Contents

Papers

Evaluation of computer-based learning material for food chemistry education ......64-82
Julia Diederan, Harry Gruppen, Rob Hartog and Alphons G. J. Voragen

Identifying pedagogical content knowledge (PCK) in the chemistry laboratory ....83-103
Janet Bond-Robinson

Conceptual understanding versus algorithmic problem solving: Further evidence from a national chemistry examination.................................................................................104-118
Dimitrios Stamovlasis, Georgios Tsaparlis, Charalambos Kamilatos, Dimitrios Papaoikonomou, and Erifyli Zarotiadou

Indexed/Abstracted in:
CHEMICAL ABSTRACTS (CA)
EDUCATIONAL RESEARCH ABSTRACTS ONLINE (ERA)
http://www.tandf.co.uk/era
Chemistry Education Research and Practice

The journals, University Chemistry Education, published by The Royal Society of Chemistry, [http://www.rsc.org/uchemed/uchemed.htm](http://www.rsc.org/uchemed/uchemed.htm) and Chemistry Education Research and Practice, published from the University of Ioannina, [http://www.uoi.gr/cerp/](http://www.uoi.gr/cerp/) have merged with effect from January 1st 2005. The new, fully electronic journal is published by The Royal Society of Chemistry under the title: Chemistry Education Research and Practice, and it will continue to be available free of charge on the Internet. There are four issues per year.

The new journal is edited by Georgios Tsaparlis (gtseper@cc.uoi.gr) and Stephen Breuer (s.breuer@lancaster.ac.uk) and intends to maintain the high standards set by its predecessors. Its editorial policy will be the following.

‘Chemistry Education Research and Practice’ is the journal for teachers, researchers and other practitioners in chemical education. It is the place to publish papers on:

- research, and reviews of research in chemical education;
- effective practice in the teaching of chemistry;
- in depth analyses of issues of direct relevance to chemical education

Contributions can take the form of full papers, preliminary communications, perspectives on methodological and other issues of research and/or practice, reviews, letters relating to articles published and other issues, and brief reports on new and original approaches to the teaching of a specific topic or concept.

The new journal welcomes contributions of the type described above; these should be sent to cerp@rsc.org
Chemistry Education Research and Practice

Editorial Board:

Norman Reid (Chair, UK)
Patrick Bailey (UK),
George Bodner, (USA)
Stephen Breuer (UK)
Onno de Jong (Netherlands)
Alex Johnstone (UK)
Bernd Ralle (Germany)
Georgios Tsaparlis (Greece)

International Advisory Panel

Liberato Cardellini (Italy)
Peter Childs (Eire)
Jan van Driel (Netherlands)
Michael Gagan (UK)
Iwona Maciejowska (Poland)
Peter Mahaffy (Canada)
Mansoor Niaz (Venezuela)
Arlene Russell (USA)
Laszlo Szepes (Hungary)
Keith Taber (UK)
David Treagust (Australia)
Uri Zoller (Israel)
Chemistry Education Research and Practice

Guidelines for Authors

 Submission of contributions

Chemistry Education Research and Practice (CERP) is the journal for teachers, researchers and other practitioners in chemical education. It is published free of charge, electronically, by The Royal Society of Chemistry, four times a year. It is the place to publish papers on:

- research, and reviews of research in chemical education;
- effective practice in the teaching of chemistry;
- in depth analyses of issues of direct relevance to chemical education

Contributions can take the form of full papers, preliminary communications, perspectives on methodological and other issues of research and/or practice, reviews, letters relating to articles published and other issues, and brief reports on new and original approaches to the teaching of a specific topic or concept.

1. The original contribution should be submitted electronically, preferably in Word for Windows format. Any associated diagrams should be attached in JPG or GIF format, if possible. Submissions should be made by e-mail as a file attachment to cerp@rsc.org or directly to the editors: Stephen Breuer at s.breuer@lancaster.ac.uk or to Georgios Tsaparlis (gtseper@cc.uoi.gr).

2. Submitted contributions are expected to fall into one of several categories (listed above). Authors are invited to suggest the category into which the work should best fit, but the editors reserve the right to assign it to a different category if that seems appropriate.

A word count (excluding references, tables, legends etc) should be included at the end of the document.

3. Presentation should be uniform throughout the article.

   Text should be typed in 12pt Times New Roman (or similar), with 1"/2.5 cm margins, double-spaced, unjustified, ranged left and not hyphenated.

   Always use an appropriate mix of upper and lower case letters: do not type words in uppercase letters either in the text or in headings. Bold or italic text and not upper case letters should be used for emphasis.

   All nomenclature and units should comply with IUPAC conventions.

   Tables and figures should be numbered consecutively as they are referred to in the text (use a separate sequence of numbers for tables and for figures). Each should have an informative title and may have a legend.
Equations should be written into the text using the word processing program, either as normal text or using the program’s equation facility.

Structures should, wherever possible, be treated as a figure and not incorporated into text.

References should be given by the name of the author (or the first author, if more than one), followed by the year of publication. If an author has more than one reference from the same year, then it should be given as Smith 2001a, Smith 2001b, etc.

Footnotes should be generally avoided and important additional information may be referenced and included in the reference list.

4. A title page must be provided, comprising:
   • an informative title;
   • authors’ names and affiliation, full postal address and e-mail; (in the case of multi-authored papers, use an asterisk to indicate one author for correspondence, and superscript a, b, etc. to indicate the associated addresses);
   • an abstract of not more than 200 words;
   • keywords identifying the main topics covered in the paper

5. Wherever possible articles should be subsectioned with headings, subheadings and sub-sub-headings. Do not go lower than sub-sub-headings. Sections should not be numbered.

The introduction should set the context for the work to be described; include references to previous related work, and outline the educational objectives.

A concluding section (which need not be headed conclusion) will include an evaluation of the extent to which educational objectives have been met. A subjective evaluation may be acceptable.

6. The formatting of references should follow the following practice:

Books and Special Publications:
Author A., (year), Title of the book italicized, Publisher, Place of publication, page no. if applicable.

Journal Articles:
Author A., Author B. and Author C., (year), Title of the article in Roman type, Full Name of the Journal Italicised, Volume no. in Bold, inclusive page numbers.

For example:


7. All contributions submitted will be refereed anonymously by two independent referees. In case of a disagreement a third referee will be consulted. The decision of the Editors on
the acceptance of articles is final.

8. Authors grant *CERP* the exclusive right to publish articles. They undertake that their article is their original work, and does not infringe the copyright of any other person, or otherwise break any obligation to, or interfere with the rights of such a person, and that it contains nothing defamatory.

9. Articles will be published on the Web in PDF and HTML formats.
Evaluation of computer-based learning material for food chemistry education

Julia Diederen,¹ Harry Gruppen,¹ Rob Hartog² and Alphons G. J. Voragen¹

¹ Wageningen University, Department of Agrotechnology and Food Sciences, Laboratory of Food Chemistry, P.O. Box 8129, 6700 EV Wageningen, The Netherlands
² Wageningen University, Wageningen Multimedia research centre
e-mail: fons.voragen@wur.nl

Received 21 September 2004, accepted 8 April 2005

Abstract: Digital exercises were designed and developed for food chemistry education. During the design process, design requirements were described for such exercises. The exercises were evaluated in three case studies, firstly to determine whether the exercises satisfy the design requirements with respect to students’ use and secondly to provide insight into the effect of the course structure and organisation on the value that the students attribute to the exercises. The results show that the exercises meet most of the design requirements. Students found the exercises clear and helpful, and most students confirmed that these exercises helped them in their preparations for their examinations. Despite this, participation in the programme was low when working on the exercises was not compulsory. The differences in evaluation results between the three studies can be explained by differences in the course structure and organisation. [Chem. Educ. Res. Pract., 2005, 6 (2), 64-82]

Keywords: food chemistry; computer based learning; learning objects; evaluation; case study; design requirements; context; learning activities; learning activity preference

Introduction

The course, Food Chemistry (6 ECTS: European Credit Transfer System), is a second year course for students in their Bachelors curriculum Food Science & Technology at Wageningen University (Wageningen, The Netherlands). In this course, students acquire basic knowledge of food chemistry, i.e. qualitative and quantitative knowledge about important (bio-)chemical reactions in food and the influence of the processing conditions on these reactions, and finally, development of laboratory skills. To facilitate this, digital exercises have been designed and developed to motivate and activate the students.

For the design of digital exercises, a design (and development) process for activating and motivating digital learning material has been developed. Diederen and co-workers (2003) described in detail in this journal the design process and, in particular, design guidelines and design requirements that are based on theories and views on learning and instruction, on food chemistry subject matter and learning goals, and on user interface design. These were followed during the design and development of a set of 106 digital exercises.

Education supported by digital learning objects is relatively new, compared to education through lectures, self-study with books, group work and laboratory classes. To usefully incorporate the digital exercises into the food chemistry course, which consists of several educational activities, one needs to answer questions such as, ‘what should be the sequential..."
order of different educational activities such as lectures and computer classes?’, ‘what is the 
effect of the number of lectures on the need for and use of digital learning objects?’ and ‘what 
may happen when there is no extrinsic motivation for the use of the digital learning objects (e.g. 
when their use is not compulsory)?’

The present article deals with the evaluation of the digital exercises in three case studies, 
each within a different course structure and organisation. The first two were carried out in the 
regular educational setting of the Food Chemistry course at Wageningen University in two 
successive years. In this setting the use of the digital exercises was compulsory in both case 
studies, but the number and type of lectures accompanying the digital exercises differed. The 
third was performed during the Food Chemistry course at Cornell University (Ithaca, NY, USA), 
in which the use of the digital exercises was optional.

The aim of these studies was twofold:
1. To determine whether the learning material satisfied the design requirements on student 
use (Diederen et al., 2003), dealing with the influence of the context, the quality and user-
friendliness of the design, the extent to which the digital exercises assist in studying the reader, 
and the additional value of the digital exercises.
2. To collect information about the possible relation between differences in the students’ 
appreciation of the digital learning material and differences in course structure and organisation.

The digital exercises

In total, 106 digital exercises were designed for food chemistry education. These exercises 
comprise interactive questions and assignments that invite students to practise on different topics 
within food chemistry. A more detailed description of the digital exercises, as well as the 
rationales for the design decisions, has been published previously (Diederen et al., 2003). In this 
section a brief description is given. The content of the exercises was based on the content of the 
reader used during the course at Wageningen University, since the digital exercises are in the 
first instance designed for this course. Since a food chemistry course at another university most 
likely differs to some extent from that of the food chemistry course at Wageningen University, it 
cannot be guaranteed that all our digital exercises are equally useful to all food chemistry 
courses.

The digital exercises have the following seven main features:
• Exercises are grouped into sequences of at most ten exercises.
• A score for each exercise and a score overview for each sequence of exercises are 
incorporated into the exercises in order to motivate the students not to guess too much and to 
repeat an exercise with a low score (maximum score is 10; minimum can even be a negative 
score).
• Exercises contain diagrams, schemes, pictures and animations.
• Exercises contain examples from industry or daily life to show the usefulness of knowing 
and understanding typical reactions in food, e.g. lipid oxidation, Maillard reactions and 
enzymatic browning.
• Within each sequence of exercises the degree of difficulty is gradually increased.
Exercises contain feedback varying from simple remarks (e.g. ‘wrong’ or ‘right’) to more
detailed hints (e.g. specific feedback according to the student’s answer) as well as different
kinds of additional information.

Exercises differ in the type of questions or assignments (e.g. multiple choice, categorising,
ranking) and cover different topics (e.g. about molecular structure, reactions, definitions,
applications).

Exercises were developed with the software program Flash (Macromedia®) and were made
accessible to the student by the web-based learning environment BlackBoard®.

Figure 1. An example of an exercise: three reactions involving quinones

An example of an exercise is shown in Figure 1. Top left the student can see that he/she is
working on exercise number 4 in a sequence of five exercises. The student can click on each of
the numbers to go to a specific exercise or on ‘score’ to go to the score overview. The score
overview shows the student the score for each exercise within this sequence. Top right the score
for the current exercise is shown. Centred at the top it is written what should be done in this
exercise. In this example, students have to complete a scheme of three different reaction
pathways of quinones by dragging the molecules on the right to the corresponding spot of one of
the pathways on the left. At the moment of the screenshot of Figure 1 the user was dragging a
quinone out of the box on the right to move it to a spot on the left. Students can ask for
information, by clicking on the ‘HINT’ button. In this particular case, on using this button the
effects of oxidation of phenolic components are explained through the process of tea
fermentation. By reading this explanation it becomes clear which reaction pathway causes
browning, formation of aromas or protein precipitation.
Evaluation method

The digital exercises were designed according to design guidelines and design requirements based on theories of learning and instruction, on subject matter and learning goals, and on theories of user interface design. Theories of learning and instruction that were used are mainly the cognitive load theory (Sweller et al., 1998 and Kirschner, 2002), motivation of students based on the ARCS-model (Attention, Relevance, Confidence, Satisfaction; Keller, 1983) and active learning (e.g. Keyser, 2000). The sources for the different design requirements have been described extensively in the previous paper (Diederen et al., 2003). The requirements define operationally the goals of the design, and the variables, to which the requirements refer, can be measured. To be able to decide whether the learning material meets the requirements, minimum criteria need to be specified for each requirement. Some requirements need to be tested by students, some by teachers and some by specific experts as user interface designers, educational specialists or food chemistry specialists. This paper deals with the requirements that need to be tested by students, which is done with case studies.

The method of evaluation of requirements on students’ use can be described as follows. The students work on the exercises and complete an evaluation questionnaire. When the results of the evaluation are equal to or higher than the criteria specified for each requirement, that requirement is judged to be satisfied. The details about how this is carried out are given later.

To investigate the influence of the course structure and organisation on the usefulness of the digital exercises for the students, the digital exercises were evaluated in different case studies.

The requirements

The requirements for the digital exercises that were evaluated in the case studies are divided into four related sets. These are listed in Table 1, together with the corresponding evaluation questions and criteria. The source for each requirement can be found in the previous paper (Diederen et al., 2003).

The four sets of requirements are:

The influence of the context on the case studies

This set of requirements takes into account the fact that the context during the case study influences the behaviour of the user (Pawson and Tilley, 1997). Results from one evaluation (case study) cannot simply be generalised. The context within each case study is defined by the prevailing conditions during the use of the learning material. Specifically for this purpose, the requirement ‘The circumstances during the evaluation are taken into account’ is described (requirement rE, Table 1). This requirement is not used during the design process, but is described as an evaluation requirement: the evaluation of the digital exercise on students’ use has to meet this requirement.

The context also includes the learning activity preferences of the students. This could influence the attitude of the students to the learning material: a student who prefers to learn by working in groups or by listening will not be attracted to the digital exercises. Also, the way the student uses the digital exercises could be influenced by the learning preferences (Vermetten et al., 2002). Our learning material aims at Bachelor students enrolled in the curriculum Food Science and Technology who have a learning preference for activities such as answering questions and studying schemes, diagrams and animations.
We realize that several other circumstances could be important to describe the context during the case studies, such as the prior knowledge of the students, the quality of the teachers, prior experiences with computer supported education for both teachers and students, the gender and age of the participants, etc. An endless list of variables could be described and analysed for each study, but we chose to limit the scope of our study.

*The quality and user-friendliness of the design*

The second set is related to requirements that deal with the way the students perceive the design of the digital exercises. The design can be judged by criteria, such as usability, clarity, and manageability. These judgements give an indication whether students are pleased with the set up of the digital exercises.

*The extent to which the digital exercises assist in studying the reader*

This set of requirements is related to the main goal of the digital learning material: facilitating the learning of food chemistry. For this, students need to understand the content of the exercises, recognize the content of the exercises in the reader, which is used in the course, and feel that it helps him to remember and understand parts of the reader after completing the exercises.

*The additional value of the digital exercises*

The fourth set of requirements is related to the expectation that digital exercises have an additional value compared to other learning materials, such as books, and other learning activities, such as lectures and laboratory classes. The digital exercises are designed to motivate the students and require their active involvement. Every opportunity to present information in the form of a diagram, scheme or animation instead of text has been seized. Therefore, it is expected that students should learn a lot in a relative short period of time. Important for this set of requirements is whether students work on the digital exercises on their own initiative (an indication that students see their value), whether students feel they learned much and whether studying the exercises contributed to their ability to pass the final examination of the course.

*The evaluation questions and criteria*

In total, twenty-six questions were developed to evaluate the requirements. For the four sets of requirements there are six, ten, four and six questions respectively (Table 1). These questions were given to the students through two different questionnaires. The first is a standard course evaluation questionnaire, which is regularly used at Wageningen University. This questionnaire concerns, amongst others, the content, the organisation, the quality of the learning materials, the quality of the teachers, the perceived value of the different educational activities and the way in which information and communication technology has been used for learning support in the course. The second is a specific questionnaire developed for the evaluation of the digital exercises. In Table 1 the questions of both questionnaires are listed, from which questions qE.1 to qE.5, q1.2, qQ.1 and q3.1 refer to the questions from the standard questionnaire.
Table 1: The sets of requirements, their evaluation criteria and the corresponding questions or statements in the questionnaires.

<table>
<thead>
<tr>
<th>Set</th>
<th>Requirement</th>
<th>Questions or statements</th>
<th>criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>The influence of the context</td>
<td>The circumstances during the evaluation are taken into account. (rE)**</td>
<td>qE.1: In my opinion other digital learning material (than the digital exercises) is valuable.</td>
<td>criteria for 5-point scale*</td>
</tr>
<tr>
<td></td>
<td>qE.2. I am satisfied with the desk-space-facilities and computer rooms during the computer classes.</td>
<td>qE.3. I am satisfied with the use of Blackboard.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>qE.4. I am satisfied with the course sequence.</td>
<td>qE.5: Overall rate of the course on a scale of 1 (poor) to 5 (excellent).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The exercises fulfil a need of most of the students in relation to their learning activity preference. (r12)</td>
<td>q12.1. I am someone who learns through (students choose maximally 3 answers from 7 learning activities).</td>
<td>A significant number (p&lt;0.05) of students like to learn by the activities ‘active answering questions’ and ‘looking at schemas/diagrams’</td>
</tr>
<tr>
<td></td>
<td>The exercises are clear. (r1)</td>
<td>q1.1. The exercises are clear (you know what you have to do).</td>
<td>criteria for 5-point scale*</td>
</tr>
<tr>
<td></td>
<td>q1.2. The formulation of the instructions of the digital learning material was clear and understandable.</td>
<td>qQ.1. The quality of the digital learning material was good.</td>
<td></td>
</tr>
<tr>
<td>The quality and user-friendliness of the design</td>
<td>Students perceive the quality of the learning material as good. (rQ)**</td>
<td>qQ.1. The quality of the digital learning material was good.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The exercises are manageable. (Contain enough hints to work through the exercises.) (r2)</td>
<td>q2.1. With those exercises that I needed hints, the hints were provided.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>q2.2. With those exercises that hints were provided, the hints were good enough to go through the exercises.</td>
<td>q10.1. I used the score as a tool that tells me how well I performed in the exercise: a low score means I did badly, a high score means I did well.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The exercises are fun to work on. (r5)</td>
<td>q5.1. The exercises are fun to work on.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Students are motivated not to guess too much. (r10)</td>
<td>q10.2. I used the score as a motivation not to guess too much (except if I did not have another choice).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Students are motivated to repeat an exercise when they do badly. (r11)</td>
<td>q11.1. When I had a low score I did the exercises again.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>q11.2. I did the exercises many times until my score for each exercise was close to 10.</td>
<td></td>
</tr>
</tbody>
</table>
Table 1 (continued): The sets of requirements, their evaluation criteria and the corresponding questions or statements in the questionnaires.

<table>
<thead>
<tr>
<th>Set</th>
<th>Requirement</th>
<th>Questions or statements</th>
<th>criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assistance in studying the reader</td>
<td>Content of an exercise is understood after completing the exercise. (r6)</td>
<td>q6.1. At the end of an exercise (after explanation by the computer) I understood the content of the exercise.</td>
<td>criteria for 5-point scale*</td>
</tr>
<tr>
<td></td>
<td>Students are able to recognise the exercises in sections of the reader. (r7)</td>
<td>q7.1. I recognised parts of the exercises in the reader.</td>
<td>criteria for 5-point scale*</td>
</tr>
<tr>
<td></td>
<td>Sections of the reader are easy to remember after completing the exercises. (r8)</td>
<td>q8.1. Parts of the reader were easy to remember after doing the exercises.</td>
<td>criteria for 5-point scale*</td>
</tr>
<tr>
<td></td>
<td>Sections of the reader are easy to understand after completing the exercises. (r9)</td>
<td>q9.1. Parts of the reader were easy to understand after doing the exercises.</td>
<td>criteria for 5-point scale*</td>
</tr>
<tr>
<td></td>
<td>The exercises are used by students on their own initiative. (r4)</td>
<td>q4.1. Tell when you worked on the exercises.</td>
<td>≥ 75% of the students worked outside reserved time and ≥ 75% worked on them more than once</td>
</tr>
<tr>
<td></td>
<td>Students perceive the exercises as valuable. (related to r3)**</td>
<td>q3.1. In my opinion the digital exercises are valuable learning material.</td>
<td>criteria for 5-point scale*</td>
</tr>
<tr>
<td></td>
<td>Students feel they learned much from doing the exercises. (r13)</td>
<td>q13.1. How much did you learn of the different parts of the Food Chemistry course? (Six parts are judged on 5-point scale from nothing to very much)</td>
<td>≥ 75% of the students answer “much” or “very much”</td>
</tr>
<tr>
<td></td>
<td>Students feel that the exercises contributed to their ability to pass the final examination successfully. (r14)***</td>
<td>q14.1A. Estimate (in percentage) how much each part of the course contributes to your ability of doing the final examination. (Your total estimate should add up to 100%). q14.2A. Estimate how much time you spent on each part of the course. (Estimate the number of hours.). q14.1B. The exercises contributed to my ability to answer questions in the final exam. q14.2B. The exercises are efficient for learning: I learned about food chemistry in a relative short period of time.</td>
<td>Average learning efficiency is ≥ 1 (learning efficiency = %contribution / %time) criteria for 5-point scale*</td>
</tr>
</tbody>
</table>

# The codes of the requirements are the same codes as used in the design process (Diederen et al., 2003).
## Requirements rE and rQ were not described during the design process. rE is an evaluation requirement (not a design requirement).
*: Average rating should be 4.0 or more AND at least 75% of the students should give a rating of 4 or 5. (5-point scale: 1=totally disagree, 2=partially disagree, 3=neutral, 4=partially agree, 5=totally agree)
**: r3 is ‘Students see the digital exercises as a valuable addition to the reader’.
***: questions for r14 differ between case studies, see Table 2
In both questionnaires, most questions are given as statements that need to be judged on a 5-point Likert scale, ranging from 1, meaning ‘strongly/totally disagree’, to 5, meaning ‘strongly/totally agree’. From the answers of all students an average judgement per statement can be calculated. A requirement is satisfied when the average rating for the accompanying statement is at least 4.0 and when at least 75% of the students rate the statement with a 4 or 5. Some requirements were not tested with 5-point scale questions, but with multiple-choice questions. Most of these requirements have also a criterion of 75% positive judgement, which is in line with the 75% criterion of 5-point Likert scale questions. See Table 1 for specific criteria.

Biases could arise from evaluation questionnaires as used for the current research. Typical method biases are for example those related to acquiescence (agree with attitude statements regardless of content), social desirability (behave in a culturally acceptable and appropriate manner), positive / negative item wording (only using positively or negatively worded statements), or common scale formats (e.g. Likert) and anchors (e.g. agree / disagree) (Podsakoff et al., 2003). It should be noted that the questionnaires are only used to compare the average attitudes of all students between case studies and to compare the average attitudes of all students with the criteria. Therefore, we think that acquiescence bias is not a big issue.

**The course structure and organisation in each of the three case studies**

The course structure and organization includes components and aspects such as kinds of course topics, number of course credits, number of lectures, number of scheduled computer classes, sequence order of lectures and computer classes, number of laboratory classes, staff/student ratio, relation between reader content and content of the digital exercises. Table 2 shows for each of the three case studies the relevant characteristics of course structure and organisation, and also the type of evaluation questionnaire used. Case studies I and II were both conducted during the Food Chemistry course at Wageningen University in the regular educational setting. The course concerned is a 6 ECTS course (168 study hours). Students had lectures and computer classes, and prior to the final examination in both case studies the course concluded with a laboratory class, all within a time-span of six weeks. After these six weeks there was a week without formal instruction, followed by an examination week. In both case studies I and II a reader, covering the course content, was available to the students. After the final examination the students were asked to fill in the two questionnaires.

With respect to the students in studies I and II, one important difference could be expected between these two groups of students: their assumed prior knowledge of food chemistry when they started working on the digital exercises. Students in case study II had fewer lectures than students in case study I and worked on the exercises on the same day as the lectures were given, which gave them no time to process newly acquired information. In fact, students in case study II were exposed to new information during the exercises. Information, which students in case study I already heard of during the lectures.

The course structure and organisation of case study III was different from the other two. This was carried out at Cornell University (Ithaca, NY, USA), with students taking the Food Chemistry course (3-credit course, comparable with 4.5-6 ECTS). For study purposes an extensive set of lecture notes, which was almost like a textbook, was given to the students. The content of these lecture notes were in some aspects different from that of the reader used at Wageningen University: topics such as water and food additives were included, while topics such as enzymes or phenolic components received less attention. Another important difference in course organisation was that the students’ use of the digital exercises was optional, while in case
This journal is © The Royal Society of Chemistry
Results and discussion

The results of the case studies are described and discussed on basis of the four sets of requirements as defined before (Table 1).

Requirements set 1: The influence of the context on the case studies

Requirements rE ‘the circumstances during the evaluation are taken into account’ and r12 ‘the exercises fulfil a need of most of the students in relation to their learning style’ are related to the influence of the context. To evaluate the influence of the context on the results of the case studies, questions related to the circumstances during the evaluation (qE.1 to qE.5) and about preferred learning activities (q12.1) were asked to the students.

The appreciation of the circumstances during the case studies

Table 3 gives the results of the answers to questions qE.1 to qE.5. These are part of the standard course evaluation questionnaire and were, therefore, not answered by the students in case study III. The responses to these questions indicate that the students in case study I liked the circumstances of the course better than the students in study II. In case study II the circumstances of the course, such as desk-space facilities and the course sequence, are viewed less favourably. It should be noted that the facilities were actually the same in both case studies. We have no satisfactory explanation for the difference. The lower appreciation of course sequence could negatively influence the response of the students to evaluation questions related to the digital exercises.

Table 3: Average response to the questions qE.1 to qE.5 in case studies I (n=28) and II (n=32-41).

<table>
<thead>
<tr>
<th>Questions</th>
<th>Case study I</th>
<th></th>
<th>Case study II</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% 4+5</td>
<td>Average response (n=28)</td>
<td>% 4+5</td>
<td>Average response (n=30-41*)</td>
</tr>
<tr>
<td>qE.1: In my opinion the digital exercises for calculations on reactions are valuable learning material.</td>
<td>74</td>
<td>4.0</td>
<td>64</td>
<td>3.7</td>
</tr>
<tr>
<td>qE.2: I am satisfied with the desk-space-facilities and computer rooms during the computer classes.</td>
<td>92</td>
<td>4.5</td>
<td>64</td>
<td>3.8</td>
</tr>
<tr>
<td>qE.3: I am satisfied with the use of Blackboard.</td>
<td>83</td>
<td>4.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>qE.4: I am satisfied with the course organisation: 1h lecture, 3h digital exercises, 1h response lecture.</td>
<td>-</td>
<td>-</td>
<td>48</td>
<td>3.6</td>
</tr>
<tr>
<td>qE.5: Overall rate for the course on a scale of 1 (poor) to 5 (excellent)</td>
<td>81</td>
<td>3.9</td>
<td>71</td>
<td>3.8</td>
</tr>
</tbody>
</table>

Answers are grouped in percentage of students who responded with a 4 or 5 (1=totally disagree, 2=partly disagree, 3=neutral, 4=partly agree, 5=totally agree) and the average response.

*: 30 students followed the course. The final examination was attended by these students and by some students who did a resit examination, and therefore more students filled in the standard course evaluation questionnaire.

The learning activity preferences

To evaluate the learning activity preference of the students in the three studies, students were asked to choose up to three options, which correspond with the activities through which they prefer to learn, from a list of the seven main learning activities in the food chemistry course of studies I and II. Their responses give an indication whether the students in the studies will like the computer-based exercises. Desired learning activity preferences for these exercises are firstly
the activity ‘active answering questions’ and secondly the activity ‘looking at schemes/diagrams’. Those students who prefer other learning activities to these two are likely to favour computerised exercises less. It should be mentioned that learning activity preference can be influenced by experience (Vermetten et al., 1999); students who find particular activities successful may come to prefer them.

If the students pick at random three learning activities out of seven, then each activity would get picked by 43% of the students. So, if an activity is chosen by more than 43% of the students, this activity is favoured by this group of students, and vice versa. To determine whether an activity is significantly in favour, the p-value is calculated for a few activities (binomial test, test proportion=0.43, SPSS 10.0). If \( p < 0.05 \) then the proportion of students that choose a learning activity is significantly higher than the test proportion.

It was noticed for all case studies that those students who learn by active answering questions also like to learn by looking at visuals, (active) reading and producing visuals (e.g. pictures, schemes) (data not given). Those who did not choose active answering questions like to listen (attending lectures) and like to learn by (active) reading. For this group of students also “explanation of my questions by others” is a frequently chosen option (data not shown).

**Requirements set 2: The quality and user-friendliness of the design of the exercises**

It is assumed that, in the perception of the students, the quality of the design of the learning material is related to characteristics such as usability, ability to motivate, clarity, manageability and ability to enjoy. Results from the questionnaire related to six requirements on how students perceive the design of the learning material are listed in Table 5.

According to the criteria for 5-point scale questions the four requirements r1, rQ, r2, and r5 are met, which means that the exercises are clear (r1), the exercises are manageable (r2), the exercises are fun to work on (r5), and the students perceive the quality of the learning material as

---

Table 4: The learning activity preference of the students in the three case studies (question q12.1). Students could choose at most three activities. For each activity the percentage of students that has chosen that activity are given.

<table>
<thead>
<tr>
<th>Activities: I am someone who learns through ...</th>
<th>Case study I (n=34)</th>
<th>Case study II (n=26**)</th>
<th>Case study III (n=16**)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>A: active answering questions</td>
<td>71 0.001</td>
<td>56 0.18</td>
<td>31 -</td>
</tr>
<tr>
<td>B: looking at schemes / diagrams / animations</td>
<td>50 0.26</td>
<td>30 -</td>
<td>56 0.21</td>
</tr>
<tr>
<td>C: (active) reading</td>
<td>56 0.09</td>
<td>59 0.09</td>
<td>50 0.37</td>
</tr>
<tr>
<td>D: listening (for example a lecture)</td>
<td>44 -</td>
<td>37 -</td>
<td>44 -</td>
</tr>
<tr>
<td>E: producing schemes / diagrams / animations</td>
<td>35 -</td>
<td>48 0.45</td>
<td>50 0.37</td>
</tr>
<tr>
<td>F: explanation of my questions by others (teacher/student)</td>
<td>32 -</td>
<td>26 -</td>
<td>25 -</td>
</tr>
<tr>
<td>G: group discussion / working in a group</td>
<td>18 -</td>
<td>26 -</td>
<td>25 -</td>
</tr>
</tbody>
</table>

*: Calculated with binomial test, test proportion=0.43.

**: Not all students had filled in this questionnaire.

---

* Calculated with hypergeometric distribution (random selection, without repetition): \( f(A) = \frac{\binom{1}{1} \binom{5}{2}}{\binom{7}{3}} = \frac{1 \cdot \frac{6!}{2!4!}}{\frac{7!}{3!4!}} = 0.43 \)

*This journal is © The Royal Society of Chemistry
good (rQ). This gives an indication whether the design and user-friendliness of the exercises is of an adequate quality for the students in each of the three case studies.

Feedback or hints were provided to make the exercises manageable. The results (Table 5) from the questionnaire for the two questions that are related to hints (q2.1 and q2.2), show that the hints are applied correctly in the digital exercises for the students in case studies I and II. The lower valuation of the hints by the students in case study III is probably related to the difference in course content. The content of the lectures and lecture notes at Cornell University varies from that of the digital exercises in such a way that these students could need some more hints to be able to answer the questions than the students at Wageningen University or possibly some different hints, since their course content was different, but there are no data that could demonstrate that these students asked for more hints.

**Table 5:** Results for the questions related to the quality of the design for case study I, II and III.

<table>
<thead>
<tr>
<th>r*</th>
<th>Question</th>
<th>Case study I</th>
<th>Case study II</th>
<th>Case study III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>% 4/5</td>
<td>average</td>
<td>% 4/5</td>
</tr>
<tr>
<td>r1</td>
<td>q1.1: exercises are clear (you know what you have to do)</td>
<td>82</td>
<td>4.3</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>q1.2: The formulation of the instructions of the digital exercises was clear and understandable.</td>
<td>90</td>
<td>4.3</td>
<td>84</td>
</tr>
<tr>
<td>rQ</td>
<td>qQ.1: The quality of the digital learning material was good.</td>
<td>95</td>
<td>4.4</td>
<td>84</td>
</tr>
<tr>
<td>r2</td>
<td>q2.1: hints are provided when needed</td>
<td>75</td>
<td>3.8</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>q2.2: the hints are useful</td>
<td>85</td>
<td>4.0</td>
<td>83</td>
</tr>
<tr>
<td>r5</td>
<td>q5.1: exercises are fun</td>
<td>88</td>
<td>4.2</td>
<td>73</td>
</tr>
<tr>
<td>r10</td>
<td>q10.1: score tells about performance</td>
<td>85</td>
<td>4.0</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>q10.2: score is motivation against guessing</td>
<td>44</td>
<td>2.9</td>
<td>53</td>
</tr>
<tr>
<td>r11</td>
<td>q11.1: low score is motivation to repeat exercise</td>
<td>79</td>
<td>3.9</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>q11.2: repeat exercises until maximum score is achieved</td>
<td>38</td>
<td>2.5</td>
<td>28</td>
</tr>
</tbody>
</table>

Given are the average rating and the percentage of students that rated the question with a 4 or 5 (answers on a 5-point scale: 1=totally disagree, 2=partly disagree, 3=neutral, 4=partly agree, 5=totally agree).

*: r=requirement

**The relation between the score and extrinsic motivation**

Each exercise is designed with a score, as described before. This score is incorporated in the exercises to motivate students: to discourage guessing and to motivate students to repeat an exercise when getting a low score. The results in Table 5 show that in case study I, the students did recognise this score as a performance-grader (q10.1), while students in case studies II and III, did not really recognize the score as a performance-grader. The same counts for the question q11.1. In case study I, around 79% of the students agreed with the statement “the score is a motivation to repeat an exercise when the score is low”. For these students, the score could be recognised as a successful tool to motivate them to perform well, but in case studies II and III fewer students agreed with this statement.

Two factors may explain the lower value of the score in case study II as compared to the score in case study I. Firstly tables 3 and 4 show that in case study II the contextual variables such as desktop facilities and computer rooms respectively learning activity preferences have lower values. Secondly the knowledge of the students prior to doing the exercises was most...
likely very much less in case study II because there were considerably fewer lectures. Due to a lower prior knowledge, students made more mistakes, resulting in a lower score. A low score is not an indication for these students that they performed poorly, but maybe it is an indication for these students that they did not have enough knowledge yet. Students in case study I maybe felt they should have enough knowledge to perform well and therefore they \textit{did} see the score as a performance grader. Students in case study III worked on the exercises of their own free will and could, therefore, not care much about a score or performance.

From the answers on the questions q10.2. and q11.2. in all three studies it is clear that the score itself was not a good motivator for students not to guess too much or to work on the exercises until they did not make any mistakes at all (Table 5). It can be concluded that the motivating factor of the score was quite variable per student: some students were motivated by the score and some were not. Just incorporating a score is certainly insufficient to induce an extrinsic motivation for the students to perform well.

\textbf{Requirements set 3: The extent to which the exercises assist in studying the reader}

The reader for the course Food Chemistry at Wageningen University contains a range of relevant chemical concepts. Since students have difficulties to recollect the facts from the reader, the digital exercises were designed in order to stimulate students to work actively with the subject matter. In other words, the digital exercises were designed in such a way that working on the exercises helps students to study the reader. Requirements r6 ‘content of the exercises is understood’, r7 ‘students are able to recognise the exercises in sections of the reader’, r8 ‘sections of the reader are easy to remember after completing the exercises’ and r9 ‘sections of the reader are easy to understand after completing the exercises’ are related to this objective. In Table 6 the results from the questionnaire for the questions related to these four requirements are listed.

<table>
<thead>
<tr>
<th>r*</th>
<th>Question</th>
<th>Case study I</th>
<th>Case study II</th>
<th>Case study III</th>
</tr>
</thead>
<tbody>
<tr>
<td>r6</td>
<td>q6.1: content is understood</td>
<td>75 4.0</td>
<td>75 4.1</td>
<td>88 4.3</td>
</tr>
<tr>
<td>r7</td>
<td>q7.1: exercises are recognised in reader</td>
<td>76 4.2</td>
<td>73 3.9</td>
<td>75 3.8</td>
</tr>
<tr>
<td>r8</td>
<td>q8.1: exercises make reader easy to remember</td>
<td>76 4.0</td>
<td>72 3.9</td>
<td>56 3.6</td>
</tr>
<tr>
<td>r9</td>
<td>q9.1: exercises make reader easy to understand</td>
<td>76 4.2</td>
<td>67 3.8</td>
<td>63 3.7</td>
</tr>
</tbody>
</table>

Given are the average rating and the percentage of students that rated the question with a 4 or 5 (answers on a 5-point scale: 1=totally disagree, 2=partly disagree, 3=neutral, 4=partly agree, 5= totally agree).

*: r=requirement

Because it is very likely that the students in case study II did have less relevant prior knowledge when they started on the exercises their learning task will have been larger: they came across more new chemical concepts during working on the digital exercises than the students in case study I. Apart from this the factors mentioned earlier such as less satisfactory desktop facilities in case study II and differences in learning activity preferences may also have been relevant.
**Requirements set 4: The additional value of the digital exercises**

The digital exercises are designed for the students as an addition to lectures and self-study activities (reading and studying the readers) in order to give them the possibility to work actively with the subject matter. Therefore, it is important to know whether students value the digital exercises as a useful addition. The requirements r3 ‘students see the digital exercises as a valuable addition’, r4 ‘the exercises are used by students on their own initiative’, r13 ‘students feel they learned much from doing the exercises’ and r14 ‘students feel that the exercises contributed to the ability to successfully pass the final examination’ are defined during the design process to ensure that the digital exercises are a valuable addition to the students.

The answers to question q3.1 confirmed that students saw the digital exercises as a valuable addition to the range of learning materials available. In the standard questionnaire students judged all parts of the course, i.e. the reader, the laboratory classes, the teachers and the computer classes. The judgement of the students for the digital exercises can, therefore, be compared with the judgement for other parts, of which the results for the reader, the laboratory part, and the teachers are also presented in Table 7. In this perspective, the judgement about the digital exercises was very high in case study I and high in case study II, indicating that students strongly valued the digital exercises.

### Table 7: Average responses for the standard course evaluation questionnaire in case studies I (n=28) and II (n=30-41).

<table>
<thead>
<tr>
<th>part of the course</th>
<th>Case study I</th>
<th>Case study II</th>
</tr>
</thead>
<tbody>
<tr>
<td>question q3.1: In my opinion the digital exercises are valuable learning material.</td>
<td>91 % 4+5</td>
<td>84 % 4+5</td>
</tr>
<tr>
<td>judgement for reader</td>
<td>3.3</td>
<td>3.5</td>
</tr>
<tr>
<td>judgement for laboratory classes</td>
<td>3.8</td>
<td>3.7</td>
</tr>
<tr>
<td>judgement for teachers (on average)</td>
<td>3.8</td>
<td>3.7</td>
</tr>
</tbody>
</table>

Answers are grouped in percentage of students who responded with a 4 or 5 (1=disagree, 2=partly disagree, 3=neutral, 4=partly agree, 5=agree) and the average response.

From the answers on questions q4.1 and q4.2 for case studies I and II (Table 8), it seems fair to conclude that requirement r4 ‘students work on the exercises on their own initiative’ was met, since 70% of the students worked on the exercises outside the computer classes and 80% of the students worked on some or all exercises more than once. Since these students worked on learning material on their own initiative (i.e. outside the computer classes), they must have seen additional value in the exercises.

The results from case study III show clearly that when the digital exercises were optional, students were less inclined to do them. Some other results strengthen this conclusion. From the twenty-five students that were enrolled in this course, five students did not use the digital exercises at all. From the twenty students who used the exercises, 25% did not enter the website for a second time. Furthermore, from the sixteen students who filled in the questionnaire 56% worked on all subjects, 13% worked only on the three main subjects (proteins, carbohydrates and lipids) and 31% worked only on the subject lipids. This preference for lipids was probably because the teacher asked them specifically to work on these lipid exercises. Although students in case study III clearly spent less time on the digital exercises, their opinion about the additional value of the digital exercises was rather positive, as will be discussed later on.
Table 8: Results for the questions q4.1 and q4.2 related to requirement r4 ‘students work on the exercises on their own initiative’ for case studies I, II and III.

<table>
<thead>
<tr>
<th>question</th>
<th>possible answers</th>
<th>case study I</th>
<th>case study II</th>
<th>case study III</th>
</tr>
</thead>
<tbody>
<tr>
<td>q4.1: Tell when you worked on the exercises.</td>
<td>During computer classes</td>
<td>100%</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Outside classes</td>
<td>70%</td>
<td>70%</td>
<td>-</td>
</tr>
<tr>
<td>q4.2: Estimate how many times you worked on the exercises.</td>
<td>Some exercises, once only</td>
<td>9%</td>
<td>10%</td>
<td>38%</td>
</tr>
<tr>
<td></td>
<td>All exercises, once only</td>
<td>12%</td>
<td>10%</td>
<td>31%</td>
</tr>
<tr>
<td></td>
<td>All at least once, some more often</td>
<td>47%</td>
<td>47%</td>
<td>13%</td>
</tr>
<tr>
<td></td>
<td>All more than once</td>
<td>32%</td>
<td>33%</td>
<td>19%</td>
</tr>
</tbody>
</table>

The additional value of the digital exercises is also related to how much students learn from the exercises and whether students feel able to successfully pass the final exam with what they learned from the exercises. For the eight different parts of the course (1. lectures, 2. digital exercises, 3. other digital learning materials, 4. reader, 5. studying own lecture notes, 6. studying notes from digital exercises, 7. laboratory classes, 8. writing report of laboratory class experiments) students were asked to tell how much they learned from these parts (question q13.1). For each part they could choose between “very much”, “much”, “reasonable amount”, “little” and “nothing”. In case studies I and II from “doing the digital exercises”, 68% and 78% of the students learned much or very much, respectively (Table 9). For requirement r13 ‘students feel they learned much from doing the exercises’, the criterion is that 75% of the students should have learned much or very much, which means that r13 is not totally met in case study I, but it is in case study II.

Table 9: Results for the part digital exercises for the question q13.1 ‘How much did you learn from the digital exercises?’.

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
<th>Percentage of students (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q13.1: How much did you learn from the digital exercises?</td>
<td>Case study I</td>
<td>Case study II</td>
</tr>
<tr>
<td>Nothing or little</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Reasonable amount</td>
<td>29</td>
<td>15</td>
</tr>
<tr>
<td>Much</td>
<td>41</td>
<td>52</td>
</tr>
<tr>
<td>Very much</td>
<td>27</td>
<td>26</td>
</tr>
</tbody>
</table>

Although the data are not separately shown in a table, it is interesting to note that for the eight different parts of the course in case study I the part ‘reading the reader’ scored the highest (76% said to have learned much or very much), followed by the digital exercises (68%). In case study II this is the other way around: the digital exercises scored the highest (78%) and the reader scored lower (59%). This difference is remarkable, especially since the time spent on each part is comparable for the two case studies.

Again, the differences between case studies I and II can be explained by the fact that students in case study II had a lower prior knowledge, when starting on the exercises, than the students in case study I. In general students in case study II had to learn more during working on the digital exercises. Students in case study I already learned about various topics during the lectures, hence the information in the digital exercises was not totally new to them.

When comparing the results for question q13.1 of the first two case studies with the results of case study III (Table 9) it is obvious that the students from Cornell University learned less.
from the exercises than the students in case studies I and II, which was to be expected. The content of the food chemistry course at Cornell University is somewhat different from the content of some of the digital exercises, which means that not all exercises are applicable to the situation of these students. In addition, the students worked on the digital exercises of their own accord, which means that most students did not work rigorously on the exercises (see also the results in Table 8). Still, 50% of the students in case study 2 said to have learned “much” or “very much”, which is a reasonably good result, in spite of the differences.

**Learning efficiency**

In case study I, the students were asked to estimate how different parts of the food chemistry course contributed to the ability to pass the final exam (q14.1A) and to estimate how much time they spent on each part (q14.2A). For the first question, students graded for the eight parts of the course (Table 10) the percentage they felt that each part contribution to their capability to pass the final examination, taking into account that the total contribution for these eight parts should add up to 100%. Students also indicated how much time (hours) they spent on each part.

From the estimated contribution and estimated time for each part a learning efficiency can be calculated. Efficiency in general is defined as the ratio of the output to the input of any system. Learning efficiency could, therefore, be defined as the ratio of the percentage contribution of a part to the ability to pass the final exam to the percentage time spent on that part. The learning efficiency for each part was calculated for each student with the following formula:

$$\text{learning efficiency}_{(\text{part})} = \frac{\% \text{ contribution} \ (\text{part})}{\% \text{ time} \ (\text{part})} = \frac{\% \text{ contribution} \ (\text{part})}{\text{time} \ (\text{part})/\text{time} \ (\text{total})}$$

In this formula:

- contribution (part) = contribution of a part to the capability to pass the final examination estimated by the student (%)
- time (part) = time spent on a part estimated by the student (hours)
- time (total) = total time spent for all parts as estimated by the student (hours)

When the learning efficiency is 1 the percentage time and the percentage contribution are equal. The higher the number the more efficiently the time is spent. The average estimated contributions and average of the total time spent on the eight parts, according to the students, are listed in Table 10, together with the average learning efficiencies. The learning efficiency is calculated for each student and then averaged. All averages have very large standard deviations.

The learning efficiency of a certain part can only be compared with other parts of the course if they have common educational goals. The first four parts in Table 10 are related to activities that are offered by the course. Of these four parts, the digital exercises and the readers have the same learning goals. The lectures are more intended for introduction and the quantitative exercises specifically deal with quantitative understanding of chemical reactions in food. The quantitative exercises are based on a problem based learning paradigm. A description of these exercises falls outside the scope of this article. The second two parts, lecture notes and notes from the exercises, are learning activities that students produce themselves. The last two parts are related to the laboratory classes. These have separate learning goals, mainly laboratory technique and report writing. The skills obtained during the laboratory classes are not examined during the
final examination and, therefore, the laboratory classes have a low learning efficiency with respect to the contribution to the final examination (Table 10).

**Table 10:** The average estimated contribution to the final examination, the average estimated percentage of time spent and average learning efficiency calculated per student for case study I (n=27).

<table>
<thead>
<tr>
<th>Parts</th>
<th>Average % contribution</th>
<th>Average % time</th>
<th>Average learning efficiency</th>
<th>% students with learning efficiency ≥ 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attending lectures</td>
<td>11</td>
<td>14</td>
<td>0.8</td>
<td>19</td>
</tr>
<tr>
<td>Completing quantitative exercises</td>
<td>9</td>
<td>8</td>
<td>1.5</td>
<td>56</td>
</tr>
<tr>
<td><strong>Doing the digital exercises</strong></td>
<td>21</td>
<td>11</td>
<td><strong>1.9</strong></td>
<td><strong>64</strong></td>
</tr>
<tr>
<td>Reading the reader</td>
<td>30</td>
<td>25</td>
<td>1.4</td>
<td>75</td>
</tr>
<tr>
<td>Studying own lecture notes</td>
<td>5</td>
<td>2</td>
<td>4.7</td>
<td>100</td>
</tr>
<tr>
<td>Studying own notes from exercises</td>
<td>7</td>
<td>2</td>
<td>5.3</td>
<td>93</td>
</tr>
<tr>
<td>Doing the practical part in the laboratory*</td>
<td>9</td>
<td>28</td>
<td>0.4</td>
<td>10</td>
</tr>
<tr>
<td>Writing the report for the practical part in laboratory*</td>
<td>7</td>
<td>9</td>
<td>1.1</td>
<td>41</td>
</tr>
</tbody>
</table>

*: The average percentage time and average efficiency are both calculated by first calculating this number for each student and then averaging this over all students.

*: The laboratory classes have their own learning goals, which are not examined during the final examination. Separate grades are credited to the laboratory part.

For quite a lot of the students (64%), an efficiency rating of the exercises is higher than 1 (average is 1.9). This indicates that requirement r14 is met, since for that the average learning efficiency should at be 1 or more. Moreover, when comparing the learning efficiency for the exercises with reading the reader, the exercises score well. Therefore, with an average learning efficiency of 1.9 and an average contribution of 21%, it can be concluded that the digital exercises certainly have an additional value for the students.

In case studies II and III students were not asked to rate the contribution. Instead, students were asked to rate the statement “The exercises contribute to my ability to answer questions from the final exam” on a 5-point scale. Unfortunately, students were not asked to rate this statement for the other parts of the course. Results are shown in Table 11, which contains the results of case studies II and III for the questions related to requirement r14.

In case study III we were only interested in the efficiency of the exercises, since other parts of the course (e.g. lectures and readers) were outside our influence. Of the students that filled in the questionnaire 50% agreed that the exercises contributed to their ability to answer questions from the final exam (Table 11). So, although the content of the exercises is derived from the reader of the course at Wageningen University, still half the students at the corresponding course at Cornell University could answer questions of their own examination with what they learned in the digital exercises.

In case study III 75% of the students agreed with the statement “The exercises are efficient for learning: I learned about Food Chemistry in a relative short period of time.” (q14.2, Table 11). This result shows that although the content of courses are not equal, the students learned in an efficient way about Food Chemistry. So, for the exercises it seems that when content is not identical to the content of the course the usefulness for examination is not 100%, but they can still help students learn about food chemistry. This implies that the digital exercises can be a valuable aid for food chemistry courses outside Wageningen University.
Table 11: Results for the questions related to requirement r14. The questions differ in the different case studies. There were open questions, multiple-choice questions and questions on a 5-point scale.

<table>
<thead>
<tr>
<th>Question / Statement</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Case study II</td>
</tr>
<tr>
<td>The exercises contribute to my ability to answer questions in the final exam.</td>
<td>3.5 (60%)*</td>
</tr>
<tr>
<td>Estimate the time spend on the digital exercises.</td>
<td>Average 15 hours</td>
</tr>
<tr>
<td>The exercises are efficient for learning: I learned about Food Chemistry in a relative short period of time.</td>
<td>-</td>
</tr>
</tbody>
</table>

*: Average of answers on a 5-point scale (1=totally disagree, 2=partly disagree, 3=neutral, 4=partly agree, 5=totally agree), with between brackets the percentage of students that chose 4 or 5 as an answer.

Case studies I and III show that the exercises were rather efficient for learning according to the students, but this efficiency as perceived by the students is not a good measure of the ability to pass the final examination successfully. Based on the case study results it will not be possible to provide evidence for the effectiveness requirements.

Conclusions

Despite the differences between the two case studies at Wageningen University and the case study at Cornell University it is clear that the digital exercises are useful to varying degrees in at least three courses with different structure and organisation. Based on the results of the evaluation process carried out in the three case studies it is concluded that most design requirements, that were described during the design process, are met. Therefore, we will continue to apply the design process for food chemistry digital learning material as described in Diederen et al., (2003). The design process and the evaluation process together are an example of a design oriented research approach in chemistry education. It has been shown in this paper that this research approach can result in useful digital learning material and in a satisfactory and clearly defined starting point for improvements of education. We expect that this research approach can also be used for development and investigation of other digital learning materials.

The evaluation results show that students’ appreciation of the same learning material is different when applied in a different course structure and/or organisation. Therefore, we have carefully described the situation of the case studies, and this is regarded as crucial for an accurate interpretation of the results. From the interpretation of the results, the following preliminary conclusions are described for further research on the effect of the on students’ appreciation.

First, when a course consists of several learning activities, the sequential order of these activities influences what students learn from each activity. This is explained by the following. The digital exercises seem to assist the students in studying the reader when the students have already gained some knowledge about food chemistry by attending the lectures. In contrast, students seem to learn more from the digital exercises without first attending several lectures. It is reasonable to assume that the knowledge of students prior to starting with the digital exercises will be influenced by previous exposure to relevant information in lectures. Also students who have less relevant prior knowledge invest more time in completing the exercises and thus are likely to learn more from working with the exercises. A second conclusion is about the usefulness of presenting a performance score to the student who does the exercises. This relates the prior knowledge of the student on the topic of the exercise to the value, which a student
attributes to the presentation of his performance score in that exercise. It states that low prior knowledge results in a low value attributed by the student to the function that presents a performance score to the student. And third, it is concluded that when the type of activities that are incorporated within the learning material matches with the learning activity preference, the student is more likely to use this learning material and vice versa. This is derived from the differences in use of the digital learning material between the case studies, which are, amongst others, explained by the idea that students will use approaches they like.

Finally the results of case study III are in keeping with the quite general belief that using digital learning material as basis for a self-study activity, which is not explicitly scheduled, will result in a low use of this digital material, even when the digital material is regarded useful.

Acknowledgements: We wish to thank the students and faculty from the Department of Food Chemistry of the Cornell Institute of Food Science, Cornell University for their participation, and Professor Dennis D. Miller in particular for his cooperation in this study.

References


Keyser M. W., (2000), Active learning and cooperative learning: understanding the difference and using both styles effectively, *Research Strategies*, 17, 35-44.


Identifying pedagogical content knowledge (PCK) in the chemistry laboratory

Janet Bond-Robinson,

Dept. of Chemistry, 1251 Wescoe Hall Drive, University of Kansas, Lawrence, KS, 66045, USA
E-mail: jrobinso@ku.edu

Received 5 January 2005, accepted 14 March 2005

Abstract: This study was carried out in the chemical teaching laboratory with new graduate students while they were guided to develop pedagogical content knowledge, PCK. PCK is expertise that demonstrates a combined knowledge of pedagogy and disciplinary subject matter; since chemistry is the discipline, the abbreviation, PChK, is used. Laboratory teaching functions for student learning entail guidance of chemical techniques, and abstract chemical concepts relevant to the lab experiment, that is, chemical explanations using concepts conceived by chemists rather than perceived, e.g., atoms and chemical bonds. Instruments were built with constructivist content and attained construct validity and internal consistency to measure teaching performance. A factor analysis reduced fifteen constructs to three forms of PChK, whose names reflect the level of chemical knowledge and pedagogical sophistication required. Mentoring activities were labeled as PChK-0. PChK-1 represents procedural knowledge to manage a chemistry laboratory. PChK-2 represents devising or generating transforming explanations connected to the students’ knowledge and previous experiences. A ‘transforming explanation’ is defined as a discipline-specific illustration of how people in that discipline think about a disciplinary process, which is linked by the explanation to students’ thinking about that same disciplinary-related process. PChK-3 guides students in chemistry-specific reasoning and generating transforming explanations for themselves. Examples of PChK-2 and PChK-3, using two chemical topics, are provided. [Chem. Educ. Res. Pract., 2005, 6 (2), 83-103]

Keywords: pedagogical content knowledge; PCK; teaching assistants; procedural knowledge; conceptual knowledge; chemical reasoning; laboratory work; practical work; tertiary level; undergraduate students; preparing future faculty; factor analysis; statistical methods.

Introduction: Pedagogical Content Knowledge (PCK) in Chemistry

The science of chemistry provides chemical knowledge, some of which is designated subject matter to be taught at primary, secondary, tertiary, and graduate levels of education. Shulman (1986) defined pedagogical content knowledge (PCK) as interwoven pedagogy and subject matter knowledge necessary for good disciplinary teaching. Pedagogical content knowledge is considered to be craft knowledge, defined as “integrated knowledge which represents the teachers’ accumulated wisdom with respect to their teaching practice. As craft knowledge guides the teachers’ actions in practice, it encompasses teachers’ knowledge and beliefs with respect to various aspects such as pedagogy, students, subject matter, and the curriculum” (van Driel et al., 1998, p. 674). Craft knowledge is acquired from prior education, the teachers’ personal backgrounds, the teaching contexts, and through experience in the ‘doing’ of teaching. Therefore, the wisdom of craft knowledge produces effective
behavior on the part of the teacher who possesses it. The military version of wisdom concerns utilization of both strategies and tactics and provides a metaphor to wisdom in teaching. Strategies involve long term directing and maneuvering forces and equipment into the most advantageous positions (Agnes, 1999); in teaching it is the directing and maneuvering of students, materials, and equipment in the classroom or laboratory, thus orchestrating and directing for the most effective environment for learning. Maintenance of the effective learning environment requires using the particular set of strategies each session; these manage the personal climate, guarantee proficient student work, enable peer learning and discussion, and provide opportunities to reason in the discipline, for examples. Tactics are skillful methods or procedures that meet local and short-range objectives. Tactics arise based on the particular topics to be taught; therefore, tactics may differ to some degree from topic to topic and change as the teacher acquires more knowledge of pedagogy, the students being taught, and the interrelationships of subject matter in the curriculum.

Craft knowledge is necessary at the tertiary level as well. Despite the absence of formal teacher training among most tertiary faculty, the craft knowledge of PCK guides the creation of a learning environment and guides a teacher’s actions in teaching a specific subject matter. Thus craft knowledge is as relevant a concept to university level faculty and teaching assistants as it is to primary and secondary faculty. In large universities graduate students instruct chemistry laboratories in the undergraduate (UG) curricula. During the times that graduate teaching assistants (GTAs) work through the challenges of teaching chemistry labs, they develop pedagogical chemical knowledge (PChK) that drives their orchestration of the environment, the interactions with students, and the tactics to address the learning of a chemical topic. We will continue to use this abbreviation when referring to chemistry teaching. Interacting with students in strategic and tactical manners that provide an effective learning environment for the majority of students is part of the challenge of professional teaching practice. Shulman (2002) explains that the learning process begins with teacher’s engagement with students: “Critical reflection on one’s practice and understanding leads to higher-order thinking in the form of a capacity to exercise judgment in the face of uncertainty and to create designs in the presence of constraints and unpredictability” (p. 38). Just as each battlefield restricts the battle in certain ways and always necessitates responses to unpredictable events, so a learning environment requires appropriate responses in that the teacher flexibly confronts problem issues that arise, judges the salient features, and makes prompt decisions for the benefit of student learning.

The chemistry teacher at tertiary levels is one who teaches organic mechanisms, stoichiometry, kinetics, various aspects of thermochemistry, and many other topics, which is why pedagogical chemical knowledge (PChK) about specific topics is an important area for further research according to other chemical educators (Geddis, 1993; van Driel et al., 1998; Bucat, 2004). Ideally, Shulman (1986) says, the teacher will transform disciplinary knowledge to encourage understanding of meaning by his/her population of students. Transformation is an explanatory process that differs from giving an explicit restatement of the chemical view of a concept or theory and expecting students to remember it. Teachers must figure out what it means to transform chemical knowledge on specific topics that explains chemistry at the level of their students. When working to transform chemical definitions to meaningful explanations, the teacher is planning the tactics to use in a specific laboratory context. In a later section we will explicate what it means to transform chemical knowledge on a specific topic.

Our settings were chemistry labs (general chemistry and organic chemistry). Our current study of GTA development investigated PChK in greater detail for two reasons. We wished to outline features of important constructivist teaching and learning practices as well as to identify critical aspects of PChK that might generalize to teaching in any science laboratory or
inquiry setting. We used factor analysis of survey data as an exploratory method to generate theory about pedagogical content knowledge. Exploratory factor analysis identifies the factor structure or model for a set of variables, which includes both establishing the number of principal underlying factors in the data and the pattern seen in the correlations of each variable to the factors identified (Kim and Mueller, 1978). The theory generated about critical aspects of PChK was then applied to our specific findings about GTA development in PChK.

**Constructivist Teaching and Learning Practices**

A model of teaching/learning as knowledge transmission/reception is the dominant mode of teaching (Gallagher and Tobin, 1987). The basis for *transmission of explicit knowledge* from expert to novice is the expectation that a learner will receive an organization of knowledge about a topic from another’s understanding into his own understanding. Unfortunately, explicit transmission of knowledge from an expert rarely leads to deep understanding in the novice (Bodner, 1986). Having a constructivist model of learning demands that a teacher encourage students’ efforts to understand the material so as to remember it. Having a model of knowledge transmission does *not* demand that teachers explicitly *push* students to reason, so this teaching practice can inadvertently reinforce memorizing rules, facts, algorithms and procedures. A mixture of transmission of knowledge, supplemented with attempts by the teacher to guide reasoning using the new information, is more likely to be effective for getting students to understand complex subject matter.

The meaning of abstract, unfamiliar concepts or scientific explanations is constructed from many forms of input. Some understanding is gained implicitly from experience. Guiding students in chemical reasoning helps them build meaning because they have to learn how concepts fit together in appropriate chemical ways. To reason effectively, a novice student will need to reorganize his own current knowledge and understanding in light of abstract chemical ideas not previously encountered or not previously understood. Explicit transmission of what we want students to know is efficient; however, when telling students ‘to know,’ teachers usually mean they want students to understand the ideas they transmitted rather than memorize what they said. Usually, in lecture situations interaction is lacking, thus it is often difficult to determine students’ current levels of comprehension. The laboratory experience, therefore, assumes special importance for helping students to think through chemical concepts and explanations. Given that human reasoning is essential to generate meaning (Mead, 1917), and since getting the meaning of an idea and how other ideas connect are the bases of understanding, opportunities for individual and group reasoning are the *essence* of an effective learning environment for people. Meaning is constructed from active learning situations such as the following: doing work; attempting to solve problems; listening to or reading relevant information when it is needed; awareness of cultural rules, norms or standards as guides; from the back and forth, give and take of a personal conversation with relative experts or with peers, or a larger discussion such as a group meeting; putting together a piece of writing; and even from conversing with oneself as in introspection.

**PChK Development**

The goal of professional teaching is that GTAs should take more professional initiative in promoting the learning of undergraduate students (UGs) in the laboratory environment. General teaching expectations and basic knowledge about teaching chemistry labs can be

---

1 A relative expert is someone who knows more than another on a topic but is not considered an expert in the field.

*Chemistry Education Research and Practice, 2005, 6 (2), 83-103*
transmitted to GTAs in the form of rules or lists of actions, such as producing a syllabus, giving a lab talk at the beginning, enforcing safety rules, and making sure students clean up after the lab. As a result of the need to gain teaching judgment, however, GTAs must adapt, reflect, and learn to appreciate and perform professional teaching strategies and tactics while they teach. In a teaching apprenticeship, modeling of teaching can occur by the instructor or advanced peers in real time or through video clips. The coaching provides specific direction and feedback to implement a high standard of work through key performance criteria. Weekly seminar discussions among peer GTAs provide communication in our teaching-apprenticeship course model. GTAs need to discuss emergent problems, problems they solved, knowledge they gained, and questions that an individual GTA wants to discuss with the group. Personal and professional discussions can create a teaching community similar to a research group, in that dialogue and feedback often foster introspection about individual work and its progress.

The survey findings of Abraham, Cracolice, Graves, Aldhamash, Kihega, Palma Gil and Varguese (1997) provided evidence that supports teaching practices that place more emphasis on concepts than procedural work in the laboratory environment. In their large study (268 American institutions responding, a 68% response) faculty were pressed to choose the most important laboratory-learning goal among the following: learning facts, laboratory skills, scientific processes, concepts or positive attitudes. Faculty chose learning concepts as the primary goal for UGs in a laboratory program. Since faculty members’ primary lab goal for UGs is learning chemical concepts, then GTAs should perform with heavy emphasis on concept teaching.

Clearly the need in laboratory teaching is a balance between managing a chemical workplace and teaching chemistry, thus balancing procedural teaching and underlying concept teaching (Bond-Robinson and Rodriques, In press). In summary of our studies, qualitative analyses of video data demonstrated the practical manner in which high performing GTAs handled the balance between the procedural and conceptual teaching. The judgment that created the balance is craft knowledge gained in practice, which we call PChK. The following is a list of strategic approaches that exemplary GTAs used.

- Exemplary GTAs utilized all the time available throughout the lab session. Generally, the first half of the lab session concerned procedural knowledge and the opening conceptual overview. The UGs and GTA worked on aspects of procedures and on progressing into the experiment.
- Exemplary GTAs orchestrated the second half of the lab as well as the beginning parts. They started working with teams and pairs, explaining and probing UGs about chemically related meanings in their work.
- The exemplary GTA was strategic in conversations with students. Sometimes the GTA utilized explicit transmission of experimental information, e.g., when asked a questions, she decided it was best to tell the students what to do; other times the GTA tried to get UGs to reason about their lab work or procedural difficulties because she decided it would benefit them more to think the problem through.
- The exemplary GTA asked each group to write their results on the board for public display.
- The exemplary GTA often drew the students together for a discussion, at which time she pointed out results, asked questions about the meaning of results, summarized class results in discussion, and gave UGs time to discuss the significance of what they found.
- Exemplary GTAs led their UGs into reasoning about how their work and results related to chemical concepts and processes rather than merely lab procedures and concrete results.
Methodology

Problem
Our problem was to understand the facets of PChK as they emerged in chemistry teaching labs. We utilized our quantitative data from undergraduate students’ (UGs’) responses about their GTA’s actions to provide the basis for determining a structure of factors in constructivist teaching practices. The purpose of a factor analysis is to discern the meaning that respondents give to items, particularly the relative meanings of all items. We asked, “How is PChK identified and classified in the work GTAs perform while teaching UGs in the laboratory?”

Summary of valid and internally consistent instruments
Two instruments were built, validated, and tested for internal consistency. The instruments that we built to measure PChK development were based on features of a constructivist-learning environment. One instrument was designed for the GTA instructors to use; the other was designed for the undergraduate students (UGs) to use in assessment of their teaching assistant. The instrument for the instructor was built first and then tested for several iterations. Then the UG assessment instrument was built from it. Both instruments contain the same twelve strategic interactions; these defined the performance criteria for GTAs. Instructors used their instrument to code remotely acquired audio-video observations of the lab section. The UGs’ instrument (shown in the Appendix) contained an extra three items that broadened for students the meanings of respect and help, explanation and student-GTA discussions of troubleshooting. Both instruments were tested with the people for whom they were designed; internal consistencies of each instrument reached high Cronbach alpha measurements of .86 (instructors) and .95 (UGs rating their GTA). Information about the content and operational definitions is found in the Appendix. A full account of the methods in the course, the manner in which we built the constructivist assessments, and a copy of the instructors’ instrument is found in Bond-Robinson and Rodriques (2005).

Data Reduction and Identification of Underlying Factors
We did a factor analysis of 245 UGs’ responses to fifteen items about their GTA, which occurred at the end of the semester in the fourth iteration of the GTA course. Exploratory factor analysis identifies the factor structure or model for a set of variables, which includes both establishing the number of principal underlying factors and the pattern seen in the factor loadings (correlations of a variable to the factors). Factor analysis assumes that variables can be observed and measured, e.g., in our case the UG survey takers responded to the constructs within the UGs’ instrument from their observations and contact with their GTA. Techniques of factor analysis also assume that the observable variables are linear combinations of some underlying unseen factors (Kim and Mueller, 1978).

The main purposes of applying a factor analysis technique to a data set are to reduce the number of variables and to classify them (Statsoft, 1984-2003). We used rotational strategies in SPSS 11.0 statistical software to get effective differentiation among the loadings. The UGs’ instrument was shown to possess construct validity when theoretically similar constructs were similar in their loadings on the underlying factors, e.g. by the similarities in loadings on factors 1 and 2 of the interactions in Table 1-A. Dissimilar constructs were shown to be dissimilar in factor loadings, such as the contrast between probing and discussion interactions in Table 1A and respect and helpfulness interaction in Table 1D.

In addition, we used factor analysis as an exploratory method to generate theory about pedagogical content knowledge.
Results of the Factor Analysis

The manner in which variables loaded on these factors differentiated aspects of GTA performance that required chemical knowledge and those that did not. Table 1 shows how the fifteen variables in 245 student responses were reduced to two factors (with eigenvalues > 1) that explained 68.2% of the variance. Factor 1 had an eigenvalue of 9.17; Factor 2's eigenvalue was 1.07. Notice that numbers in the table illustrate that all the variables loaded on both factors to a greater or lesser degree. The analysis data also allows us to give meaning to the factor loadings. Analysis of the variables with the highest loadings on Factor 1 appear in Table 1-A and 1-B. Variables that load most heavily on Factor 2 begin with Respect 1, shown in Table 1-D. After much deliberation of the loadings, we concluded that the variables that loaded more highly on Factor 1 seem to be involved in purposefully teaching chemistry, albeit with differing strategies. We labeled Factor 1, therefore, ‘pedagogical chemical knowledge’. The highest loading variables on Factor 2 identified it as mentoring or advising activities sensitive to the students (Table 1-D). We labeled Factor 2 ‘responsive mentoring’.

The fifteen constructs from the UGs’ instrument are ordered in Table 1 by the magnitude of ‘loading’ on Factor 1. Variables in Table 1-A show similar magnitudes of loading on Factor 1 and similar loadings on Factor 2; these variables show covariation with each other; so we grouped them together and gave them a label. For example, those in Table 1-A were labeled ‘prompts conceptual thinking in students’. Note that Table 1-A further identifies these constructs as a form of PC\(hK\); those meanings will be discussed later. Another factor loading pattern is shown in Table 1-E: The variable of interaction, the last entry in Factor 1, had significant loading on both factor 1 (.527) and on Factor 2 (.526). This result was interpreted to mean that interaction is an important component of both factors.
Table 1 Factor Analysis on GTAs’ teaching (parts A-E). Each part shows labels given to actions with similar loadings on Factors 1 and 2. Extraction method was ‘Principal Component Analysis. Rotation Method: Varimax’. The two factors (Eigen value >1.0) explained 68.2% of the variance.

### A.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fac. 1</th>
<th>Fac. 2</th>
<th>Operational Definition</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Troubleshoot With Peers</td>
<td>.815</td>
<td>.233</td>
<td>Encourages us to discuss procedural problems together as a team.</td>
<td>Prompts peer conceptual thinking about experiment</td>
</tr>
<tr>
<td>Facilitates Reasoning</td>
<td>.800</td>
<td>.262</td>
<td>Stimulates our team to discuss chemical concepts underlying the experiment.</td>
<td>Prompts peer thinking about underlying concepts</td>
</tr>
<tr>
<td>Prompts Reasoning</td>
<td>.788</td>
<td>.292</td>
<td>Prompts me to think about chemical concepts when I ask questions.</td>
<td>Prompts thinking about underlying concepts</td>
</tr>
<tr>
<td>Troubleshoot By Reasoning</td>
<td>.750</td>
<td>.249</td>
<td>Encourages me to think through problems when mistakes are made rather than just telling me what to do.</td>
<td>Prompts thinking of acts or underlying concepts</td>
</tr>
</tbody>
</table>

Summary In general, the GTA uses pedagogical content knowledge in these actions taken with UGs, indicating they are taking initiative to prompt their students to reason through the procedural or concept-related subject matter.

### B.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fac 1</th>
<th>Fac.2</th>
<th>Operational Definition</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Links Concepts</td>
<td>.719</td>
<td>.407</td>
<td>Links chemical concepts to lab procedure so that I can understand them.</td>
<td>Initiates thinking about underlying concepts</td>
</tr>
<tr>
<td>Concrete Explanation</td>
<td>.710</td>
<td>.403</td>
<td>Discussions of chemical concepts with me occur at the level of my knowledge and previous experiences</td>
<td>Initiates thinking about underlying concepts</td>
</tr>
</tbody>
</table>

Summary In general, GTA uses pedagogical content knowledge to explain the procedural or concept-related lab work and link it to abstract chemical concepts in a manner that is related to their students’ current understandings.
### C.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fac. 1</th>
<th>Fac. 2</th>
<th>Operational Definition</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advice</td>
<td>.705</td>
<td>.521</td>
<td>Comments are helpful to my work during lab.</td>
<td>Initiates Help</td>
</tr>
<tr>
<td>Guidance</td>
<td>.690</td>
<td>.453</td>
<td>Comments on problems given at the level of my knowledge.</td>
<td>Initiates Help</td>
</tr>
<tr>
<td>Aware</td>
<td>.654</td>
<td>.410</td>
<td>Notices I’m having difficulties and helps, even if I do not ask.</td>
<td>Initiates Help</td>
</tr>
<tr>
<td>Short Talks</td>
<td>.630</td>
<td>.441</td>
<td>Short talks are clear and help me understand the experiment.</td>
<td>Initiates Help</td>
</tr>
</tbody>
</table>

**Summary**
In general, GTA uses pedagogical content knowledge to advice and guide students about the lab procedures, techniques, instruments and necessary calculations.

### D.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fac. 1</th>
<th>Fac. 2</th>
<th>Operational Definition</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respect 1</td>
<td>.299</td>
<td>.876</td>
<td>Is respectful of me as a person.</td>
<td>Responsive Mentor</td>
</tr>
<tr>
<td>Respect 2</td>
<td>.377</td>
<td>.800</td>
<td>Is respectful of my knowledge and ability to learn.</td>
<td>Responsive Mentor</td>
</tr>
<tr>
<td>Helpful</td>
<td>.434</td>
<td>.709</td>
<td>Helps me when I ask for help.</td>
<td>Responsive Mentor</td>
</tr>
<tr>
<td>Safety</td>
<td>.186</td>
<td>.686</td>
<td>Models safety and other rules and enforces them.</td>
<td>Responsive Mentor</td>
</tr>
</tbody>
</table>

**Summary**
In general, GTA uses no content knowledge, indicating they are respectful of students and helpful. While we understand that predicting safe conditions requires chemical knowledge, we assume that students believed the lab is safe or it would not be part of the curriculum.

### E.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fac. 1</th>
<th>Fac. 2</th>
<th>Operational Definition</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction</td>
<td>.526</td>
<td>.627</td>
<td>Interacts with us throughout the lab.</td>
<td>Fits strongly into both factors; thus a part of Mentor, PChK -1, 2, &amp; 3.</td>
</tr>
</tbody>
</table>
**Data Classification**

Table 2 Two ways to classify from factor analysis results. Based on a semantic differentiation scale from 1: ‘very poor or never occurred’ to 5: as ‘very good or occurred very often’.

<table>
<thead>
<tr>
<th>Class A: PChK oriented</th>
<th>Class B: Function oriented</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument Mentor &amp; Requires Chemical Knowledge</td>
<td>Instrument Chemical Manager &amp; Chemical Concept Teacher</td>
</tr>
<tr>
<td>UGs’ 4.59 4.11</td>
<td>UGs’ 4.29 4.21</td>
</tr>
<tr>
<td>Instructors’ 4.67 3.44</td>
<td>Instructors’ 4.34 3.22</td>
</tr>
</tbody>
</table>

We found two useful ways to classify these constructs from Table 1. In Table 2 we described them as Class A and Class B. Our discussion of pedagogical chemical knowledge will involve only Class A. (Class B was used with GTAs to emphasize their functions as procedural manager and chemistry teacher. Table 3 provides a summary of each section of Table 1 classified into four forms of PChK.

Table 3 Summaries of Forms of Pedagogical Chemical Knowledge (PChK)

<table>
<thead>
<tr>
<th>Form PChK</th>
<th>Table 1 Part of GTA Function</th>
<th>Knowledge Requirement</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>PChK-0 I-D</td>
<td>Management of the Laboratory Environment</td>
<td>Mentoring that does not require chemical knowledge.</td>
<td>Interacts with students; Helpfulness; Respects students’ abilities to learn</td>
</tr>
<tr>
<td>PChK-1 I-C</td>
<td>Management of Chemical Laboratory Environment</td>
<td>General procedural knowledge of chemical lab work; Specific technique, procedures, calculations, and safety knowledge of each lab investigation</td>
<td>Models and enforces safety precautions; Demonstrates techniques; Troubleshoots lab problems; Gives guidance to students</td>
</tr>
<tr>
<td>PChK-2 I-B</td>
<td>Teaching Chemical Concepts</td>
<td>Understanding chemical topics and concepts in order to transform them to make sense to students (which is dependent on student knowledge)</td>
<td>Correlates macro-level events with nano-level processes; Chooses examples wisely; Links chemical symbols, math variables, and nano-level processes together</td>
</tr>
<tr>
<td>PChK-3 I-A</td>
<td>Teaching Chemical Concepts</td>
<td>Flexible knowledge to probe and guide student’s reasoning as well as confidence in knowledge and role so as to direct the learning environment</td>
<td>Uses questioning strategies to probe conceptual reasoning; Gives occasional directed guidance; also directs students to work through questions or procedural problems with each other</td>
</tr>
</tbody>
</table>
Discussion of Results Associated with Research and Practice

There is a need for a theoretical basis of pedagogical content knowledge, particularly in the work of transforming chemical knowledge (Bucat, 2004). Perhaps these definitions of a variety of forms of PChK will provide that theoretical basis. Table 3 gives a short overview on the forms of PChK that we determined in chemistry lab. Therefore, we call this knowledge and performance pedagogical chemical knowledge, PChK.

Analyzing Pedagogical Content Knowledge

Difficulty in demonstrating the chemical knowledge and pedagogical sophistication of the forms of PChK was judged by the frequency of performance. The aspects of PChK-3 performance were seen the least often, and PChK-1 interactions were observed most often. The amount of chemical understanding and pedagogical sophistication required of the GTA for effective interactions increases from interactions showing PChK-1 to those showing PChK-3. High PChK-2 performers worked to attain sensitivity to student knowledge because it permitted their explanations to work better with their UGs. To perform with PChK-3, knowledge must be well organized and flexibly applicable to guide students’ work effectively and facilitate UGs in mechanical reasoning with components of the experiment and facilitate conceptual reasoning with the underlying concepts that the lab illustrates. Helping students to reason requires the knowledge base of PChK-2 as well as that of PChK-3. Further, interactions of PChK-3 require that GTAs take control of the learning environment in a professional manner. Final results clearly showed that the GTA instructors needed to put more developmental emphases on PChK-2 and PChK-3 interactions so that GTAs would meet the instructors’ higher standard of chemistry teaching.

PChK-0

Any knowledge possessed and acted upon by the chemistry GTAs that did not require chemical knowledge was labeled PChK-0. These are the interactions we described as responsive mentoring, which are important pedagogical actions. The UGs gave GTAs high marks for PChK-0. This is not surprising since new teachers often believe that their primary role is to have cooperative and friendly relationships with students (Geddis, 1993).

PChK-1

The PChK-1 performers required understanding of chemistry at the lab level to give general procedural guidance and give directed advice. See Table 3 for specific interactions. GTAs demonstrated techniques and provided relevant, helpful advice on the work progress during the lab session. The interactions of PChK-1 correspond most closely to the UG’s procedural emphases in that UGs wish to get the lab experiment started quickly and finished quickly (Malina and Nakhleh, 2003). Much of the knowledge employed in PChK-1 may have been generated from the GTA’s own UG lab experiences because the teaching model in those labs was more likely to have been one primarily of transmitting procedural knowledge (Abraham et al., 1997, Hilosky et al., 1998). Thus, modeling of PChK-1 on past laboratory educational experiences when they were UGs may account for the ease of GTAs’ understanding and the frequency of exercising PChK-1 interactions. Evidence supporting this claim is that international GTAs, who often have strong theoretical knowledge but less UG laboratory experience (Tanner et al., 1993), have more difficulty acquiring PChK-1 than domestic GTAs (Bond-Robinson and Rodriguez, In press). The UGs’ and instructors’ ratings indicated that the most complicated and difficult aspect of PChK-1 to execute effectively was
preparation and delivery of an appropriate talk. These were far more challenging to learn that it was for GTAs to provide guidance or give specific pertinent advice as the lab proceeded.

**PChK-2**

Performance showing PChK-2 required groundwork ahead of time, which is an essential aspect of transforming subject matter into forms for student consumption. As they prepared for lab, GTAs had to identify underlying abstract chemical concepts that related to the lab investigations if they were to be effective in helping students understand abstract concepts that underlie the lab investigation. The following GTA difficulties emerged during seminar discussions and videotaped teaching observations. The value of identifying concepts was not evident to most GTAs at first; some never understood the need to identify them because they were not mentioned in the lab procedure. Facilitating UGs to connect these underlying abstract concepts with their lab work required that GTAs make some decisions ahead of time about ways to transform subject matters, e.g., by identifying clear examples; relating the reality of the laboratory to the atomic/molecular level; devising analogies from a familiar idea to a chemical one; or putting together mathematical variables to explain chemical processes. Even if the GTA identified the abstract concepts, it was very challenging to produce explanations at the level of UGs’ knowledge. GTAs had to figure out the extent of their UGs’ knowledge and abilities to reason, and doing so demanded some tactics. Transforming chemical knowledge was difficult for all and totally neglected by some GTAs. In summary, many actions indicating PChK-2, e.g., linking the lab to the abstract concepts of lecture, and making effective explanations at the level of UGs that were more than restatements of the chemical view, were weakly executed or non-existent after one semester. The weak performance here is one reason that instructors rated GTAs as barely above mediocre in chemistry teaching (See Table 2.).

**PChK-3**

The PChK-3 interactions performed by a GTA promoted reasoning between the GTA and UG(s) or among UGs. The PChK-3 required more aggressive objectives and positioning as well as the ability for a GTA to flexibly exploit his or her chemical knowledge as needed. The challenge GTAs found in understanding and executing PChK-3 was the other reason for their mediocrity in chemistry teaching. Some GTAs were uncomfortable ‘butting into’ the UGs’ workspaces. Many GTAs responded with inertia when the situation called for beneficial prompting of reasoning by using generic and directed questioning strategies (Davis, 2003). A generic question is not focused on a specific answer; instead its purpose is to encourage students to think and articulate, e.g., “Tell me about what you just finished.” Use of directed questions that asked about specific concepts was a more familiar technique, but GTAs generally performed directed questioning in only limited fashions. When a GTA did attempt to guide a group’s thinking in a troubleshooting process, doing so utilized the GTA’s own mechanical knowledge and confidence. Some GTAs never encouraged a group of UGs to discuss a problem or question among themselves. Our data showed that the interactions involved in PChK-3 were more difficult for new GTAs to perform than those of PChK-2.

**Implications For Teaching and Learning**

One of the ways to clarify scientific explanations is to be as precise as possible about what they are. In the literature of science education, scientific explanations are discussed under many different banners: as theories, as models, as argument patterns, as analogies, as mathematical equations, as exemplars. As far as the classification hierarchy goes,
‘explanation’ is the generic, overarching term. Descriptors, such as theories, models, analogies, etc., interact as sub-categories under the category of explanation.

What does it mean to transform chemical subject matter?

I have used the term, transformation, as others do (e.g., Bucat, 2004), because it is a part of Shulman's description of what PCK looks like. Shulman describes the kind of explanations that effective teachers make as transformations of subject matter knowledge, which are appropriate to the level and specific characteristics of their student populations. This kind of explanation is named, a transforming explanation, in that this kind of explanation makes the representational meaning explicit that chemists have for chemical ideas by connecting them to the macroscopic level of the students. Since transforming explanations are appropriate to the level and specific characteristics of their student populations, it is important to know how students think about scientific explanations. A great number of studies in science education have examined the relationship of students’ explanations of various scientific phenomena with those of scientists, but none have examined UGs’ views of what a scientific explanation actually is. I did a study with an undergraduate researcher to examine general chemistry students’ views of scientific explanations (Bond-Robinson, 2004, Bond-Robinson and Harrington, 2004). We found that students’ responses indicated they had actually answered two different questions: (1) aspects of science-course topics they thought needed to have an explanation, and (2) characteristics that make an explanation effective for them personally. They wanted explanations about facts of the discipline as they were related to natural phenomena; procedural knowledge such as how, when, or what to do in assignments or problem solving; and applications to their own lives from topics under study. Further, students expected clear explanations of what they needed to know. Characteristics of good explanations were closely linked to their current knowledge and relevant to them personally. Good explanations were simple as opposed to complex or technical; good explanations were concrete as opposed to abstract. We found their expectations about the nature of scientific explanations in a science course to be pragmatic, personal, daunting for teachers to meet, and fairly unrelated to the theories in science. The study of science produces explanations, which tend to be theoretical in modern science; these explanations are not concrete, simple, or necessarily at the level of any student since scientific explanations are generated by scientists for other scientists. Commonly, teachers believe that their job is to show students how scientists have explained particular phenomena. Unless the scientific explanation is fairly radically transformed, the students will not recognize the statement as an explanation for them. For example teachers expect to explain the theories and models of their science, which are representational explanatory tools for scientists. Students did not find it necessary to reason like a chemist to gain effective chemical explanations. A specific definition for transforming explanations could be the aspects of chemistry that are not obvious to students, aspects of chemical reasoning that require a teacher to demonstrate and for which students need help to utilize. A transforming explanation is a chemistry-specific illustration of how chemists think about a chemical process, which is linked by the explanation to students’ thinking of that same chemical process. In chemistry students’ thinking, the process is likely to be at a macroscopic level of their personal experience with objects or events associated with a chemical process, supplemented by textbook and class-taught ideas, facts, and pictures. Students may not accurately distinguish between a chemical process and a conceptual object. We will discuss the student’s failure to distinguish between them later.

Chemical Explanations and ‘Transforming Explanations’ of Chemistry for Students

What explanations does the science of chemistry provide? What are these representational tools? In chemistry chemists discuss the occurrences of chemical reactions using atomic and
kinetic theories and those of thermochemistry. Chemists explain why and how those reactions occur with accompanying stoichiometry at the invisible level of atoms, ions, and molecules. Understanding the chemical meaning of a reaction is to correlate the visible chemical change with a mental model of atoms, ions, and molecules reacting in a nanoscopic world. Further, chemists see a process occurring over time rather than a static conclusion.

The chemical explanations must be transformed to the level of the student population, which will be a ‘transforming explanation.’ Note that understanding stoichiometry requires that students acquire a mental model of a process in motion, visible chemical change, and connections to the nanoscopic world of atomic and molecular interactions. Transforming explanations have the power to guide students in developing an appropriate mental model. Students of chemistry must learn to express this chemical process symbolically in an appropriate chemical sentence, keeping track of correct chemical formulas and conservation of mass. Finally, they must apply proportional reasoning so they can utilize mathematical symbolism to calculate the magnitude of reactants required to form a designated magnitude of one of the products.

Figure 1 shows a tetrahedral model of chemical reasoning, which deals with the ramifications of the stoichiometry example and reflects how chemists routinely move among representations of each modality as they solve problems. The tetrahedron is an adaptation of Johnstone’s two-dimensional triangular model (1991) that shows sub-microscopic, symbolic, and macroscopic representations of chemistry at each corner. Bucat (2004) emphasized that the meaning of chemical representations has to be acquired by students new to the field of chemistry. Teachers must teach the representations that chemists use in their work, which are highly varied even when related to the same process (as shown in the stoichiometry example). Some nanoscopic representations are formulas for compounds as well as structural 2-D and 3-D structural representations.

**Figure 1** Five Modes of Representational Reasoning in Chemical Work

Chemists use the term, mechanism, to describe the step-by-step process in which atoms, molecules, ions and electrons interact to complete a reaction at the nanoscopic level. A reaction can be viewed as a system of working parts (Agnes, 1999), which brings to mind the substrates, equipment, reagents, and methods that led to the reaction mixture. As a result, mechanistic systems exist at both the nanoscopic and macroscopic levels of interaction. It may be that students do not expect to learn, or do not realize that they need to learn, explanations that have layers of complexity. Therefore, part of the difficulty in learning
chemistry may be that students do not realize that explanation in chemistry requires using the multiple ways in which chemists reason about a chemical process. Transforming explanations have the capacity to illustrate the true nature of chemistry. Since students’ experiences exist at the macroscopic level, effective transforming explanations must repeatedly come back to the macro level of the chemical process under study. In the past we have called this kind of transforming explanation, ‘making chemistry relevant to students’. The tetrahedral model of representational reasoning allows chemists to become more systematic when designing transforming explanations for their students. When making a transforming explanation teachers must start from what students already know and find ways to reasonably connect them to another chemical representation of that process under study. Thus, transforming explanations occur one at a time.

How to Transform Chemical Knowledge (PChK-2)

The nature of transforming explanations clearly depends directly on the disciplinary content as is illustrated with chemistry. In chemistry, the specific definition of transformation of chemical subject matter knowledge utilizes the representational reasoning of chemists. The analogies, theories, exemplars, models and mathematical equations will all be closely related to the same representational reasoning patterns that were illustrated in the tetrahedral model.

Our grounded definition of PChK-2 requires that teachers link the concrete environment of the lab with abstract concepts and representations of concepts occurring in chemical processes. We hypothesize from Figure 1 that transforming explanations in chemistry require linking macroscopic matter and events to nanoscopic particles and processes and their corresponding chemical symbolic representations. Some transforming explanations additionally require mathematical symbolism in equations. All these transforming explanations are examples of reasoning. Thus, Figure 1 with its five modes of representational reasoning illustrates that teachers can make explicit explanations in a variety of ways. For example, chemical explanations of phenomena are usually given at the nanoscopic level in the textbook. Using Figure 1 as a template, the teacher can transform the nanoscopic explanation in at least four ways: (1) remind students of the relevant objects and materials that are present in the environment; (2) illustrate the macro mechanical system pictorially as it happened or will happen; (3) write the chemical symbols that represent everything that occurred in the chemical process; and (4) introduce the mathematical relationship among variables that shows the cause and effect relationships, and which quantifies the chemical process. Crucial in transforming explanations is that students’ attention is directed to specific features of each type of chemical representation. Many teachers readily illustrate some of these representational differences and how chemists use them. The difference in what I am proposing is that using the tetrahedral model about how chemists reason leads to systematic generation of transforming explanations each time we teach a new topic.

Graduate students have knowledge about chemical topics, but the vast majority does not have knowledge about what Bucat (2004) calls “particular teaching and learning demands” (p. 217). New teachers must learn to transform subject matter to produce effective explanations. In Table 4 we utilize the tetrahedral model of representational reasoning and describe the teaching demands of a common lab topic that many of our GTAs taught recently, the enthalpy of phase changes. We show how the teaching/learning demands relate to differing representations integral to these five modes of reasoning. In an experiment investigating enthalpy of phase changes, the UGs would benefit from transforming the mechanics of the calorimetry system to specific relationships among matter and energy. For example, a reasonable connection exists between the $\Delta T$ in an equation about heat of phase change and the perceived change in temperature of water in their calorimeters, which is sensed or seen instrumentally as a moving line on a real time graph. The amount of
temperature change can be related to a conceived concept that chemists use to describe how different materials transfer heat differently, specific heat capacity. Another transforming explanation connects the change of temperature to the change in kinetic energy of moving water molecules and molecular motion in phase changes. Reasoning about the kinetic energy leads readily into reasoning specifically about intact molecules (using structural formulas) to avoid the conception that change of phase involves breaking of intramolecular bonds. Individual transforming explanations, such as these made over time by an instructor, enable students to learn how to reason at the unseen nanoscopic level where most chemical explanations of phenomena transpire.

Why must teachers promote chemical reasoning by their students using PChK-3?

Teaching science as static facts, static representations, or static models requires no reasoning. Producing explanations that make chemistry understandable to students requires that the teacher show them how to reason like a chemist. Chemists reason with the varied forms of representations chemists use. Teaching students the representations of chemistry is getting students to ‘see’ things in motion the way that chemists do. Reasoning with chemical representations includes specially designed words that represent scientific concepts created as tools to conceive (Blumer, 1931). For example atoms, electricity, and mass are such conceptions, which we could call theoretical objects. Theoretical objects are components of chemical models. The complexity and special definitions for students to understand are challenging. Without conceptual understanding many students do not distinguish between similar theoretical concepts, such as between heat and temperature (Carson and Watson, 1999; Harrison et al., 1999; Greenbowe and Meltzer, 2003). Clearly, students can benefit from transforming explanations that show how verbal, theoretical vocabulary is connected to concrete instances in the lab, i.e., a transforming explanation from explanatory theoretical relationships to relationships among concrete objects and visible processes.

Mechanical reasoning in the tetrahedral model is our addition to a portrayal of chemical thinking. Chemical thoughts are not static facts or representations. Chemical thinking describes chemical processes or mechanical systems that chemists build. It has been pointed out in the cognitive literature (Chi, 1992) that matter and processes are often confused in students’ natural classifications of chemical phenomena. Producing a transforming explanation requires teachers to help students separate chemical processes from chemical objects, for example distinguishing between the chemical conception of movement of energy and the conception of a material quantity of heat (such as 19th century caloric theory). Mechanical reasoning describes a sequence of cause and effect relationships. Models are one type of mechanical explanation. General classroom usage of the term, model, may also be confused between an object and a process. The word model implies a representation that is a standard of excellence, or a small copy of an existing object such as model car; additionally a model is also defined as a guide or plan to be followed, (Agnes, 1999) such as a blueprint. The latter of these connotations fits the scientific usage of models best whereas the model car is probably the more common connotation of model for students. The solar system model of the atom is not important for how the atom looks; it is important for what it reveals about how an atom works. The chemist visualizes an atomic model as a working model. I argue that a scientific model illustrates how theoretical objects work; a scientific model is not a static representation. Therefore, teaching students the representations of chemistry is getting students to ‘see’ things in motion the way that chemists do. Teaching science as static facts or static models requires little in the way of reasoning. Therefore, transforming explanations made to students in chemistry often involve mechanical reasoning of a sequence of cause and effect relationships in chemical processes with theoretical objects or in a mechanical system built with macroscopic objects.
How to Promote Reasoning Among Chemistry Students (PChK-3)

Modeling reasoning by making transforming explanations to students at their levels of understanding is the first step toward getting students to reason by making transforming explanations themselves. Thus PChK-2 precedes PChK-3. The tetrahedral model can be used as an instrument for teachers to reason among chemical representations with their students; but the teacher has to be modeling this kind of reasoning explicitly in order for students to acquire the pattern and the habit. Following exemplary modeling the tetrahedral figure as a tool is at least as significant for students to use similarly and explicitly. The action and thinking required to write out each type of representational usage, to draw the representations of each type, and to reason through each of them allow students to visualize the similarity of meaning among representations and gradually acquire understanding of the uniqueness of meaning shown by the features of certain representations. Bucat (2004) worries about limiting complex understandings when he describes “statements subject to shallow interpretation” (pp. 217-218) that lead to the superficial understandings of chemical phenomena. Visualizing multiple modes of representations and reasoning from one to another requires that students perform chemical reasoning, which leads them into deeper understanding from a chemical point of view. Ideally, learning to make a transforming explanation one at a time, i.e., from one mode of representation to another, avoids overloading functional working memory.

Generalizability of PChK to Science Labs and Inquiry

Although this study investigated lab teaching by graduate students in general chemistry and organic chemistry, the three forms of PChK have broader applicability to PCK in the sciences. For example, the progression of knowledge building and performance that revealed itself among graduate students in chemistry is relevant to teaching assistants in other science laboratories. The instructors’ and UGs’ instruments would fit any science lab setting. Secondly, ‘transforming explanation,’ defined as a discipline-specific illustration of how people in that discipline think about a disciplinary process, which is linked by the explanation to students’ thinking about that same disciplinary-related process, is generalizable as written to all forms of PCK.

The development of graduate students parallels many aspects of novice classroom teachers’ development as they teach students in secondary and primary schools. Strategic interactions were adapted from constructivist teaching inventories used at the secondary level that were not chemistry specific. These twelve interactions are ideal ways to facilitate and assess the quality and frequency of these interactions in the classroom as teachers attempt to implement and facilitate student inquiry activities.

Results and implications from this work can apply directly and beneficially to pre-service science teachers under the conditions that a progression of development is encouraged. Currently, there is little emphasis on teaching prospective teachers about the day-to-day mechanisms to facilitate disciplinary knowledge construction by students. Emphases on questioning strategies, (e.g., Davis, 2003; Penick, 1996), have been useful, but they are not usually seen in a constellation of twelve interactions between the teacher and students. In addition, this study showed that a natural progression exists from the mentoring practices of PChK-0 to the procedural knowledge of PChK-1. It is a greater challenge for teachers to transform their own subject matter knowledge on the relevant science topics (PCK-2). Finally, it is a further significant transition to learning how to orchestrate and promote students’ to reason in a disciplinary fashion, which is a manifestation of PCK-3.
**Missing Forms of PChK Not Found in the Laboratory**

The laboratory-teaching assistants do *not* design the laboratory program; faculty members and laboratory coordinators do. Thus, a significant portion of teacher expertise is absent in laboratory teaching. Therefore, we propose the identity of PChK-4 that supplies that missing expertise in a chemistry classroom. Generic PCK-4 is demonstrated when teachers design effective day-to-day curriculum, activities and resources for their population of students. This labeling of practical, useful, and strategic design knowledge as PCK-4 is appropriate, not because it is the last identified, but because the demonstrated work of PCK-4 requires the knowledge of procedural work at the macro level, transforming explanations of topics in effective explanations in many reasoning modes, and direction of an effective learning environment that promotes reasoning among students, thus PCK-1 through PCK-3. Therefore, laboratory teaching is an appropriate *prerequisite situation* for new teachers to experience the responsibilities for student learning. They begin teaching practice by acquiring PCK-1, PCK-2, and PCK–3 before the necessity arises to learn how to design day-to-day disciplinary curriculum and activities. We hypothesize that a PCK-5 construct could exist that represents the craft knowledge necessary to prepare pre-service and in-service science teachers for directing and facilitating deeper science learning among their students which require PCK-0 to PCK-4.

**Conclusions**

The difficulty in acquisition of forms of PChK occurs in the order: of PChK-3 > PChK-2 > PChK-1 > PChK-0. The interactions of PChK-1 overwhelmingly occur at the level of the macroscopic environment. The PChK-1 interactions involve generally operating at the tangible, observable level of concrete objects. Table 4 gives macroscopic examples of reasoning that occurred in a lab experiment about enthalpy of phase changes. In addition to macroscopic sensing and availability of tangible objects, macro-mechanical reasoning illustrates procedures carried out by UGs. The obvious macroscopic nature of PChK-1 interactions is another reason that GTAs perform them at a higher frequency than interactions of PChK-2 and PChK–3. The obvious nature of PChK-1 interactions also explains why the UG labs that GTAs experienced were likely to have modeled procedural macroscopic types of interactions between teacher and students. Therefore, the new teachers’ difficulties in performing with PChK-2 and PChK-3 are due to at least two factors: infrequent modeling of constructivist teaching practices experienced as students, and the nature of chemistry’s explanations of phenomena that describe theoretical objects (scientifically conceived concepts) interacting through conceptualized processes at the nanoscopic level.
Table 4 Modes of chemical reasoning in a laboratory experiment: enthalpy of phase changes

<table>
<thead>
<tr>
<th>MODE OF REASONING</th>
<th>EXAMPLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>NANO REPRESENTATION</td>
<td>Molecules and their activities, changes in kinetic energy of molecules, changes in intermolecular forces among molecules; latent heat energy used to change intermolecular forces instead of changing temperature. These are examples of theoretical objects.</td>
</tr>
</tbody>
</table>
| MACRO REPRESENTATION | Finding and touching equipment  
Seeing melting, vaporizing, and subliming  
Physically sensing temperature and its changes  
Graphically seeing temperature change as generated by probe & software  
Words and pictures from reading about the lab experiment  
Seeing and hearing directions and pictures on chalk board |
| MECHANICAL REPRESENTATION | Mechanical Work at Macro Level:  
Make a calorimetry system to generate desired data  
Measuring the mass of a substance  
Combining two substances at different temperatures  
Mechanical processes at Nano Level:  
Physical process of phase changes at the level of moving molecules  
Mechanical processes expressed at Mathematical Level:  
Relationship of change in temperature to the heat capacity of a particular substance  
Relationship of heat gain to enthalpy  
Relationship of Law of Conservation of Energy to their experiment |
| CHEMICAL REPRESENTATION | $N_2(l)$  
$CO_2(s)$  
$H_2O(s)$ |
| MATHEMATICAL REPRESENTATION | $\Delta H$, $s$, $\Delta T$, $q$;  
$q = m \cdot s \cdot \Delta T$ |

Acknowledgements

To the National Science Foundation (REC 0093319 to Janet Bond-Robinson) for funding the most recent research; To Dr. Kyungmoon Jeon of Gwangju National University, Korea, for discussions about factor analysis; To Dr. Romola B. Rodriques for instructing GTAs and using the instruments.
References


Bond-Robinson J. and Rodrigues R.B., (In press), The science of design research used to catalyze graduate TA performance, *Journal of Chemical Education*.


Appendix

Additional Information about the Instruments’ Content and Operational Definitions

The content of the instruments contained strategic interactions necessary for a high-performing laboratory learning environment. The content of the instruments came from several sources. One source was the educational reform literature (Ausubel, 1968, Bruner, 1966, Dewey, 1916, Novak and Gowin, 1984). Another source was a valid constructivist observation instrument, the ESTEEM model, used in science classes (Burry-Stock and Oxford, 1994; Enger and Yager, 1998; Yager and Weld, 1999; Burry-Stock et al., 2000) that had been used by the author. The original ESTEEM instrument was not evaluated for the format, tasks, and time period in a chemistry laboratory setting; therefore, items relevant for teaching in the chemistry laboratory were added, e.g., enforcing safety, giving an opening talk, and specific procedural guidance in the laboratory. The first instrument for the instructors to use contained twenty-seven items as a result of the three content influences of the reform literature, the ESTEEM instrument, and the items specific for the chemistry laboratory. The items on the instrument are considered strategic interactions that, when executed well, lead to an effective laboratory learning environment for students. Thus, the purpose of designing this instrument was to pinpoint the strategic interactions, which then become the criteria for measuring a GTA’s teaching performance. The quality scale runs from very poor (rating of 1) to very good (rating of 5). Alternatively, a frequency scale ran from ‘never’ to ‘very often’. Content and construct validities were independently established over four iterations of the course. Validity of the instrument contents was thus based on the literature about characteristics of a constructivist-learning environment and the content of practical work done in a laboratory setting.

Construct validity was evaluated in the new setting with the population of students who enroll in laboratories as well as the very small group of two instructors who assessed GTA performance. First, construct validity depended upon consistent measurability of definitions that were created to describe each interaction. If an operational definition did not encourage clear measurement, the definition for that interaction was changed or dropped, thus increasing
the overall construct validity of the instrument. Over four iterations of the course operational definitions were modified, replaced with ones that could be recognized and measured more consistently, or dropped entirely. The original twenty-seven interactions were honed to the current set of twelve that could be reliably measured. See the first column of Table A-1 for the twelve interactions. Constructs are also referred to as variables.

**The Instrument Survey Given to Undergraduate Students in the Laboratory**

Data derived from the UGs’ instrument below was used as the basis for the factor analysis performed to determine the forms of \( PChK \).

### Table A-1 Undergraduate Students’ Instrument: Operational definitions in relation to the twelve strategic interactions

<table>
<thead>
<tr>
<th>Strategic Interaction</th>
<th>UGs’ Instrument #</th>
<th>CONSTRUCT (variable)</th>
<th>OPERATIONAL DEFINITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1</td>
<td>Directions</td>
<td>My TA’s short talks are helpful to my understanding.</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Safety</td>
<td>My TA models safety and other lab rules and enforces them.</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>Interaction</td>
<td>My TA interacts with me throughout the lab.</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>Helpful</td>
<td>My TA helps me when I ask for help.</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>Respect 1</td>
<td>My TA is respectful of me as a person.</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>Respect 2</td>
<td>My TA is respectful of my knowledge and ability to learn.</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>Trouble 1</td>
<td>My TA encourages us to discuss our procedural problems together as a team.</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>Trouble 2</td>
<td>My TA encourages us to think through problems or mistakes rather than telling us what to do or doing it for us.</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>Guidance</td>
<td>My TA’s comments on troubleshooting are relevant and at the level of my knowledge.</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>Awareness</td>
<td>My TA helps me when I am having difficulty with the experiment, even if I do not ask.</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>Advice</td>
<td>My TA’s comments are helpful to my work as the lab progresses.</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>Links Concepts</td>
<td>My TA links underlying chemical concepts from the lecture to the lab experiment to help me understand how they connect.</td>
</tr>
<tr>
<td>10</td>
<td>13</td>
<td>Prompts Ss</td>
<td>My TA prompts me about underlying concepts of the experiment when I ask questions.</td>
</tr>
<tr>
<td>9</td>
<td>14</td>
<td>Explanations</td>
<td>My TA explains the chemical concepts underlying the experiment with me.</td>
</tr>
<tr>
<td>11</td>
<td>15</td>
<td>Discussions</td>
<td>My TA stimulates us to discuss the concepts underlying the experiment with each other.</td>
</tr>
</tbody>
</table>
Conceptual understanding versus algorithmic problem solving: Further evidence from a national chemistry examination

Dimitrios Stamovlasis, Georgios Tsaparlis*, Charalambos Kamilatos, Dimitrios Papaoikonomou, and Erifyli Zarotiadou

University of Ioannina, Department of Chemistry, GR-45 10 Ioannina, Greece

Received 12 April 2002, accepted 2 March 2005

Abstract: Following our previous paper (Chem. Educator, 2004, 9, 398-405), we analyze further the results of a national examination from the perspective of conceptual learning versus algorithmic problem solving. Detailed achievement data were studied for a sample of 499 eleventh-grade students (age about 17), who were following various branches or streams leading to all kinds of higher-education studies in Greece (the ‘Positive’, the ‘Theoretical’, and the ‘Technological’ Branches). Using qualitative criteria, we distinguished the questions into: (i) simple knowledge-recall, (ii) conceptual, and (iii) well-practiced (algorithmic), stoichiometric, exercises. The latter could further be divided into simple and more demanding ones. As in the previous study, this categorization was also supported by statistical principal component analysis, but this time a marginal structure was extracted, because (possibly) of the limited number and the low difficulty of the postulated conceptual questions. The interest of the study lies mainly in the comparison among the different branches, with the students of the Positive Branch demonstrating the highest mean scores. In addition, students’ thinking was categorized according to Nakhleh’s scheme. The Positive Branch had the highest number of students with algorithmic and with conceptual ability, but all branches had about equal share of students high only in conceptual ability. [Chem. Educ. Res. Pract., 2005, 6 (2), 104-118]

Keywords: conceptual understanding; algorithmic problem solving; national examinations; principal components analysis; Nakhleh’s scheme.

Introduction

School chemistry, as any other school subject, is an internationally more or less settled subject with respect to its content and objectives. In upper secondary school, chemistry is usually presented as a simplified and reduced version of college general chemistry. Researchers in science education contend that such a situation and process is not correct: in school we do not aim at educating future scientists, but to develop instead a scientific literacy which is sufficient to make the future citizen able to understand and participate in decision making on crucial social and economic matters (National Education Standards, 1993). A main feature of the way school chemistry is taught and tested all over the world is that the emphasis is often placed on learning rules and algorithms that enable students to respond with success to examination questions, including relatively complicated computational ‘problems’/exercises. What happens, however, in the case of conceptual questions, even apparently simple ones? Further, are all students, irrespective of their aspirations for higher studies, demonstrating the same interest, inclination, and ability in the various kinds of test questions?

In a recent previous paper (Stamovlasis et al., 2004), we analyzed the results of the Greek
National Examination from the perspective of conceptual understanding versus algorithmic problem solving. Detailed achievement data in the special subject ‘Chemistry for the Positive Branch’ (see below) were studied for a sample of 647 eleventh-grade students (age about 17) who were oriented toward science, engineering or medical subjects (following the Positive Branch or Stream). It was demonstrated that principal component analysis (PCA) could serve as a tool for scrutinizing the items of examination papers in chemistry. Further, national, large-scale examinations provide reliable data that are appropriate for such an analysis. PCA distinguished between conceptual questions and the computational, well-practiced (algorithmic) questions. Some more demanding computational questions (requiring analysis and synthesis) shared some common space with the conceptual questions. On the other hand, the easy recall and simple application-of-knowledge questions were separated out from all other questions. The above conclusions were also supported by Multivariate Analysis of Variance (MANOVA). Achievement was at about the same level in the conceptual and the more demanding algorithmic questions. Finally, a scheme suggested by Nakhleh (1993) was also used to categorize the students according to the various categories of algorithmic versus conceptual thinking.

In this paper, we carry further the analysis of the results of the Greek National Examination in Greece from the same perspective, this time by considering the course ‘Chemistry for General Education’, taken by all eleventh-grade students (age about 17), irrespective of their inclination for specific subjects. The content of the course was mainly from organic chemistry. We repeat here that this examination was the first given after an educational reform in which, for the first time in the past thirty years or so, some of the questions required some form of conceptual understanding; in former cases, the dominant character of the examination questions was equally distributed between knowledge recall and algorithmic exercises. Both of these abilities (recall and algorithmic) were well practiced both within and outside of school. In contrast, the students did not have previous special training in manipulating conceptual questions in the specific domain of organic chemistry.

Formal statistical data revealed that ‘Chemistry for General Education’, among nineteen general and specialized nationally tested subjects (the latter including ‘Chemistry for the Positive Branch’) was the subject with the largest difference between oral and written marks, the former being allocated by the class teachers, while the latter coming from the objective national examination: a 30.3% difference. This was apparently due to the difficulty of the chemistry paper.

The rationale for dividing examination questions into conceptual and algorithmic/computational ones, together with the categorization scheme of Nakhleh (1993), which is also used in this paper, have been discussed and reviewed in our previous paper (Stamovlasis et al., 2004). Here we only need to repeat and stress three points: (a) that various authors have used various methods to categorize questions as being algorithmic or as requiring conceptual understanding; (b) conceptual questions have been associated with higher-order cognitive skills (HOCs), and algorithmic questions with lower-order cognitive skills (LOCS) (Zoller, et al., 1995; Zoller & Tsaparlis, 1997); (c) the degree to which an examination item is categorized as requiring conceptual thinking is, to some extent, a function of the students’ background and the sort of teaching they have been exposed to in class (Niaz, 1995).

A question that requires just LOCS for some students may require a shift to HOCs for others in a different context. It is therefore possible that questions (or some of them) that are categorized here as conceptual might be considered by students with a different background from that of the ones in our study not as conceptual, but as requiring just knowledge. On the other hand, computational questions may require for their answer not just the use of algorithms, but also conceptual understanding and critical thinking. As such, their relation
with conceptual questions may be not dichotomous (Niaz, 1995). In this work, of necessity, questions were identified as conceptual according to the operational definition of Zoller and Tsaparlis (Zoller & Tsaparlis, 1997; Tsaparlis & Zoller, 2003). This assignment was further checked by proper statistical analysis (see below).

**Method**

**The sample**

Our sample consisted of 499 students (age about 17), who took the Greek National Examinations for the Eleventh Grade. All the students came from various urban schools and were representative of the student population in urban areas in Greece. The large size of the sample makes it very likely that it was a homogeneous sample, even if students came from many different schools. The examinations took place at the end of the school year in June 1998-99, and were part of the university placement examinations that started in eleventh grade and were completed at the end of twelfth grade. Chemistry (organic chemistry) was one of the examined subjects for all students.

All students took the same courses up to the end of tenth grade. Starting at eleventh grade, students had to follow one of three branches (or streams): The ‘Positive’ Branch (PB), the ‘Theoretical’ Branch (ThB), and the ‘Technological’ Branch (TB). The PB is for students who want to study science, engineering and related applied subjects, or medicine, or related subjects. The ThB is for students who want to study literature, law, humanities etc. Finally, the TB leads to the same studies as the PB, except for medical subjects, but attracted weaker students.

At the outset, it must be noted that the weakest students joined the TB, while the academically stronger students separated into the two other branches: those with an interest in the humanities, etc. went into the ThB and those with an interest in science, medicine and engineering into the PB. These latter ones got extra instruction in chemistry, above that received by the other two branches (see Stamovlasis et al., 2004). Note that many schools had only two branches (ThB and PB) running together, while in a smaller number of schools all three branches were running.

All students took the same end-of-year examination in (organic) chemistry, and it is the results of this examination that are analyzed in this paper. In addition, the students in the PB took a further examination in (general/physical) chemistry, assessing the further instruction they received (the results of this examination have been analyzed in Stamovlasis et al., 2004).

The numbers of students in our sample were distributed as follows: PB, $N = 234$, ThB, $N = 172$, and TB, $N = 93$.

**The test**

The organic chemistry paper consisted of four parts: Part 1 contained largely recall, fixed-response questions, while Part 2 included both knowledge and a couple of simple conceptual questions (questions 2.3.a and 2.3.b). Parts 3 and 4 were more difficult, both involving stoichiometric calculations. The test and the marks allocated to each question and sub-question are included in Appendix 1. The test was constructed by a special committee of the Greek Ministry of Education. The present authors were not involved in the examination construction procedure, but two were involved in the grading procedure. Note that the national examination of Greece is similar to the examinations used for student selection for higher education, such as SAT (in the U.S.) and GCE (in England and Wales).

To facilitate analysis, a further distinction of the questions was made after agreement among the researchers. In Part 1, we grouped together the multiple-choice recall questions...
1.1-1.4, and the fill-in questions on organic reactions 1.6 and 1.7, but we kept separate the open (knowledge-recall) question 1.5, and the one-to-one correspondence (monomers-polymers) 1.8. In Part 2, question 2.1 was of the right-wrong type, with explanation, while question 2.2 asked for two organic reactions; of particular interest is question 2.3, consisting of two conceptual sub-questions 2.3.a and 2.3.b which deal with the exhaust gases of cars with or without catalytic converters. In Part 3, questions 3.a and 3.b demanded numerical stoichiometric calculations on the fermentation of sugar (in grape must) to ethanol (chemical equation given). Question 3.c provides the second example of a simple conceptual question: students had to judge and explain whether the mass of produced wine would be different from that of the must. Finally, in Part 4 all three questions (4.a, 4.b, and 4.c) involved demanding stoichiometric computations that also required knowledge of organic reactions.

An alternative, independent/objective evaluation/categorization of the questions of the test can be done by using statistical principal component analysis (PCA). This can lead to a reduction in the number of variables and to the detection of structure in the relationships between variables (Anderson, 1984; Stamovlasis et al., 2004). Using Kaiser’s criterion (Kaiser, 1958) and the scree test (Catell, 1966) we arrived at a marginal structure, which is in agreement with our qualitative categorization. The marginal structure may be attributed to the limited number and the low difficulty of the postulated conceptual questions. The results of this analysis are given in Appendix 2.

The duration of the examination was three hours. The papers that supplied the data for this study were marked by two chemistry teachers who are among the authors of this paper (E.Z. and D.P.). The marking procedure was similar to the one reported in our previous study (Stamovlasis et al., 2004).

Research questions
The following research questions were asked:

1. Was the postulated categorization of questions (conceptual versus algorithmic) supported by the data?
2. How were students, assigned as conceptual thinkers, distributed amongst the various branches? Were we right to assume that students with both high conceptual and algorithmic thinking would be found mainly in the PB? Were we right to assume that students with algorithmic thinking were expected to be distributed mainly in the PB and the TB?
3. Is Nakhleh’s scheme operating in our case?
4. Is competence in algorithmic problem solving connected with competence in conceptual problem solving (and vice versa)?

In addition, we wanted to see what differences in achievement were to be found among the three branches. We expected, of course, that the PB would have the highest achievement.

Results and Discussion

Achievement
The mean score of the whole sample ($N = 499$) in the whole paper was about 50% [49.4%, with standard deviation (SD) 26.6%]. This should be contrasted with the mean score of the PB in the special advanced examination in general chemistry ($N = 647$, $M = 68.0\%$, $SD = 26.2\%$ - see Stamovlasis et al. 2004), demonstrating the distinctive features of the organic chemistry paper. A steady decline of mean scores is observed in going from Part 1 of the test (80.4%, SD 17.0) to Part 2 (50.0%, SD 29.6), to Part 3 (36.1%, SD 37.1), and to
Table 1. Mean scores (%) in the questions of the test that are of greatest interest to this study (standard deviations in parentheses).

<table>
<thead>
<tr>
<th>Question</th>
<th>Total Sample ((N = 499))</th>
<th>Positive Branch ((N = 234))</th>
<th>Theoretical Branch ((N = 172))</th>
<th>Technological Branch ((N = 93))</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.a</td>
<td>40.4 (45.2)</td>
<td>52.9 (46.0)</td>
<td>31.4 (42.2)</td>
<td>25.9 (40.5)</td>
</tr>
<tr>
<td>2.3.b</td>
<td>54.5 (42.6)</td>
<td>60.0 (42.5)</td>
<td>51.4 (42.8)</td>
<td>46.4 (40.7)</td>
</tr>
<tr>
<td>3.a</td>
<td>44.3 (44.0)</td>
<td>60.6 (43.2)</td>
<td>33.0 (40.6)</td>
<td>24.2 (37.3)</td>
</tr>
<tr>
<td>3.b</td>
<td>31.2 (42.6)</td>
<td>47.1 (46.0)</td>
<td>20.9 (36.5)</td>
<td>10.2 (26.2)</td>
</tr>
<tr>
<td>3.c</td>
<td>29.7 (43.6)</td>
<td>37.9 (46.9)</td>
<td>22.8 (39.5)</td>
<td>22.0 (38.7)</td>
</tr>
<tr>
<td>4.a</td>
<td>31.9 (42.4)</td>
<td>42.3 (46.4)</td>
<td>26.9 (38.2)</td>
<td>15.1 (30.8)</td>
</tr>
<tr>
<td>4.b</td>
<td>34.4 (43.0)</td>
<td>46.0 (45.7)</td>
<td>28.9 (39.6)</td>
<td>15.5 (32.4)</td>
</tr>
<tr>
<td>4.c</td>
<td>21.5 (33.1)</td>
<td>31.9 (38.3)</td>
<td>15.4 (26.6)</td>
<td>6.5 (18.2)</td>
</tr>
</tbody>
</table>

Part 4 (29.0%, SD 35.2). As expected, the PB had a highest mean score (59.0%, SD 27.5), while intermediate was the ThB (44.1%, SD 23.7) and lowest the TB (35.1%, SD 19.4). The differences are statistically significant, as shown by one-way ANOVA for independent samples.

Table 1 contains detailed mean score results on the various questions that were of the greatest interest to this work. One-way ANOVA for dependent samples gives (through the Tukey test) the following critical values of differences of means for statistical significance at \(p = 0.05\): 8.6 for the total sample, 13.3 for the PB, 13.4 for the ThB, and 16.3 for the TB.

The two simple conceptual questions 2.3.a and 2.3.b proved hard for many students. More demanding was 2.3.a (40.4%), because of the chemical equation, while in 2.3.b the mean score was moderate (54.5%).

The mean scores were low in questions 3.a and 3.b: 44.3% versus 31.2%. Question 3.b required 3.a for its answer, and that is the reason for the much lower scores in 3.b than in 3.a. Question 3.c provides an example of a question that could be treated either through stoichiometric calculations (by using the previously obtained masses of alcohol and sugar from questions 3.a and 3.b) or as a simple qualitative question. Recall that the question asked students to decide and explain whether the mass of produced wine would be different from that of the grape must. Working from the chemical equation, given that one of the products (carbon dioxide) is a gas, it is an easy qualitative conclusion that the mass of the produced wine is less than that of the must. A possible explanation for the low mean mark (29.7%) could be that many students had difficulty carrying out the stoichiometric calculations or to draw the qualitative inference and reach the correct conclusion. Students of the PB were better at stoichiometric calculations, so it is likely that these students may have mainly worked with calculations rather than qualitative reasoning. This assumption is supported by the fact that there was a large drop in the mean score in 3.c compared to 3.a and 3b for the PB: 37.9 versus 60.6 and 47.1%. In contrast, in the ThB, the mean score in 3.c was about the same with 3.b: 22.8 versus 20.9%. In addition, in the TB the score in 3.c (22.0%) was similar to that in 3.a (24.2%) and much higher than in 3b: (10.2%).

Finally, in Part 4 we observe low mean scores: 31.9% (in 4.a), 34.4 (in 4.b), 21.5 (in 4.c). These questions involve demanding stoichiometric computations that required knowledge of organic reactions. The lowest score of 4.c might be caused by a misconception generated by the textbook (Kapetanou & Mavropoulos, 1998) that the students used. This book stated that “if a brown-red solution of bromine in carbon tetrachloride is added to an unsaturated compound containing a double bond, then the bromine solution is decolorized”. Of course,
this statement assumes that stoichiometric amounts of the compounds would react, or that an excess of the unsaturated compound would be added. But using the data provided with the problem, the correct answer is that the bromine solution is not decolorized because it was present in excess!

In all the questions studied here (Table 1), the PB had higher achievement than the ThB, and the ThB higher than the TB. Statistical comparison of the three branches by means of one-way ANOVA for independent samples and use of Tukey and Schaffe’s tests confirms that in all questions (except 2.3.b where the difference between PB and ThB is not significant), the PB had statistically significant superior performance at $p = 0.01$ to both the ThB and the TB. On the other hand, the ThB had significantly superior performance to the TB in questions 3.b, 4.a, 4.b, and 4.c (at $p = 0.05$).

**Conceptual versus Algorithmic**

Figure 1 shows percentage mean scores of the total sample as well as of the separate branches in the sets of conceptual (2.3.a and 2.3.b), demanding algorithmic (3.a., 3.b, 4.a, 4.b, and 4.c) and simple algorithmic (1.6-1.8, 2.1 and 2.2) questions. The simple algorithmic questions were easy and produced the highest scores. On the other hand, in the case of the demanding algorithmic questions we had an escalation of complexity, so the achievement here was lower than in the two conceptual questions. The latter questions were obviously not very demanding. This sets, of course, a limitation to this study: had there been more demanding conceptual questions, the expectation is that the achievement in them would have dropped. In the relevant literature there is ample evidence of such a reversal. Note that in the case of the PB, the difference in the mean scores in the two kinds of questions is significantly smaller. This is due to the fact that in this branch we had much higher proportion of students with both abilities high (see below – Table 3).

The two bipolar dimensions, Conceptual and Algorithmic, were used to describe subject achievement. The students were divided into four categories based on their conceptual understanding and their algorithmic problem solving ability, according to Nakhleh (1993). Criteria for categorization were their performance on the questions that required conceptual understanding.
understanding (2.3.a and 2.3b) and those demanding algorithmic problem solving ability (3.a, 3.b, 4.a, 4.b, and 4.c). In this way, each individual was placed into one of four categories (see Table 2): High Algorithmic/High Conceptual (A1C1), High Algorithmic/Low Conceptual (A1C0), Low Algorithmic/High Conceptual (A0C1), and Low Algorithmic/Low Conceptual (A0C0).

In order to assign each subject to the proper bipolar category, the following statistical criteria were used: average score (M), standard deviation (SD) and confidence limits (CL) at significance level \( p=0.05 \) were calculated. All subjects with scores greater than \( M + CL \) were considered ‘High’, and all subjects with scores lower than \( M – CL \) were considered ‘Low’. In this way, a zone of subjects with scores \( M – CL \leq \text{score} \leq M + CL \) was excluded; these subjects are grouped in a fifth separate category with the code A*C*. The categorization and the frequency distributions are shown in Table 2. It is observed that that about half (52.1%) of the subjects demonstrated low conceptual thinking abilities (C0) and about half also (55.1%) demonstrated low algorithmic abilities (A0), while 43.1% were weak in both these abilities (A0B0).

**Table 2.** The categorization and the frequency distribution of Algorithmic versus Conceptual abilities.

<table>
<thead>
<tr>
<th>Category</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1C1</td>
<td>24.7%</td>
</tr>
<tr>
<td>A1C0</td>
<td>9.0%</td>
</tr>
<tr>
<td>A1</td>
<td>33.7%</td>
</tr>
<tr>
<td>A0C1</td>
<td>12.0%</td>
</tr>
<tr>
<td>A0C0</td>
<td>43.1%</td>
</tr>
<tr>
<td>A0</td>
<td>55.1%</td>
</tr>
<tr>
<td>C1</td>
<td>36.7%</td>
</tr>
<tr>
<td>C0</td>
<td>52.1%</td>
</tr>
<tr>
<td>Totals</td>
<td>88.8%</td>
</tr>
</tbody>
</table>

Note that the missing 11.2% belongs to the A*C* category that was excluded from the categorization.

Of great interest is the distribution of the five groups amongst the three branches. Table 3 has the relevant data, together with the scores. As expected, the PB had by far the highest proportion of students high in both abilities. In addition, all three branches were about equivalent in the share of students who were high in conceptual ability only, while the PB and the ThB had about the same proportion of students high in algorithmic ability only. The PB had the highest number of students with algorithmic ability (irrespective of the conceptual ability): 45.7% versus 27.3% for the ThB and 15.1% for the TB. Similarly, the PB had also much higher number of students with conceptual ability (irrespective of the algorithmic ability): 47.9% versus 27.9% and 24.8%.

Next, we compare the mean total scores of the five categories among the three branches. One-way ANOVA for independent samples and use of Tukey and Schaffe’s tests shows that in most cases the differences are statistically significant. In almost all cases, the PB had higher scores than the ThB, and the ThB higher than the TB. These differences could be attributed on the one hand to different characteristics of the students who opted for one branch or another, and on the other to the different educational experiences of the students in the three branches that made them more or less prepared for the examinations. This issue is addressed in more detail in the Concluding Comments.

Finally, we compare the corresponding distribution and achievement of the students of the PB in the two examinations (organic chemistry for general education and special examination in general chemistry), taking into account that we have different samples of the PB in the two cases. With regard to the distribution in the Nakheleh table (Table 3 in this paper, and Table 5 in Stamovlasis et al., 2004), we observe a reversal of the figures resulting from the two examinations. This can be attributed to the special features of the organic chemistry paper as
Table 3. Frequency distribution among categories and branches, and corresponding mean scores in the whole examination paper.

<table>
<thead>
<tr>
<th>Positive Branch*</th>
<th>Theoretical Branch</th>
<th>Technological Branch</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency (%)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>A1C1</td>
<td>35.9 (11.3)</td>
<td>86.9 (11.3)</td>
</tr>
<tr>
<td>A1C0</td>
<td>9.8 (11.5)</td>
<td>69.4 (11.5)</td>
</tr>
<tr>
<td>A0C1</td>
<td>12.0 (12.9)</td>
<td>49.5 (12.9)</td>
</tr>
<tr>
<td>A0C0</td>
<td>30.8 (11.4)</td>
<td>28.9 (11.4)</td>
</tr>
<tr>
<td>A<em>C</em></td>
<td>11.5 (20.9)</td>
<td>53.1 (20.9)</td>
</tr>
</tbody>
</table>

* For the purpose of comparison, the corresponding mean scores of the students of the Positive Branch in the advanced examination was: A1C1: M = 91.0, SD = 8.9; A1C0: M = 67.8, SD = 11.8; A0C1: M = 70.2, SD = 15.0; A0C0: M = 34.6, SD = 15.4; A*C*: M = 71.3, SD = 15.9 (see Table 6 in Stamovlasis et al., 2004).

stated in the Introduction. The data for comparison of the achievements are given in Table 3 and as a footnote to it (see also Table 6 in Stamovlasis et al, 2004). We observe that the achievement of the A1C1, A1C0, and A0C0 students were similar in the two examinations: 86.9 in organic chemistry versus 91.0 in general chemistry for A1C1; 69.4 versus 67.8 for A1C0; 28.9 versus 34.6 for A0C0. In contrast, the A0C1 and the A*C* students scored much higher in the general chemistry examination (70.2 versus 49.5 for the A0C1; 71.3 versus 53.1 for the A*C*). Note that, using the t-statistic, with the exception of the A1C0 case, all other differences are statistically significant. Our results seem to reinforce Nakhleh’s categorization of students. On the other hand, it is important to note that, although the view of Mason, Shell, and Crawley (1997) that it appears to be rare or unusual that a student who understands the concepts lacks the ability to deal with an algorithmic problem may be potentially correct, in practice we do encounter such students; they are the ones with a lack of interest and practice in mathematical manipulations.

A two-way MANOVA for the Conceptual and Algorithmic Dimensions
An experimental design with interaction was carried out to investigate the main effects and the possible interactions between the two variables (a two-way Multivariate Analysis of Variance). Table 4 shows that in all cases the main effects are statistically significant, and in most cases they explain a considerable portion of the variance. On the other hand, the interactions are statistically significant in the case of the most demanding Parts 3 and 4 of the test, but explain only a small portion of variance. In addition, we carried out a three-way ANOVA with the Conceptual and Algorithmic Dimensions and the three branches as independent variables and the total score as dependent variable. The interaction Algorithmic Dimension × Branch had a p = 0.161, explaining 1.4% of the variance, while for the interaction Conceptual Dimension × Branch p = 0.361 and variance explained = 0.9%. These results show no interaction among the branches and the Conceptual/Algorithmic Dimensions with regard to the total score. This is an overall characteristic, while statistically significant differences among the branches do exist (see relevant discussion of the data in Table 3).
Table 4. Main effects and interactions of Conceptual and Algorithmic Dimensions in a two-way MANOVA for the four parts and the total examination paper.

<table>
<thead>
<tr>
<th></th>
<th>Algorithmic</th>
<th>Conceptual</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Main effects</td>
<td>Main effects</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.000</td>
<td>0.000</td>
<td>0.216NS</td>
</tr>
<tr>
<td></td>
<td>75.7</td>
<td>25.4</td>
<td>1.2</td>
</tr>
<tr>
<td>Part 1</td>
<td>0.000</td>
<td>0.020</td>
<td>0.659 NS</td>
</tr>
<tr>
<td></td>
<td>25.1</td>
<td>1.6</td>
<td>0.5</td>
</tr>
<tr>
<td>Part 2</td>
<td>0.000</td>
<td>0.000</td>
<td>0.621NS</td>
</tr>
<tr>
<td></td>
<td>35.3</td>
<td>27.4</td>
<td>0.5</td>
</tr>
<tr>
<td>Part 3</td>
<td>0.000</td>
<td>0.000</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>64.8</td>
<td>20.7</td>
<td>2.7</td>
</tr>
<tr>
<td>Part 4</td>
<td>0.000</td>
<td>0.003</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>63.1</td>
<td>2.4</td>
<td>4.1</td>
</tr>
</tbody>
</table>

Concluding comments and recommendations

1. **Was the postulated categorization of questions (conceptual versus algorithmic) supported by the data?** In line with the findings of our previous study (Stamovlasis et al., 2004), the classification of the questions that came out from principal component analysis (PCA) (see Appendix 2) is in agreement with the classification that was based on their nature (knowledge recall, simple algorithmic, demanding algorithmic, or conceptual understanding) as identified by the researchers. The statistical treatment with PCA gave a marginal structure and might be due to the small number of the conceptual questions and their low difficulty. The above conclusion was further reinforced by multivariate analysis of variance (MANOVA), which showed small interactions between the conceptual and the algorithmic dimensions for the most demanding Parts 3 and 4 of the test.

2. **Distribution of conceptual thinkers in the various branches.** The Positive Branch (PB) had the highest number of students with algorithmic ability (irrespective of the conceptual ability), as well as a much higher number of students with conceptual ability (irrespective of the algorithmic ability). An important finding was that all three branches were about equal in the share of students who were high only in conceptual ability. Finally, the PB and the Theoretical Branch (ThB) had about the same proportion of students high in algorithmic ability.

3. **Achievement in the three branches.** As expected, the PB had the highest mean achievement, intermediate was the ThB and lowest the Technological Branch (TB).

4. **The operation of Nakhleh’s Scheme.** The model of student categorization suggested by Nakhleh (1993) was found to be operating. It is interesting to note that there is a considerable number of students from the ThB (27.3%) who demonstrated high conceptual abilities. These may be the corresponding second-tier students mentioned by Nakhleh. It is then the instructor’s responsibility to make chemistry more interesting and attract these students to science. In addition, a three-way ANOVA showed no significant interaction between the branches and the conceptual/algorithmic dimensions.

5. **Is competence in algorithmic problem solving connected with competence in conceptual problem solving (and vice versa)?** In agreement with our previous study (Stamovlasis et
al., 2004), the statistical analysis for comparison and interaction between the conceptual and the algorithmic questions of the test supported the independence between the conceptual dimension and algorithmic dimension. We could conclude then that competence in algorithmic problem solving may be independent of competence in conceptual questions. The interpretation of the statistical analysis is not that the two abilities cannot coexist in the same person, but that the level of performance in one dimension does not depend on the level of performance in the other dimension. Put another way, one dimension does not presuppose the other, that is, the algorithmic problem-solving ability does not presuppose conceptual understanding, and vice versa.

Of particular importance is to examine the cause(s) for the differences among the three branches:

(i) Did the differences reflect the characteristics of the students who opted for one branch or another? or

(ii) Were there different educational experiences for the students of the three branches that made them more or less prepared for the examinations?

Surely, the students of the different branches had different characteristics. For instance, those of the ThB on the average liked the humanities more, but did not like mathematics or science; recall also that the students of the TB were, on average, weaker students. On the other hand, the students of the three branches had different educational experiences: they followed a number of common courses, plus some special courses depending on the branch. Regarding chemistry, all students had a common (easier) chemistry course (“Chemistry for General Education”), but students of the PB had an additional (more advanced/harder) chemistry course (“Chemistry for the Positive Branch”). Though the contents of the two courses did not overlap (one dealt with organic chemistry, the other with general/physical chemistry), students of the PB were surely more attracted to and more experienced in chemistry, so that it is very likely that they had developed a ‘chemical-type thinking’ to a greater degree. (They had no more experience in organic chemistry at the stage under consideration). Finally, although all students had received training in dealing with knowledge questions and algorithmic exercises within and outside the school, it is true that students in the PB had paid more attention to chemistry (and possibly be given more attention by instructors) than the students in the other two branches.

In conclusion, this study has provided further evidence about the extent of differences between conceptual understanding and algorithmic problem solving, being in agreement and reinforcing the findings of our previous similar study (Stamovlasis et al., 2004). While it was found that a considerable number of students did lack one or both of these abilities, it was encouraging to find that about a quarter of our sample demonstrated both. At this point, however, the limitations of the present study (in addition to the different student characteristics and educational experiences in the three branches discussed above) should be re-emphasized. While the algorithmic questions of the test, dealing with stoichiometric calculations in organic chemistry, were numerous and had a graded difficulty, including some demanding problems/exercises, the conceptual questions of the test were limited and not very demanding. It is very likely that with more demanding conceptual questions, the proportion of students who could deal with them would have shown further decline. In any case, and although our investigation was necessarily contextually and locally bound, the results of the national examination have provided further evidence supporting the distinction and different nature of algorithmic and conceptual questions.

Turning to the implications of this work, these cannot be different from those of similar previous work (see Stamovlasis et al., 2004). Taking into account that lack of understanding makes conceptual questions difficult for most students, teachers and schoolbook authors
should place emphasis on providing students with an understanding of chemistry (Gillespie, 1997). In addition, all students, but especially those experiencing difficulty with conceptual questions, must continually be given practice, encouragement, and support for dealing with such questions, with the aims both to improve their capabilities and develop their confidence. Finally, combined HOCS and LOCS-type, formal and informal, examinations and tests are needed for challenging and fostering students to develop their HOCS capacity (Zoller, 1993). A proper balance of the two types of questions should be included; otherwise, students may skip the demanding items (if there are few) or have a massive failure and disappointment (if there are too many).

Note: At present, the structure of the upper secondary school system in Greece (that is, the distinction into the three branches or streams, Theoretical, Positive, and Technological) remains, but there has been a dramatic shift of students from the Positive to the Technological Branch, because the latter, while leading to the same higher-education institutions as the Positive Branch (except bio- and medical subjects), does not include chemistry and biology among its four special subjects, but two ‘softer’ courses, one on computer science and another on business administration. Note however, that since the time of the examination analyzed in this paper, many changes have been introduced in the number of examinations and course requirements.

References

Nakhleh M.B. (1993), Are our students conceptual thinkers or algorithmic problem solvers?, *Journal of Chemical Education, 70*, 52-55.
Appendix 1: The organic chemistry paper for eleventh-grade
National Greek Examination (June 1999)

PART 1 (25 marks)

Q-1.1. The general formula $C_nH_{2n}$ $(n \geq 2)$ applies to: (a) all non-cyclic hydrocarbons; (b) alkanes; (c) alkenes; (d) alkynes. (3 marks)

Q-1.2. Organic compound $\text{CH}_3\text{CH(OH)CH}_3$ is named: (a) propanol; (b) methylethylether; (c) propanal; (d) 2-propanol. (3 marks)

Q-1.3. The products of complete burning of ethanol are: (a) $\text{CO}_2$, $\text{O}_2$, and $\text{H}_2$; (b) $\text{CO}_2$ and $\text{H}_2\text{O}$; (c) $\text{CO}$ and $\text{H}_2\text{O}$; (d) $\text{C}$, $\text{CO}$, $\text{CO}_2$, and $\text{H}_2\text{O}$. (3 marks)

Q-1.4. Benzene is: (a) a saturated hydrocarbon; (b) an aromatic hydrocarbon; (c) an unsaturated hydrocarbon with two double bonds; (d) an unsaturated hydrocarbon with one triple bond. (3 marks)

Q-1.5. (a) Which phenomenon is called isomerism? (b) List by name the kinds of structural isomerism. (3 marks)

Q-1.6. Esterification is called the reaction between ............... and ................... to form .................. and ................... The reverse reaction to esterification is named .................. (3 marks)

Q-1.7. Of the carbonyl compounds, .................... are oxidized by Tollen’s reagent (ammoniacal solution of silver nitrate), forming a .................. mirror, while their isomeric ....................... do not react. (3 marks)

Q-1.8. For each monomeric species in (I) there is a correspondent polymeric species in (II). Write down these correspondences. [Each time, write down the capital letter for each species in (I), and next to it write down the corresponding lower-case letter in (II).] (4 marks)

(I) A: $\text{CH}_2=\text{CH}_2$; B: $\text{HC}=\text{CH}$; C: $\text{CH}_2=\text{CHCl}$; D: $\text{CH}_2=\text{CHCH}_3$.

(II) a: PVC; b: bakelite; c: polyethylene; d: polypropylene; e: benzene.

PART 2 (25 marks)

Q-2.1. Explain whether or not the following statements are correct or wrong (9 marks):

(a) The molecular formula of a compound supplies more information than its structural formula.

(b) There are three organic compounds that correspond to the formula $\text{C}_3\text{H}_8\text{O}$.

(c) The compound with structural formula $\text{CH}_2=\text{CH-CH}_2\text{CH}_2\text{OH}$ is named 1-buten-4-ol according to the IUPAC rules.

Q-2.2. Fill-in the following chemical equations for the reactions (8 marks):

(a) $\text{CaC}_2 + \text{H}_2\text{O} \rightarrow$  ( b) $\text{HC}=\text{CH} + \text{H}_2\text{O} \xrightarrow{\text{H}_2\text{SO}_4 / \text{HgSO}_4}$

Q-2.3. On checking the exhaust gases of two cars A and B, it was found that the exhaust gases of A contain: $\text{CO}_2$, water vapour, $\text{CO}$, hydrocarbons ($\text{C}_8\text{H}_{18}$) and nitrogen oxides, while the exhaust gases of B contain $\text{CO}_2$, water vapour, and $\text{N}_2$ only.

2.3.a. Write the chemical equation for the reaction that explains the absence of hydrocarbons in the exhaust gases of car B. What catalyst is required for this reaction? (5 marks)

2.3.b. In which case (car A or car B) is it possible that volatile lead compounds may be detected? Explain. (3 marks)

PART 3 (25 marks)

A vessel contains an amount of must, which undergoes wine fermentation:

\[
\text{zymase} \quad \text{C}_6\text{H}_{12}\text{O}_6 \rightarrow 2\text{CH}_3\text{CH}_2\text{OH} + 2\text{CO}_2
\]

After completion of the fermentation, 200 L of wine 11.5° (11.5% v/v) were produced. The density of ethanol is $d = 0.8 \text{ g / mL}$.
3.a. Calculate the volume and the mass of the alcohol produced. (10 marks)
3.b. Calculate the mass of the sugar \((C_6H_{12}O_6)\) that underwent fermentation. (10 marks)
3.c. If you compared the (initial) mass of must with the mass of produced wine, would you find any difference? Explain. (5 marks)

Atomic weights:  
\[
\begin{align*}
\text{C} & : 12, \\
\text{H} & : 1, \\
\text{O} & : 16.
\end{align*}
\]

PART 4 (25 marks)
1.2 mol bromoethane \((C_2H_5Br)\) is divided into three equal parts.
4.a. The first part of the bromoethane is dissolved to anhydrous ether, and an excess of sodium is added to the solution, producing an alkane \((A)\). Calculate the mass of the produced alkane \((A)\). (8 marks)
4.b. The second part of the bromoethane reacts completely with AgOH, and organic compound \((B)\) is produced. Then excess of sodium reacts with \(B\), when a new organic compound \((C)\) is produced and a gas \((D)\) is released. Calculate the volume of gas \(D\) under standard temperature and pressure (stp). (8 marks)
4.c. The third part of bromoethane reacts with excess of alcoholic (ethanoate) solution of KOH, and a gaseous hydrocarbon is produced. This gaseous hydrocarbon is then bubbled into a 1 L solution of bromine in tetrachloromethane \((\text{Br}_2/\text{CCl}_4)\) 8% w/v. Will the bromine solution be decolorized? (9 marks)

Atomic weights:  
\[
\begin{align*}
\text{C} & : 12, \\
\text{H} & : 1, \\
\text{O} & : 16, \\
\text{Br} & : 80.
\end{align*}
\]
Appendix 2: A Principal Component Analysis of the questions

The data was treated with Principal Component Analysis (PCA) to reduce the number of variables and to detect structure in the relationships between variables, that is to classify variables (Anderson, 1984). We arrived at such a classification by looking at the correlation between variables and the factors (or ‘new’ variables) as they are extracted from the analysis; these correlations are also called factor loadings. Variables highly correlated with different factors are classified into different classes or categories. This pattern is referred as a simple structure.

In the literature, there are two criteria concerning the number of factors to be retained. According to Kaiser’s criterion (Kaiser, 1958), one should retain only factors with eigenvalues greater than 1. The scree test, proposed by Catell (Catell, 1966), is a graphical method, using the plot of the factor eigenvalues (Figure 2). The aim is to find the place where the smooth decrease of eigenvalues appears to level out at the right of the plot. In addition the Varimax rotation method was used, which leads to a pattern of loadings on each factor that is as diverse as possible, leading itself to easier interpretation.

Despite the fact that two of the fist three eigenvalues are marginally below unity, we find it useful to maintain the three factors. Note that the three factors explain significant portion of the variance. In addition the values of the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO Test) and Bartlett's Test of Sphericity (see also Stamovlasis et al., 2004) were 0.958 and <0.001 (at least) respectively (using the SPSS 10.1 software). A value of the KMO test close to unity supports the usefulness of PCA; on the other hand, the value of Bartlett's Test is the significance level for rejecting the hypothesis that the initial variables are not related. Thus we are justified for using PCA as an analytical tool. Table 5 gives the results of the PCA for the whole sample with total explained variance of 62.2%. The classification of the questions that came out from PCA is in agreement with classification based on their nature as judged by the researchers (knowledge recall, simple algorithmic, demanding algorithmic, or conceptual understanding).

Factor 3 loads on knowledge-recall questions 1.1-1.4 and 1.5 that required just rote learning. This is also shown by the fact that in these questions we had the highest mean scores: 93.3% in 1.1-1.4 and 85.2% in 1.5.

Factor 1 loads on the questions that demanded applying of algorithms: 1.6-1.8, 2.1, 2.2, 3a, 3b, 4a, 4b, 4c. Questions 1.6 and 1.7 demanded knowledge of organic reactions, and thus achievement was moderate (M = 56.1%). Question 1.8 involved correspondence between monomers and polymers (M = 74.5). Question 2.1 was about molecular formulas, of the right-wrong type with explanation, so achievement was moderate (M = 56.2%); similarly question 2.2 referred to two organic reactions, and led also to moderate achievement (M = 51.7%). The chemical equation that enters question 2.3.a justifies its partial algorithmic character and the considerable loading (0.41) by (algorithmic) Factor 1, hence its much lower loading by (conceptual) Factor 2 (0.60) compared to 2.3.b (0.80).

Finally, Factor 2 loads mainly on conceptual questions 2.3.a and 2.3.b, and partially to 3.c. Question 3.c. does not belong to a single factor but is divided between factors 1 and 2; this is explained by taking into account that 3.c could be treated both as a conceptual and an algorithmic question (see above).

The classification of the questions of interest to this work was not altered on extracting just two factors, which show a clear separation between the algorithmic and the conceptual questions.
**Figure 2.** Plot of eigenvalues used for the scree test.

![Plot of eigenvalues used for the scree test](image)

**Table 5.** Principal Component analysis. Total sample \(N = 499\).

<table>
<thead>
<tr>
<th></th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Eigenvalues</strong></td>
<td>6.728</td>
<td>0.995</td>
<td>0.981</td>
</tr>
<tr>
<td>% Variance explained</td>
<td>48.1</td>
<td>7.1</td>
<td>7.0</td>
</tr>
<tr>
<td>% Cumulative variance</td>
<td>48.1</td>
<td>55.2</td>
<td>62.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1-1.4</td>
<td>0.21</td>
<td>-0.03</td>
<td><strong>0.79</strong></td>
</tr>
<tr>
<td>1.5</td>
<td>0.11</td>
<td>0.27</td>
<td><strong>0.67</strong></td>
</tr>
<tr>
<td>1.6-1.7</td>
<td><strong>0.64</strong></td>
<td>0.24</td>
<td>0.38</td>
</tr>
<tr>
<td>1.8</td>
<td><strong>0.71</strong></td>
<td>-0.13</td>
<td>0.17</td>
</tr>
<tr>
<td>2.1</td>
<td><strong>0.62</strong></td>
<td>0.35</td>
<td>0.19</td>
</tr>
<tr>
<td>2.2</td>
<td><strong>0.70</strong></td>
<td>0.28</td>
<td>0.32</td>
</tr>
<tr>
<td>2.3.a</td>
<td>0.41</td>
<td><strong>0.60</strong></td>
<td>0.13</td>
</tr>
<tr>
<td>2.3.b</td>
<td>0.07</td>
<td><strong>0.80</strong></td>
<td>0.17</td>
</tr>
<tr>
<td>3.a</td>
<td><strong>0.68</strong></td>
<td>0.42</td>
<td>0.20</td>
</tr>
<tr>
<td>3.b</td>
<td><strong>0.71</strong></td>
<td>0.45</td>
<td>0.08</td>
</tr>
<tr>
<td>3.c</td>
<td>0.30</td>
<td>0.54</td>
<td>0.02</td>
</tr>
<tr>
<td>4.a</td>
<td><strong>0.78</strong></td>
<td>0.24</td>
<td>0.11</td>
</tr>
<tr>
<td>4.b</td>
<td><strong>0.82</strong></td>
<td>0.30</td>
<td>0.11</td>
</tr>
<tr>
<td>4.c</td>
<td><strong>0.77</strong></td>
<td>0.31</td>
<td>0.07</td>
</tr>
</tbody>
</table>

* Varimax normalized; Factor loadings \(\geq 0.60\) are shown in bold.