DOCUMENT RESUME

ED 348 036 IR 015 773

AUTHOR Verhagen, P. W.
TITLE Preferred Length of Video Segments in Interactive Video Programs.
PUB DATE Feb 92
NOTE 27p.; In: Proceedings of Selected Research and Development Presentations at the Convention of the Association for Educational Communications and Technology and Sponsored by the Research and Theory Division; see IR 015 706.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Freshmen; Computer Assisted Instruction; Field Dependence Independence; Foreign Countries; Graphs; Higher Education; *Instructional Design; *Interactive Video; *Learner Controlled Instruction; Predictor Variables; Pretests Posttests; *Program Length; *Retention (Psychology); Student Characteristics; *Time Factors (Learning); Verbal Ability
IDENTIFIERS Netherlands

ABSTRACT

This study investigated questions related to the length of video segments in interactive video programs: (1) the preferred segment length if learners decide how much information they want presented before stopping to answer question; (2) the relationship between segment length and direct recall of factual information when segment length is self-chosen and when segment length is fixed; (3) the relationship between segment length and delayed recall of factual information when segment length is self-chosen and when segment length is fixed; and (4) effects of fatigue on self-chosen segment length. Subjects were 235 freshmen at a university in the Netherlands. Five experimental conditions were used: variable, in which subjects determined the length of each segment; crossed, the same as variable with information elements presented in a different order; linear, which allowed no stopping; short-long, fixed segment length with short segments followed by long; and long-short, fixed segment length with long segments followed by short. Data on field dependence independence, verbal ability, imagery-based strategies for storing information, and retention were gathered by testing; logs of interactive video sessions provided data on relevant time intervals, starting and stopping points, and questions answered correctly. Wide variances in the results of data analyses led to the conclusion that the main message may be that quality communication can be accomplished in many forms, whereby different audiovisual formats and segment lengths may appear feasible as long as learners start their tasks with a realistic expectancy of the demand characteristics and be motivated to tune their mental effort accordingly. (8 figures, 17 references) (MES)

ERIC

U S DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

ED 348 036

Title:

Preferred Length of Video Segments in Interactive Video Programs

Author:

P. W. Verhagen

IR015773

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Michael Simonson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

ERIC

PREFERRED LENGTH OF VIDEO SEGMENTS IN INTERACTIVE VIDEO PROGRAMS.

P.W. Verhagen

University of Twente
Department of Education
Division of Educational Instrumentation
P.O. Box 217
7500 AE Enschede
The Netherlands

Introduction

Interactive video may be conceived as the combination of the interactive capabilities of the computer and the audiovisual power of video. This combination is generally considered as a powerful medium for delivering instruction, nowadays often incorporated in so-called multi-media systems. This article focuses on one design issue with respect to the video component.

The Problem

An essential property of adding computer control to video programs is the possibility to stop for questions or excercises, or to give feedback to students about their progress, at any desired moment. A simple but fundamental question is: What is the optimum moment to stop for the purpose of achieving maximum learning? The literature shows some ideas with respect to this question (see for instance Bork, 1987; Laurillard, 1987; Phillips, Hannafin & Tripp, 1988; Schaffer and Hannafin, 1986). One question is whether preferred segment length plays a role in this. This paper attempts to clarify this issue in the context of a larger study about optimal segment length (Verhagen, in preparation). To make the research manageable, choices were made that narrow the problem of that study down to the following central question:

" What is the optimum length of well designed audiovisual segments to present factual information via an interactive video program to learners who possess certain characteristics?"

The exact meaning of this question is specified as follows:
- The measure for "length of segment" is the number of information elements in one segment. The term "information element" is defined as the smallest meaningful unit of language to state a name, a label or a single proposition, according to Gagné's definition of verbal information (Gagné, 1985).
- "Well designed" is operationalized in terms of the approval of the audiovisual material by an expert jury, in combination with successful formative testing of the material under development.

- "Factual information" means verbalizable information about facts, concepts, procedures and principles to be stored in memory on the recall level (see also Gagné, 1985; or Merrill, 1983; Romizowski, 1986; Wager & Gagné, 1988).
- The learner characteristics involved concern verbal ability; the ability to process pictorial information; a cognitive style aspect (field dependance); and sex. These are briefly described below.

One term cannot yet be defined, and that is what to understand by "optimum" length of a segment. It is part of the study to find out what has to be regarded as optimum. In this paper, one aspect is elaborated: the extent to which learner preference influences the way in which optimum segment length has to be regarded.

With respect to recall, two situations have to be taken into account:

a. Segment length and direct recall of information
In general, video segments serve to initially supply learners with information as a part of a learning experience that may further consist of rehearsal or application of that information, with or without the demand to demonstrate the ability to recall the presented information by answering questions. The ability of learners to recall the presented information directly after the completion of a video segment can be considered as a measure of the effectiveness of the initial presentation. The better this is the case, the quicker a learner may proceed in an instructional situation and the shorter the needed time on task will be and thus the more efficient the learning will take place, at least as far as concerns this factor. This efficiency may be dependent on segment length.

b. Segment length and delayed recall of information
After a learner has worked through all segments of a program, the effectiveness of the learning session as a whole may be measured from his or her ability to answer questions about the program after some time has elapsed. This effectiveness may also appear to be dependent on segment length.

The relation between segment length and direct recall may or may not tally with the relation between segment length and delayed recall. On one hand, correct direct recall may be a proof of successful information transfer, which may be confirmed by a high level of delayed recall. On the other hand; in a learning situation, mistakes with respect to direct recall will in most cases be remediated by rehearsing the missed information in one form or another. Rehearsal supports retention. Incorrect direct recall may thus cause rehearsal which leads to better delayed recall than correct direct recall, be it at the expense of increased time on task. However, there is reason to strive for maximum correct direct recall: rehearsal as a consequence of the frustrating experience of failing to answer questions correctly violates rules for motivational instruction. Keller for instance argues that instruction should be such that learners can develop confidence in their possibilities to fulfill assigned learning tasks successfully (see Keller, 1983; Keller & Kop, 1987; Keller & Suzuki, 1988). Rehearsal should therefore not be the consequence of failure by the learner, but should be carefully planned and integrated in the learning situation as an activity that appears logical to the student and may thus be easily accepted. It is possible that preferred segment length automatically provides the answer to this problem, if learners are in the position to make themselves the trade-off between segment length and failure to answer questions about the presented information. The experiment that is decribed below is designed to study this possibility.

Before turning to a description of the experiment it has to be emphasized that the study has a non-instructional character. The study is primarily designed to find rules about segment length for the purpose of the initial presentation of information as a component of one instructional pattern or another. The role of such presentations within instructional patterns is in principle not considered. It was however attempted to support the validity of the segments used in the experiments as potential building blocks in instruction by using a pattern of presentation, questioning, remediation, and feedback that resembles tutorial instruction.

Self-chosen segment length is one possibility, program-controlled segment length another. The desirability of self-chosen segment length has to be valued against its effectiveness for learning compared to program-determined fixed segment length.

A further point is whether segment length can be independent of time on task in the sense that long learning sessions may cause fatique effects that affect optimal segment length (after some time students may probably need shorter segments than when they are fresh).

With the former, the following research questions are listed:

1. What is the preferred segment length if learners have to decide for themselves how much information (how many information elements) they want to have presented to them before they stop to answer questions about that information?

2. Which is the relationship between segment length and direct recall of factual information presented by these segments
   a. when segment length is self-chosen?
   b. when segment length is fixed?

3. Which is the relationship between segment length and delayed recall of factual information presented by these segments
   a. when segment length is self-chosen?
   b. when segment length is fixed?

4. Does fatigue effect affect self-chosen segment length?

Research Method

In the study, subjects attended two experimental sessions. During the first, they mainly worked with an experimental videodisc program, followed by a posttest. This session took one morning or one afternoon. The second session was used for a retention test and for tests to score some variables with respect to individual characteristics. The second session took about two hours.

The experimental video program.

For the study, an experimental videodisc program about cheesemaking has been produced. This program contains 252 information elements that are presented

through video with off-screen narration. Together they form a connected discourse of 36 minutes if the program is played linearly without stopping.
According to the earlier definition, an information element is operationalized as one uninterrupted statement of the narrator about which one factual question can be put. This means that in the experiments with one question per information element in principle 252 questions could be asked about the content of the program. However, in eight cases this would have been too artificial because of the simplicity of the text concerned. In practice 244 questions were put to the subjects. The 252 information elements are distributed over 150 camera positions (shots).

The computer program is used to control the system so that the video program can be broken down into various sequences of segments. Thereby the video can only be started or stopped between two information elements. This causes video segments of the program always to consist of whole numbers of information elements. The way in which the program is broken down into segments depends on the different experimental conditions that are described below .

The video program has two parts: (a) an introduction of 4.5 minutes (information elements 1-33; 31 questions), and (b) a main program of 31.5 minutes (information elements 34-252; 219 elements with 213 accompanying questions).

In all conditions, the introduction was used twice. The first time the introduction was presented as an ordinary linear video program to give an overview of the cheesemaking process. This prepares the subjects for the presentation style and the kind of information in the main program (in which the cheesemaking process is treated in much more detail).

The second time the introduction was used to give the subjects the opportunity to experience the way of working with the main program according to the experimental condition in which he or she was placed.

It is believed that by this approach the subjects were well-prepared to enter the main program (all data about working with the program were collected during work with the main program).

Experimental conditions

Five experimental conditions were used:

VAR:
In the so-called variable condition, the subjects determined the length of each segment. The basic procedure was as follows:
- In the ready position to begin a new segment, the screen showed a video still with the superimposed message "click the mouse". This video still was the first frame of the information element with which the segment would start.
- After clicking, the program started playing until the subject decided that it was time to ask for it to stop by again clicking a mouse button. The program then stopped at the end of the information element in which the stop was requested. This defines a segment: if the playing started with element i, and the program stopped at the end of element i+k, a segment containing k+1 information elements was created.

- The subject then answered all k+1 questions about the segment he/she just saw, one question after the other. All questions were open questions that required short sentences, single words, or numbers to be typed in as answers.
- Next, all questions were one by one repeated corresponding to the order of the information elements in the segment together with the answers of the subjects and feedback about right or wrong. Right answers were reinforced by a complete sentence that restated the correct answer. If a question was encountered that was answered wrongly, the feedback informed the subject of this fact followed by the message "click the mouse to repeat the related video fragment". After clicking, this video fragment was repeated followed by a second attempt to answer the question. Also after this second time feedback about right or wrong was provided, this time together with the right answer regardless of whether the second answer was right or wrong.
- After all questions of the segment were reviewed, the subject was allowed to continue the program beginning with element i+k+1, the information element that followed the last element of the just-completed segment[1].

In this way, the subject divided the video program into segments by the repetition of watching, stopping, answering questions, getting feedback (with built-in repetition of video parts with respect to missed questions), and starting the next segment.

The VAR condition is the main condition used to answer the research questions about preferred segment length (Research Questions 1, 2a and 3a).

CROSSED:
The same as VAR, with the difference that first, the second half of the main program was presented (information elements 145-252), and then the first half (information elements 34-144). This was possible because the subjects entered the main program after the introduction (with the overview of the cheesemaking process) had been studied twice. With that background it was not an unnatural experience to start in the middle of the main program.
This condition was used to find out whether fatique effects would affect the results (Research Question 4).

LIN:
In this condition no stopping was allowed. The subjects watched the complete main program without stopping and then had to answer the 213 questions of the main program as if they had divided the program into just one long segment. This condition was made to compare linear use of the video program with interactive use. It was meant as a control condition relative to information overload if the experimental program was not segmented.

SHORT-LONG (SL):
In this condition, segment length was fixed. The first half of the program was worked through in 23 steps (short segments), the second half in 5 steps (long segments). Except for fixed length, this condition functioned in the same way as

1) At this place -- where a segment just has been completed -- subjects could also have a break and leave their chairs to relax in an adjacent room. To do so they were required to click the right-hand mouse button, which caused the message "pause" to be displayed and by which the logging of time was switched to the logging of the duration of the interval. To continue after a pause, subjects had to hit <ENTER> after which the video still of the first element of the next segment was displayed together with the message "click the mouse" as described above.

the VAR condition. The SL condition was used to show how recall of facts was affected by fixing the segment length (Research Questions 2b and 3b).

LONG-SHORT (LS):
The same as the former, with the difference that in this case the long segments came first. This condition was made to balance the former.

Subjects

Subjects were 235 university freshmen of several technical and social science departments at a single university in the Netherlands. Age, sex, university department, and year of study were registered. The subjects were randomly distributed over the experimental conditions with the restriction that in the first phase of the experiment, only the conditions LIN, VAR and CROSSED existed. Students of the Department of Education participated in that phase only. In total, 3 subjects were removed from the data analysis because of false data due to obvious atypical behaviour or as a consequence of technical failures of the equipment during the experiment. The resulting distribution is presented in Table 1.

Table 1: Distribution of the subjects

	LIN	VAR	CROSSED	LS	SL	Total
TO:	22	27	22	0	0	71
PAPP:	12	11	11	10	10	54
BETA:	17	30	16	22	22	107
Total:	51	68	49	32	32	232

Legenda (between brackets: numbers):

TO : Students of the Department of Education (the Dutch name is Toegepaste Onderwijskunde: TO)

PAPP : Students of the Faculty of Public Administration and Public Policy

BETA : Students of the Departments of Computer Science (30), Physics (5), Mathematics (4), Chemical Engineering (21), Electrical Engineering (20), Mechanical Engineering (15), and Business Administration (12)

Subject Characteristics

The following tests were administered, in all cases selected because of possible influence of the pertaining variable on self-chosen segment length:

Group Embedded Figures Test (GEFT): The GEFT was used as this test appears to correlate to the cognitive-style dimension "field dependence" (Witkin, Oltman, Raskin, & Karp, 1971). In the experiment a possible relationship between segment length and field dependence was assumed, as field-independent subjects may have more possibilities than field-dependent subjects to see through the information structure of the experimental video program and tune their stopping places accordingly. Field-dependent subjects may, on the other hand, have problems with the fact that in the VAR and CROSSED conditions the decision where to stop has to be taken by themselves without any guidance from the program (apart from the fact that they know that all presented factual information has to be stored in memory for later recall). This may lead them to play safe by stopping frequently, probably more frequently than field-independent subjects might do.

CLOZE-test: Part of the experimental task of the subjects was to try to remember as much as possible of the "verbal information" that is presented by the video segments. This performance may be influenced by verbal ability. Subjects were ranked in this respect by their scores on a CLOZE test. The CLOZE test used consisted of two short texts in which every fifth word was replaced by an empty space. The total amount of deletions was 91. Subjects had to try to reconstruct the original texts by filling in the right words.

STAR-PLUS-test: This concerns a modification of the familiar sentence-picture comprehension and verification task of Chase and Clark (1972). Pezdek, Simon, Stoeckert, and Kiely (1987) showed that in their experiments good television comprehenders were more likely to utilize an imagery-based strategy for storing information. This result was the reason to test the subjects in the present study with the STAR-PLUS-test, to be able to determine whether the cognitive trait it measures influences preferred segment length[2].

Experimental setting

Three adjacent rooms were used, situated on the first floor of one of the university buildings.

The first room was used for administration by the experimental leader on duty (the researcher or student assistants) and as a relaxation room during the experiments. Also the programr of the project found his workplace here and was continuously available for technical assistance. The subjects were received in this room and used the room to take pauses while working with the experimental program (see the description of the procedure below). Coffee, tea and -- upon request -- hot chocolate were provided. In a corner, special arrangements were made to run the STAR-PLUS test that was administered during the second session. At the back of the room were located the entrances of the other two rooms used in the experiment.

The second room was used for all the paper-and-pencil testing. This test room was equiped with four small tables and four chairs: three for subjects and one for the experimental leader. (No more than three subjects could participate in the experiment at one time, as this is the number of interactive video sets that was available for the experiments.)

The actual experiment took place in the third room. For that purpose, this room was equipped with three interactive video sets each containing an MS-DOS personal computer (Olivetti M240) with a built-in video overlay board (Cameron

2) A detailed description of this test goes beyond the purpose of this paper. A complete description is given by Verhagen (in preparation).

MCS), a keyboard and a mouse (Microsoft); a color monitor (Barco CM33); a videodisc player (Philips VP406 or VP410); and headphones (Sennheiser HD-414). Each set was placed on a table with a surface of 1.2 by 1.2 meters. The tables were placed so that the distance between subjects was about 2.5 meters (which respects their needs for privacy according to ergonomic criteria, Oborne, 1987, p199). Screens between the tables prevented subjects to look at the video monitor of a neighbor. The subjects faced windows with dark blue but slightly transparent curtains that allowed some visual contact with a tree-rich parking lot on the university campus. It is believed that this prevented arousal of claustrophobic feelings. Except for the blue curtains and blue adjustable chairs, the dominant colors in the room were shades of grey. The lighting level was adjusted to balance the luminence of the video monitors. It is believed that all these measures together created a friendly atmosphere in which subjects felt at ease.

Procedure

The procedure was the same in all conditions. The experiment took place in two sessions. Session 1 was prepared by switching on the equipment one-half hour before the arrival of the subjects (to reach a stable and qualitatively good colour picture on the rather old video monitors used) and by feeding data about the expected subjects into the computer (name, subject number, experimental condition, age, sex, university department, year of study). After loading these data, the video monitor turned to black, after which the interactive-video sets stood waiting for the subjects.

The subjects arrived at 8:45 (if they participated in a morning session) or at 13:15 (for an afternoon session). (They had been reminded of the first session with a letter that also emphasized that they were expected to appear with a clear mind after a good night's sleep. This letter was sent about one week before the first session.) The session started in the test room with a few words of welcome and a short explanation of the aim of the experiment and the role of the subjects. After this, a pretest was administered that contained just one assignment: "Write down anything you know about the making of cheese". The introductory words and the pretest took together about 10 to 12 minutes (no one subject appeared to be able to describe anything substantial about the cheese-making process).

Next, the subjects were invited to enter the room with the interactive video sets. They were instructed to adjust their chairs to reach a comfortable seating posture and to read a one-page instruction in a version that suited the experimental condition to which they were assigned. This instruction repeated some of the general remarks from the oral instroduction and explained how to work with the equipment. After having observed that everyone had finished reading, the experimental leader asked whether there were questions and answered these in case there were. Next, the subjects were invited to wear the headphones and the experimental leader started the interactive video program by typing a password.

The program instructed the subjects that they were about to watch an introduction about cheesemaking of about 4.5 minutes, which would play without stopping after clicking the mouse. The subjects then watched the introduction which automatically stopped after 4.5 minutes on a freeze frame. Near the bottom of the screen a message was diplayed asking the student to click the mouse to continue. After this, the program instructed the subjects that the introduction would be repeated but that this time they had to work in the manner in which they were expected to work with the main program that followed the introduction.

The subjects then worked through the introduction in one of the ways described before (dependent on the experimental condition). In the VAR and CROSSED conditions, the experimental leader took the opportunity to demonstrate to the subjects what was expected, by showing how to control segment length for a first short segment with a length of three information elements. By working through the introduction the second time, the subjects thus practiced their experimental task. At the end of the introduction the system ceased responding to the subjects, giving instead a message to ask the experimental leader for assistance. The experimental leader then had a short individual conversation with each subject to find out whether he or she understood the task, and gave additional information or reinforcement if necessary. After this, the main program was put into position by again typing a password. At this point, about 20 to 25 minutes had passed since the introduction had been started for the first time.

The subjects worked through the main program, which typically took about 2.5 hours, ranging from less than 2 hours to more than 4.5 hours. Subjects in the LIN condition often needed more time than the others, subjects in the SL and LS conditions less. At the end the program notified the subject to ask the experimental leader what to do next. The subject then was invited to return to the test room and two posttests were administered. The first one consisted of open questions to measure recall, the second one of multiple choice questions to measure recognition. To the second one a questionnaire was attached that asked the subjects about their strategy for deciding at which segment lengths to stop (VAR and CROSSED condition only), what they thought of program-controlled fixed segment lengths (SL and LS conditions only), and their opinion of the program (all conditions)[3]. Taking the tests took about 25 to 30 minutes.

Last, the subjects were sent home with three messages: (a) Reappear in the second session with a clear mind as you did this first time; (b) do not study anything about cheesemaking between this session and the next one; and (c) do not tell your fellow students what has happened here, because they may be in the sample of subjects and every subject should start the experiment without preknowledge of the experimental task.

Session 2 took place three weeks after Session 1 and was completely devoted to testing subjects. The morning session of Session 2 started at 9:15, the afternoon session at 13:45. For most tests, the test room was used. As the second session started later than the first one and was shorter (about two hours in total), all tests could be administered while subjects in session 1 were busy with the interactive video program. The tests were administered in the following order:
- The Group Embedded Figures Test (GEFT)
- A retention test with open questions.
- A retention test multiple-choice questions.
- The STAR-PLUS test (in the first room using the special equipment that was installed there for this purpose).
- The CLOZE test.

3) The results of the questionnaire are discussed elsewhere (Verhagen, in preparation).

ERIC

Data analysis.

During their work with the experimental video program, all the actions of the subjects were logged. As a result, the following data were available:
1. All relevant time intervals (watching video, answering questions, pauses, etc.);
2. Starting and stopping points of segments (and thus number of segments and number of information elements per segment);
3. Questions answered correctly the first time.
4. Questions answered correctly the second time (after repetition of the video fragment that followed negative feedback about incorrect first answers).

Next to the data from the program, the scores of all tests were available. All scores were organized in such a way that answers to the questions of the post- and retention tests can be directly related to the pertinent element numbers of the main program.

Results

Background variables

Of the background variables "Age", "Sex", "University Department", and "Year of Study", only Sex and Department appeared to correlate with the research results. However, the number of male and female students was not equal between the different departments. The majority of the subjects from the Department of Education were women, while the technical departments were represented by nearly only male subjects. A multivariate regression analysis revealed that differences between fields of study (educational science, public administration & public policy, technical engineering) explain some of the differences in experimental results, while sex differences do not[4]. Sex differences are therefore not further taken into account in this paper but the differences between the departments are. As can be seen from Table 1, TO students did not participate in the SL and LS conditions, while subjects from the other department are rather equally distributed over all conditions.

Self-chosen segment length

Subjects appeared to differ substantially with respect to the mean of their self-chosen segment lengths. This is displayed in Table 2 and Figure 1.

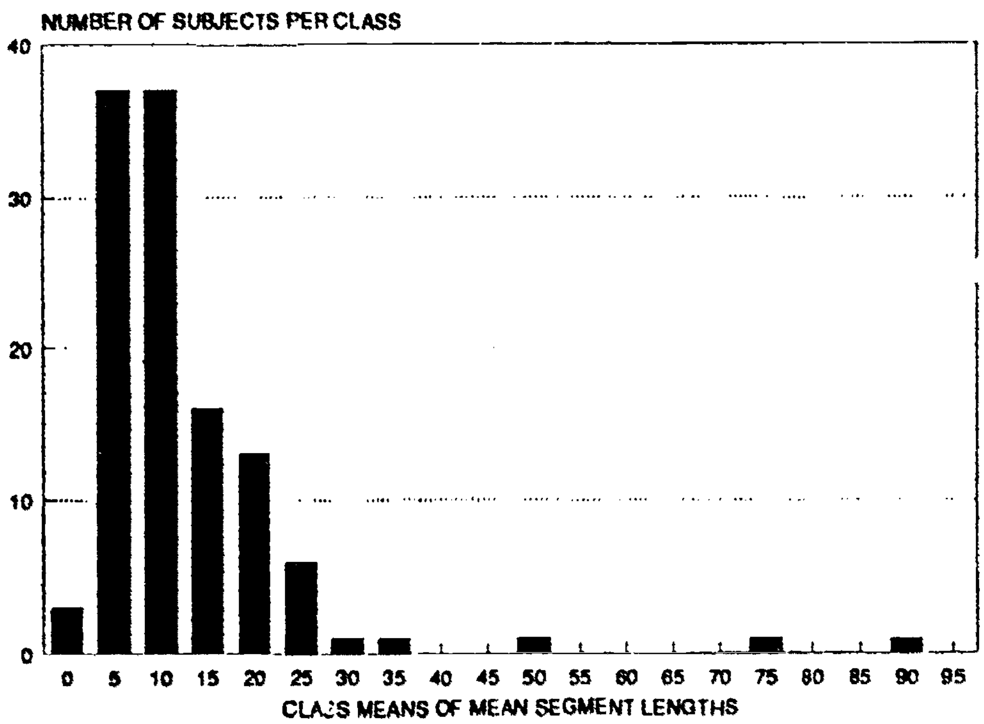
---

4) More details can be found with Verhagen (in preparation).

Table 2: Self-Chosen Segment Length and University Department

Department	Mean number of information elements per segment	Standard deviation
TO:	7.42	5.34
PAPP:	13.87	6.67
BETA:	17.66	15.78
Total:	12.70	11.77

Legenda:
TO : Department of Education (the Dutch name is Toegepaste Onderwijskunde: TO)
PAPP : Faculty of Public Administration and Public Policy
BETA : Departments of Computer Science, Physics, Mathematics, Chemical Engineering, Electrical Engineering, Mechanical Engineering, and Business Administration



VAR and CROSSED conditions pooled.
Overall mean segment length: 12.19 information elements; standard deviation: 11.77.
Minimum: 2.19 information elements; maximum: 87.50 information elements.

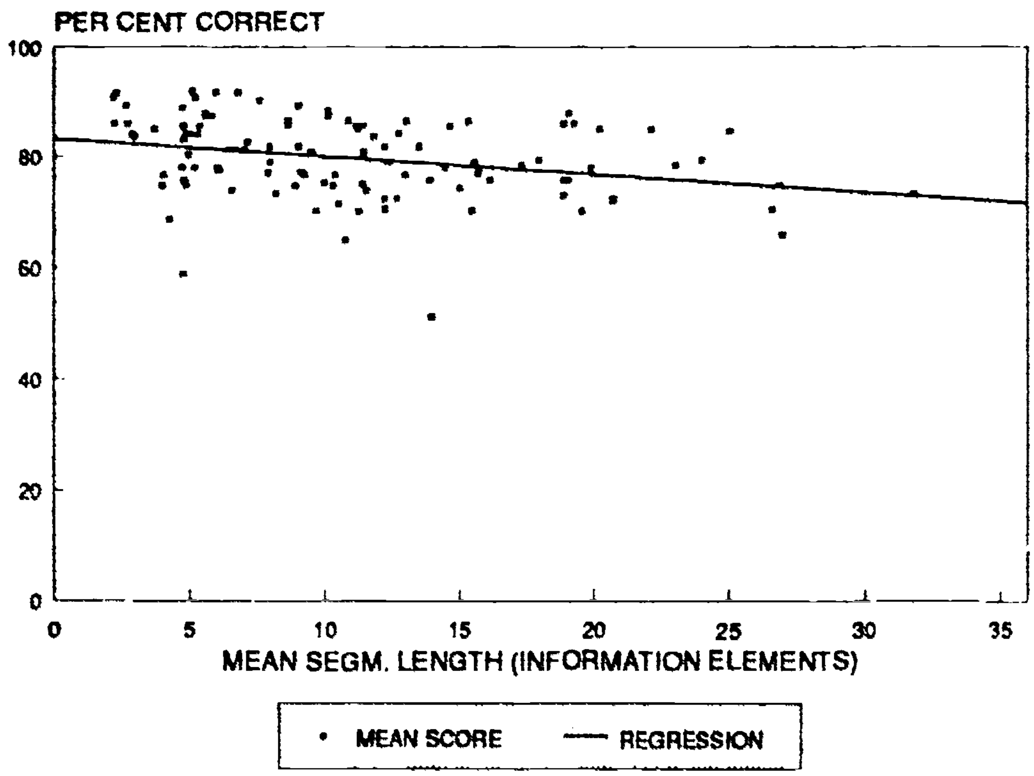
Figure 1: Distribution of means of self-chosen number of information elements per segment

The differences between the mean segment lengths of the students from the different departments appears to be significant ($Chi^2 = 38.61$; $p<.0001$). Because of the fact that no TO students participated in the SL and LS conditions (Table 1) and the fact that TO students appear to choose shorter segment lengths than students from other departments (Table 2), TO students are only included in analyses where the VAR, CROSSED or LIN conditions are considered. On all occasions where also the SL and LS conditions are at stake, the TO-students are thus left apart.

Although Table 2 shows that the mean segment length of PAPP students and Bèta students are different too, it is assumed that no reason exists to treat PAPP students separately because in all conditions about an equal number of randomly assigned PAPP students were represented.

Within-program performance as a function of self-chosen segment length.

Figure 2 shows the performance of subjects while working with the program as a function of the mean of self-chosen segment lengths.



VAR and CROSSED conditions pooled;
data from all subjects with mean self-chosen segment length < 36 information elements.

Figure 2: Within-program performance as a function of mean segment length

The figure shows for every mean segment length the mean percentage of correct answers to the questions within the program when they were posed for the first time. Only subjects with mean segment lengths shorter or equal to 36 information elements per segment are included. The three subjects with a longer mean segment length (see Figure 1) are considered as outlayers[5].

The regression line for all subjects that describes the relation between the mean of self-chosen segment length in the 'VAR' and 'CROSSED' conditions and the mean percentage of correct answers within the program follows from the equation[6,7]:

$$(MEAN\_SCORE)_{self\text{-}chosen\ length} = 85.10 - .36*(MEAN\_SEGMENT\_LENGTH\_)$$

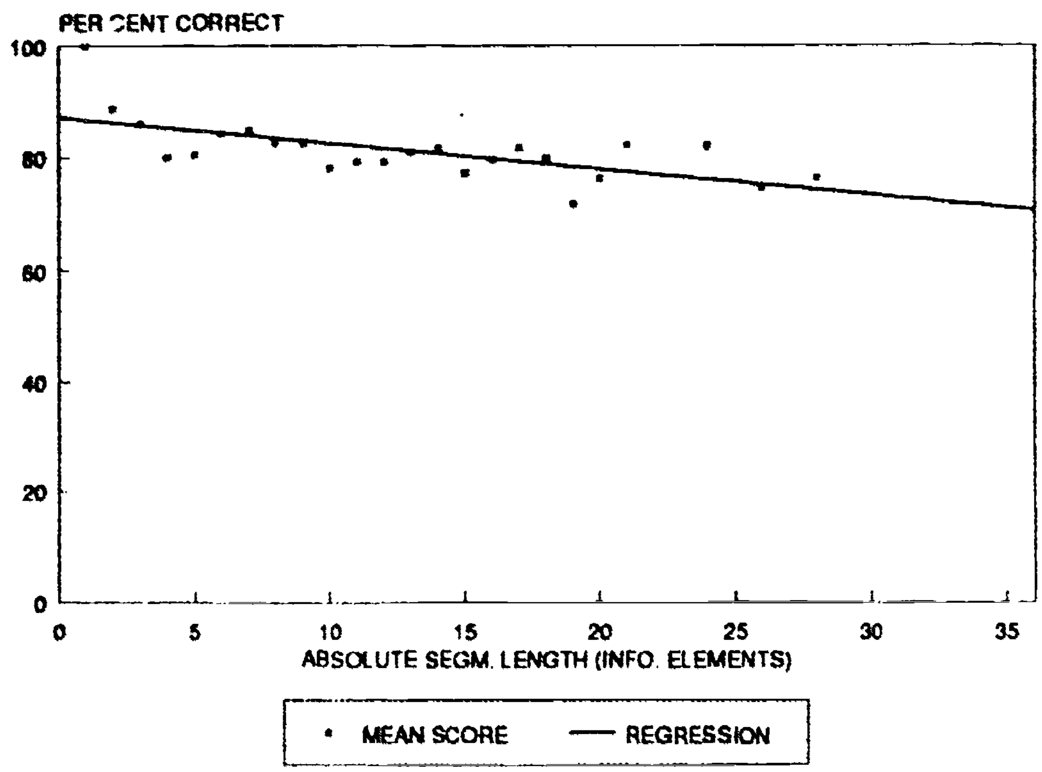
The regression analysis explains only 1.5% of the variance in the $MEAN\_SCORE_{self\ chosen\ length}$ variable.

The values on the X-axis of Figure 2 are means of segment lengths that may be composed of contributions of a wide variety of segment lengths dependent on the variance per subject. The next figure (Figure 3) shows the percentage of correct answers within the program as a function of absolute segment length.

5) The chosen criterion is that outlayers have mean segment lengths that are longer than the overall mean segment length plus two times the standard deviation. The exact measure is 35.73 (which is 12.19 plus 2*11.77)

6) In this and the following equations, "SCORE" means "PERCENTAGE OF CORRECT ANSWERS".

7) "MEAN_SEGMENT_LENGTH" means that the values on the x-axis represent the means of self-chosen segment lengths per subject.



VAR and CROSSED conditions pooled;
only data form subjects that occurred at least 10 times are used in this analysis.

Figure 3: Within-program performance as a function of absolute segment length

To arrive at this figure all segments of one length were pooled, regardless of with which subject a segment was found. Only segment lengths for which at least 10 observations were available were included in the analysis. Per segment length the mean was computed. The resulting values were plotted in the figure with the related regression line. The resulting regression line for all subjects in the 'VAR&CROSSED' combination follows from the equation[8]:

$$\{MEAN\_SCORE\}_{absolute\ length} = 85.48 - .36 * \{ABSOLUTE\_SEGMENT\_LENGTH\}$$

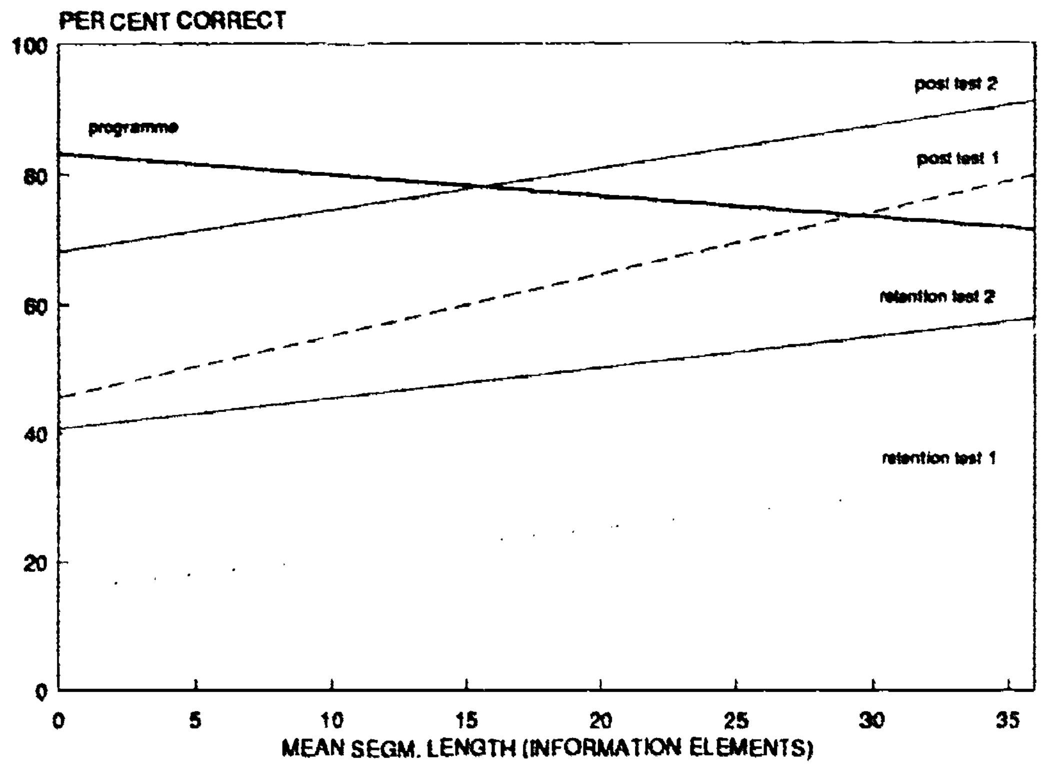
This regression analysis explains 45.9% of the variance in the $MEAN\_SCORE_{absolute\ length}$ variable.

The relation between self-chosen segment length and performance on post tests and retention tests

Figure 4 shows the relation between self-chosen mean segment length and test performance. Only regression lines are displayed. As is clear from the figure, subjects with longer self-chosen segment lengths performed better than subjects who chose shorter ones. The explained variances of the posttests and retention

8) See footnote 6.

tests are, however, limited to 18% (Posttest 1), 12% (Posttest 2), 9% (Retention Test 1), and 6% (Retention Test 2).

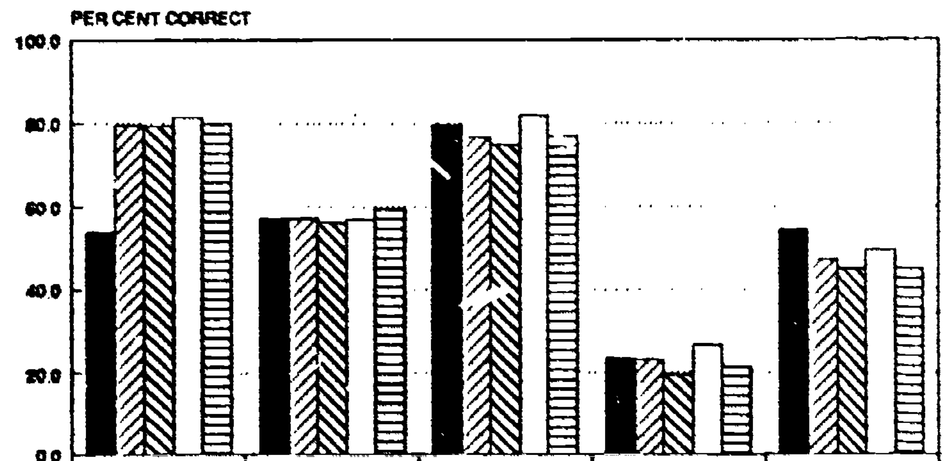


Programme = score on questions within the program.
Data from all students in the VAR and CROSSED conditions.

Figure 4: Test Performance as a Function of Mean Segment Length

Differences in test performance between all conditions

Figures 5, 6, and 7 and Tables 3, 4, and 5 show for all conditions the performance while working with the program and the test results on the posttests and the retention tests. This is done for the program as a whole as well as for the first and second halves of the program separately. This was done to see whether the different treatments per half in the VAR, CROSSED, SL, and LS conditions affected the results.

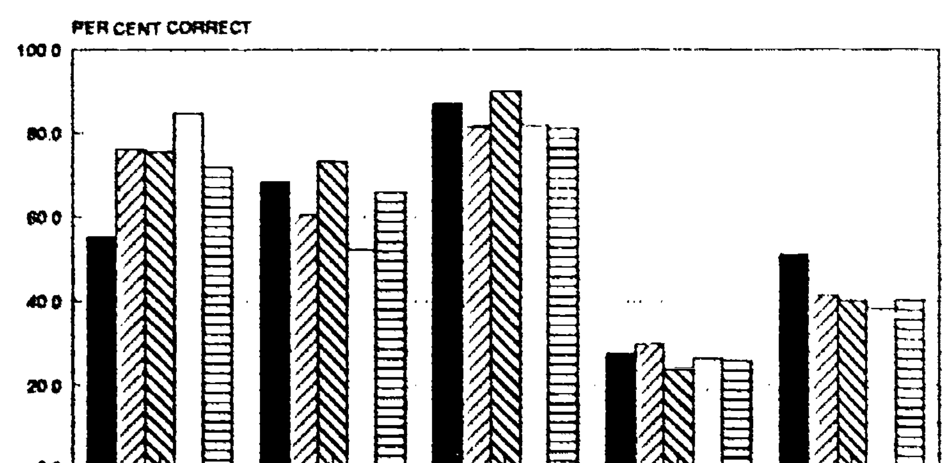


	PROGRAMME	POSTTEST 1	POSTTEST 2	RET. TEST 1	RET. TEST 2
LINEAR	53 8	57 2	79 4	23 4	54 4
VARIABLE	79 6	57 2	76 5	23 0	47 2
CROSSED	79 3	56 2	74 8	19.4	44 8
SHORT-LONG	81 4	56 9	81.9	26 7	49 6
LONG-SHORT	79 8	59 4	76 7	21.5	44 9

LINEAR VARIABLE CROSSED
SHORT-LONG LONG-SHORT

Programme = score on questions within the program. Data from all students except TO.

Figure 5: Test Performance by Condition (Whole Program)

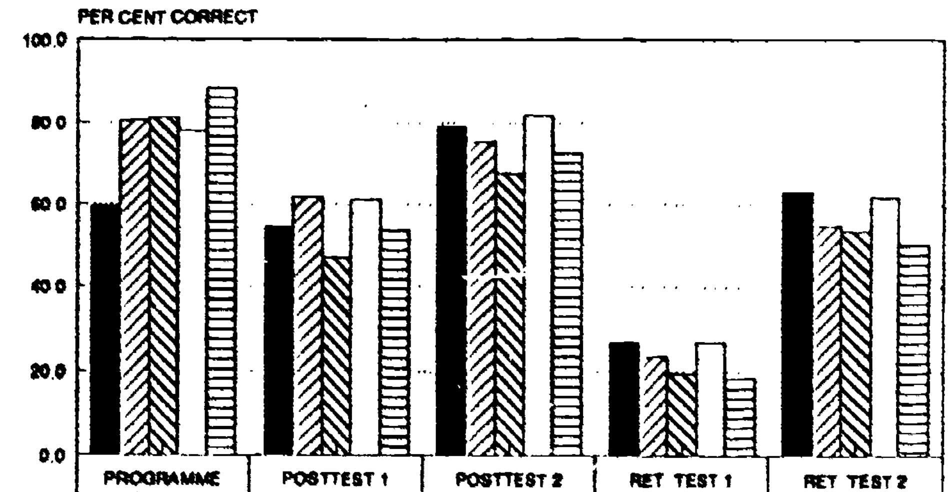


	PROGRAMME	POSTTEST 1	POSTTEST 2	RET. TEST 1	RET. TEST 2
LINEAR	55 1	68 3	86 9	27 7	51 2
VARIABLE	76 0	60 5	81 5	29 9	41 3
CROSSED	75 4	73 2	90 0	23 9	39 9
SHORT-LONG	84 6	52 3	81 9	26 4	38 3
LONG-SHORT	71 8	66 0	80 9	25 8	40 1

LINEAR VARIABLE CROSSED
SHORT-LONG LONG-SHORT

Programme = score on questions within the program. Data from all students except TO.

Figure 6: Test Performance by Condition (First Half of Program)



	PROGRAMME	POSTTEST 1	POSTTEST 2	RET TEST 1	RET TEST 2
LINEAR	59.4	54.4	79.0	26.8	62.9
VARIABLE	80.6	61.7	75.4	23.5	54.8
CROSSED	81.3	46.8	67.4	19.5	53.3
SHORT-LONG	78.0	61.0	81.9	26.9	61.6
LONG-SHORT	88.3	53.5	72.5	18.4	50.0

LINEAR VARIABLE CROSSED
SHORT-LONG LONG-SHORT

Programme = score on questions within the program. Data from all students except TO.

Figure 7: Test Performance by Condition (Second Half of Program)

Table 3: Performance Within the Program Compared (all conditions)

Conditions	PROGRAM Whole	Part 1	Part 2
LIN < VAR, CROSSED, SL, LS	**	**	**
SL > VAR, CROSSED		**	
SL > LS		**	
SL < LS			**
LS > VAR, CROSSED			**
LS < VAR		*	

Legend:
The table shows differences in performance level between the conditions with respect to answering questions within the program for the first time. A' questions were open questions. Differences were evaluated with a Chi^2 test. Onl; ignificant differences are indicated. Part 1 means first half of the program; Part 2 means second half.
*: Level of significance < .05; **: Level of significance < .01

ERIC

Table 4: Performance on Posttests Compared (all conditions)

Conditions	POSTTEST 1 Whole	POSTTEST 1 Part 1	POSTTEST 1 Part 2	POSTTEST 2 Whole	POSTTEST 2 Part 1	POSTTEST 2 Part 2
LIN > VAR		*				
LIN < VAR			**			
LIN > CROSSED			**			**
LIN > SL		**				
LIN < SL			*			
VAR > CROSSED			**			**
VAR < CROSSED		**			**	
VAR > LS			**			
CROSSED > SL		**			**	
CROSSED < SL			**			**
CROSSED > LS					*	
CROSSED < LS			*			
SL > LS			*			*
LS > SL		**				

Legend:
The table shows differences in performance level between the conditions with respect to post-test scores. Posttest 1 contained open questions that referred to 56 of the information elements of the program. Posttest 2 contained 20 multiple choice questions. Differences were evaluated with a Chi^2 test. Only significant differences are indicated. Part 1 means first half of the program; Part 2 means second half.
*: Level of significance < .05; **: Level of significance < .01

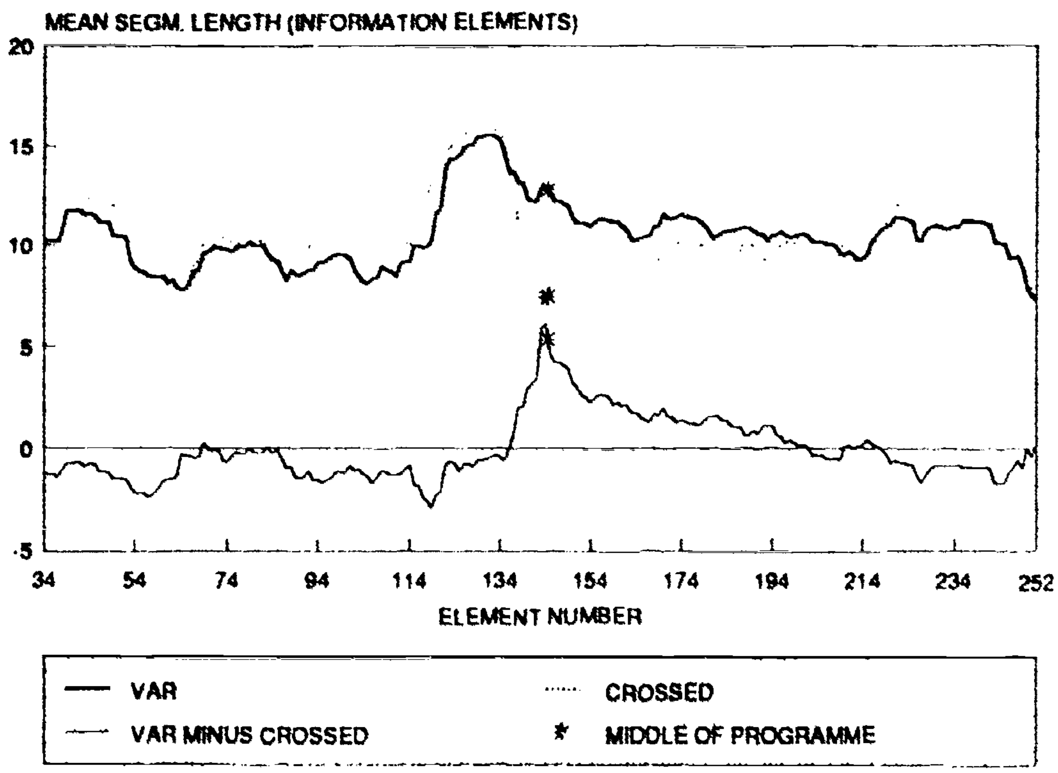
Table 5: Performance on Retention Tests Compared (all conditions)

Conditions	RETENTION TEST 1 Whole	RETENTION TEST 1 Part 1	RETENTION TEST 1 Part 2	RETENTION TEST 2 Whole	RETENTION TEST 2 Part 1	RETENTION TEST 2 Part 2
LIN > VAR				*	*	
LIN > CROSSED	*		**	**	*	*
LIN > SL					*	
LIN > LS			*	**	*	**
VAR > CROSSED	*	*				
CROSSED < SL	*		**			*
CROSSED > LS			**			**
SL > LS			**			**

The table shows differences in performance level between the conditions with respect to retention-test scores. Retention Test 1 contained open questions that referred to 81 of the information elements of the program. Retention Test 2 contained 24 multiple choice questions. Differences were evaluated with a Chi^2 test. Only significant differences are indicated. Part 1 means first half of the program; part 2 means second half.
*: Level of significance < .05; **: Level of significance < .01

Mean segment length as a function of the position in the program

Figure 8 shows the mean segment length as a function of the position in the program. Each information element was part of one segment per subject. This means that every information element was part of as many segments as there were subjects. In Figure 8, the values of Mean Segment Length for each Element Number were computed per element as the mean length of all segments of which an element was a part. This was done for the VAR and CROSSED conditions separately. To be able to observe trends (for instance: Do subjects choose longer segments near the end or just the reverse?), only the data of subjects with 12 segments or more are used in this figure. Figure 8 is also used to discuss fatique effects (Research Question 4).



Only data of subjects with 12 segments or more were used in this analysis.

Figure 8: Mean Segment Length by Element Number

Discussion

Preferred segment length

The first research question asked which segment length learners might prefer if they can choose segment length for themselves. With the present group of subjects the answer is that self-chosen segment length varies over a wide range (Figure 1). Apart from a few extreme scores, mean segments length varies from about 2 to 36 information elements. Although segment length is defined in terms of semantically relevant information elements, it is illustrative to mention the duration of the different segments as they were apparently preferred. The main program has a length of 31.5 minutes and contains 219 information elements.

This leads to a mean length of 8.63 seconds of one information element. The mean of self-chosen segment length thus ranges from about 18 seconds to 311 seconds. Mere preference seems thus to rule segment length.

Segment length and direct recall

The second research question was whether a relation exists between segment length and direct recall and if it makes a difference whether segment length is self-chosen or program controlled. Self-chosen segment length is discussed first.

Figure 2 shows that the mean of self-chosen segment length has only a weak relationship with direct recall. According to the regression line in Figure 2 subjects with a mean segment length of 2 information elements answered about 84% of within-program questions correctly the first time, while subjects with a mean segment length of 36 information elements answered about 72% of these questions correctly. This is a relatively small difference given the difference in mean segment length.

It is, however, true that the regression line from Figure 2 explains only 2.5% of the variance of the performance on within-programme questions. It is therefore interesting to observe that in the given situation Figure 3 suggests that the mean of all scores per absolute segment length is a reasonable predictor of learner performance for that segment length (with an explained variance of almost 46%). The similarity between the lines of Figures 2 and 3 is thereby striking. The small slope of the regression lines reinforces the idea that subjects know for themselves how much information they can handle in one time and decide what segment length to choose accordingly.

The greater precision of absolute segment length may be caused by the fact that subjects make conscious decisions about segment length on a segment-by-segment basis. This may cause much variance in the segments lengths of individual subjects (of which Figures 1 and 2 are derived). It may, however, cause limited variance in performance per segment length if each subject optimally tunes his or her attention, ability, and mental effort to each segment that is presented and chooses the length of that segment accordingly. It has to be noted that subjects with better memories have thereby obviously more possibilities to decide what segment length they prefer. The data suggest that in this respect technical students have the most room to manoeuvre (see Table 2).

The question whether the within-program performance of subjects in the variable conditions differs from the performance in the conditions with fixed segment length can be discussed on the basis of Figures 5, 6, and 7 and Table 3. The VAR and CROSSED conditions show no significant differences in this respect. The subjects in the LIN condition appear however to perform worse than all other conditions (around 55% while all other conditions are doing around 75% to 80%). Answering 213 questions in one time about 31.5 minutes of video obviously is a difficult job.

If the first and second half of the program are considered separately, the SL and LS conditions appear to differ from the VAR and CROSSED conditions and from each other (Figures 6 and 7, and Table 3). SL subjects did better during the first half of the program than subjects in all other conditions, while LS subjects did better during the second half. It is clear that the SL and LS subjects experienced the short segments as relatively easy. The long segments were more difficult for them. LS subjects did significantly worse than the SL subjects during the first half of the program and SL subjects did significantly worse than the LS subjects during the second half. From all other differences, only the different performance of LS subjects and VAR subjects during the first half of the program was significant; the LS subjects did worse than the VAR subjects.

The order of magnitude of the significant differences between the variable conditions and the SL and LS conditions is not impressive. The differences are similar to the differences within the variable conditions according to Figures 2 and 3. The relation between segment length and performance level is obviously rather weak, regardless of the fact whether segment length was self-chosen or program controlled (under the condition that program-controlled segment length falls within the range of self-chosen segment lengths).

Segment length and delayed recall

The third research question asked whether a relation exists between segment length and delayed recall and if it makes a difference whether segment length is self-chosen or program controlled.

Figure 4 shows that subjects with longer mean self-chosen segment length performed better on the post tests and retention tests than subjects who chose shorter segments. This may be caused because these subjects have better memories or because they just preferred longer segments and invested more mental effort to make that choice succesful (not to miss too many questions within the program), which resulted in deeper processing that perhaps paid itself back on the different post- and retention tests. A combination of these two causes is likely.

From Figures 5, 6, and 7 and from Tables 4 and 5 it is apparent that the LIN condition did relatively well as far as the posttests and the retention tests are concerned. One possible explanation for this could be that this is a consequence of the relatively bad performance within the program, which yielded extra exercise due to repetitions of video fragments. This appears, however, not to be true. An extra analysis showed that repetition of video was negatively correlated with posttest and retention-test performance in all conditions (more details can be found in Verhagen, in preparation). It appeared that answering the questions within the program correctly at the first attempt was a good predictor of correctly answering questions about the pertinent information elements in later tests.

An alternative explanation of the good performance of LIN subjects might be that the LIN subjects invested more mental effort, as they knew that they would be questioned about the complete program. Similar to the assumption about the better performance of subjects with longer mean segments in the variable conditions (Figure 4), here also this extra mental effort could have caused better remembering. This idea is reinforced by the fact that the SL and LS conditions yielded relatively low posttest scores for the program halves in which the within-program task was easy (where subjects worked with short segments).

The effect of mental effort has been discussed by Salomon (1984). According to Salomon, the Amount of Invested Mental Effort (AIME) is related to the

Perceived Demand Characteristics (PDC) of a task, and the Perceived Self Efficacy (PSE) for that task. PDC, PSE and AIME can be balanced in many ways, yielding different levels of learning. The performance level while answering questions that a subject considers to be acceptable thereby forms one factor that can be balanced with preferred segment length in the variable conditions. The fact that mean segment length differs considerably in these conditions means that different subjects made different trade-offs with respect to their memory capacity and their willingness to invest mental effort. It may however be that in general subjects in the variable conditions preferred a relatively relaxed approach to their task and thus choosed segment lengths that allowed reasonable performance levels with a moderate AIME. This could explain why LIN subjects, who faced harder PDCs, performed relatively well on the different post- and retention tests. They tuned their AIME to the task and as the task was difficult, their AIME was substantial which caused relatively good storage in memory.

Similar reasoning can be applied to the SL and LS conditions. In these conditions the posttest questions and the retention-test questions were answered relatively well for those halves of the program where the subjects worked with long segments (which was relatively difficult) and relatively badly for the halves where they worked with short segments (and felt no need to invest much mental effort).

In all, the results suggest that designers have much freedom in choosing segment lengths when developing interactive video programs, and also if they decide to use segments with fixed lengths. The point is to inform learners of the PDC of each segment ahead and motivate them to tune their AIME to the needed level. For this, in principle, many audiovisual communication options are available (see again Verhagen, in preparation, for a more detailed discussion of this issue).

A last observation with respect to delayed recall is that there appear to be recency effects. This follows from the differences between the scores on the two halves of the program of the VAR and CROSSED conditions in particular, especially where the posttests are concerned. As is apparent from Figures 6 and 7 and from Table 4, the VAR subjects scored better than CROSSED subjects on posttest questions with respect to the second half of the program and worse on questions about the first half, while the scores on the posttests about the program as a whole (Figure 5) showed no noticable differences. VAR subjects saw the second half of the program later than the first half and the CROSSED subjects just the other way around. The mean-time difference between the two halves was about one hour and 15 minutes. The most probable explanation for the different test results is therefore that forgetting starts immediately after learning and that after 75 minutes the forgetting curve is already important enough to cause the observed differences.

Information load as a function of the position in the progam

Under the assumption that the influence of individual differences of subjects will disappear if the mean segment length for each information element in Figure 8 are obtained as the mean of many subjects, the "ideal" curve would be a straight horizontal line. In that case form and content of the program would have no influence on preferred segment length in the sense that the information load of the program would then be perceived to be constant through out.

This appears not to be the case. This is reason to analyse more carefully various formal featu:es of the program such as factors like complexity of narration, mutual influence of picture and sound, use of super-imposed texts, and information load due to technical terms. The analysis of these factors goes beyond the scope of this paper. More information can be found with Verhagen (in preparation). A notable aspect is the occurrence of the peak before the middle of the program. It concerns the only part of the program with an animation sequence. This sequence has however also some repetition of content elements in it. The conclusion that the animation sequence is perceived as more easy as the normal video sequences can thus not be drawn without further analysis.

A clear trend that shows whether subjects decided to take longer segments towards the end of the program or on the contrary chosed shorter ones did not emerge.

Fatigue effects

The last research question was whether fatique effects affect self-chosen segment length. Figure 8 showes that there appears to be no reason to take these into account. The curves of the VAR and CROSSED conditions are very similar with the exception of the dip in the curve of the CROSSED condition round the middle of the program. There is, however, an easy explanation for this. The middle of the program was the forced end of the program for the CROSSED subjects. They thus could not choose the length of their last segment freely. This last segment appeared often to be relatively short, which caused the mean segment length to drop on that place (an effect that can also be observed at the end of the program for both conditions, where the VAR subjects stopped and the CROSSED subjects were half way). Similarly, in the CROSSED condition elements just after the middle could not be part of segments that began before the middle, causing these elements to be part of relatively many short segments. Apart from this, segment length appears not to be substantially different between the VAR and CROSSED conditions. Given the mean time difference of 75 minutes with which VAR or CROSSED conditions arrived at a similar point in the program, it seems that fatigue effects did not affect segment length.

Conclusion

The main results of the study are:

1. Self-chosen mean segment length varies from less than two information elements up to more than 36 information elements, with only a small increase in the amount of wrong answers to questions within the program. Subjects seem to adjust the amount of information they choose to get in one time to their memory capacities and to their willingness to invest mental effort.
2. If a difficult task is expected, subjects seem to adapt themselves by investing more effort. This could be the reason why the subjects in the LIN condition performed relatively well on posttests and retention tests and why subjects in the SL and LS conditions did better in this respect when questions related to that half of the program in which they worked with long segments.
3. No differences are observed between the VAR and the CROSSED conditions which can be attributed to fatique effects. Posttest scores show a recency effect, suggesting that the time difference between VAR and CROSSED of

about 75 minutes yields so much forgetting about the program half that was studied first, that this already affects posttest performance.

The wide variety of preferred length yields little guidance for designers who design interactive video programs with fixed-length segments. The main message may be that quality communication can be accomplished in many forms, whereby different audiovisual formats and segment lengths may appear feasible as long as learners start their tasks with a realistic expectancy of the demand characteristics and can be motivated to tune their mental effort accordingly. A further discussion goes beyond the purpose of this paper. Detailed discussion and further analyses are reported elsewhere (Verhagen, in preparation).

References

Bork, A. (1987). Lessons from computer-based learning. In: D. Laurillard (Ed.), *Interactive media: Working methods and practical applications.* (pp.28-43). Chichester: Ellis Horwood.

Chase, W.G., & Clark, H.H. (1972). Mental operations in the comparison of sentences and pictures. In L. Gregg (Ed.), *Cognition in learning and memory.* New York: Wiley.

Gagné, R.M. (1985). *The conditions of learning and theory of instruction.* (Fourth edition.) New York: Holt-Saunders International.

Keller, J.M., & Kopp, T. (1987). Application of the ARCS model of motivational design. In C.M. Reigeluth (Ed.), *Instructional theories in action: Lessons illustrating selected theories and models.* (pp.289-320). Hillsdale, NJ: Lawrence Erlbaum Associates.

Keller, J.M. & Suzuki, K. (1988). Use of the ARCS Motivation Model in Courseware Design. In D.H. Jonassen (Ed.): *Instructional Designs for Microcomputer.* (pp.401-434). Hillsdale, NJ: Lawrence Erlbaum Associates.

Keller, J.M. (1983). Motivational design of instruction. In C.M. Reigeluth (Ed.), *Instructional-design theories and models: An overview of their current status.* (pp.383-434). Hillsdale, NJ: Lawrence Erlbaum Associates.

Laurillard, D. (1987). Pedagogical design for interactive video. In D. Laurillard (Ed.), *Interactive media: Working methods and practical applications.* (pp.74-90). Chichester: Ellis Horwood.

Merrill, M.D. (1983). Component Display Theory. In C.M. Reigeluth (Ed.), *Instructional design theories and models: An overview of their current status.* (pp.279-333). Hillsdale, NJ: Lawrence Erlbaum Associates.

Oborne, D.J. (1987). *Ergonomics at work.* Chichester: John Wiley & Sons.

Pezdek, K., Simon, S., Stoeckert, J., & Kiely, J. (1987) Individual differences in television comprehension. *Memory & Cognition, 15* (5), 428-435.

Phillips, T., Hannfin, M., & Tripp, S. (1988). The effects of practice and orienting activities on learning from interactive video. *Educational Communications and Technology Journal, 36*, 93-102.

Romiszowski, A.J. (1986). *Developing auto-instructional materials*. London: Kogan Page.

Salomon, G., (1984). Television is "easy" and print is "tough": The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology, 76*, (4), 647-658.

Schaffer, L., & Hannafin, M.J. (1986). The effects of progressively enriched interactivity on learning from interactive video. *Educational Communication and Technolcgy Jo:arnal, 34*, 89-96.

Verhagen, P.W. (in preparation). *Length of segments in interactive video programs*. Dissertation. Enschede: University of Twente.

Wager, W., & Gagné, R.M. (1988). Designing computer-aided instruction. In D.H. Jonassen (Ed.), *Instructional designs for microcomputer courseware*. (pp.35-60). Hillsdale, New Jersey: Lawrence Erlbaum.

Witkin, H.A., Oltman, P.K., Raskin, E., & Karp, S.A. *A manual for the Embedded Figures Tests*. Palo Alto, California: Consulting Psychologists Press, Inc.