in favor of girls (American ones), as the policy of eliminating gender-related items may actually introduce bias in the opposite direction. (Since true gender differences are unknown, there is no way to gauge whether the elimination of such items really reduces gender bias.)

Another possible reason for an artifactual difference between the Israeli and American results is between-country variability in gender differences with regard to test-taking strategy, particularly guessing. Because more guessing necessarily results in higher mean test scores, gender differences in mean achievement can reflect, to an unknown extent, differences in the degree of guessing. Thus, the higher mean scores obtained by boys on the mathematical tests in this study may reflect their greater tendency to guess. Accordingly, the larger gender gap found here may reflect more substantial gender differences in guessing among Israeli children. While guessing cannot usually be directly measured, the results of this study concerning item skipping indirectly support such an explanation: uniformly, across tests and grades, boys tended to skip fewer items than girls. While this could be attributed to ability differences, under the assumption that examinees omit items they do not know, it can also be ascribed to differential tendencies to guess. This interpretation is supported by previous Israeli findings concerning gender differences in omission (e.g., Ben-Shakhar & Sinai, 1991; Gafni & Melamed, 1990; Zohar, 1990).

A second interpretation of the disparity between our findings and the American ones assumes that both sets of results are valid and that the difference between them is a real one. Obviously, there are many possible reasons for this difference. We shall focus on one: the cultural gap between North America and Israel regarding the issue of gender equality, expectations from boys and girls, and the opportunities offered to each gender, as reflected in the attitudes of relevant agencies (e.g., family, school, media). Thus, the gender gap in math found in this study may reflect a more traditional—i.e., differential, sexist and stereotyped—attitude of Israeli society. This possibility is especially plausible in view of the demographic composition of the Jewish population in Israel and particularly in Jerusalem: it includes a high percentage (about 50%) of students whose families immigrated from Arab countries in North Africa and the Middle East, as well as a relatively large proportion of children who attend religious schools (the two characteristics overlap considerably). It is to be expected that such differential expectations (and treatment) of the two genders would be particularly salient in the realm of mathematics. If this explanation is true, then modernization should lead to a progressive decrease in the math gender gap in Israel. The smaller size of gender differences found in this study, as compared to Lieblich's (1983) results of fifteen years earlier, supports this prediction.

In contrast to the realm of mathematics, we found a virtual lack of gender differences in verbal and spatial ability. In fact, given the boys' higher tendency to guess, the results for some subtests can be interpreted as indicating a slight advantage for girls. In this respect, the present findings support recent studies performed in other countries, mainly the United States, which pointed to a clear tendency towards the disappearance of the gender gap (Feingold, 1988, 1992; Hyde & Linn, 1988; Linn & Hyde, 1989; Linn & Petersen, 1985). On the other hand, they are in sharp contrast to the previous results of Lieblich (1983), which may point to a substantial narrowing of the gender gap in Israel since the early seventies. An alternative explanation lies in the difference between the tests used in each study: the verbal scale of the WISC-R test (used by Lieblich) contains tests of general knowledge, arithmetic and social understanding, which require abilities beyond verbal ones. With regard to the spatial realm, all the tests included in our study (unlike the corresponding WISC-R subtests, such as "Block Design") require finding some rule and thus permit verbal mediation. The results of our research may therefore be specific to tasks requiring complex manipulations of visually presented two-dimensional information, whose solution may be achieved by multiple strategies, including verbal ones (Halpern, 1989; Linn & Petersen, 1985). This argument receives some support from earlier findings that boys did better than girls on three-dimensional rotations, while no significant differences were found for two-dimensional rotations, in which it is easier to employ verbal mediation (Linn & Petersen, 1985; Zohar, 1990). A more unequivocal answer can be obtained by examining gender differences on the latest Hebrew version of the WISC-R (WISC-R95; Cahan, Hazani, Wolf, Peyser & Shimborsky, in press), a task that is currently under way.

With respect to score variance, consistent gender differences were found across grades for 11 of the 12 subtests. In each of these tests the variance for boys exceeded that for girls by 10%-20%. These results are in accord with Maccoby and Jacklin's (1974) conclusions, as well as with the more recent results reported by Feingold (1992), based on the national norms of several aptitude tests in the U.S. They provide support for the hypothesis of greater male variability in another culture, at a relatively early age and using different tests. Moreover, the boys' greater variability in our study is unqualified by domain: it characterizes *all* tests, including the verbal ones. It is worth noting that the gender differences in variability found in this study, unlike others (see Feingold, 1992), cannot be attributed to differential drop out of boys and girls. As noted earlier, grades 4 through 6 in Israel are characterized by universal school attendance.

This is not to say that the gender differences in test score variability found in this study, as well as others, are necessarily representative of "true" differences in *intelligence* variability. Rather, they may result from gender differences in other factors affecting test behavior and test scores, such as guessing strategies. Guessing introduces random error and, therefore, artifactually increases the variance of *observed scores*. Hence, gender differences in test score variability may reflect, to an unknown extent, gender differences in guessing rather than true differences in the variability of cognitive ability. The plausibility of this interpretation is increased by the present finding that girls have a greater tendency to omit items, not only on mathematical tests (in which a noticeable difference was found in the level of the two genders), but rather, to various degrees, on all the subtests.

Thus, not only does the present study provide a reliable estimation of gender differences in mean cognitive ability and variability among Israeli school age children, but it also offers an empirical examination of gender differences in test-taking style, along with their linkage to gender differences in mean achievement and variability. Differential guessing tendencies provide an interesting alternative hypothesis to gender and cross-cultural differences. In addition, the girls' lower tendency to guess suggests that interventions aimed at raising their scores in intelligence tests ought to relate to their test-taking style and strengthen their belief in their capacities. In other words, it is likely that the performance of girls will improve if they are encouraged to dare to guess.

Before closing, it should be noted that the present findings refer to a specific age range (9-12) and may not be generalizable to older or younger children. Further research is needed to consider that question. Research on adolescents would be particularly welcome in light of the dramatic changes characteristic of the teenage years, as well as the fact that most gender differences revealed in American studies were found in this or older age groups.

APPENDIX

Description of the 12 Tests

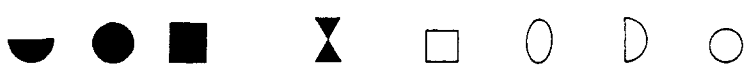
1. *Verbal Classification* (Adapted from the Cognitive Abilities Tests, Levels A-E)
   Each item presents 3 to 4 words belonging to a specific group. Respondents are required to identify what the words have in common (i.e., what type of group they belong to) and to select the word that belongs to the same group from the list of alternatives.

Example: Cat mouse bear

1. rose 2. lion 3. run 4. hungry 5. brown

2. *Figure Classification* (from the Cognitive Abilities Tests, Levels A-E) The task is the same as in the Verbal Classification Test, but involves figures instead of words.

Example: 1 2 3 4 5



3. *Verbal Analogies* (adapted from the Cognitive Abilities Tests, Levels A-E)
   Respondents are presented with one pair of words and one single word. After identifying the relationship between the given pair,

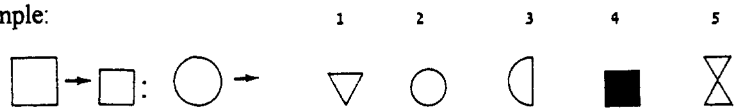
they are required to choose the word in the list of alternatives whose relationship with the single word is the same as the relationship between the words in the pair.

Example: sock ----> foot glove -----> ?

1. head 2. jacket 3. hand 4. wool 5. winter

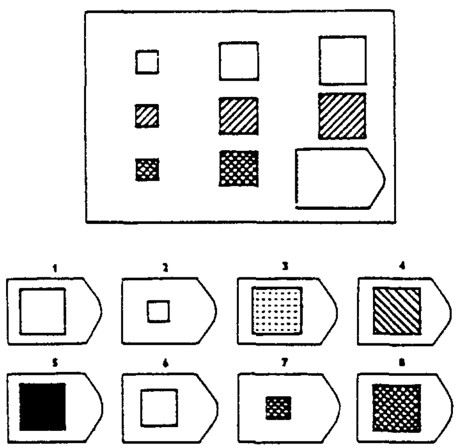
4. *Figure Analogies* (from the Cognitive Abilities Tests, Levels A-E) The task is the same as in the Verbal Analogies test, but involves figures instead of words.

Example:



5. *Progressive Matrices* (from Parts C and D of the Standard Progressive Matrices Test)
Each item in this test contains 8 figures arranged in three rows and three columns. Each row and each column constitute a series. One figure is missing. Respondents are required to identify the pattern of the series, and to choose the missing figure from the list of alternatives.

Example:



6. *Vocabulary* (from the MILTA Test)
A word is presented in each item, and respondents are required to choose a synonym of that word from the list of alternatives.

Example: Path: 1. sign 2. mountain 3. road

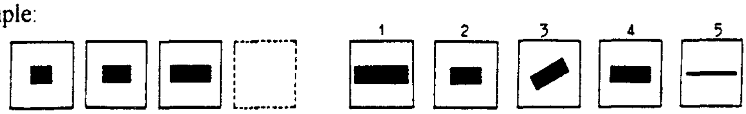
7. *Number Series* (from the MILTA Test)
Respondents are presented with a series of numbers. They are re-

quired to identify the pattern of the series and to add an appropriate number to it from the list of alternatives.

Example: _ 7 9 11 13 15 6 8 3 2 5
(1) (2) (3) (4) (5)

8. *Figure Series* (taken from the Culture Fair Intelligence Test, Scale 2, Form A)
The task is the same as in the Number Series test, except that it refers to figures that vary according to a set pattern.

Example:

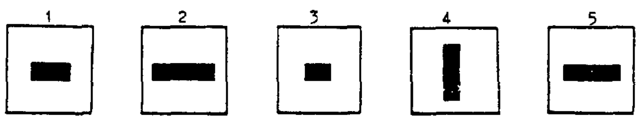


9. *Verbal Oddities* (from the MILTA test)
Five words are presented in each item. Four of these have common characteristics; one does not belong. Respondents are required to flag the word that is an exception to the group.

Example: 1. shirt 2. jacket 3. undershirt 4. handkerchief 5. pants

10. *Figural Oddities* (from the Culture Fair Intelligence Test, Scale 2, Form A; originally titled "Classifications")

Example:



11. *Word Arithmetic Problems* (from the MILTA test)
Respondents are required to solve word problems involving: addition, subtraction, multiplication, division, fractions, distance problems, capacity, and area.

Example:

Mother gave 9 candies to each child. If there were 8 children, how many candies did she give out altogether?

1. 63 2. 54 3. 72 4. 81 5. 98

12. *Sentence Completion* (from the MILTA test)
Every item consists of a sentence with one or two words missing. Respondents are required to complete the sentence with the most appropriate word or word combination appearing in the list of alternatives.

Example:

_______________ people walk on sidewalks in the city.

1. Lonely 2. Careful 3. Nice 4. Tired 5. Young

REFERENCES

Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematics ability: Fact or artifact? *Science, 210*, 1262-1264.

Benbow, C. P., and Stanley, J. C. (1982). Consequences in high school and college of sex differences in mathematical reasoning ability: A longitudinal perspective. *American Educational Research Journal, 19*, 598-622.

Benbow, C. P., and Stanley, J. C. (1983). Sex differences in mathematical reasoning ability: More facts. *Science, 222*, 1029-1031.

Ben-Shakhar, G., & Sinai, Y. (1991). Gender differences in multiple-choice tests: The role of differential guessing tendencies. *Journal of Educational Measurement, 28*, 23-35.

Born, M. P., Bleichrodt, N., & Van Der Flier, H. (1987). Cross-cultural comparisons of sex-related differences in intelligence tests: A meta-analysis. *Journal of Cross-cultural Psychology, 18*, 283-314.

Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development, 60*, 1239-1249.

Cahan, S., Hazani, J., Wolf, N., Peyser, M., & Shimborsky, G. (in press). WISC-R95, New Israeli norms for the Wechsler Intelligence Scale for Children. Jerusalem: Psychological Service of the Israeli Ministry of Education (Hebrew).

Cattell, R. B., & Cattell, A. K. S. (1965). *Culture fair intelligence test: Scale 2*. Champaign, IL: Institute for Personality and Ability Testing.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Cohen, N. (in preparation). *The effects of chronological age and schooling experience on IQ scores*. Ph.D dissertation, the Hebrew University of Jerusalem.

Cole, N. S. (1990, April). *Gender differences and admission policies: Implications for test use*. Paper presented at the annual meeting of the American Educational Research Association.

Doolittle, A. E., & Cleary, T. A. (1987). Gender-based differential item performance in mathematics achievement items. *Journal of Educational Measurement, 24*, 2, 157-166.

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 43*, 95-103.

Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research, 62*, 61-82.

Feingold, A. (1994). Gender differences in variability in intellectual abilities: A cross-cultural perspective. *Sex Roles, 30*, 81-92.

Friedman, L. (1989). Mathematics and the gender gap: A meta-analysis of recent studies on sex differences in mathematical tasks. *Review of Educational Research, 59*, 2, 185-213.

Gafni, N. (1978). The relation between ethnic origin, social class and gender in the level and structure of intelligence. Master's thesis, the Hebrew University of Jerusalem (Hebrew).

Gafni, N., & Melamed, A. (1990). *Differential tendencies to guess as a function of gender, test language and test date*. Report No. 115. Jerusalem: National Center for Testing and Evaluation (Hebrew).

Halpern, D. F. (1989). The disappearance of cognitive gender differences: What you see depends on where you look. *American Psychologist, 44*, 1156-1158.

Hyde, J. S. (1981). How large are cognitive differences? A meta-analysis using omega^2 and d. *American Psychologist, 36*, 892-901.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*, 1, 53-69.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107,* 139-155.

Jacklin, C. N. (1989). Female and male: Issues of gender. *American Psychologist, 44,* 2, 127-133.

Lieblich, A. (1983). The structure of IQ among ethnic groups and minorities in Israel. In: M. Nisan & U. Last (Eds.), *Education and psychology*. Jerusalem: School of Education, the Hebrew University (Hebrew).

Lieblich, A., Ninio, A., & Ben-Shachar-Segev, N. (1976). *WISC-R: Manual to the School-Age Hebrew Version of the Wechsler Intelligence Test*. Jerusalem: Institute for Developmental Psychology, the Hebrew University & The Psychological and Counseling Services, Ministry of Education and Culture (Hebrew).

Linn, M. C., & Hyde, J. S. (1989). Gender, mathematics, and science. *Educational Researcher, 18,* 17-27.

Linn, M. C., & Petersen, A. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56,* 1479-1498.

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

Nevo, B. (1986). *Gender differences in level on university entrance examinations*. Lecture delivered at the National Council for the Advancement of Women, Jerusalem (Hebrew).

Olson, J. F., & Fennesy, L. (1990). *An investigation of gender differences due to test-taking instructions*. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Ortar, G., & Shachor, A. (1980). *Milta: Test battery for ages 9-18*. Jerusalem: Ministry of Education and Culture (Hebrew).

Raven, J. C., Court, J. H., & Raven, J. (1975). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. London: Lewis.

Rosenthal, R., & Rubin, D. B. (1982). Further meta-analytic procedures for assessing cognitive gender differences. *Journal of Educational Psychology, 74,* 708-712.

Thorndike, R. L., & Hagen, E. (1971). *Cognitive abilities test: Form 1, Levels A-H, Grades 3-12*. Boston: Houghton Mifflin.

Zeidner, M. (1986). Sex differences in scholastic aptitude: The Israeli scene. *Personality & Individual Differences, 6,* 847-852.

Zohar, A. (1990). *Mathematical reasoning ability: Its structure and selected aspects of its genetic inheritance*. Ph.D dissertation, the Hebrew University of Jerusalem (Hebrew).