CRITICAL THINKING IN PHYSIOLOGY: A *REASON!*ABLE APPROACH

[1.]Van Gelder, T., [2.]Williams, N., [2.]Di Nicolantonio, R. and [2.]Kemm, R.

[1]Department of Philosophy, Faculty of Arts
University of Melbourne

Email: tgelder@ariel.unimelb.edu.au

[2]Department of Physiology, Faculty of Medicine, Dentistry and Health Sciences
University of Melbourne

Abstract

To help improve their critical thinking skills, undergraduate science students used a new software package, Reason!, to assist them identify and evaluate arguments. At the outset students could not reliably identify and evaluate the reasoning in a short scientific text. Pre- and post-testing revealed a small but significant improvement in reasoning.

Keywords

Critical thinking, reason, biomedical sciences, physiology.

1. Introduction

1.1 The need for a changed educational approach

In teaching we aspire to help students not only to acquire a working knowledge of their content discipline but also to develop essential thinking skills. In scientific disciplines such as physiology, one of these skills is the ability to read relevant pieces of the scientific literature and to make informed judgements about the quality of evidence and reasoning they contain.

How good are our students at this kind of critical thinking? Informal observations suggest that most have underdeveloped skills, and many have serious difficulties. Very few have had any systematic training in argument analysis and evaluation. As a result they approach texts in a haphazard way and rarely reach a sound assessment of the quality of the reasoning and the reliability of the conclusions. In this respect, our students appear to be quite representative of the population at large, as described in other research. For example, in a very intensive study Deanna Kuhn found that a majority of people have no reliable grip on even the most basic skills of argumentation about causal issues (e.g., Kuhn, 1991).

Does training in biomedical sciences help improve critical thinking abilities? Ideally we could answer this question by reference to studies based on a rigorous program of pre- and post testing. Such studies are few and far between. Arguably the most reliable best data comes from large-scale studies of the development of critical thinking skills at university. These indicate that over four years of undergraduate education critical thinking skills do improve about one standard deviation over controls, though most of that improvement happens in the first year. Surprisingly, they found no significant correlation between improvement and the number of science courses taken (Pascarella, 1989; Pascarella, 1991).

Traditionally, education in physiology has contained little or no focus on general critical thinking skills. The hope seems to be that students will somehow just pick up these skills in the course of their studies. In later years, they may benefit from informal mentoring processes. Under such circumstances, it is hardly surprising that students emerge with inadequate abilities.

1.2 Incorporating critical thinking into a science curriculum

The present study aimed at improving the general critical thinking skills of our students using a new stand-alone computer software package *Reason!* This program is designed to guide and scaffold students in the complex processes involved in identifying and evaluating the central argument in some text, and to help them acquire general transferable skills in this area.

The study reports the first component of a long term multi-stage approach to this issue. These are

i) to identify and evaluate conclusions, premises and assumptions in a scientific argument,

ii) to integrate those skills into their private learning and their group learning of their Physiology using the scientific and medical literature and

iii) to be able to discriminate important new knowledge in the context of present understandings (e.g. justifying what is novel, seminal or controversial).

This paper describes the first stage of the project. Students in 3rd year physiology participated in laboratory sessions using the Reason! package, and were pre- and post-tested. Our goals in this exercise were to

i) Introduce students to the process and benefits of this approach.

ii) produce significant and rigorously demonstrable improvements in student reasoning skills; and

iii) make formal and informal assessments of the efficacy of the intervention and the quality of our evaluation method.

2. The Reason! Package

The software we used was an adapted version of a package developed at the University of Melbourne for use in teaching reasoning skills in the Faculty of Arts. As in most universities, the philosophy department is the one place where reasoning is taught as a subject in its own right. Experience in that department suggested that traditional lecture-based teaching methods do little to improve critical thinking skills (van Gelder, 1998). In response they have been developing the *Reason!* software and an entire course package built around it.

Reason! is designed to perform two major functions. First, it guides and scaffolds students in the process of analysing and critically evaluating the main argument in some text. At the highest level, this process can be broken down into the following major phases and steps:

Phase	**Major Steps**
Analysis	Identify the main conclusion.

	Identify key premises (evidence) Identify key assumptions. Determine the structure of reasoning.
Evaluation	Assess the reliability of premises. Assess the strength of inferences Assess the overall quality of support provided for the conclusion.
Reporting	Summarise and report analysis and evaluation.

In *Reason!* all work in analysis and evaluation phases is represented on a hierarchical "argument tree" on the central workspace (see Fig 1). Each box contains a claim which is either a conclusion or a premise; arrows represent inferences; horizontal lines represent premises "working together" forming single complex reasons; and colours represent the user's evaluative judgements.

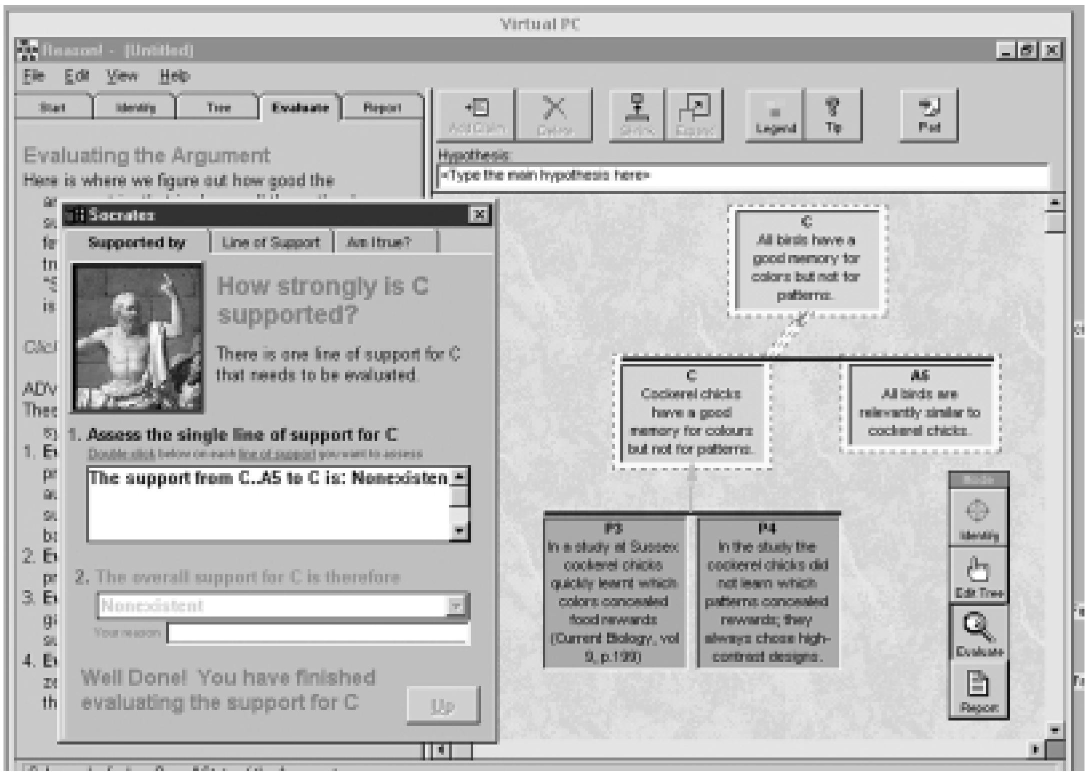

Figure 1: Screenshot of the *Reason!* package.

On the right is the workspace with a argument tree representing the reasoning in a text, with colours corresponding to evaluations. On the left is Socrates. Users click anywhere on the tree and Socrates pops up and asks them to address the most important evaluative issue at that point. Behind Socrates are panels containing general guidance.

Guidance comes in two primary forms. First, a panel on the left contains general instructions pertaining to each major phase. Second, in the evaluation phase students can click anywhere on the argument tree and a character, Socrates, pops up and draws their attention to the most pertinent question arising at that point.

The second major function of the package is to enhance student's general argument analysis and evaluation skills, so that eventually they just "know what to do" and no longer need *Reason's!* guidance and scaffolding. Critical thinking is a very difficult skill to master, and as with most skills, mastery only comes through long hours of quality practice (Ericsson, 1994). We call this the practice hypothesis. *Reason!* is designed to be an environment in which large amounts of quality practice can take place, and consequently a tool for testing the practice hypothesis.

One advantage of the *Reason!* package is that the reasoning and the student's evaluations of it are laid out in a very clear and perspicacious way. This enables the student to understand the reasoning better, and the instructor to see at a glance how the student is treating the problem and what feedback might be appropriate.

Reason! was developed with funding from the University of Melbourne and was first deployed in teaching in philosophy in 1998. The basic package is "generic," applying to any kind of reasoning (outside formal languages such as mathematics). We made a series of small changes to adapt the package to the specific context of biomedical sciences.

3. Integrating *Reason!* into teaching

Students in 3rd year physiology undertook a six hour program of three lab sessions using *Reason!* as the major vehicle for learning reasoning processes and recording their analyses of selected texts. Students worked in groups of 3 per computer. We believe that, working this way, students learn more rapidly as they feed off each other's knowledge and ideas.
Students were taken through three pieces of writing:

i) a short, relatively simple text not requiring any specialised physiological knowledge. The goal here was to learn how to use the Reason! package, and the basics of argument analysis and evaluation.

ii) a letter from the journal *Nature* dealing with the relationship between testosterone levels and lifespan (Nieschlag, 1993). This text was chosen as an engaging and realistic sample of the kind of material students ought to be able to critically assess in their research activities.

iii) a single complex sentence of physiology where the students had to review the principles learned from the package such that the skills could be carried with them to a research situation.

Colour coded

Chicks have a good memory for colours but not patterns, according to biologists in Britain. They say this may provide clues to the roles that different types of plumage play in birds.

Daniel Osorio at Sussex University in Brighton and his colleagues trained cockerel chicks to peck food crumbs from containers with patterns printed on their sides. Chicks quickly learnt to recognise which colours concealed a food reward. Given a choice of patterns rather than colours, however, the birds always chose high-contrast designs, even when these had never contained food (Current Biology, vol 9, p 199).

"We're using chicks as a model for all birds," says Osorio. The chicks' preferences suggest that birds evolved contrast patterns in their plumage to attract attention in displays, while colours may convey specific messages.

From *New Scientist*, 27 February 1999

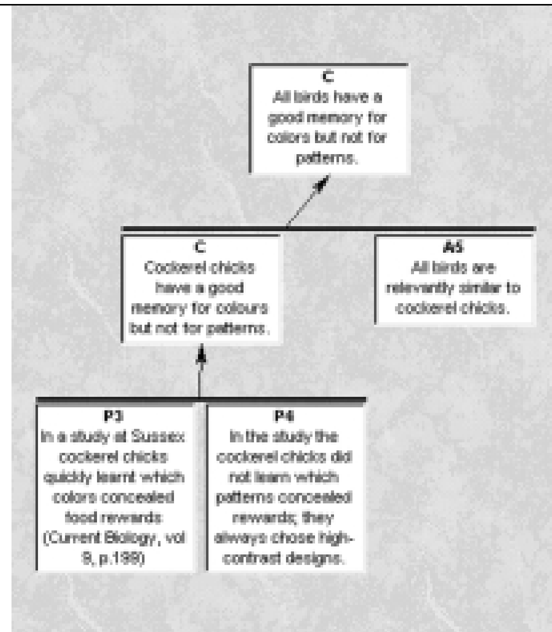

Figure 2: On the left is one of the texts used in the training program. One the right is analysis of the central argument in *Reason!* format.

4. Evaluation

Students were pre- and post-tested with specially designed tests. Each test had two components. In Part A students were required to read a short text involving scientific reasoning, to identify the main conclusion, the evidence, any hidden assumptions, and to evaluate the quality of the reasoning. Written answers were recorded on the test sheet.

Text to be analysed: Epidermal growth factor (EGF) is a protein of 6 kilodaltons and is best characterised as an epithelial cell mitogen. EGF is present in most human fluids at concentrations ranging from 1-800 ng/ml. The concentration in blood is less than 1 mg per ml and is below the sensitivity of specific assays. The administration of EGF to animals produces a number of dramatic effects associated with enhanced proliferation of epithelial cells. In the skin, cell proliferation is increased, leading to a more rapid rate of differentiation. By comparison cells, become cancerous in the trachea. Further, EGF can have both stimulatory and inhibitory effects in the same organ. Organ culture studies on tooth morphogenesis shows that EGF stimulates cell proliferation in the dental epithelium but inhibits the proliferation in dental mesenchyne. These findings suggest that the more physiologically relevant source of EGF is likely to be localised, with its effects being modulated depending on the associations with nearby cells.

Attempt #1: The major claim is that EGF is produced locally with effects modulated by association with nearby cells. The supporting data is that the blood concentration of EGF is almost zero, yet it is said to be in high concentrations in other body fluids. Examples of differences between organs is given. The author correctly deduces that the only explanation for such differences is that each organ must be independently regulated

Figure 3a: An example of text used for a multi-rating question, as in Part B, with an example of only one of the attempts to be rated

	Excellent	V.Good	Good	Fair	Poor	Awful
Identification of the main conclusion	❑	❑	❑	❑	❑	❑
Overall—how good is the	❑	❑	❑	❑	❑	❑

Figure 3b: An example of a multi-rating question, as used in Part B referring to the attempt in Figure 3a.

Part B consisted of a three-part "multi-rating" question. It is well known that both written answer questions and multiple choice questions have distinctive advantages and drawbacks (see, e.g., Brown *et al* 1997, ch.6). In their recent book *The Definition and Assessment of Critical Thinking* (1998), Fisher & Scriven mount a sustained defence of multi-rating questions as the optimal compromise, offering the best of both worlds. One of our goals in this stage of our project was to investigate whether this approach would in fact be suitable for our pre- and post-testing.

In the multi-rating question, students were provided with three different attempts to analyse and evaluate the same text they had just completed in Part A. They were then asked to rate or grade each attempt (hence the term "multi-rating").

Such questions can be scored like multiple choice questions, but in theory they to require the students to utilise skills more closely resembling those actually required in real critical thinking situations.

We used a crossover design with two tests which were identical in all respects, except they were based on different texts. A comparison of results for the two tests suggest the texts were of equal difficulty.

5. Results

In part A of the test, we found that post-test scores were 18% higher after the *Reason!*-based workshop. This improvement however was only of borderline significance (Paired $t=1.74$, $p=0.09$). In Part B we found no significant difference in student grades was found ($t=0.8$, $p>0.4$). Both Part A and B were assessed at the same time. These results suggest an improvement in written reasoning tasks, but not of the judgemental skills involved in the ranking of given interpretations. It remains to be determined whether this improvement in Part A performance results from the learning aspects of the *Reason!* package *per se*, or is a product of the normal educational development that might be expected to occur over a semester of tertiary education.

We also analysed the data in order to evaluate the reliability of the assessment procedure. The scores of Part A of the assessment taken before the instruction using *Reason!* were compared with the scores obtained from the same form of assessment taken after the instruction period. A significant correlation ($r=0.26$, $df=45$, $p<0.05$) was found between pre- and post-*Reason!* scores for Part A of student assignments (Figure 3). By contrast, no significant linear relationship was observed between their performances of the Part 2 task ($r=0.04$, $p>0.05$, Figure 2). This suggests either some form of “reproducibility” in the cognitive skills of students in the written task as opposed to the ranking task or alternatively, that open-ended questions discriminate reasoning tasks in a superior fashion to that in ranking pre-selected responses.

In addition to the pre- and post-testing we conducted small focus group sessions to gauge student impressions. This feedback indicated that students recognised the value of the introduction of this new element to the curriculum as well as the *Reason!* package as a learning tool.

Figure 4: Correlations between Pre- and Post-*Reason!* student scores for the Part A (upper panel) and Part B (lower panel) sections of the student assessment.

6. Discussion

The test results underscored our initial suspicion that many students have a great deal of difficulty with critical thinking. They could not reliably identify the main conclusion of a short scientific text; most could not understand and re-present the structure of reasoning involved and most had very little idea how to go about systematically evaluating that reasoning. These observations reinforced our sense of the importance of continuing to improve our instructional strategies to better enhance critical thinking abilities.

The data is encouraging in the context of our informal observations of students' activity. While using *Reason!* the students were clearly thinking in a more systematic and probing way than they would otherwise. They were able to see more of the complexity of the reasoning, and more possible problems with the conclusion. The modest nature of the detected improvement is broadly consistent with the "practice hypothesis," the main hypothesis from cognitive science governing our approach. Critical thinking is an advanced cognitive *skill*, and a pre-requisite for significant improvement is large amounts of quality practice. There is unlikely to be any instructional "magic bullet" able to produce dramatic improvements in a short period. The challenge is to transform this improved performance while using *Reason!* into a critical "frame of mind" which will be applied habitually and reflexively in their other activities.

The multi-rating questions were not useful in this stage of the project. Students showed no improvement on the multi-rating part of the test (Part B). This in itself is not necessarily an indictment of the assessment method; it is possible that students really were no better off, and the slight improvement on Part A was a statistical artefact. However the fact that there was also no significant correlation between pre- and post scores on Part B is more worrying. An underlying cause of the failure of the multi-rating questions appears to be that they are asking students to make very holistic judgements of quality. One insight that has emerged from this stage of the project is that student performance on critical thinking tasks is roughly proportional to the level of detail of the questions they are being asked to address. That is, global or holistic judgements are very difficult and student performance asymptotes at random guessing. However the more the analytic and evaluative tasks are broken down into simpler discrete components, the more reliable their judgements become. Moreover, one reason people in general and students in particular are so bad at critical thinking is that they do not habitually—indeed, they do not know how to—break a holistic issue such as "how good is the reasoning in this text" down into simpler questions such as "what is the conclusion?" "what are the premises?" etc..

Another possibility is that students were not sufficiently motivated to think carefully in making the kind of demanding evaluative judgements required in the multi-rating questions. After all, the pre- and post-tests were not part of their official course assessment and the materials used were directly relevant to their assessment.

7. Conclusion

Improving students' critical thinking skills ought to be a high priority for courses in the sciences. In Stage 1 of our project, evaluation revealed a statistically significant improvement in reasoning skills after a short period of intensive exposure to the *Reason!* software package. From this and other informal observations we cautiously infer that a well-designed program using interactive software such as *Reason!* can make a valuable contribution to improving critical thinking skills. On this basis we are now extending the use of the package vertically through the three years of their study of Physiology providing our graduates with the ability to discriminate what is novel, seminal or controversial in their chosen field of study.

We are presently testing our procedure to better assess learning outcomes and optimise our use of *Reason!*, and to provide guidance for its incorporation into other biomedical science courses, including Genetics and Biochemistry.

8. References

Brown, G., Bull, J., & Pendlebury, M. (1997) *Assessing Student Learning in Higher Education*. London: Routledge.

Ericsson, K. A., & Charness, N. (1994) Expert performance. *American Psychologist*, **49**, 725-747.

Kuhn, D. (1991) *The Skills of Argument*. Cambridge: Cambridge University Press.

Pascarella, E. (1989) The development of critical thinking: Does college make a difference. *Journal of College Student Development*, **30**, 19-26.

Pascarella, E. T., & Terenzini, P. T. (1991) *How College Affects Students: Findings and Insights from Twenty Years of Research*. San Francisco: Jossey-Bass.

Fisher, A., & Scriven, M. (1998) *Critical Thinking: Its Definition and Assessment*. EdgePress.

Nieschlag, E., Nieschlag, S., & Behre, H. (1993) Lifespan and testosterone. *Nature*, **366**, 18 November, p.215.

van Gelder, T. J. (1998) Penicillin for the mind? Reason, education and cognitive science. Preprint No. 1/98, University of Melbourne Department of Philosophy.