An instructional environment for learning to solve legal cases: PROSA

Muntjewerff, A.J.

Citation for published version (APA):
Muntjewerff, A. J. (2001). An instructional environment for learning to solve legal cases: PROSA.

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 7

Evaluating PROSA

7.1 Introduction

Once designed and built, the next phase of the life cycle of PROSA consists of assessing the (intended) effects of PROSA. This chapter is on evaluating the effectiveness of PROSA. We follow the distinction between validation and evaluation of instructional materials as described by Romiszowski (1986), whereby evaluation is further distinguished as either developmental testing or field testing. To summarize:

- validation
- evaluation
  - developmental testing
  - field testing

The term ‘validation’ means the inspection of the materials. The essence of materials validation is to check if the ‘content’ agrees with what was intended. This check can be performed using either a quantitative technique as convergent measurement or a more qualitative approach by having experts assess the materials. For validating the materials in PROSA we asked two expert to assess the materials. The first expert checked the subject matter content for correctness and completeness. The expert worked as a researcher and teacher at the department of administrative law at the Faculty of Law of the University of Amsterdam. A second expert of the same department who teaches the subject matter validated the legal cases designed for PROSA.

---

1 In Dutch: leerstoelgroep Bestuursrecht.
On the basis of their comments on content, correctness and relevance (minor) changes were made in some descriptions of domain concepts and in some legal solutions belonging to the legal cases.

The evaluation of instructional materials on the other hand involved two stages of verifying the effectiveness, efficiency and suitability of the materials. Evaluation is described by Romiszowski as:

[...] measuring the results achieved by the materials when used by the learners (Romiszowski, 1986, p. 399).

The distinction between developmental testing and field testing is made based on the difference in function these tests have and based on the different types of information gathered by these tests. Both tests also differ in the procedures to be followed. Developmental testing is concerned with evaluation of the materials during the development of the materials. Field testing is concerned with evaluation of the materials during their actual use in teaching.

Developmental testing is similar to “formative” evaluation where the objective is to check for unexpected effects (errors) and to obtain suggestions for improving effectiveness by trying out the system.

The function of the developmental testing session is to get to know the target population better, improve one’s ability of communicating effectively with the target population and ‘catching the big problems’ in the material at the earliest possible opportunity (Romiszowski, 1986, p. 412).

From a methodological, life cycle and product oriented view, developmental testing is a preparatory phase to optimize the likelihood to obtain success in the next phase: the proof of the pudding. Field testing is the evaluation of the materials in the real-life situation in which the materials are intended to be used\(^2\). A field testing session is carried out to provide information on what the students have learned as a result of using the materials that were designed. This is carried out by measuring exactly what the students have learned as a result of using the materials.

\(^2\) Although it may be argued that in our field test the emphasis is on the testing character more than on the field character, the fact is that in our test the students are working with PROSA in the educational context and setting for which it is planned.
The field-testing [...] seeks to supply information on how the materials function in the reality for which they were planned (Romiszowski. 1986, p. 412).

The materials evaluation of PROSA involves both a developmental test and a field test. The developmental test was carried out during the design and implementation of the computer program. This developmental test is described in paragraph 7.2.

A field test was carried out with a complete version of PROSA to determine if law students learn what they are expected to learn from working with PROSA. In the field test we return to our leading research questions introduced in Chapter 1 that read:

- Will it prove helpful to law students to present them with an instructional environment for solving legal cases which, rather than presenting an explicit method for legal case solving, offers guided access to the subject matter content?
- Will it prove helpful to law students to present them with an instructional environment for solving legal cases that serves as an external memory?

This field test is described in paragraph 7.3.

7.2 Developmental Test

Potential defects, errors and malfunctioning of PROSA were checked by members of the target population. We describe the objectives, the design of the test and the subjects involved. The subsection closes with presenting the results.

7.2.1 Objectives

Two, small, developmental tests were carried out. The first one was carried out early in the construction process to get a coarse picture of the usability of the user interface and the functionality.
Chapter 7 Evaluating PROSA

Here one student played the role of informant. The second test was carried out with a first complete version to test the acceptability of the design and to correct any defects and weaknesses in the program. Here a small number of students, 7 altogether, played the role of informants.

7.2.2 Experimental Design

In the first developmental test one student worked with a first version of the computer program PROSA in the presence of the experimenter. This version of PROSA was still incomplete, its basic design was implemented, but only a few cases could be run. The student was asked to comment on appearance and content and to formulate a list of wishes. The session was a highly interactive one, in which the experimenter (designer) probed the student actively for information and arguments.

In the second test, 7 students worked with a complete version of the computer program PROSA. The experimenter was present to monitor the behavior of the student, to intervene when necessary, and to encourage comments. The comments, suggestions and questions put forward by the student during her/his session with PROSA were recorded on tape. Each session concerned one student who had to solve legal cases in PROSA. A short introduction about the objective of PROSA, how to operate the computer program and how to solve a legal case in PROSA were given, after which the student started to solve legal cases in PROSA. Each session closed with the following set of questions to be answered by the student.

- what do you think about the ease of operation?
- could you easily find what you were looking for?
- what do you think of the interaction?
- what do you think of the cases?
- what do you think of the legal case solving?
- was it easy for you to do the things you wanted to do?
- was it easy for you to see everything?
- is there anything you want to have included?
- is there anything you want to have altered?
- what do you like about the computer program?
- what do you dislike about the computer program?
7.2.3 Subjects

In the first developmental test the subject was a law student of the Faculty of Law at the University of Amsterdam. The main requirement for selecting the student was that he recently followed a course on administrative law, this being the subject matter domain in PROSA.

In the second developmental test the subjects were 7 law students of the Faculty of Law at the University of Amsterdam.

A notice with a short description of the study was put in the University of Amsterdam weekly (Folia) and was circulated in the Faculty of Law. Students were also recruited from the course on administrative law given by the department of administrative law. Participation in the test was voluntary.

The main requirement for selecting a student was that she followed or had followed a course on administrative law, being the domain of practice in PROSA.

7.2.4 Results

The first developmental test was carried out to get feedback as early as possible on the user interface. The student worked with the computer program under construction and was asked to comment.

Although only a test version, the overall opinion of the student on PROSA was very positive. The student was very enthusiastic about the arrangement of the screen with the legal case window on one side, the legal rules window on the other end and the construct legal solution window in the middle. He experienced working this way as very supportive and "natural". This session with PROSA also showed the importance of incorporating an introduction to PROSA. Besides information on how to work with the program the introduction should contain information on what the student will learn from working with PROSA.

The second developmental test was carried out to evaluate the acceptability of the system and to find deficiencies. The students had to work with PROSA for three hours and, although they were allowed to take a break whenever they wanted one, they all worked continuously. The students were stimulated to ask questions and to ask for help whenever they needed it. They were asked to comment on the ease of operating the computer program as well as on the content of PROSA. The comments were divided into two categories.
Chapter 7 Evaluating PROSA

The first category lists comments and suggestions with regard to the appearance of the computer program and how to operate it, the second category sees to comments with regard to the content of the program. All comments and suggestions made pointed in the same direction showing that there is a large consensus among the students about PROSA. All students were very positive in their comments and said they liked to work with PROSA. They did not experience working with it as boring or annoying. All students stated that the most important aspect of PROSA is that the computer program forces you to work in a systematic way to be able to construct a complete legal solution. All students stated that PROSA is very useful in preparing oneself for an exam. They all experienced it as an important supplement to existing educational practice.

7.2.4.1 “Look and feel” and operation

All students reacted very positive to the design of the screen. They liked the arrangement of the different parts on the screen, their colors and the buttons. They commented on it as very appealing and challenging. All students were given a short introduction on how to work with PROSA and all but one student found the computer program easy to operate. This student stated in fact that she herself had no problems, but that she expected other students to have problems working with PROSA. All but this one student could operate the computer program with ease after one or two cases. All students commented positive on the large screen option (very convenient), the list of concepts, especially the references to applicable articles, and the ‘match article component - fact’ option. Four students wanted to add some argumentation to their solution either by typing in text or by dragging and dropping text from the list of concepts. However, this need faded out the longer the students worked with PROSA.

7.2.4.2 Content

The students stated that the legal cases presented by PROSA were largely similar to what was presented in the course and in the examination they had taken as part of their study program.

---

3 Two of the seven students, preparing themselves for the administrative law exam in April, came back to work for another three hours with PROSA.
Four students commented that when selecting an article from a statute in the legal rules window to copy it to the construct legal solution window, the article number was not copied with it. The students suggested to add the article number because this makes it easier to refer to the article in the statute, or to find related articles later on in the solution construction process. All students found the list of concepts particularly useful and suggested a number of concepts to be added to the list. The developmental test also identified some technical errors, for example problems with the drag and drop functionality. On the basis of the outcomes of this developmental test it appeared that no major adaptations or changes to PROSA’s content and design were necessary, while some technical errors were repaired.

7.3 Field Test

To test whether PROSA really is effective, an experimental study was performed. Two criterion tests were used. The first one consisted of tests designed by ourselves and being completely representative for PROSA’s content. The second consisted of the official examination. The course on administrative law is officially completed with an exam. This exam consists of legal cases to be solved by the student. Because in the exam the student has to construct legal solutions for the specific problem situations presented, this appeared to be a proper setting for testing. Therefore our field test consists in fact of two experiments in which the same students (subjects) participated. The main hypothesis for these experimental studies is:

- PROSA helps students to improve their legal case solving performance.

To be able to test our main hypothesis we designed two experiments. Experiment I is designed to test the effectiveness of PROSA as such, where experiment II is designed to test the effectiveness of PROSA in relation to the official exam. We first describe the design, the planning, the subjects and the results of experiment I. This is followed by a description of experiment II.
7.3.1 Experiment I

To test our main hypothesis that PROSA helps students to improve their legal case solving performance we tested the effectiveness of PROSA as such.

7.3.1.1 Experimental design

Subjects were distributed over an experimental and a control group. The experimental group received the PROSA training between the pre-test and the post-test. The control group did not (see Fig. 7.1).

<table>
<thead>
<tr>
<th></th>
<th>pre-test</th>
<th>post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>experimental group</strong></td>
<td>$O_1$</td>
<td>PROSA</td>
</tr>
<tr>
<td><strong>control group</strong></td>
<td>$O_3$</td>
<td></td>
</tr>
</tbody>
</table>

Figure 7.1: Experimental design Experiment I. $O =$ Observation (data gathering), PROSA = experimental treatment, $t_1 =$ time 1 (August 9, 1999), $t_2 =$ time 2 (August 16, 1999).

The tests

All tests (pre and post) consisted of two parts. Part one contained the legal case solving test. Both the experimental and the control group received five legal cases in the domain of administrative law. Both groups had to write down their solutions and hand them in to the experimenter. During the test the students were allowed to use their law books$^4$.

The second part of the test contained a multiple-choice test. The participants in both groups had to hand in their answers to 15 multiple choice questions on administrative law. Both the experimental group and the control group performed the same tests. For all tests subjects had two hours altogether to complete their work.

$^4$ In Dutch: wetboeken.
Experimental treatment: PROSA
The experimental treatment consisted of working with PROSA and minimally solving six cases\(^5\). The student had to fill in a form on which she had to keep record of the hours per session she had worked with PROSA\(^6\). We also gathered data on students’ opinions about PROSA. At the end of the experimental treatment the student had to fill in a questionnaire where she could state how much she liked working with PROSA, how useful she thought the program to be and to what degree she thought she has improved in solving administrative law cases. The questionnaire also allowed the student to write down her comments on PROSA\(^7\).

7.3.1.2 Operationalisation and hypotheses

Our main hypothesis stating that PROSA helps students to improve their legal case solving performance is first operationalized in the following hypothesis:

- students who worked with PROSA improved their legal case solving performance.

Translated in testable hypotheses this means:

\[ h_0 = \text{students do not improve their legal case solving performance} \]
\[ h_1 = \text{students improve their legal case solving performance} \]

The control group is used to be sure that the results can be attributed to the experimental treatment.

---

5 At the time of the experiment PROSA contained 25 legal cases.
6 The students in the experimental group were allowed to work with PROSA during a period of one week. They could go to the computer room where PROSA was installed on the machines and start working. The experimenter was present all the time in case technical problems or other difficulties might arise. There was one student who worked only for 3 hours, there were two students who worked 11 and 12 hours respectively. The mean for the experimental group is 7 hours and 25 minutes.
7 See Appendix F.
Our control hypothesis therefore was:

- students who did not work with PROSA did not improve their legal case solving performance.

Translated in testable hypotheses this means:

\[ h_0 = \text{students improve their legal case solving performance} \]
\[ h_1 = \text{students do not improve their legal case solving performance} \]

To check whether the experimental group and the control group are interchangeable there should also be no significant difference between the scores on the pre-test for both groups \((O_1 = O_2)\).

### 7.3.1.3 Planning and subjects

The population from which we took our sample consisted of law students of the Faculty of Law at the University of Amsterdam who were preparing themselves for the administrative law exam to be held on the 25th of August 1999. Because the students had to sign up for the exam before the 8th of July we were able to contact the students. All 460 students who had signed up received a letter in the second week of July with an invitation to participate in the field test.

In this letter the purpose of the test was described, testing the instructional effectiveness of PROSA, as well as the indication of a reason for the students to participate, being extra practice in solving cases in the domain of administrative law as a preparation for the coming exam.

The students were invited to attend a first session, the pre-test 8. At this first meeting 43 students showed up 9. During this meeting students signed up for participating either in the experimental group or in the control group.

---

8 The pre-test took place on the 9th of August 1999. The experimental treatment was scheduled from 10 to 13 August 1999. The post-test was scheduled on the 16th of August 1999.

9 The number of responses is in accordance with the standard percentage of responses on surveys being 10%. Taking into account that it was also the holiday season we were rather happy with this number.
Table 7.4 shows the partition of the number of tries. We restricted our selection to those students who tried and failed no more than two times.

Table 7.4: Number of tries.

Table 7.5 shows the partitioning of the grades on the last exam.
Based on these findings we may conclude that both groups are notable similar.

### 7.3.1.4 Materials

The materials used in the pre-test and the post-test were legal cases to be solved by the student and multiple choice questions to be answered by the student. The legal cases presented in the pre-test and the post-test were a selection of cases designed for PROSA, but not yet implemented.
An Instructional Environment for Learning to Solve Legal Cases

The topics of the cases in both the pre-test and the post-test were the same, as was the order of the topics.

- case 1 interested party
- case 2 order
- case 3 administrative authority
- case 4 appeal
- case 5 objection

As an example we show the interested party case used in the pre-test (see Fig. 7.2).

The soccer club SHO wants to move to municipal grounds at the border of the buildings of the municipality of Oud Beijerland. Major and aldermen propose to the city council to rent the municipal grounds to SHO for a ten year period. The city council agrees with this proposal and instructs the Major of Oud Beijerland to make an agreement with the soccer club.

On the 2nd of March 1998 the written agreement is made between the municipality Oud Beijerland and SHO.

Wim van Prooijen is the owner of a piece of land with an agrarian use. This piece of land is situated next to the piece of land where SHO will move to. Wim van Prooijen was engaged in negotiations with the municipality about purchasing this particular piece of land. Wim van Prooijen wants to use the land to expand his agrarian farm. Wim van Prooijen is furious.

Is Wim van Prooijen an interested party according to the Gala?

Figure 7.2: The first test case ‘soccer club SHO’.

The multiple choice questions were based on the textbook used in the course on Administrative law and concerned the major concepts in the domain, asking either for the correct description of a concept or the correct instance (Nicolai et al., 1997). As an example we show a multiple choice question used in the post-test (see Fig. 7.3).
The Governor of Zeeland is

a. a natural person
b. a legal person in private law
c. a legal person in public law
d. a body of a legal person in public law

Figure 7.3: Example of a mc question.

The materials in the experimental setting were the legal cases to be solved in PROSA.

The students had to solve five legal cases in the pre-test and five legal cases in the post-test. Students’ legal case solving performance is assessed on the basis of the legal solution they produce. Therefore the students’ legal solutions were graded on structure and completeness. The correct solution for a particular case was used as the grading model for that case. The correct solution indicated 2 points for each necessary component (including the right answer to the question), 1 point for the correct place of the component, no grades for a missing component, a correct component in the wrong place or a wrong answer, and minus 1 point for each unnecessary component. This grading procedure is illustrated using the correct legal solution of the ‘soccer club SHO’ case (see Fig. 7.4). Each legal solution of the student was compared with the model solution. This resulted in a student score for each legal case and the sum of scores for all five cases. The sum was used to compare the results within groups and between groups (see Appendix B1).
Soccer club SHO legal solution (total score 36)

Section 1:2 subsection 1 of the Gala
Interested party means the person whose interest is directly affected by an order. (2) (1)

the person = Wim van Prooijen (2) (1)

Section 1:3 subsection 1 of the Gala
Order means a written ruling of an administrative authority constituting a legal act under public law. (2) (1)

a written ruling = the written agreement (2) (1)

Section 1:1 subsection 1 of the Gala
Administrative authority means:
(a) a body of a legal person which has been established under public law, or
(b) another person or body which is invested with some public authority. (2) (1)

Section 2:1 subsection 1 of the Civil Code
The state, the provinces, the municipalities, the district water boards and also all bodies with statutory powers under the Constitution possess legal personality. (2) (1)

Section 6 of the Municipal Corporations Act
Each municipality has a city council, Mayor and Aldermen and a mayor. (2) (1)

a body of a legal person which has been established under public law = de mayor of Oud Beijerland (2) (1)

legal act = the written agreement is made between the municipality Oud Beijerland and SHO (2) (1)

the legal act has to be under public law, however that is not the case it is under private law (2) (1)

legal act under public law is not the rent agreement (2) (1)

Therefore Wim van Prooijen is not an interested party in according to the Gala. (2) (1)
Chapter 7 Evaluating PROSA

7.3.1.5 Main results Experiment I

We carried out the experiment to test our main hypothesis that PROSA helps students to improve their legal case solving performance. We predicted that students who worked with PROSA would show an improvement in their legal case solving performance. Here we present the results.

![Graph showing comparison between experimental and control group improvement]

Figure 7.5: Experimental and control group improvement.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>case solving score pre-test mean</th>
<th>Std. Deviation</th>
<th>case solving score post-test mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>experimental group</td>
<td>13</td>
<td>42.92</td>
<td>16.54</td>
<td>76.46</td>
<td>21.65</td>
</tr>
<tr>
<td>control group</td>
<td>13</td>
<td>43.31</td>
<td>16.91</td>
<td>38.92</td>
<td>17.36</td>
</tr>
</tbody>
</table>

Table 7.6: Legal case solving results experiment I.
Our prediction was confirmed. When we look at the means for the pre- and post-tests we see that the pre-test means of both groups are more or less the same, where in the post-test the mean for the experimental group is twice the mean of the control group.\footnote{We also see that there is a large standard deviation. This indicates that there are major differences in scores between students.}

PROSA is aimed at improving the skill in solving legal cases, not in acquiring new legal concepts. In order to check whether the (large) effect we found for the experimental group was not (partially) due to some “Hawthorn” effect (or other general, not identified, nor predicted factors) we also had a look at the data from the multiple choice tests. Here we predict of course no effects for both groups. The results are presented here.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>topic score pre-test mean</th>
<th>Std. Deviation</th>
<th>topic score post-test mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>experimental group</td>
<td>13</td>
<td>11.69</td>
<td>1.25</td>
<td>11.46</td>
<td>1.71</td>
</tr>
<tr>
<td>control group</td>
<td>13</td>
<td>11.23</td>
<td>1.69</td>
<td>12.23</td>
<td>1.54</td>
</tr>
</tbody>
</table>

Table 7.7: Multiple choice results experiment I.

Our prediction was confirmed. When we look at the means for the pre- and post-tests we see that the pre-test means of both groups are more or less the same, where in the post-test the mean for the control group is even somewhat higher than the mean for the experimental group.

### 7.3.1.6 Tests

To find out if the results we found were significant we selected a series of tests. Because the predicted effect is an interaction effect we did not use a variance analysis (VA). VA is not particularly suitable for testing an interaction effect. We therefore selected the Paired Samples T-Test and the Independent Samples T-Test. We performed a Paired Samples T-Test where we tested the null hypothesis that the mean of differences equals 0. We used an Independent Samples T-Test to determine the equality of means.
Chapter 7 Evaluating PROSA

VA also has relatively strong assumptions regarding the normal distribution of the population, therefore we also used the (non-parametric) Wilcoxon Signed Ranks Test where the hypothesis of no improvement was tested. We first present the \( t \)-tests outcomes and Wilcoxon outcomes of the legal case solving results. These are followed by the presentation of the \( t \)-tests outcomes and Wilcoxon outcomes of the multiple choice results.

<table>
<thead>
<tr>
<th>comparing legal case solving results</th>
<th>( t )</th>
<th>( p ) (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>experimental group pre-post(^{13})</td>
<td>-4.549</td>
<td>0.001</td>
</tr>
<tr>
<td>control group pre-post(^{14})</td>
<td>1.566</td>
<td>0.143</td>
</tr>
<tr>
<td>experimental group - control group pre-pre(^{15})</td>
<td>-0.059</td>
<td>0.954</td>
</tr>
</tbody>
</table>

Table 7.8: \( t \)-tests on legal case solving results.

<table>
<thead>
<tr>
<th>comparing legal case solving results</th>
<th>( p ) (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>experimental group pre-post(^{16})</td>
<td>0.002</td>
</tr>
<tr>
<td>control group pre-post(^{17})</td>
<td>0.152</td>
</tr>
</tbody>
</table>

Table 7.9: Wilcoxon tests on legal case solving results.

<table>
<thead>
<tr>
<th>comparing multiple choice results</th>
<th>( t )</th>
<th>( p ) (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>experimental group pre-post(^{18})</td>
<td>0.507</td>
<td>0.621</td>
</tr>
<tr>
<td>control group pre-post(^{19})</td>
<td>-2.142</td>
<td>0.053</td>
</tr>
<tr>
<td>experimental group - control group pre-pre(^{20})</td>
<td>0.791</td>
<td>0.401</td>
</tr>
</tbody>
</table>

Table 7.10: \( t \)-tests on multiple choice results.

\(^{13}\) See Appendix C2.1.
\(^{14}\) See Appendix C2.2.
\(^{15}\) See Appendix C3.1.
\(^{16}\) See Appendix C1.1.
\(^{17}\) See Appendix C1.2.
\(^{18}\) See Appendix C2.3.
\(^{19}\) See Appendix C2.4.
\(^{20}\) See Appendix C3.2.
An Instructional Environment for Learning to Solve Legal Cases

PROSA

<table>
<thead>
<tr>
<th>comparing multiple choice results</th>
<th>p (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>experimental group pre-post</td>
<td>0.567</td>
</tr>
<tr>
<td>control group pre-post</td>
<td>0.053</td>
</tr>
</tbody>
</table>

Table 7.11: Wilcoxon tests on Multiple Choice Results.

7.3.1.7 Conclusions experiment I

The legal case solving results show a huge effect of PROSA. There is an obvious and significant improvement on the post-test for the experimental group. The multiple choice results show that there is no improvement in availability of knowledge on topics for the experimental group. This indicates that the effect we found can entirely be attributed to an improvement in legal case solving skills thanks to PROSA. PROSA improves students legal case solving skills.

If we look at the correlation coefficients we see a high correlation for the control group between the two tests (Table 7.12). It would be normal to have a similar correlation for the experimental group as well. Normally good students benefit more from support than poor students, so the differences between both groups of students will increase on the post test results. In our study all students in the experimental group, but one, improve after working with PROSA. However, it appears that students who have a low score on the pre-test improve to a larger degree than students who had higher scores on the pre-test (see Appendix B1). This bizarre effect, which by the way cannot be attributed to 'the ceiling effect', requires further study. PROSA helps extremely well, however, it seems unpredictable how much one will benefit from working with PROSA.

<table>
<thead>
<tr>
<th>comparing legal case solving tests</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>experimental group pre-post</td>
<td>0.050</td>
</tr>
<tr>
<td>control group pre-post</td>
<td>0.827</td>
</tr>
</tbody>
</table>

Table 7.12: Correlation coefficient between the two tests.

---

21 See Appendix C1.3.
22 See Appendix C1.4.
Chapter 7 Evaluating PROSA

7.3.2 Experiment II

In Experiment II we tested the same main hypothesis as in Experiment I, i.e. whether PROSA improves solving cases in administrative law. Now the criterion consisted of the scores on the official examination. We ‘re-used’ the subjects from Experiment I, whereby those who participated in the control group also got the opportunity to train themselves by using PROSA.

7.3.2.1 Experimental design

In our second experiment we reused the experimental group and the control group. The students in the control group worked with PROSA after the post-test. However, we did not test if they improved their performance after working with PROSA. We only had them work with PROSA to make it possible to compare the results of students who worked with PROSA with the results of the other students who did participate in the exam, but who did not work with PROSA (see Fig. 7.6 and Fig. 7.7).

<table>
<thead>
<tr>
<th>PROSA students</th>
<th>pre-test $t_1$</th>
<th>post-test $t_2$</th>
<th>exam $t_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>experimental</td>
<td>$O_1$</td>
<td>PROSA $O_2$</td>
<td>$O_5$</td>
</tr>
<tr>
<td>control</td>
<td>$O_3$</td>
<td>$O_4$</td>
<td>PROSA</td>
</tr>
<tr>
<td>Non PROSA students</td>
<td></td>
<td></td>
<td>$O_6$</td>
</tr>
</tbody>
</table>

Figure 7.6: Experimental design Experiment I and Experiment II. $O$ = Observation (data gathering), PROSA = Experimental treatment, $t_1$ = time 1 (August 9, 1999), $t_2$ = time 2 (August 16, 1999), $t_3$ = time 3 (August 25, 1999).
7.3.2.2 Operationalisation and hypotheses

We operationalized our main hypothesis that PROSA helps students to improve their legal case solving performance in the following hypothesis:

- students who worked with PROSA performed better on the official exam than students who did not work with PROSA.

Translated in testable hypotheses this means:

\[ h_0 = \text{PROSA students did not perform better on the exam than Non PROSA students} \]
\[ h_1 = \text{PROSA students performed better on the exam than Non PROSA PROSA students} \]

7.3.2.3 Planning & subjects

The official administrative law exam took place on the 25th of August 1999. The exam consisted of four cases. The students had to solve the four legal cases within three hours. The students were allowed to use their law books, case law materials and the doctrine. The examiners corrected the exams resulting in a score for each student. The subjects were law students of the Faculty of Law at the University of Amsterdam who showed up at the administrative law exam.
Chapter 7 Evaluating PROSA

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students who signed up for the exam</td>
<td>460</td>
</tr>
<tr>
<td>Students who showed up at the exam</td>
<td>334</td>
</tr>
<tr>
<td>PROSA students</td>
<td>36</td>
</tr>
<tr>
<td>Non PROSA students</td>
<td>298</td>
</tr>
</tbody>
</table>

Table 7.13: Subjects.

In our first experiment 22 students had signed up for the experimental group and 21 for the control group. The experimental group worked with PROSA being the experimental treatment in experiment I. The control group worked with PROSA after the post-test of the first experiment. So both groups worked with PROSA. However, the number of students in the experimental group that actually appeared at the experimental treatment was 21, were 17 students of the control group showed up at their opportunity to work with PROSA. This resulted in a total number of 38 students who worked with PROSA. However, although 38 students had worked with PROSA only 26 of them participated in all parts of the first experiment. There were two PROSA students who did not participate in the exam bringing the number of PROSA students who participated in the exam to 36. Specifying the PROSA students who took the exam shows:

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROSA experimental group</td>
<td>13</td>
</tr>
<tr>
<td>PROSA control group</td>
<td>12</td>
</tr>
<tr>
<td>PROSA others</td>
<td>11</td>
</tr>
<tr>
<td>Non PROSA students</td>
<td>298</td>
</tr>
</tbody>
</table>

Table 7.14: PROSA and Non PROSA student numbers.

23 Students in the control group could work with PROSA from 17 to 20 August 1999. In this period students also had to prepare themselves for the exams in August, so we expected students to drop out.
7.3.2.4 Materials

The materials in the exam were legal cases designed by the teachers of the administrative law course and handed out on paper to be solved by the students. The official administrative law exam consisted of four cases, being a short situation description, and a series of questions related to this case to be answered by the student. Here we show one of the cases in the exam (see Fig. 7.8).

Already for many months Greet Bokkestein quarrels with the Municipal Social Service. She blames the Service that they have wrongfully withdrawn her benefit. For some weeks she takes up her daily position at the counter of the Social Service that is seated in a building of the municipality. She directs abusive language, words such as ‘thieves’ and ‘plebeians’, to the civil servants and annoys visitors. On the 15th of July 1999 Greet Bokkestein is issued a letter in which the municipality refuses her access to the Municipal Social Services building.

Is there a legal provision to make on the basis of the Gala against the ruling in the letter?

Figure 7.8: Case in the August exam.

For each case the maximum score was indicated and this score was split in sub scores for each question belonging to the case (e.g. for the first case a student could score 10 points, being 6 points for question 1 and 4 points for question 2). The examiners used a correction model for scoring the students answers. The model contained the elements in an answer that were recognized and the related score, correct alternatives and, for some answers, frequent mistakes. An answer had to be consistent and without any grammatical errors.

7.3.2.5 Main results Experiment II

We carried out the experiment to test our main hypothesis that PROSA helps students to improve their legal case solving performance.
Chapter 7 Evaluating PROSA

We predicted that students who worked with PROSA would perform better on the official exam than students who did not work with PROSA. Here we present the results.

<table>
<thead>
<tr>
<th>N</th>
<th>Exam score mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>5.22</td>
<td>1.44</td>
</tr>
<tr>
<td>298</td>
<td>5.04</td>
<td>1.35</td>
</tr>
</tbody>
</table>

Table 7.15: Exam score means.

Our prediction was not confirmed. PROSA students did not perform better on the official exam than Non PROSA students (see Appendix D). An Independent Samples T-Test showed that there is no significant difference in exam scores between the PROSA and Non PROSA students (Table 7.16).

<table>
<thead>
<tr>
<th>comparing exam score means</th>
<th>t</th>
<th>p (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROSA -NON PROSA</td>
<td>-0.734</td>
<td>0.463 (445)</td>
</tr>
</tbody>
</table>

Table 7.16: t-test exam score means.

7.3.2.6 Summary Experiment I and Experiment II

We performed two experiments to test the effectiveness of PROSA. The first experiment tested the effectiveness of PROSA as such. The test results show a huge and remarkable difference on legal case solving performance between students who worked with PROSA and students who did not work with PROSA. This indicates the effectiveness of PROSA. PROSA helps to improve students' legal case solving skills.

However, the second experiment in which the effectiveness of PROSA was tested in relation to the official exam did not show any significant difference on exam results between students who worked with PROSA and students who did not work with PROSA.

We see ourselves confronted with apparently contradictory results. Besides the fact that we were somewhat disappointed by these results, we were very much interested what may have caused these outcomes.

24 Because there were no differences between the three PROSA groups we joined them.
An Instructional Environment for Learning to Solve Legal Cases

PROSA

We will therefore discuss these apparently contradictory findings in more detail in the next section to find a plausible explanation, even if this is a post hoc one.

7.3.3 Post Hoc Analysis

The contradictory results may be explained by one or more of the following hypotheses. For some of these we have further data available.

- the PROSA population differs from the Non PROSA population
- the cases in PROSA (like the cases in our tests) differ from the cases in the exam
- the grading of the PROSA solutions differs from the grading of the exam solutions

We will discuss these various issues that may have contributed to these apparently contradictory results in an effort to try to solve this exam mystery.

7.3.3.1 The PROSA population differs from the exam population

We do not have any directly relevant data available here. The results of the first try of the administrative law exam in April were that about 90% of the participants failed the exam. The next opportunity for taking the exam was in August. The majority of the students who failed in April took this exam in August. Among these students were the 43 students that assigned for the PROSA evaluation. The August exam results of the entire population fall in the same range as the results of the PROSA students. Therefore we state that the PROSA population mirrors the entire population. It could be that students who were willing to spent extra time participating in an evaluation study are either more motivated to succeed or are more insecure about their abilities, or both, in comparison with the other students. However, we do not have any further evidence.
Chapter 7 Evaluating PROSA

7.3.3.2 The cases in PROSA differ from the cases in the exam

Basically there are no major differences between the cases used in the exam and the cases used in the evaluation study with regard to the situation description and the question, although in the exam there were two or more, separate, questions related to one situation description, where in PROSA each situation description is combined with only one question. However, there are two main issues with regard to the cases used in PROSA and in the examination that may explain the results. For one there may be a difference in case topics. The second issue is related to the type of question related to the cases. It may be that there are items in the exam that do not require a complete legal solution.

Topics

At the time of the experiment there were 25 cases available in PROSA with the following topics and difficulty levels.

<table>
<thead>
<tr>
<th>topic</th>
<th>easy</th>
<th>medium</th>
<th>difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>interested party</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>order</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>administrative authority</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>appeal</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>objection</td>
<td>3</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 7.17: Cases in PROSA by topic and difficulty level.

Therefore students had more practice with the topics 'interested party' and 'order' than with the topics 'administrative authority', 'appeal' and 'objection'. The cases used in the legal case solving pre-test and the legal case solving post-test were comparable to the PROSA cases being designed for implementation in PROSA later on.

A closer look at which test items make the least difference for the post-test in Experiment I shows that they were concerned with administrative authority (case 3) and with appeal (case 4) (see Appendix E1). This may indicate that the students did not get enough practice in PROSA on these topics, resulting in moderate improvement on cases with these topics.

The exam cases, however, cover a larger range of topics than were available in PROSA at the time of the experiment.
Comments made by students in the opinion questionnaire also showed that they wanted more cases included both on existing topics and on other administrative law topics.

Taking into account that students could score up to 50 points for the exam a division in points for topics available in PROSA and topics not (yet) available in PROSA reveals the following picture (Table 7.18).

<table>
<thead>
<tr>
<th></th>
<th>points topic in PROSA</th>
<th>points topic not in PROSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>case I</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>case II</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>case III</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>case IV</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>total</td>
<td>13</td>
<td>37</td>
</tr>
</tbody>
</table>

Table 7.18: Points per topic for the cases that were part of the official exam.

This table indicates that a student can, for example, score 10 points on case I. However, there are different topics in the case, where only 4 points are related to topics available in PROSA at the time of the experiment. Therefore we decided to have a closer look at the scores on the cases with a topic that was available in PROSA. To check the random selection from the entire population we made an inventory of the exam results for each group. Taking the exam score mean per group shows the following picture (Table 7.19)\(^{25}\).

\(^{25}\) To be able to compare the scores on the cases in the exam that match the topics in PROSA we took the exams of the students who worked with PROSA and we took a random selection of exams of students who did not work with PROSA. Although 36 students had worked with PROSA we restricted the selection to the 26 students who participated in all our tests. However, in the experimental group three exams were not available and one student in the control group did not participate in the exam, so in the end only 22 exams were available. The exams of 3 students were not available to us because these students were engaged in legal proceedings concerning the examination. The random selection out of the total of 298 students who did not work with PROSA also consisted of 22 exams, however, on closer inspection we had to leave out one exam, because the exam number could not be traced back to a score.
Chapter 7 Evaluating PROSA

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Exam score mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROSA students</td>
<td>22</td>
<td>5.32</td>
<td>1.32</td>
</tr>
<tr>
<td>Non PROSA students</td>
<td>21</td>
<td>4.76</td>
<td>1.55</td>
</tr>
</tbody>
</table>

Table 7.19: Exam score results.

This difference in means is not significant (see Table 7.20).

<table>
<thead>
<tr>
<th>comparing exam score means</th>
<th>t</th>
<th>p (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROSA - NON PROSA</td>
<td>-1.269</td>
<td>0.211</td>
</tr>
</tbody>
</table>

Table 7.20: t-test exam score results.

A post hoc analysis of the data only for the items with a topic that was also available in PROSA\textsuperscript{26} shows a small difference (Table 7.21).

<table>
<thead>
<tr>
<th>score case I/2</th>
<th>N</th>
<th>case score mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROSA</td>
<td>22</td>
<td>1.09</td>
<td>1.34</td>
</tr>
<tr>
<td>NON PROSA</td>
<td>21</td>
<td>1.71</td>
<td>1.59</td>
</tr>
<tr>
<td>score case II/3</td>
<td>N</td>
<td>case score mean</td>
<td>Std. Deviation</td>
</tr>
<tr>
<td>PROSA</td>
<td>22</td>
<td>2.32</td>
<td>1.17</td>
</tr>
<tr>
<td>NON PROSA</td>
<td>21</td>
<td>2.05</td>
<td>1.24</td>
</tr>
<tr>
<td>score case III/6</td>
<td>N</td>
<td>case score mean</td>
<td>Std. Deviation</td>
</tr>
<tr>
<td>PROSA</td>
<td>22</td>
<td>2.91</td>
<td>0.81</td>
</tr>
<tr>
<td>NON PROSA</td>
<td>21</td>
<td>2.76</td>
<td>1.76</td>
</tr>
</tbody>
</table>

Table 7.21: Group mean per case.

However, the differences are not significant (Table 7.22).

<table>
<thead>
<tr>
<th>score case I/2 PROSA - NON PROSA</th>
<th>t</th>
<th>p (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.394</td>
<td>0.171</td>
</tr>
<tr>
<td>score case II/3 PROSA - NON PROSA</td>
<td>-0.735</td>
<td>0.467</td>
</tr>
<tr>
<td>score case III/6 PROSA - NON PROSA</td>
<td>-0.355</td>
<td>0.724</td>
</tr>
</tbody>
</table>

Table 7.22: t-test group mean per case.

There are no indications that (a difference in case) topics can explain our results.

\textsuperscript{26} Case I question 2, Case II question 3 and Case III question 6.
Legal case question
It may also be that there are items in the exam that do not require a complete legal solution. It appeared that the cases used in the exam sometimes only use the situation descriptions as an introduction to a question that refers to a 'remember concept' learning outcome instead of a 'use procedure' learning outcome. This has consequences for the content and format of the answer. For instance, in the Greet Bokkestein case, the answer to the question (e.g. Is there a legal provision to make on the basis of the GALA against the ruling in the letter) as stated in the correction model is the classification of an instance as a concept (e.g. the refusal of access to a building with a public purpose may be classified as an order in the meaning of the Gala), not a full reasoning about the case, i.e. arriving at a conclusion on the basis of legal argument. As we stated before, PROSA specifically trains the legal case solving skill and as we also showed there will be no improvement on concept questions. Therefore we actually should remove the questions that test the availability of knowledge of concepts before comparing the scores. When we have a look at the means of the score per group per case question that requires more less a full reasoning we get the following overview (Table 7.23)\textsuperscript{27}.

<table>
<thead>
<tr>
<th>Case</th>
<th>PROSA</th>
<th>NON PROSA</th>
<th>PROSA</th>
<th>NON PROSA</th>
<th>PROSA</th>
<th>NON PROSA</th>
<th>PROSA</th>
<th>NON PROSA</th>
<th>PROSA</th>
<th>NON PROSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>I/1</td>
<td>22</td>
<td>2.45</td>
<td>1.97</td>
<td>21.23</td>
<td>1.24</td>
<td>2.05</td>
<td>1.17</td>
<td>2.32</td>
<td>1.17</td>
<td>2.05</td>
</tr>
<tr>
<td>II/3</td>
<td>22</td>
<td>2.32</td>
<td>1.77</td>
<td>21.20</td>
<td>1.12</td>
<td>2.05</td>
<td>1.24</td>
<td>2.32</td>
<td>1.24</td>
<td>2.05</td>
</tr>
<tr>
<td>III/6</td>
<td>22</td>
<td>1.64</td>
<td>1.18</td>
<td>21.40</td>
<td>1.06</td>
<td>1.40</td>
<td>1.06</td>
<td>1.09</td>
<td>1.02</td>
<td>0.95</td>
</tr>
<tr>
<td>III/7</td>
<td>22</td>
<td>1.09</td>
<td>1.02</td>
<td>21.04</td>
<td>0.95</td>
<td>0.43</td>
<td>0.95</td>
<td>0.43</td>
<td>0.95</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Table 7.23: Group mean per case.

\textsuperscript{27} Case I question 1, Case II, question 3, Case III question 6 and 7, Case IV question 9.
Chapter 7 Evaluating PROSA

There is only a significant difference on case IV/9 (Table 7.24).

<table>
<thead>
<tr>
<th>Score case</th>
<th>PROSA - NON PROSA</th>
<th>t</th>
<th>p (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I/1</td>
<td>PROSA - NON PROSA</td>
<td>-0.042</td>
<td>0.967</td>
</tr>
<tr>
<td>II/3</td>
<td>PROSA - NON PROSA</td>
<td>-0.735</td>
<td>0.467</td>
</tr>
<tr>
<td>III/6</td>
<td>PROSA - NON PROSA</td>
<td>-0.355</td>
<td>0.724</td>
</tr>
<tr>
<td>III/7</td>
<td>PROSA - NON PROSA</td>
<td>-1.441</td>
<td>0.157</td>
</tr>
<tr>
<td>IV/9</td>
<td>PROSA - NON PROSA</td>
<td>-2.421</td>
<td>0.020</td>
</tr>
</tbody>
</table>

There are no indications that (a difference in case) questions can explain our results. When we look at the case questions where both the topic is available in PROSA and the question requires a reasoning we are left with only II/3 and III/6. However, results on both questions do not show any significant difference between PROSA and Non PROSA students.

7.3.3.3 The grading of legal solutions

There is a major difference in the way PROSA assesses a legal solution as complete and correct, compared to the examiners way. The grading model for the exam does not present the correct legal solution as a full reasoning, but only depicts separate elements that are remunerated; however, the full answer should be consistent. The grading model also states elements that are considered as incorrect. It could be that “the PROSA effect” remains hidden due to this differences in grading.

We decided to have a closer look at the content of the legal solutions constructed by the students. We wanted to know if students who worked with PROSA produced better structured and more complete legal solutions than students who did not work with PROSA. We re-scored the legal solution of a case in the exam following our grading principles. The correct solution for the case was used as the grading model for that case. Using our grading model of the solution we then scored the solutions of the students.

We took the following situation description from the exam and the related question, and reconstructed the legal solution (see Figs. 7.7, 7.8 and 7.9).
An Instructional Environment for Learning to Solve Legal Cases

PROSA

Already for many months Greet Bokkestein quarrels with the Municipal Social Service. She blames the Service that they have wrongfully withdrawn her benefit. For some weeks she takes up her daily position at the counter of the Social Service that is seated in a building of the municipality. She directs abusive language, words such as ‘thieves’ and ‘plebeians’, to the civil servants and annoys visitors.

On the 15th of July 1999 Greet Bokkestein is issued a letter in which the municipality refuses her access to the Municipal Social Services building.

Figure 7.9: Situation description.

Is there a legal provision to make on the basis of the Gala against the ruling in the letter?

Figure 7.10: Related question.
Chapter 7 Evaluating PROSA

GALA Article 8:1 section 1. An interested party may appeal to the court against an order. (2) (1)

GALA Article 1:2 section 1. Interested party means: the person whose interest is directly affected by an order. (2) (1)

the person = Greet Bokkestein (2) (1)

GALA Article 1:3 section 1. Order means: a written ruling of an administrative authority constituting a legal act under public law. (2) (1)

a written ruling = a letter (2) (1)

GALA Article 1:1 section 1. Administrative authority means:
(a) an authority of a legal person which has been established under public law, or
(b) another person or body corporate which is invested with any public authority (2) (1)

Civil Code Book 2 Article 1 section 1. The State, the provinces, the municipalities, the district water boards and also all bodies with statutory powers under the Constitution, have legal personality. (2) (1)

Municipality Act Article 6. Municipality Act Article 6 Each municipality consists of a city council, a Major and Aldermen and a Major. (2) (1)

a legal act (2) (1)

under public law (2) (1)

constituting a legal act under public law = refuses her access to the Municipal Social Services building (2) (1)

There is no legal provision to make on the basis of the Gala against the ruling in the letter. (2) (1)

Figure 7.11: Graded legal solution.
An Instructional Environment for Learning to Solve Legal Cases

PROSA

We again took the exams of the students who worked with PROSA and the random selection of exams of students who did not work with PROSA. For each student her legal solution was re-graded resulting in an individual score for this case. The scores for each group were added resulting in a sum for each group. The maximum score for the legal solution is 36 points. If we have a look at the mean for each group we get the following picture (Table 7.25).

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>re-graded mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROSA students</td>
<td>22</td>
<td>17.59</td>
<td>7.64</td>
</tr>
<tr>
<td>Non PROSA students</td>
<td>21</td>
<td>8.14</td>
<td>6.10</td>
</tr>
</tbody>
</table>

Table 7.25: Re-graded scores mean.

The remarkable and significant (see Table 7.26) effect of PROSA re-appears after re-grading the legal solution.

<table>
<thead>
<tr>
<th>comparing re-graded score means</th>
<th>t</th>
<th>p (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROSA - NON PROSA</td>
<td>-4.466</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 7.26: t-test re-graded scores mean.

There is a strong indication that the apparently contradictory results can be explained by the fact that the legal solutions in the official exam are graded in a different way than they are graded in PROSA. PROSA apparently has effects independent of the initial starting position of the student. To find out if we may attribute the apparently contradictory results solely to the different ways of grading we predicted a strong relation between the post-test score of the experimental group in the first experiment and the PROSA exam score, where all other relations between test scores should be low or non existent (see Tables 7.27, 7.28, 7.29).
Chapter 7 Evaluating PROSA

<table>
<thead>
<tr>
<th>Comparing legal case solving results</th>
<th>Correlation</th>
<th>p (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental group pre-test - teacher score</td>
<td>0.057</td>
<td>0.875</td>
</tr>
<tr>
<td>Experimental group pre-test - PROSA score</td>
<td>0.079</td>
<td>0.828</td>
</tr>
<tr>
<td>Experimental group post-test - teacher score</td>
<td>0.390</td>
<td>0.265</td>
</tr>
<tr>
<td>Experimental group post-test - PROSA score</td>
<td>0.849</td>
<td>0.002</td>
</tr>
<tr>
<td>Experimental group teacher score - PROSA score</td>
<td>0.328</td>
<td>0.355</td>
</tr>
</tbody>
</table>

Table 7.27: Correlation coefficient between tests for the experimental group.

<table>
<thead>
<tr>
<th>Comparing legal case solving results</th>
<th>Correlation</th>
<th>p (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group pre-test - teacher score</td>
<td>-0.417</td>
<td>0.178</td>
</tr>
<tr>
<td>Control group pre-test - PROSA score</td>
<td>-0.104</td>
<td>0.747</td>
</tr>
<tr>
<td>Control group post-test - teacher score</td>
<td>-0.295</td>
<td>0.351</td>
</tr>
<tr>
<td>Control group post-test - PROSA score</td>
<td>-0.350</td>
<td>0.265</td>
</tr>
<tr>
<td>Control group teacher score - PROSA score</td>
<td>0.182</td>
<td>0.571</td>
</tr>
</tbody>
</table>

Table 7.28: Correlation coefficient between tests for the control group.

<table>
<thead>
<tr>
<th>Comparing legal case solving results</th>
<th>Correlation</th>
<th>p (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non PROSA teacher score - PROSA score</td>
<td>-0.052</td>
<td>0.821</td>
</tr>
</tbody>
</table>

Table 7.28: Correlation coefficient between tests for the Non PROSA group.

Our prediction was confirmed. The experimental group post-test score compared to the experimental group PROSA score on the exam shows a high (0.849) and significant (0.002) correlation, where all other relations of scores show a low, or even negative correlation, and no significance. This indicates that the effect can be attributed to PROSA. We may conclude that the differences in results are solely caused by the different ways of grading between teachers and PROSA. Students who work with PROSA improve their legal case solving performance. That is, they show higher scores on tests after working with PROSA. There is, what we call, a "PROSA effect".

---

28 See Appendix E2.1.
29 See Appendix E2.2.
30 See Appendix E2.3.
However, there are two comments we have to make. In experiment II it was shown that students who worked with PROSA did not improve their exam scores. Our post hoc analysis on the grading models that were used in PROSA and in the exam showed that after re-grading the exam the PROSA-way the PROSA effect reappeared. What we call the "scoring effect" was predicted and confirmed in a comparison between post-test scores and PROSA scores on the exam for the experimental group (see Table 7.27, the correlation coefficient being 0.849). This indicates once more that working with PROSA has effect on students' (exam) scores when you assess their legal solution the PROSA way.

In subsection 7.3.1.7 we found a low correlation (0.050) between the pre- and post-test scores of the experimental group. This was unexpected, because it means that the usually positive correlation between initial performance and learning effect did not appear, e.g. the best performers usually profit most form training. PROSA appears to have a strong effect independent of the initial performance level of the student. There is even a slight, insignificant, negative correlation found for the scores of the control group where the scores on the post-test were compared with their scores on the exam in the re-graded case (see Table 7.28, the correlation coefficient being -0.350). For the control group their post-test can be seen as a pre-test when scores are compared with the PROSA scores on the exam, because the control group worked with PROSA after the post-test and before the exam. We did not perform a real post-test for the control group. If we had it could have provided more evidence for the PROSA effect being independent of the students’ initial level of performance.

In summary we can explain all these results by stating that:

1. PROSA has a strong effect, but this effect is independent of the students' initial level of performance, and
2. The effect is dependent on the way the students' solutions are scored. Our system of scoring gives a strong effect, while the way teachers score the same solutions gives no effect at all. Both ways of scoring are so different that no correlation is found between the official scores and those done the "PROSA way".

288
Chapter 7 Evaluating PROSA

7.3.3.4 Conclusions post hoc analyses

These post hoc analyses reveal that only one of three hypotheses explain our apparently contradictory results. The effect of PROSA re-appeared after re-grading the legal solution the PROSA way. We therefore conclude that the results from Experiment II were distorted by this difference in grading.

7.4 Conclusion

In this chapter we evaluated PROSA. A developmental test was carried out to identify and to correct any weaknesses in PROSA as early as possible in the construction process. The outcome of the developmental test was twofold. For one it resulted in a list with (technical) errors to be repaired. The other, more important, result was that it showed the acceptability of the design for the target population. Students not only liked to work with PROSA, they also experienced working with it as very natural. Although the PROSA interface is rather static, it did not seem to bother the students. They did not experience working with PROSA as boring. They also stated that they liked the way in which the screen organizes the components and elements in legal case solving. After completing a full working version of the program, the field test was carried out to determine the instructional effectiveness of PROSA. Our main hypothesis was that PROSA helps students to improve their legal case solving performance. We tested the effectiveness of PROSA as such. Our prediction that students who worked with PROSA would improve their legal case solving performance was confirmed on the basis of these test results. Results showed a striking improvement, the experimental group almost doubled their scores, while the control group showed no improvement at all. We also investigated whether using PROSA had similar effects on the official exam. Our prediction that students who worked with PROSA performed better on the exam was not confirmed by our test results. However, a post hoc analysis of our data showed that the difference between our and the examiners' grading of a legal solution could explain these apparently contradictory results. A re-grading example did show higher scores on an exam case for students who worked with PROSA compared to scores of the students who had not worked with PROSA.
An Instructional Environment for Learning to Solve Legal Cases

PROSA

However, it also revealed that there is a problem with the way legal solutions are scored. We will address this problem in more detail in the next chapter, Chapter 8.

Students who worked with PROSA had a very positive opinion on the appearance and the content of PROSA as well as on the usefulness of PROSA for training the legal case solving skill. They also valued PROSA as a tool to prepare oneself for exams.

Based on our findings we conclude that PROSA is specifically effective for learning to construct a complete and correct legal solution.

PROSA seems to succeed in communicating the idea of a complete legal solution to the student. This completeness can probably be attributed to the fact that the instructional environment serves as an external memory.

Legal education covers a broad range of subjects from introducing subject matter content to learning to defend a case before the court. Within this spectrum PROSA may contribute its specific functionality to help students to learn to solve a legal case in such a way that it results in a complete and correct legal solution.