

Individual Differences in Students' Retention of Knowledge and Conceptual Structures Learned in University and High School Courses: The Case of Test Anxiety

MOSHE NAVEH-BENJAMIN¹* and HAGIT LAVI¹

¹*Ben-Gurion University of the Negev, Israel*

WILBERT J. MCKEACHIE² and YI-GUANG LIN²

²*The University of Michigan, USA*

SUMMARY

The purpose of this research was to assess individual differences in students' retention of knowledge several years after studying the material. Assessment of retention of materials as a function of students' test anxiety can allow one to evaluate whether high test-anxious students' original deficient academic performance and organization of the materials are due to a retrieval deficit or a deficit in learning and knowledge organization. In two studies, students with different test-anxiety levels completed tasks that enabled us to evaluate both students' levels of knowledge and their cognitive organization of the materials. The tasks were administered either at the end of the course, or at different retention intervals up to 7 years after the end of learning. Results indicated that whereas high test-anxious students tested at the end of the courses performed worse than other students on tests of knowledge and cognitive organization, high test-anxious students tested at various retention intervals after the courses performed as well as other students. The theoretical and practical implications of these results are discussed. © 1997 John Wiley & Sons, Ltd.

Appl. Cognit. Psychol. 11: 507-526 (1997)

No. of Figures: 2 No. of Tables: 7 No. of References: 37

The question regarding what is retained from school learning, although extremely important, has not been, until recently, the subject of extensive research. In the last 10 years there has been an increase in research on this topic, both by educators and by cognitive psychologists. This research indicates a gradual decline in memory for

The research reported in this paper was partially supported by a grant (number G008690010) from the Office of Educational Research and Improvement (OERI), Department of Education, and by a grant from the U.S.-Israel Binational Science Foundation (grant number 87-00100). The opinions expressed therein do not necessarily reflect the position or policy of OERI or BSF and no official endorsement should be inferred.

*Correspondence to: Moshe Naveh-Benjamin, Department of Behavioral Sciences, Ben-Gurion University of the Negev, Beer-Sheva, 84105, Israel.

Contract grant sponsors: Office of Educational Research and Improvement, Department of Education, and U.S.-Israel Binational Science Foundation; Contract grant numbers: G008690010 and 87-00100

In summary, to evaluate the above questions, in study 1, a test of the knowledge of the materials studied and a task to assess students' organization of course material (the FITS technique) were administered to students between 0 and 3 years after the end of a university course. In study 2, a test on the knowledge of the materials and a different task to assess students' organization of the materials (the 'ordered-tree' technique) were administered to high school students at the end of their studies and to other students up to 7 years after the end of high school.

STUDY 1

Study 1 measured students' knowledge and cognitive organization of materials learned in an introductory-level university course.

Method

Subjects

Two hundred and ten University of Michigan students (65% females), enrolled between 1982 and 1987 in an elective 'Learning to Learn' course, participated in the study as the 'experimental participants'. Most of these students were freshmen or sophomores who had taken the course because of their interest in the subject matter. A control group, composed of 28 students who had similar characteristics to those of the experimental subjects, but who had not attended the 'Learning to Learn' course, was used.

Design

One independent variable was retention interval, defined as time elapsed since the end of the course. Each student was tested at one of three retention intervals: within 2 weeks after the end of course-related activities when high test-anxious students, who did not know their course grades yet, are likely to still be anxious (retention interval of 0); 10–14 months after the end of the course (retention interval of 12 months); and 30–40 months after the end of the course (retention interval of 36 months). The second independent variable was the level of test anxiety reported by the student. The distribution of students in each combination of retention interval and test anxiety is presented in Table 1. The third independent variable was the amount of use (rehearsal) of course material after the end of the course, and the fourth independent variable was the student's GPA (Grade Point Average), which was used as an index of ability.

A cross-sectional method was used in which each student was tested at only one of the retention interval.

Materials

Test anxiety. This was evaluated using parts of the Worry–Emotionality scale (Liebert and Morris, 1967). The scale contains items dealing with worry about tests and emotional reactions to tests. Although the Worry and Emotionality scales are highly correlated, the Worry items have been more highly related to decrements in test performance. Students responded to five items (four of which were worry items) on a 5-point scale ranging from not at all (1) to very strong (5). An exemplary item

was: 'I
remem
satisfac
the Sta
three g
scored
anxiety
test-an
group.
combin

Use of
questio
Benjar
(frequ
use of
materi
'used v
to Lea
course
month
scale f
rehear
partici

Grade
studer
GPAs
(0.81)

Me
perfor
task i
the cc

Table 1. Distribution of students in each combination of retention interval and test anxiety group (study 1)

Retention interval (in months)	Test anxiety		
	Low	Medium	High
0	30	22	22
12	16	23	24
36	23	26	24
Control	9	8	11

was: 'While being in an evaluative situation I am also anxious that I can not remember facts which I know'. In our previous studies, these five items have shown satisfactory reliability (test-retest, 0.85) and validity (concurrent validity of 0.70 with the State-Trait Anxiety questionnaire, Spielberger, 1983). Students were divided into three groups according to their responses to the anxiety questionnaire. Students who scored above the 67th percentile in the sample were designated as the high test-anxiety group, and students below the 33rd percentile were designated as the low test-anxiety group. The remaining students were assigned to the medium anxiety group. The distribution of gender (2/3 females) was about the same in each combination of test anxiety and retention interval.

Use of knowledge (re-exposure to materials). This was evaluated using a short questionnaire used successfully in the past (test-retest reliability of 0.90, Naveh-Benjamin, 1988), which assesses quantitative indices of use of the materials studied (frequency and recency) on a 5-point scale (frequency of use, from 'very infrequent use of the materials after the end of the course' (1), to 'very frequent use of the materials' (5), and recency of use on a 3-point scale from 'not used recently' (1), to 'used very recently' (3). Examples of these items are: (a) 'Since finishing the Learning to Learn course, have you read or used things you studied or learned in it, in other courses?'; (b) 'Have you read any Learning to Learn course materials in the last month?' From these scales one overall index of rehearsal was composed on a 3-point scale from 1 ('low and non-recent rehearsal of the materials') to 3 ('high and recent rehearsal of the materials'). The mean rehearsal score for the experimental participants was 1.96 (SD, 0.76; range 1-3).

Grade point average (GPA). This was used as an index of ability and was based on students' performance up to the time at which they participated in the study. Mean GPAs for the experimental and the control groups on a 3-point scale were 1.97 (SD, 0.81) and 2.05 (SD, 0.68), respectively [$t(236) = 0.50, p > 0.05$].

Measurements of the dependent variables were obtained from students' performance on a test of the content covered in the course (knowledge) and on a task intended to reveal aspects of the cognitive structure (knowledge organization) of the course: the 'Fill-in-the-Structure' (FITS) task.

Knowledge test. This was composed of four short-answer type questions. The questions sampled materials from different topics studied in the course and covered major issues discussed. Examples of questions used are:

- (a) Jim is studying for his final examination in chemistry. Describe two self-monitoring strategies that would be helpful to him.
- (b) One of the points made in the course is that intelligence is not fixed. Explain why this should be true.

Two trained judges, who were blind with respect to subjects' identity, scored each question on a scale from 0 to 10. Their average score for each student was used in the analyses.

Cognitive structure test—The 'fill-in-the-structure' (FITS) task. The instructor's tree-like representation of course materials, based on his perception of the structure of the course, is presented in Figure 1. Students received this instructor's representation with 16 of the 28 concepts missing from different levels of the structure (Figure 2). The missing concepts appeared at the bottom of the page intermixed with distractor concepts. The students' task was to choose the appropriate concepts and place them in their proper positions in the structure. Students were not familiar with the FITS task prior to its administration in the study. Details about scoring this task are presented in the Results section.

This task has previously been shown to require appropriate knowledge of both horizontal and vertical relationships between concepts in the course (e.g., Naveh-Benjamin, 1988; Naveh-Benjamin *et al.*, 1995). It provides both useful information about students' knowledge structures at the beginning of the course (reflecting stereotypes, prior knowledge, and linguistic relationships among concepts, Naveh-Benjamin, McKeachie and Lin, 1989), and also indices of the development of these structures throughout university courses (reflecting both general and specific knowledge). Changes in cognitive structures as measured by the technique were shown to be related to students' performance in courses in various academic disciplines (e.g., biology, sociology, English, and psychology).

Procedure

The 'Learning to Learn' course is an introductory psychology course at the University of Michigan. It was taught with an emphasis on information processing, in which the various topics are related to methods of student learning. The course was taught with minimal changes from year to year by the same instructor. Each student in the study completed in a 40-minute session a questionnaire that included questions about his or her rehearsal of the material after the end of the course. In addition, the course knowledge test and the FITS task to infer students' cognitive structures were administered. The results of these tests did not count for grading purpose for those students tested at the 0 retention interval. To avoid carry-over effects, half of the students in each retention interval received the knowledge test first, while the other half received the cognitive structure task first. Participants did not know during the course that they would be tested later. Information about test anxiety was collected at the last class meeting or right after the final examination in the course. This made the test anxiety instrument a specific measure of students'

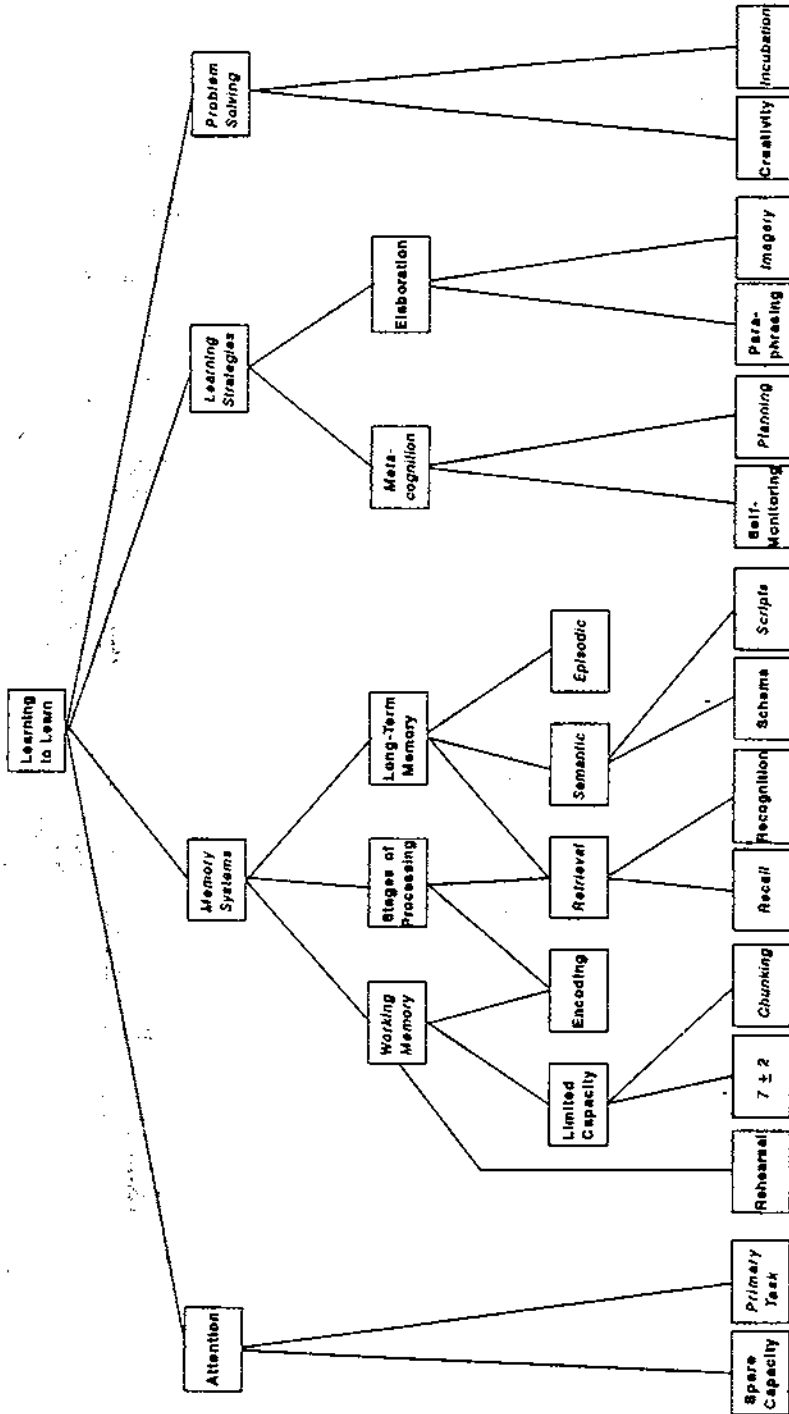


Figure 1. Instructor's cognitive structure representation.

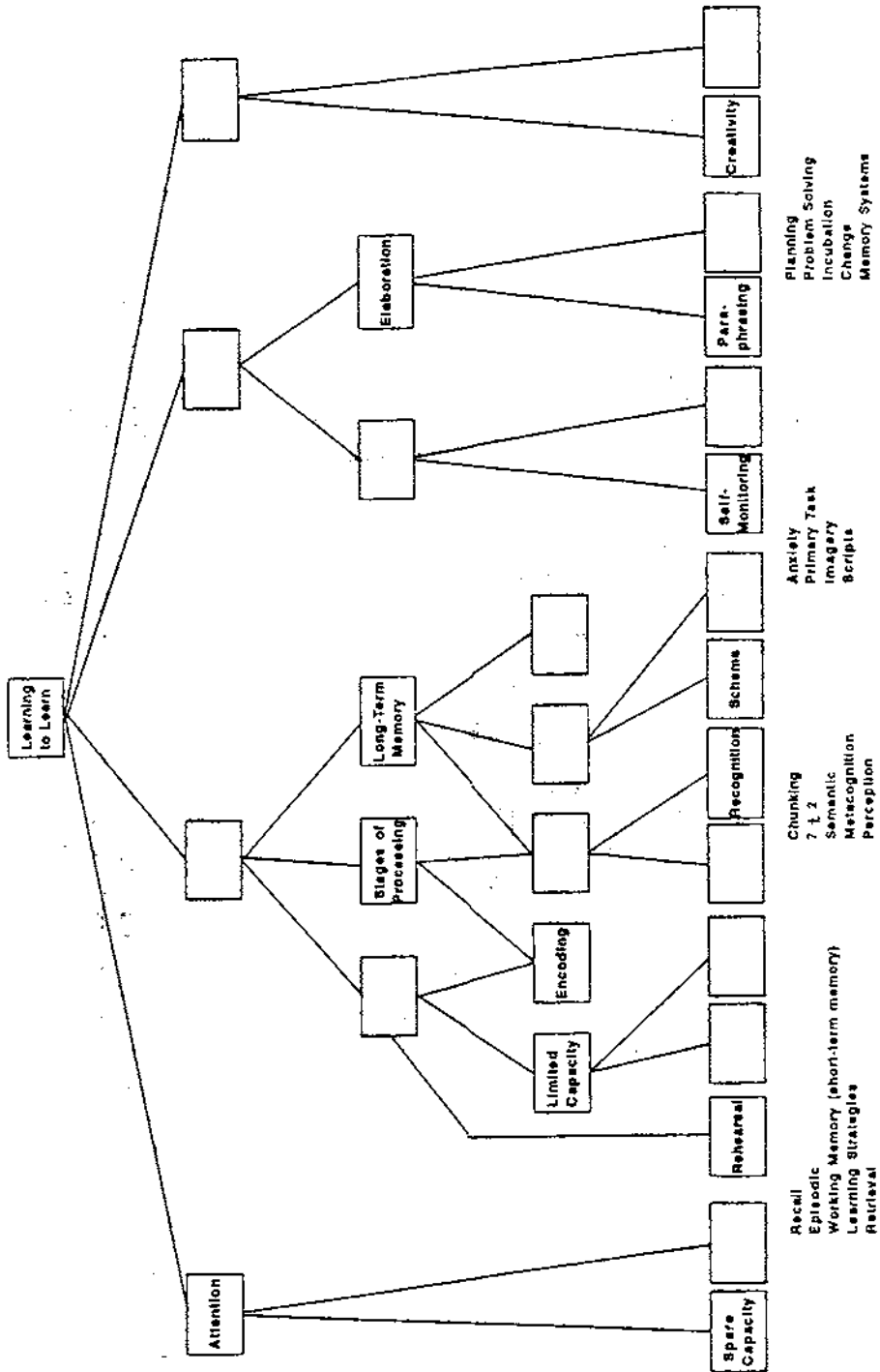


Figure 2 Fill-in-the-Structure task (FITS)

anxiety
are g
questi
rating

Result

As B
exper
the co
to the
mater
know
[F(2,3
effect
high
perce
effect
anot
[F(2,
result
the
[F(2,
of G
meas
the c

Si
perfo
in al
stud
much
are

T
assu
calc
cov
cov
non
viol
C
dep
and
and
tha
wer
var
I
kno

© 19

anxiety in this course, though this is not crucial since state and trait anxiety measures are generally correlated. Finally, since some students received the test anxiety questionnaire before taking the final examination, it is unlikely that their test anxiety ratings reflected only a subjective view of their performance in the test.

Results and discussion

As Bahrck (1984) pointed out, in ecological research of this type we have neither experimental control over the level of use (rehearsal) of the material after the end of the course, nor over the random assignment of students with different ability levels to the different groups. In a preliminary analysis we evaluated the influence of use of materials and GPA on our measures of performance. First, for students' course knowledge index, a one-way ANOVA showed the effect of use to be significant [$F(2,207) = 6.47$; $p < 0.05$; $MSe = 3.75$], and another one-way ANOVA showed the effect of GPA to be significant [$F(2,207) = 7.53$; $p < 0.05$; $MSe = 3.71$]. In both cases, higher use and GPA levels resulted in better performance. Second, for students' percentage of correct responses in the FITS task, a one-way ANOVA showed the effect of use to be marginally significant [$F(2,207) = 2.36$; $p < 0.10$; $MSe = 0.028$], and another one-way ANOVA showed the effect of GPA to be significant [$F(2,207) = 6.51$; $p < 0.05$; $MSe = 0.027$]. In both cases, higher use and GPA levels resulted in better performance. Finally, for students' absolute vertical deviations in the FITS task, a one-way ANOVA showed no significant effect of use [$F(2,207) = 0.22$; ns; $MSe = 0.14$], and another one-way ANOVA showed the effect of GPA also to be non-significant [$F(2,207) = 0.57$; ns; $MSe = 0.15$]. Apparently, the measure of vertical deviations was not as sensitive to the effects of these variables as the other measures.

Since the influences of rehearsal and GPA on some of the measures of performance were detected in preliminary analyses, we controlled for these effects in all the relevant analyses that follow by using the knowledge use reported by the students and their GPA as covariates. Since the patterns of performance were not much different for adjusted and unadjusted scores, all the following means presented are adjusted to reflect the covariates.

To allow the use of ANCOVA, we checked whether there were no violations of the assumption regarding the homogeneity of the regression slopes. To do so, we calculated for each test-anxiety group the regression slope relating each of the covariates to each of the dependent measures. The resulting slopes (for each covariate) for each test-anxiety group were compared to each other using *t*-tests and none of the comparisons was statistically significant ($p > 0.05$), indicating no violation of the above assumption and allowing the use of ANCOVA.

One final set of preliminary analyses evaluated the correlations among the dependent measures used in this study. These were 0.64 between the knowledge test and percentage correct, -0.24 between the knowledge test and the error measures, and -0.38 between percentage correct and error (the negative sign reflects the fact that a larger error measure indicates lack of knowledge). These correlations, which were all significant ($p < 0.05$), are of medium range and indicate some common variability but also some unique characteristics of each measure.

Let us look at students' course knowledge. Table 2 presents mean score on the knowledge test as a function of retention interval and level of test anxiety (scale

0–10). An analysis of covariance (ANCOVA) was performed with retention interval and test anxiety as the independent variables, and amount of rehearsal and GPA as the covariates (variance accounted for by the independent variables and the covariates was 49.3%). The ANCOVA showed a significant decrease in performance as a function of the retention interval [$F(2,199) = 69.82; p < 0.01$], no significant effect of test anxiety [$F(2,199) = 0.56; ns$], and relevant to our hypotheses, a significant effect of the interaction between retention interval and test anxiety [$F(4,199) = 3.13; p < 0.05; MSe = 2.99$]. Contrasts using the Duncan post-hoc procedure indicated that while high test-anxious students' performance at a retention interval of 0 was somewhat lower than the average of the other two anxiety groups ($p < 0.10$), there were no significant differences between high test-anxious students and others at the other retention intervals.

Secondly, let us look at students' cognitive structure changes. The FITS task resulted in two indices regarding students' performances. The first provides information regarding correct answers: it is the percentage of correct filled-in concepts. A correct answer was defined as locating the appropriate concept in its proper position according to the instructor's structure (Chunking and 7 plus/minus 2 were interchangeable). Table 3 presents the overall percentage of correct answers in the FITS task for the low, medium, and high test-anxious students, as a function of the retention interval.

Table 2. Scores on the knowledge test as a function of retention interval and test anxiety level (study 1)

Retention interval (in months)	Test anxiety					
	Low		Medium		High	
	M	SD	M	SD	M	SD
0	7.28	2.14	6.68	1.71	5.96	1.85
12	4.95	1.74	4.20	1.72	4.41	1.74
36	2.77	1.58	3.21	1.54	3.82	1.32
Control	0.86	0.79	0.50	0.49	0.86	0.65

Table 3. Percentage of correct answers in the FITS task as a function of retention interval and test anxiety level (study 1)

Retention interval (in months)	Test anxiety					
	Low		Medium		High	
	M	SD	M	SD	M	SD
0	78.4	14.1	74.4	16.3	67.3	14.7
12	63.8	12.2	62.4	16.5	61.5	14.8
36	46.3	12.1	53.7	13.9	55.4	20.2
Control	32.9	14.9	29.6	12.5	34.6	12.4

An
anxi
covari
was 3
functi
anxi
retent
Co
anxi
of th
betw
contr
differ
orga
reten
simil
orga
Ot
char
et a
misp
place
betw
also
some
relat
wher
they
inde
usin
betw
the i
be
supp
are
indi
T
anx
test
the
cov
per
sigr
sigr
MS
V
rete
(p

An analysis of covariance was performed with retention interval and level of test anxiety as the independent variable, and the amount of rehearsal and GPA as the covariates (variance accounted for by the independent variables and the covariates was 32.8%). The ANCOVA showed a significant decrease in performance as a function of retention interval [$F(2,199) = 35.77; p < 0.01$], no significant effect of test anxiety [$F(2,199) = 0.34; ns$], and more importantly, a significant interaction of retention interval and test anxiety [$F(4,199) = 2.75; p < 0.05; MSe = 2.36$].

Contrasts using the Duncan post-hoc procedure indicated that while high test-anxious students' performance at a retention interval of 0 was lower than the average of the other two anxiety groups ($p < 0.05$), there were no significant differences between high test-anxious students and others at the other retention intervals. These contrasts indicate that at the end of the course (retention interval of 0), there were differences between students, where high test-anxious students had the lowest organization score, indicating poorer organization of the structure. At the other retention intervals of 24 and 36 months, however, high test-anxious students had a similar overall score to that of the other two groups, indicating as good an organization of the structure as that of other students (and even slightly better).

Other indices of students' cognitive structure organization include several characteristics of the errors made by the students (for details, see Naveh-Benjamin *et al.*, 1995). One representative measure of these errors reflects vertical misplacements. These can occur within the same hierarchical level, when a student places a concept in the wrong location but still within the same hierarchical level, or between different levels, when a concept is placed not only in the wrong location but also in an inappropriate hierarchical level. Errors within a level may still indicate some similarity to the instructor's structure, in terms of the hierarchical relationships, whereas errors between levels may indicate hierarchical dissimilarity, where students perceive certain concepts as either more general or more specific than they really are. For each student we computed, only for the mislocated concepts, an index of absolute deviations of vertical misplacements. This measure is based on using for each mistaken placement of a concept, the distance in hierarchical levels between its actual placement and where it was supposed to be placed according to the instructor's structure. For example, if the concept 'Recall', which is supposed to be placed in the bottom (fourth) level, is placed where 'Working Memory' is supposed to be (second level), the score will be 2.0 (4 - 2). These absolute deviations are then averaged across all errors of a given subject. A high degree of misplacement indicates a lack of hierarchical information.

Table 4 presents these deviations as a function of retention interval and test anxiety. An analysis of covariance was performed with retention interval and level of test anxiety as the independent variables, and the amount of rehearsal and GPA as the covariates (variance accounted for by the independent variables and the covariates was 6.9%). There was a significant increase in deviations (poorer performance) as a function of the retention interval [$F(2,199) = 3.04; p < 0.05$], no significant effect of test anxiety [$F(2,199) = 1.61; ns$], and more importantly, a significant interaction of retention interval and test anxiety [$F(4,199) = 6.10; p < 0.01; MSe = 0.16$].

With Duncan's post-hoc procedure, high test-anxious students' deviations at a retention interval of 0 were larger than the average of the other two anxiety groups ($p < 0.05$); these differences disappear at the other retention intervals where there

Table 4. Absolute vertical deviations in the FITS task as a function of retention interval and test anxiety level (study 1)

Retention interval (in months)	Test anxiety					
	Low		Medium		High	
	M	SD	M	SD	M	SD
0	0.23	0.37	0.53	0.58	0.75	0.46
12	0.54	0.52	0.38	0.33	0.50	0.32
36	0.70	0.30	0.61	0.37	0.63	0.33
Control	0.85	0.27	0.86	0.36	0.97	0.63

were no significant differences in deviations between high test-anxious students and others.

Overall, high test-anxious students were initially disadvantaged at the end of the course with respect to their knowledge of the materials and their cognitive structure organization. High test-anxious students assessed 1–3 years after the end of the course performed as well as other students on both the knowledge test and the cognitive structure organization task. Note that for all indices of knowledge and organization, the lack of differences between high test-anxious students and others at later retention intervals cannot be explained by floor effects. Using the Duncan post-hoc procedure, performance in each of the measures at a retention interval of 36 months was significantly better than that of the control group ($p < 0.05$).

Interestingly, in contrast to the significant interaction effects between retention interval and test anxiety, there was no interaction between retention interval and GPA. For both the knowledge test measure and the two cognitive structures measures, ANCOVAs, with retention interval and GPA as the independent variables and rehearsal and test anxiety as the covariates, indicated no significant interactions ($p > 0.10$). Apparently, forgetting in this study did not vary as a function of GPA, though, as mentioned above, overall, students with a better GPA performed better on these measures.

The purpose of study 2 was to evaluate whether these results could be replicated in a different subject area, using a different student population, and a different cognitive structure task.

STUDY 2

This study employed materials related to the subject of literature studied at high school during the past 7 years. During this period, the materials in literature were taught with minimal changes from year to year. The cognitive structure task used was the 'ordered-tree' technique (Naveh-Benjamin *et al.*, 1986), which is an indirect task to evaluate students' organization of the materials.

Method

Subjects

Two hundred and fifty-eight Israeli students (75% females and ages between 17 and 27) who had studied the subject of literature for a final comprehensive examination at the end of high school participated in this study.

Design

One independent variable was retention interval, defined as time elapsed since the end of high school. Each student was tested at one of four retention intervals: close to the end of high school (retention interval of 0 years); 3 years after the end of high school; 5 years after the end of high school; and 7 years after the end of high school¹. The second independent variable was the level of test anxiety reported by the students. The distribution of students among the different retention intervals and test anxiety levels appears in Table 5. There were no differences among these groups with respect to distribution of occupation, and university studies after high school, as revealed by their reports in the questionnaires administered. The third independent variable was the amount of use (rehearsal) of the material after the end of high school. The fourth independent variable was their GPA in their final high-school comprehensive examinations. As in study 1, a cross-sectional method was used in which each student was tested at only one of the retention intervals.

Materials

Test anxiety. This was evaluated using parts of the worry-emotionality scale (Liebert and Morris, 1967) as in study 1. Students were divided into three groups according to their responses to the anxiety questionnaire. Students who scored above the 70th percentile were assigned to the high test-anxiety group, and those who scored below the 30th percentile were assigned to the low anxiety group. The rest of the students were assigned to the middle test-anxiety group. The slight differences in the cut-off points for test anxiety used in this study and those used in study 1 are related to slight differences in the distribution of test anxiety in the two studies. Note that, although students reported retrospectively about their test-anxiety level while studying and taking the final tests on literature materials in high school, we have

Table 5. Distribution of students in each combination of retention interval and test anxiety group (study 2)

Retention interval (in years)	Test anxiety		
	Low	Medium	High
0	14	20	16
3	27	34	21
5	16	30	22
7	20	21	17

¹Results regarding the performance of control subjects used in this study could not be analysed due to technical problems that arose at administration time.

found in the past that these retrospective reports are quite reliable and valid (Naveh-Benjamin, 1992). As in study 1, the distribution of gender (3/4 females) was about the same in each combination of test anxiety and retention interval.

Use of knowledge (re-exposure to materials). The level of rehearsal score was determined as in study 1. The mean score was 1.68 (SD, 0.82; range 1–3).

GPA. This was based on students' reports of their final high-school comprehensive examination grades. Although we did not check these reports, previous research (e.g., Bahrick and Hall, 1991) indicates that students' retrospective reports about their high-school grades are reliable. The mean score on a 3-point scale was 1.97 (SD, 0.67; range 1–3).

Measurements of the dependent variables were obtained from students' performance on a knowledge test of the content covered in the materials studied for the final high-school comprehensive examination in literature, and on a task intended to reveal knowledge of the cognitive structure of the course content—the 'ordered-tree' task.

Knowledge test. This was composed of ten short-answer type questions. The questions sampled materials from different topics studied and covered major issues discussed. Examples of questions used are:

- (a) What are the foundations of the short story?
- (b) What are the three types of prose in literature? Give an example of each.
- (c) What are three differences between medieval and modern poetry? A trained judge who was blind with respect to subjects' identity scored each question on a scale from 0 to 10.

Cognitive structure test—The 'ordered-tree' task. The tree-like representation of literature materials studied for the final comprehensive examination in high school was based on the judgment of four senior teachers in high school who had taught literature for many years. There were some small differences in the way the materials were taught in high school during these years, but to make sure that students at all retention intervals were exposed to the concepts, the structure included only those topics and concepts that had been taught and used in all of the last 7 years.

The task was presented in a four-page booklet. Each page had the same set of 19 concepts chosen by the teaching staff as important and representative of the major content area covered during the study period. The 19 concepts, taken from the above tree-like representation of literature materials, appeared in a matrix, and 19 vertical blanks were provided on the right side of the page, where the concepts were to be listed by the students. The order of the concepts in the matrices was different on each page to avoid the effects of response set. As mentioned earlier, this task has been used in the past (Naveh-Benjamin *et al.*, 1986, 1989) and was shown to be reliable and valid.

Students were then asked to arrange the 19 concepts in a vertical order so that concepts closely related in meaning in terms of the materials learned would appear

close to each other. This was performed four times. In the first and last trials, students could start with any one of the concepts (uncued trials) and in the second and the third trials they had to start with different designated concepts (cued trials) in order to break up tendencies to simply memorize a particular order. Between each of the trials there was a few minutes interval during which students were involved in unrelated activity. Although the original Reitman and Rueter (1981) technique used many trials, we demonstrated in our previous research (Naveh-Benjamin *et al.*, 1986, 1987, 1989) that inducing the knowledge structure from four data sets resulted in reliable cognitive structures. Details about the derived measures of the knowledge structure are presented later, in the Results section.

Procedure

Each student in the study completed a questionnaire that included questions about his or her test anxiety and about the use of the materials after the end of high school. In addition, the knowledge test and the 'ordered-tree' task were administered in a session that took about 60 minutes. The results of these tests did not count for grading purposes for those students tested at the 0 retention interval. To avoid carry-over effects, half of the students in each retention interval received the knowledge test first, while the other half received the cognitive structure task first.

Results and discussion

As in study 1, in preliminary analyses we evaluated the influence of use of knowledge and GPA on our measures of performance. First, knowledge use affected literature knowledge significantly [$F(2,255)=7.59$; $p<0.05$; $MSe=10.96$], and so did GPA [$F(2,255)=7.57$; $p<0.05$; $MSe=10.97$]. In both cases, higher knowledge use and GPA levels resulted in better performance. Second, rehearsal did not affect performance on the cognitive structure task [$F(2,255)=1.66$; ns; $MSe=0.032$], but GPA did [$F(2,255)=3.69$; $p<0.05$; $MSe=0.033$].

Thus, as in study 1, we controlled for these effects in all the relevant analyses that follow by using the knowledge use reported by the students and their GPA as covariates. Since the patterns of performance in this study were not much different for adjusted and unadjusted scores, the adjusted means are presented.

To allow the use of ANCOVA, we checked, as in study 1, whether there were violations of the assumption regarding the homogeneity of the regression slopes. To do so, we calculated for each test-anxiety group the regression slope relating each of the covariates to each of the dependent measures. The resulting slopes (for each covariate) for each test-anxiety group were compared to each other using *t*-tests and none of the comparisons was statistically significant ($p>0.05$), indicating no violation of the homogeneity of the regression slopes assumption and allowing the use of ANCOVA. In one further preliminary analysis we computed the correlation between the two dependent measures, which was 0.23 ($p<0.05$), reflecting some common variability with some unique characteristics of each.

Table 6 contains mean scores on the knowledge test as a function of retention interval and level of test anxiety (scale 0–10). An analysis of covariance (ANCOVA) was performed with retention interval and test anxiety as the independent variables and amount of knowledge use and GPA as the covariates (variance accounted for by the independent variables and the covariates was 26.3%). There was a significant

decrease in performance as a function of retention interval [$F(3,244)=25.17$; $p<0.01$], no significant effect of test anxiety [$F(2,244)=1.08$; ns], and no significant interaction between retention interval and test anxiety [$F(6,244)=1.20$; ns; $MSe=10.55$].

Using Reitman and Rueter's (1981) algorithm, we obtained an 'ordered-tree' cognitive structure representation for each student. We then compared each student's tree with the instructor's (the one produced by the literature expert teachers), using McKeithen, Reitman, Rueter and Hirtle's (1981) measure of similarities between trees (based on the proportions of chunks common to the two trees relative to all chunks of the two trees). The range of this index is from 0.0 (lack of similarity) to 1.0 (identity).

Table 7 contains the mean similarity measures for the low, medium, and high test-anxious students as a function of the retention interval. An analysis of covariance was performed with retention interval and level of test anxiety as the independent variables and the amount of rehearsal and GPA as the covariates. There was a significant decrease in similarity as a function of retention interval [$F(3,244)=3.10$; $p<0.05$], no significant effect of test anxiety [$F(2,244)=1.61$; ns], and more importantly, a significant interaction of retention interval and test anxiety [$F(6,244)=2.31$; $p<0.05$; $MSe=0.03$]. Using the Duncan post-hoc procedure, the high test-anxious students' similarity measure at a retention interval of 0 was lower than the average of the other two anxiety groups ($p<0.05$). There were no significant differences between high test-anxious students and others in the similarity measure at the other retention intervals².

Overall, there was a trend for high test-anxious students to have a deficit in their knowledge of the materials and a clear deficit in their cognitive structure organization at 0 retention interval. At later retention intervals, knowledge and cognitive organization of the materials of high test-anxious students were

Table 6. Scores on the knowledge test as a function of retention interval and test anxiety level (study 2)

Retention interval (in years)	Test anxiety					
	Low		Medium		High	
	M	SD	M	SD	M	SD
0	8.42	2.48	7.15	3.37	5.42	3.62
3	3.62	4.20	2.13	3.57	2.52	3.71
5	1.43	2.94	1.94	3.41	2.16	3.27
7	1.25	2.68	1.43	2.41	2.21	3.60

²It is unlikely that the lack of differences between high test-anxious students and others is due to floor effects. Although we could not use in study 2 the results of the control group, the results of study 1, which employed a control group, indicated above-floor performance and very similar patterns to those obtained in this study. In addition, the lack of differences between high test-anxious students and others was evident, not only in the retention interval of 7 years, but also in retention intervals of 3 and 5 years where performance was clearly above-floor. Finally, only 22% of the subjects had lower than 0.20 similarity score indicating that most subjects had better than 0.00 performance. A *t*-test showed high test-anxious students to have a significantly better than chance (0.00) performance ($p<0.05$).

Table 7. Similarity of students' cognitive structure to the instructor's structure in the 'ordered-tree' task as a function of retention interval and test anxiety level (study 2)

Retention interval (in years)	Test anxiety					
	Low		Medium		High	
	M	SD	M	SD	M	SD
0	0.41	0.19	0.43	0.16	0.23	0.20
3	0.30	0.17	0.22	0.19	0.30	0.14
5	0.29	0.21	0.26	0.16	0.23	0.22
7	0.25	0.16	0.24	0.18	0.22	0.14

comparable to those of other students. Again, as in study 1, this was in contrast to the interaction between retention interval and GPA. For both the knowledge test measure and the cognitive structures similarity measure, ANCOVAs, with retention interval and GPA as the independent variables and rehearsal and test anxiety as the covariates, indicated no significant interactions ($p > 0.10$). Apparently, forgetting in this study did not vary as a function of GPA, though, as mentioned above, overall, students with a better GPA performed better on these measures.

GENERAL DISCUSSION

High test-anxious students were disadvantaged at the end of their courses. Their initial deficiency in knowledge and in cognitive structure organization of the materials is consistent with previous results obtained by Benjamin *et al.* (1981), and by Naveh-Benjamin *et al.* (1987). The results of study 2 are particularly noteworthy, confirming the test-anxiety effects on cognitive structure organization with high-school students who had not been studied previously. More interesting, high test-anxious students were not disadvantaged in their ability to answer questions about the materials and in their organization of course materials some time after course completion. In study 1, they did not show a disadvantage after 1–3 years after the end of the course, and in study 2, they did not show such a disadvantage 3–7 years after the end of the study period. These results were obtained in both studies, reflecting their replicability across materials studied, educational context, and student population.

Note that, although the effect of the interaction of test anxiety with retention interval on the knowledge measure in study 2 was not statistically significant, the trends in both studies were in the same direction, indicating the replicability of the data. Furthermore, specific comparisons of high test-anxious students with others at each of the retention intervals converged on the same results in both studies.

The results obtained for GPA were quite different. Though GPA affected overall performance, that is, those students with better GPAs performed better on the various tasks, it did not interact with retention interval as did test anxiety. Students with high GPAs showed a trend in forgetting similar to that of students with lower GPAs.

Theoretically, these results can be interpreted as consistent with the retrieval-deficit hypothesis as the locus of high test-anxious deficiencies at the end of a course. According to this hypothesis, when test-anxious students are tested years after a course, there is no reason for them to be anxious, hence no interfering thoughts block their retrieval of relevant information. As a result, they do as well as other students on long-term retention tests. The current results are not compatible with the learning-deficit hypothesis suggested as an explanation of high test-anxious students' 0-retention interval performance. According to this perspective, the lower performance of high test-anxious students reflects less effective learning and organization of the material during the course. If that were the case, however, we would have expected these students to show poor performance relative to other students some time after the end of the course. They did not.

The current results regarding the retention of knowledge as a function of test anxiety are also compatible with some laboratory results. These results had shown that while materials learned under high arousal tend to be remembered less well immediately, later on, its memorization level is relatively high in comparison to materials learned under low arousal conditions. This is due to better consolidation processes over time of the materials learned ('Kleinsmith effect'; Kleinsmith and Kaplan, 1963, 1964). Though these studies have used very simple materials (nonsense syllables or word-digit pairs), much shorter retention intervals (up to 1 week), and an arousal level that was induced rather than used as an individual difference characteristic as here (but see Howarth and Eysenck, 1968), and their results not always replicated (see Keppel's, 1984, review), the trends obtained are quite similar.

Note that our results on the effects of test anxiety on long-term forgetting are not necessarily incompatible with classical (e.g., Underwood, 1964) and more recent research on forgetting (e.g., Bahrlick, 1984; Slamecka and McElree, 1983), which document that forgetting rates are similar for better and poorer learners. In our case, the differential forgetting rates obtained may be based on the fact that initial performance at the end of the course does not reflect high test-anxious students' actual knowledge, which is higher but masked by retrieval problems. This initial depressed performance will cause the forgetting rates to look slower for high test-anxious students.

From a practical point of view, the results are quite clear: in the long run, high test-anxious students seem to possess cognitive structure organization and the ability to apply it to answer questions about the materials studied in a formal setting, that is as good as that of other students. This might reassure educators bothered by the immediate instructional implications of high test anxiety. High test-anxious students' long-term retention of knowledge structure of the materials and its use is not as poor as previously indicated by studies that concentrated more on the immediate effects of anxiety on performance.

Use of knowledge (rehearsal) and ability level (GPA) were related to retention as indexed by at least some of the measures used. These results are similar to the ones obtained by Bahrlick and Hall (1991) who showed rehearsal and original learning to be related to algebra and geometry performance (though rehearsal did not interact with retention interval), but different from those obtained by Bahrlick (1984), who showed no effects of rehearsal on foreign language retention. One reason for the discrepancy in the results could be the overall level of rehearsal. Both in the current study and in that by Bahrlick and Hall (1991), the overall level of rehearsal indicated

that subj
interval a
overall re

Finally
differenc
anxious
the end
forgettin
anxiety,
high test
results,
forgettin
retentio
individu
be guid
differen
along t

Bahrlick
Span
Bahrlick
cont
Benjar
in in
Conw
kno
Jou
Conw
acq
Desid
Jou
Furu
lea
How
Jou
Kepp
(E
Klein
int
Klein
sy
Lieb
a
Mc
in
I
Mc
e

that subjects engaged in activities related to the materials during the retention interval and hence it affected performance. Bahrick's (1984) data suggest very little overall rehearsal activity, and hence no effects on retention.

Finally, the answer to the general question regarding the effects of individual differences on retention of materials studied in school is not simple. High test-anxious students, as those with low ability, though deficient in their performance at the end of the study phase, do not, as might be expected, demonstrate more forgetting of the materials over time, relative to other students. In the case of test anxiety, there are indications of the opposite: forgetting appears to be slower for high test-anxious students who initially demonstrate lower performance. The current results, coupled with results mentioned earlier that show no effect of initial ability on forgetting, imply that there may not be a uniform effect of individual differences on retention, and certainly not one indicating relative better performance over time to individuals who perform better initially. Rather, predictions of these effects should be guided by an informed theoretical analysis regarding the loci of the original differences between individuals, coupled with analysis of the processes that occur along the retention interval.

REFERENCES

- Bahrick, H. P. (1984) Semantic memory content in permastore: fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General*, **113**, 1-35.
- Bahrick, H. P. and Hall, L. K. (1991). Lifetime maintenance of high school mathematics content. *Journal of Experimental Psychology: General*, **120**, 20-33.
- Benjamin, M., McKeachie, W. J., Lin, Y-G. and Holinger, D. P. (1981). Test anxiety: deficits in information processing. *Journal of Educational Psychology*, **73**(6), 816-824.
- Conway, M. A., Cohen, G. and Stanhope, N. (1991). On the very long-term memory of knowledge acquired through formal education: twelve years of cognitive psychology. *Journal of Experimental Psychology: General*, **120**, 395-409.
- Conway, M. A., Cohen, G. and Stanhope, N. (1992). Very long-term memory of knowledge acquired at school and university. *Applied Cognitive Psychology*, **6**, 467-482.
- Desiderato, O. and Koskinen, P. (1969). Anxiety, study habits, and academic achievement. *Journal of Counseling Psychology*, **16**, 162-165.
- Furukawa, J. M. (1977). Cognitive processing capacity and learning-mode effects in prose learning. *Journal of Educational Psychology*, **69**, 736-743.
- Howarth, E. and Eysenck, H. J. (1968). Extraversion, arousal, and paired-associate learning. *Journal of Experimental Research in Personality*, **3**, 114-116.
- Keppel, G. (1984). Consolidation and forgetting theory. In H. Weingartner and E. S. Parker (Eds.), *Memory Consolidation* (pp. 149-161). Hillsdale, NJ: Erlbaum.
- Kleinsmith, L. and Kaplan, S. (1963). Paired-associate learning as a function of arousal and interpolated interval. *Journal of Experimental Psychology*, **65**, 190-193.
- Kleinsmith, L. J. and Kaplan, S. (1964). Interaction of arousal and recall interval in nonsense syllable paired-associate learning. *Journal of Experimental Psychology*, **67**, 124-126.
- Liebert, R. N. and Morris, L. W. (1967). Cognitive and emotional components of test anxiety: a distinction and some additional data. *Psychological Reports*, **20**, 975-978.
- McKeachie, W. J. (1984). Does anxiety disrupt information processing or does poor information processing lead to anxiety? *International Review of Applied Psychology*, **33**, 187-203.
- McKeachie, W. J., Pollie, D. and Spiesman, J. (1955). Relieving anxiety in classroom examinations. *Journal of Abnormal and Social Psychology*, **50**, 93-98.

- McKeithen, K. B., Reitman, J. S., Rueter, H. H. and Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, **13**, 307-325.
- Mueller, J. H. (1980). Test anxiety and the encoding and retrieval of information. In I. Sarason (Ed.), *Test anxiety: theory, research and applications* (pp. 63-86). Hillsdale, NJ: Erlbaum.
- Naveh-Benjamin, M. (1988). Retention of cognitive structures learned in university courses. In M. M. Gruneberg, P. Morris and R. Sykes (Eds.), *Practical aspects of memory: current research and issues*, vol. 2 (pp. 383-388). Chichester: Wiley.
- Naveh-Benjamin, M. (1991). A comparison of training programs intended for different types of test-anxious students: further support for an information processing model. *Journal of Educational Psychology*, **83**, 134-139.
- Naveh-Benjamin, M. (1992). Students' retrospective reports about test anxiety. Unpublished manuscript.
- Naveh-Benjamin, M., McKeachie, W. J., Lin, Y-G. and Tucker, D. (1986). Inferring students' knowledge structures and their development using the "ordered-tree" technique. *Journal of Educational Psychology*, **78**, 130-140.
- Naveh-Benjamin, M., McKeachie, W. J. and Lin, Y-G. (1987). Two types of test-anxious students: support for an information processing model. *Journal of Educational Psychology*, **79**, 131-136.
- Naveh-Benjamin, M., McKeachie, W. J. and Lin, Y-G. (1989). Use of the "ordered tree" technique to assess students' initial knowledge and conceptual learning during an academic course. *Teaching of Psychology*, **16**, 182-187.
- Naveh-Benjamin, M., Lin, Y-G. and McKeachie, W. J. (1995). Inferring students' cognitive structures and their development using the "fill-in-the-structure (FITS) technique. In P. D. Nichols, S. F. Chipman and R. L. Brennan (Eds.), *Cognitively Diagnostic Assessment* (pp. 279-304). Hillsdale, NJ: Erlbaum.
- Reitman, J. S. and Rueter, H. H. (1981). Organization revealed by recall orders and confirmed by pauses. *Cognitive Psychology*, **12**, 554-581.
- Semb, G. B. and Ellis, J. A. (1994). Knowledge taught in school: what is remembered? *Review of Educational Research*, **64**, 253-286.
- Slamecka, N. J. and McElree, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **9**, 384-397.
- Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory (STAI)*. Palo Alto, CA: Consulting Psychologists Press.
- Stanhope, N., Cohen, G. and Conway, M. (1993). Very long-term retention of a novel. *Applied Cognitive Psychology*, **7**, 239-256.
- Tobias, S. (1985). Test anxiety: Interference, defective skills, and cognitive capacity. *Educational Psychologist*, **20**, 135-142.
- Tobias, S. (1992). The impact of test anxiety on cognition in school learning. In K. A. Hagtvet and T. Backer Johnsen (Eds.), *Advances in test anxiety research* (vol. 7, pp. 18-31). Amsterdam: Swets & Zeitlinger.
- Tyron, G. S. (1980). The measurement and treatment of test anxiety. *Review of Educational Research*, **50**, 343-372.
- Underwood, B. J. (1964). Degree of learning and the measurement of forgetting. *Journal of Verbal Learning and Verbal Behavior*, **3**, 112-129.
- Wine, J. D. (1971). Test anxiety and the direction of attention. In I. G. Sarason (Ed.), *Test anxiety: theory, research and application* (pp. 349-378). Hillsdale, NJ: Erlbaum.