Teaching critical appraisal to students in the Behavioural and Life Sciences

Anna Phillips

This study aimed to examine the concept of critical thinking, and implement and evaluate seminars to teach undergraduates critical appraisal skills. Participants were 140 undergraduate students on a Behavioural Medicine module. Four seminars were designed to teach and reinforce critical thinking with source material including research proposals and journal articles. Two assessments relying on critical thinking were linked to the seminars. Teaching method effectiveness was examined using questionnaire evaluations and comparison of current students' coursework grades with previous years'. Students reported finding the seminars helpful, and their self-ratings of critical appraisal skills improved from pre- to post-seminar. For the first piece of coursework, there was an improvement in the mean grade from the previous year, but no improvement on two years previously. There was no overall improvement in grade for the second coursework. The issues surrounding critical appraisal teaching and assessment are discussed. Future studies could investigate whether longer-term follow-up yields evidence of improvement in objectively measured critical ability.

Keywords: *critical appraisal; questionnaires; coursework assessment; seminars.*

CRITICAL THINKING or critical appraisal is 'purposeful, reasoned, and goal-directed thinking… involved in solving problems, formulating inferences... and making decisions' (Halpern, 1998). It is considered not only to be a key skill, but an indispensible part of education (Norris, 1985). From a philosophical standpoint, it has been argued that for students to fulfil their right to question, challenge, demand justification, and ultimately make choices, this right must be recognised, and this skill be taught (Siegel, 1980). Indeed, this is a skill required throughout life but also in many areas of employment post-university. Consequently, during the course of undergraduate education, and arguably earlier, students should begin to learn and develop this skill. In addition, in many university psychology or related degree courses, completion of a final year project is required including a literature review entailing both synthesis and critical appraisal of relevant published source material.

The cognitive psychology approach to critical thinking is to delineate the set of operations and procedures that it involves in order to distinguish critical thinking from other thinking such as creative thinking (Huitt, 1998). For example, that critical thinking is an 'active, systematic process of understanding and evaluating arguments' (Mayer & Goodchild, 1990). Behavioural psychology has developed this further and identified the subtasks associated with achieving outcomes such as critical thinking ability, and how behaviour can be shaped through experience and practice towards such achievement (Huitt & Hummel, 2006). For example, giving students the opportunity to practice critical thinking skills and reinforcing their experience with positive feedback could be considered a behavioural approach to learning. Further, individuals specialising in critical thinking in different content areas would contend that critical thinking is best developed with actual specific content rather than being taught the skills separately (Huitt, 1998).

The skill of evaluation is a higher order thinking skill, and the highest in terms of the originally defined educational objectives in Bloom's taxonomy (Bloom, 1956), with synthesis just below it. In revised versions of

the taxonomy, the ability to create and generate new ideas or ways of viewing the world has been placed higher up the taxonomy, with evaluation in second place (Anderson & Krathwhol, 2001). In contrast, others have suggested that evaluation and synthesis are two different but equivalent skills. It has been argued that evaluation is the skill of analysing, assessing and making logical judgements, often what is considered critical thinking, whereas synthesis is examining parts and putting material together in a new way, what might be considered creative thinking (Huitt, 1992). However, both critical and creative thinking have been shown to be crucial to successful problem solving (Duemler & Mayer, 1998). The importance of integrating creative thinking into critical appraisal has also been emphasised by others (e.g. Bonk & Smith, 1998).

The present study aimed to examine a variety of conceptualisations of critical thinking and approaches to teaching it within the psychology and life sciences educational literature, and apply the principles from these studies in a set of seminars aiming to teach students to critically assess scientific literature. A further aim was to examine the effectiveness of such teaching methods in improving students' critical thinking performance.

A task force designed to propose a framework for critical thinking in nursing education drew on existing literature suggesting that it involves elements of interpretation, analysis, evaluation, inference, explanation, and self-regulation of one's own thinking (Dexter et al., 1997). They also emphasised the importance of logic, i.e. the ability of students to critique their own logic and spot the mistakes (Dexter et al., 1997). Others have defined critical appraisal skills as analysing, applying standards, discriminating, information seeking, logical reasoning, and predicting and transforming knowledge (Scheffer & Rubenfeld, 2000). However, it has also been suggested that there is no particular right way to categorise or define the dimensions of critical appraisal (Dexter et al., 1997).

Whatever the consensus on the exact components of critical appraisal, teaching critical appraisal to university students is a challenge, as prior experience with such skills training may vary, leaving some students with relatively poor reasoning and problem-solving skills (Klimoviene et al., 2006). It is also clear that many university degree courses do not employ specific teaching of critical appraisal skills but rely rather on accidental learning throughout the degree programme (Castle, 2006). It has also been suggested that what is required is more transparency in the links between what is being taught and its assessment (Biggs, 2000). This principle was applied to radiography students, who were required to rate their own critical appraisal skills on the basis of what they had been taught throughout their degree using a modified version of the Critical Thinking Questionnaire (California Academic Press), scores on which were then compared to their written assessment grades (Castle, 2006). However, although students' rated their abilities highly, their assessed work showed 'little attempt to use critical discussion…' (Castle, 2006). Consequently, it was proposed that lecturers should align coursework and teaching with learning outcomes, so that students were aware that they were being taught critical appraisal, and only to assess one aspect of critical appraisal at a time.

A systematic review of the effectiveness of critical appraisal teaching for health care workers revealed that from 137 articles on this topic, many did not assess the effectiveness of this teaching (Hyde et al., 2000). In addition, several studies only implemented critical appraisal teaching with a small number of students, therefore making it difficult to generalise their results to the general student population. These are common limitations, observed previously in similar systematic reviews (see, for example, Audet et al., 1993). Of the 16 articles which did assess the impact of teaching critical appraisal, it was found that most of the studies (N=14) showed a benefit, mainly in

terms of participants attitudes, although this may reflect 'desirable responding', and improved knowledge and skills (Hyde et al., 2000). A previous systematic review of courses teaching critical appraisal to medical students showed an overall significant increase in students' knowledge, but not necessarily an application of this knowledge to critical reviewing of the literature (Norman & Shannon, 1998). In addition, when these courses were used with medical residents, the gains in knowledge were smaller, and although participants reported reading literature more critically, this was not reflected in objective tests of their ability (Norman & Shannon, 1998). Similarly, another systematic review concluded that the evidence for an improvement in knowledge was weak, and that 'the ability of participants to appraise evidence critically… was not convincing' (Taylor et al., 2000).

It has been suggested that students can improve their critical appraisal skills but that this requires the opportunity to practice these skills (Bensley & Haynes, 1995). This also fits with the behavioural psychology approach to critical thinking (Huitt & Hummel, 2006). Perhaps longer-term training is necessary in order to improve critical appraisal. In one study, critical interactive thinking exercises were implemented over three years at two universities in Michigan and Columbia, which included the writing of a composition addressing a specific problem or little understood phenomenon, preparation of the defence of a hypothesis, presentation of the arguments supporting the hypothesis to a peer group, and group assessment of the quality of the arguments made (Peters et al., 2002). Although students were initially apprehensive, it was reported that the majority eventually reported the experience as being a positive one which enhanced their critical appraisal skills (Peters et al., 2002). However, the outcome in terms of actual improved performance on assessments is not discussed by the authors.

It could be that particular aspects of teaching critical appraisal help to improve this skill in students, for example, the integration of group-based learning. In a study reported above, working in groups with peers improved students' experience of learning critical appraisal (Peters et al., 2002), and this has been deemed to be an important strategy in teaching radiography students how to think critically (Kowalczyk & Leggett, 2005). Similarly, simple factors such as the frequency of attendance at critical appraisal courses and credits contingent on attendance were thought to be associated with the relative success critical thinking teaching in medical students and residents (Norman and Shannon, 1998).

Consequently, new approaches to teaching critical appraisal should incorporate elements of: transparency between skills teaching, learning outcomes, and assessment; opportunity for practice of the skills being taught; supportive group learning; and incentives to attend the teaching sessions, for example, teaching sessions linked to assessments. In addition, given the paucity of studies incorporating any assessment, particularly the quantitative objective assessment of the effectiveness of such teaching, it would seem important to examine the effectiveness of a new approach using measurements of both students' own views of the helpfulness of the skills teaching, ratings of their own critical appraisal ability, and their objective assessment grades. Further, much of the previous literature in this area has focussed on health sciences and medical students, and only small numbers of these, thus it would seem important to incorporate and assess this type of teaching within the teaching of aspects of psychology and in other undergraduate settings with a relatively large group of students.

The present study aimed to examine the effectiveness of a set of teaching methods focused on teaching undergraduate Sport and Exercise Sciences students skills of critical appraisal as part of a Behavioural

Medicine/Health Psychology module. The methods consisted of four seminars. Different types of source material, including faux research proposals and real journal articles, were used in order to gradually increase in the depth of critical appraisal required and give multiple opportunities for practice. Seminars were planned with opportunity for group discussion and facilitator-led discussion in order to ease students into this skill with opportunities for peer learning. Coursework relying on critical appraisal skills were introduced part of the way through the teaching strategy once students had had the opportunity to practice these skills in a non-assessed context. The effectiveness of these teaching methods was examined using student evaluation via qualitative questionnaires and rating scales, and the quantitative comparison of current students' coursework grades with previous years' grades. It was expected that students would generally rate the teaching methods as helpful, and that their coursework grades would improve following the implementation of these methods.

Method

Participants/Design

Participants were 140 second-year Sport and Exercise Sciences and Biosciences students opting to take a Behavioural Medicine/Health Psychology module which commenced October 2007. This module is a basic introduction to elements of health psychology, with a particular focus on the immune and cardiovascular systems. Each student attended four seminars as part of this course and completed questionnaire measures at the seminars and two pieces of coursework subsequently. Of the 113 who returned the initial questionnaires and/or completed demographic details, 38 were male and 75 were female with an average age of 19.5 (SD=0.75). The majority (88 per cent) of participants were white, with seven identifying themselves as Asian, and two as mixed race. Seven reported that they had dyslexia or a related learning difficulty.

Seminar teaching

Each seminar extended the skills taught in the previous seminar, and the subject matter was linked to that covered in the lectures during the module. The second seminar was linked to the first piece of coursework required for the module, and the linked third and fourth seminars were associated with the second piece of coursework.

In seminar one, students were informed that they were to be taught the skill of critical appraisal, which would be valuable for their coursework and in the wider context of university and future work. First, the seminar leader outlined basic research designs, sampling methods, and pitfalls to look out for in scientific research designs. This was interactive using brainstorming of examples of types of research design, sampling methods, and pitfalls. Following this, students were asked to imagine themselves in the position of a research council with the power to review research proposals and decide whether or not to grant money to the proposed projects. They were split into four groups and each group was assigned a different research proposal. The proposals were not real and were designed to include some key flaws, but were based on real-life research questions in the area of Behavioural Medicine/Health Psychology, for example, examining the influence of spirituality on mortality rates in adults. The groups were given 30 minutes to make a recommendation of whether to reject, accept with major revision, accept with minor revision, or accept without changes; they were also to discuss the flaws of their proposal in readiness to present this to the whole group. The seminar leader circulated between the groups to help with questions and to facilitate deeper discussion where necessary. The groups were then asked to imagine themselves as the researchers presenting the proposal in front of them and asked, on the basis of the criticisms they had been discussing, to redesign and improve on the original proposal. Following 30 minutes of discussion, each group presented their

original proposal and the criticisms they had raised along with their new and revised proposal. Whole group participation was encouraged, and the seminar leader emphasised key points and added any that each group had missed. The whole group voted on the recommendation that this new proposal should receive before proceeding to the next small group presentation. At the end of the presentations, the seminar leader reiterated the importance of students being critically minded in their reading of research, and informed them that the next seminar would be similar, but based upon a published research article.

The second seminar commenced with an introduction by the seminar leader, explaining that they would be building on their critical appraisal skills by examining a real research article reporting on a study which, for example, examined the impact of caregiving on the antibody response to vaccination. They were told that this article and seminar would be the basis for the first piece of coursework. The seminar leader then asked the group to help with brainstorming ideas about what to look for when critically reading a journal article, and gave examples to illustrate points such as 'what an hypothesis is'. The students were then split into three groups. Each group had to discuss a section of the article; the introduction and participants, the methodology, or the results and discussion. The groups were encouraged to pick out the strengths and weaknesses of their section to later present to the whole group. Following each group presentation and whole group discussion, they were then asked to identify the key criticisms of the article as a whole. Finally, the seminar leader outlined the coursework requirements and reminded the students that there were related lectures and associated reading lists which would help with their understanding of the general area of the article.

Seminar three was introduced as part of a pair of seminars aimed to help the students further develop their critical skills, and lead into the second piece of coursework. It was explained that the format of group discussions and presentations would be the same as previously. Students were introduced to their task of critically appraising two articles which seemingly examined the same research question but found different results. Their goal was to critically discuss how the studies differed and why these differences might lead to different findings. Before splitting into groups, the seminar leader encouraged the students to brainstorm how articles might differ in terms of the participants, the methodology, and the results and discussion. Examples of potential differences were given, such as population age, timings of measurement in experimental studies, and types of statistical tests used. The students were then split into three groups for discussion and presentation preparation as before. This time the group topics were; population, method, or results and discussion, and were general, not based on the actual articles, so consisted of brainstorming things to look for and comparisons to make between the articles. The ideas generated by each group were presented and then discussed by the whole group. The seminar leader then informed the group that they needed to independently read the articles and make comparisons on the basis of these items that they had identified as important to examine. They were asked to come to the seminar the following week having read the articles and prepared a comparison table between the two.

In the final seminar the students as a group recapped on the details of each of the studies they had read. They were then split into three groups to discuss differences between the articles in terms of; the population, the methodology, or the results and discussion. Each group was asked to discuss why these differences might have led to different findings, and identify two most likely reasons for the different findings. Facilitated discussion continued as previously, and was followed by small group presentations and full group discussion. The students were again reminded the students that there were related lectures and associ-

ated reading lists which would help with their understanding of the general area of the article. The coursework requirements were then outlined and there was an opportunity for questions. An example of the types of articles used is: two studies examining the effects of psychological stress on the antibody response to vaccination.

Coursework

The first piece of coursework was a two-page journal article review based on the critical appraisal of the journal article in seminar two. Students were informed in the seminar and on a handout what the review should comprise. It was recommended to students that the review commence with a brief introductory paragraph on the general area of the article, a concise summary of its methods and findings and implications. The main body of the review should be a critical discussion of the strengths, limitations and weaknesses of the study itself, both in terms of its methodology, analyses conducted and the conclusions drawn from the data. Students were told that the criticisms made should be about the key major flaws in the article, although other more minor criticisms could be discussed, and that the majority of marks would be given to this critical analysis part of the review. They were also asked to discuss why the main flaws are an issue, and make suggestions regarding their influence on the results or overall message of the article. Additionally, they could also make recommendations for changes which would improve the quality of the study. This coursework was assessed using a specific cover sheet with weighted marks assigned to each section of the coursework.

The second piece of coursework, based on seminars three and four, was a four-page critical review of the two journal articles which examined the same research question but had different findings. The format required was as above, and students were informed about this through seminar handouts. Students were told that the majority of marks would be assigned to the critical discussion of the main potential reasons for why the results of the two studies differed. They were recommended to briefly first identify differences between the studies that they did not think led to the differences in results, and explaining why they did not think these differences were important, then move on to the key differences that were likely to explain the differing results. It was emphasised that when identifying the key differences between the papers, the students need to think and justify in their reviews exactly how these differences might have led to different findings, rather than just listing them. The coursework was assessed with weighted marks as previously.

Questionnaires

In order to assess students' perceptions of the seminars, and of their own critical abilities, three simple questionnaires were designed. Due to the likelihood of social desirability bias, and in order to reduce this as much as possible, the questionnaires were completely anonymous, so answers across the sessions were not able to be matched up to the same participants. The first two questionnaires collected basic demographic information and then asked participants to rate, on a seven-point Likert scale, their pre-seminar critical appraisal ability. At the end of the seminars, the students were requested to complete the remainder of the questionnaire which asked; which parts of the seminar they had found to be the most interesting, to rank the parts in order of usefulness, to again rate their critical appraisal skills, answer whether or not they thought the skills the seminar had taught would be useful for their coursework and more generally, and finally to provide any further comments they had about the seminars.

A single feedback questionnaire was used for seminars three and four, as these were linked sessions. This questionnaire was of a more qualitative nature, as it was felt that this might lead to more honest responses rather than the potentially induced response of admitting to an increase in critical appraisal

skills such as might be engendered by the initial feedback questionnaires. This questionnaire was also targeted specifically at the learning outcomes the seminar was designed to meet. The questionnaire asked first what the students thought the aims of the seminars were. It then requested them to indicate on five-point rating scales the extent to which they thought the seminar helped them: understand what to look out for when critically appraising articles; understand the content of the articles; understand the key differences between the articles; argue critically about how differences between studies could lead to different findings; and understand what was required for the coursework. They were also asked to indicate to what extent they found the seminars to be interesting and enjoyable, and to write any further comments or suggestions. It was felt that the previous two feedback questionnaires dealt with which aspects of the format of the seminars the students found useful or interesting, i.e. the group presentations, feedback, etc., so the final questionnaire did not include these aspects.

Procedure

Participants were asked to sign up to four mandatory seminars as part of the Behavioural Medicine/Health Psychology module. These consisted of two separate sessions in the first semester and two linked sessions in the second semester. Seminars in semester one were two weeks apart, but those in semester two were one week apart. All seminars were two hours long and based in a seminar room with capacity for 35 students. Seminars were run in time slots that did not overlap with other teaching, generally 3.00 p.m. to 5.00 p.m., but students had the option of which slot to sign to attend with a maximum number of 30 students to be booked into each slot. Each seminar was repeated five times and was delivered by one of three postgraduate seminar leaders who had received training from the author in advance. The postgraduate seminar leaders were also provided with seminar teacher's notes and were peer-reviewed if they were new to this teaching.

Prior to each seminar, a handout about what the seminar would entail, the learning outcomes, and the coursework requirements, where appropriate, was placed on WebCT, a web-based intranet tool, for students to read. Links to any reading matter required for the seminars were placed on these handouts. The seminar handouts for the latter three seminars also included details of the coursework assignments, and guidelines for their completion.

At the start of seminars one and two, participants were presented with the feedback questionnaires in order to complete their demographic details and the pre-seminar rating of their critical appraisal skills. They were asked to complete the remainder of the questionnaire at the end of the session. For the latter two seminars, the feedback questionnaire was handed out at the close of the final seminar. All questionnaires were collected by the seminar leader at the end of the session. The author also collected data on which seminars were taught by which postgraduate seminar leader, in order to examine the effects of this, as a covariate, upon the mean ratings.

Data analysis

Data from the feedback questionnaires and students' coursework grades were entered into the statistical package SPSS version 15. Summary data for the demographic and categorical questions were collected through simple frequency counts and descriptive statistics. Comparison of students' ratings of their critical appraisal ability pre-seminar and post-seminar for seminars one and two was conducted using univariate Analysis of Variance (ANOVA). Analysis of covariance (ANCOVA) was used to assess whether the change in students ratings was influenced by the postgraduate seminar leader, sex, ethnicity, or dyslexia status. Students' written comments were also collated as string variables within SPSS. Average coursework grades for each piece of coursework were

compared between the present year and previous year using between-subject ANOVAs. Change in coursework grades across the year was examined using repeated measures ANOVA. Analysis of the change in grades over time by the type of student was assessed using a mixed model repeated measures ANOVA with type of student as the between-subject variable. Variations in degrees of freedom reflect occasional missing data.

Results

Feedback Questionnaires

Eighty-one per cent of the students on the Behavioural Medicine module attended seminar one and completed a feedback questionnaire. The demographic details are described above. Students' mean (SD) pre-seminar rating of their critical appraisal was 3.4 (1.06), and the post-rating was 5.0 (0.89), representing a significant increase in students' ratings of their own critical ability from pre- to post- seminar, $F(1,107)=305.76$, $p<.001$, $\eta^2=.741$. This increase in ratings was not significantly influenced by postgraduate demonstrator, sex, or dyslexia status, although Asian students rated their abilities as significantly higher overall than white or mixed race students, $F(2,102)=3.40$, $p=.04$, $\eta^2=.062$. The students did not tend to rate one particular aspect of the seminar as the most interesting, but ticked all of the sections they found to be interesting, consequently the frequencies for each aspect of the seminar do not add up to 100 per cent. Sixty-six per cent of the students found the critical appraisal of a grant proposal to be the most interesting part of the seminar, 54 per cent stated that the development of a revised proposal was one of the most interesting aspects. Eleven, 10, and six per cent found the feedback, group presentation, and introduction sections to be interesting, respectively. Students were asked to rank the sections of the seminar in the order of most to least useful. The critical appraisal of a research proposal was rated as the most useful receiving first place ranking from 72 per cent of the participants. The development of a new proposal was rated by the majority of students as the second most useful section, receiving 59 per cent of the votes for second place as well as 17 per cent rating it as the most useful section. The majority of participants (39 per cent) rated the introduction to research design to be in the middle in terms of usefulness, and the group presentations were rated as the second least useful aspect by the majority (35 per cent). The feedback at the end was rated by most (50 per cent) as the least useful part of the seminar. One-hundred-and-six (96 per cent) students felt that the seminar would help them in their coursework, and 99 (90 per cent) agreed that it would help them in future academic and vocational work. Finally, students wrote the following positive and negative comments about the seminar, and suggestions for improvement.

> *'Good group work – good test cases, lots to discuss and develop, made us consider lots of effects.'*
>
> *'Good seminar – very helpful.'*
>
> *'Good that we were split into groups and worked with people we didn't really know.'*
>
> *'It was a great way of interacting with others, and sharing people's ideas.'*
>
> *'More feedback on critical analysis.'*
>
> *'Slightly shorter/more engaging introduction, e.g. just asking questions of an unwarmed up audience does not get a good response.'*

Ninety-five per cent of the students attended seminar two, although only 96 (69 per cent) returned the feedback questionnaire. The respondents mean age was 19.5 (0.76) years, and 62 (65 per cent) were female. Five (five per cent) respondents were Asian, 87 (93 per cent) were white, and two (two per cent) described themselves as mixed race. Five (five per cent) reported having dyslexia or a related condition. Students' mean (SD) pre-seminar rating of their critical appraisal was 3.8 (0.95), and the post-rating was 4.9 (0.89). This was a significant increase in students' ratings of their own critical ability from pre- to post- seminar, $F(1,93)=241.83$, $p<.001$, $\eta^2=.722$. This increase in ratings was not

significantly influenced by postgraduate demonstrator, sex, ethnicity, or dyslexia status. In addition, the pre-seminar two rating was significantly higher than the previous pre-seminar rating for seminar one, $F(1,92)=6.35$, $p=.01$, $\eta^2=.065$, although not as high as their post-rating for seminar one. However, the post-seminar rating for seminar two was not significantly different from the post-seminar rating for seminar one. Regarding how interesting the students found each section of seminar two, again students did not pick out a particular aspect that they found to be the most interesting, but ticked all the parts they had found interesting. Twelve (13 per cent) rated the introduction as interesting; 61 (64 per cent) the critical appraisal of the article; 39 (41 per cent) the help they got from the demonstrator during the group work; 16 (17 per cent) the group presentations; and 25 (26 per cent) the feedback session. When the students were asked to rank the elements of the seminar in the order of usefulness, the critical appraisal of the journal article was rated as the most useful receiving first place ranking from 36 per cent of the participants. Receiving help from the postgraduate demonstrator during the group work was rated by the majority of students as the second most useful section, receiving 34 per cent of the votes for second place as well as 35 per cent rating it as the most useful section. The majority of participants (25 per cent) rated the feedback on the presentations to be in the middle in terms of usefulness, and the group presentations were rated as the second least useful aspect by the majority (24 per cent). The introduction section was rated by most (40 per cent) as the least useful part of the seminar. All of the respondents agreed that the seminar would help them with their coursework, and 83 (90 per cent) agreed that it would help with future academic and vocational work. Finally, students wrote positive and negative comments about the seminar, and suggestions for improvement. These consisted of statements that the seminar was *'Excellent!'*, *'Good and very helpful for the coursework'* in terms of a *'...clearer idea on how to do well on the coursework'*, *'Good step-by-step take through of what we need'*, but that *'group size was too large' ... 'meaning productivity was difficult in the time we had.'*

Ninety-five (68 per cent) students completed feedback forms at seminar four. Of these, 35 (38 per cent) were males, five per cent were Asian, 93 per cent were white, and two individuals classified their ethnicity as mixed race. Two students (two per cent) reported being dyslexic and the mean (SD) age was 19.9 (0.76) years. The responses to the question about the aims of seminars three and four resulted in many different responses. These were most commonly 'to help with the coursework' or 'to help us critically analyse journal articles'. Students also rated the extent to which the seminar helped them to: know what to look out for when appraising articles; understand the content of the journal articles; pick out the key differences between the articles; critically argue how certain differences might have resulted in different findings; and understand what was required for the coursework. The ratings were made on five-point scales from very well (4) to not at all (0). The overall mean (SD) ratings are shown in Table 1, which show that although the students rated the seminars as moderately successful in teaching them the skills they were intended to teach, they did not necessarily find them interesting and enjoyable. Seminars taught by one particular demonstrator received significantly higher ratings than those taught by another of the three demonstrators, $F(2,91)=3.04$ to 7.13, $p<.05$, $\eta^2=.063$ to $.134$, with the exception of ratings for understanding the coursework. However, the mean differences in ratings between postgraduate demonstrators were relatively small, the largest being a difference of 0.87 for ratings of how well the seminars taught the skill of critical argument. Males also gave higher ratings than females in response to the question about how well the seminars taught the skill of critical argu-

Table 1: Mean (SD) ratings for Seminars three and four.

Extent to which the seminars helped you to...	*Mean*	*SD*
Know what to look out for when appraising journal articles	2.8	.73
Understand the content of the articles	2.7	.76
Understand the key differences between the articles	2.8	.87
Critically argue how differences might have led to different results	2.6	.88
Understand the coursework	2.8	.85
Extent to which you found the seminars interesting and enjoyable	2.2	.87

ment, $F(1,91)=4.00$, $p=.04$, $\eta^2=.042$, and how interesting and enjoyable they found the seminars, $F(1,90)=4.54$, $p=.04$, $\eta^2=.048$. There were no differences in ratings for students of different ethnicities, or between those with and without dyslexia. Finally, 15 students also added suggestions in the section provided for other comments regarding ways to improve the seminars. The more helpful ones suggested that seminar four was more effective and perceived as more useful than seminar three.

Coursework

Coursework grades for the current and previous academic years were transformed into numerical scores and compared. The descriptive statistics for the journal article review and critical review for the previous two years and the present year are displayed in Table 2. When the journal article review coursework grades were compared, there was a significant increase from the previous year, $F(1,237)=5.20$, $p=.02$, $\eta^2=.021$. However, the previous year (2006–2007) were a notoriously difficult group and their grades are also significantly lower than the year prior to that, $F(1,149)=11.04$, $p=.001$, $\eta^2=.069$. Consequently, the current year was also compared to two years previously; there was no significant difference in the grades for the journal article review, $F(1,184)=2.02$, $p=.16$, $\eta^2=.011$. The differences in grade for each year group on the journal article review are shown in Figure 1, overleaf.

The descriptive statistics for the second piece of coursework, the critical comparison of two papers, are also displayed in Table 2 overall for each year. When comparing the current grades to those from previous years, there was no significant difference, between the current year and 2006–07, $F(1,233)=2.67$, $p=.10$, $\eta^2=.011$, although the current year appear to have slightly lower scores. However, comparison between the current year and 2005–06 did reveal a significant difference, $F(1,181)=4.96$, $p=.03$, $\eta^2=.027$; scores from the current year were lower than those from 2005–06. The grades for this piece of coursework for the current and previous years are

Table 2: Mean (SD) for the coursework for each year group with and without Biosciences students.

	Journal Article Review		Critical Review of Two Papers	
	Mean	*SD*	*Mean*	*SD*
2005–06	61.2	8.21	60.5	8.62
2006–07	56.4	8.41	59.1	11.03
2007–08	59.0	9.36	56.8	10.44
2007–08 Sportex	60.9	9.36	56.9	10.52
2007–08 Biosciences	53.5	8.28	56.2	10.3

Figure 1: Journal article review coursework scores by year group.

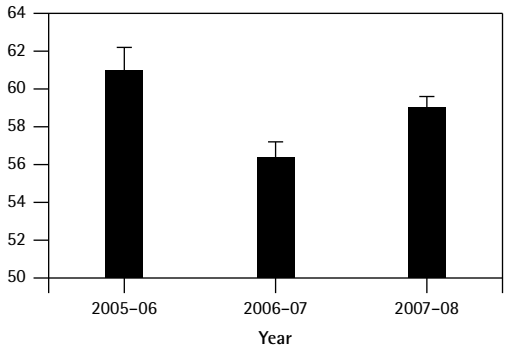
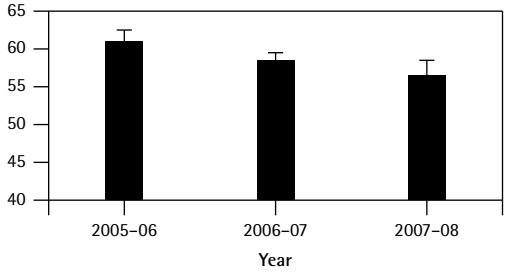


Figure 2: Critical review of two articles coursework scores by year group.



shown in Figure 2. When the grades for the first and second coursework were compared, there was a significant decrease over time, $F(1,133)=5.02$, $p=.03$, $\eta^2=.036$, as can be seen in Table 2.

Due to the 2007–08 intake of 40 Biosciences students who are not familiar with the type of assessment used on this module, a separate sensitivity analysis was conducted in which the coursework grades are summarised for Sport & Exercise Sciences and Biosciences students separately (see Table 2). There was a significant effect of student type, with Biosciences students scoring lower overall across both pieces of coursework, $F(1,132)=7.58$, $p=.007$. In addition, a significant interaction effect was also observed between the change in grades over time and the type of student, $F(1,132)=10.67$, $p=.001$, such that Sportex students' grades significantly decreased from the first to the second coursework, but Biosciences students' grades significantly improved (see Table 2).

Discussion

Generally the students seemed to enjoy the seminars and find them useful. Although they enjoy the introduction and summing up parts of the seminars less than the group work and feedback, they still rated these aspects as important. A positive finding was that, on the whole, students' perceptions of the seminar teaching was not associated with their own gender, age, ethnicity, dyslexia status, or the particular postgraduate teacher. In addition, students' ratings of their critical appraisal skills improved from pre- to post-seminar for the first two seminars, and the mean rating was higher at the start of seminar two in comparison to seminar one. Students also found the critical appraisal,

group work, and feedback aspects of the seminars to be particularly useful and helpful. Most students agreed that the seminars were effective, and all agreed that the seminars would help them with the coursework. For the journal article review coursework, there was an improvement in the mean grade from the previous year, but no change in comparison with two years previous. For the critical comparison of two articles, there was no improvement in coursework grade. In fact the mean was significantly lower than that from two years previous. Sportex students' grades declined from the first to the second piece of coursework, but the Biosciences students' grades improved. Overall this amounted to a significant decrease in scores between the assessments.

The unexpected lack of relationship between students' objective assessment grades and their positive evaluation of their critical appraisal skills and the change in this due to the seminars could reflect a general tendency of students to overestimate their critical appraisal abilities. This mismatch has been observed by others who have attempted to teach critical thinking to student radiographers (Castle, 2006). Another explanation and potential limitation of the present study is that the generally positive nature of the students' responding on the seminar evaluation questionnaires could be, in part, due to socially desirable responding such that the students knew the aim of the seminars and responded in the way they felt the researcher wanted. Socially desirable responses have been cited previously as an explanation for the lack of relation between students' self-report of their critical abilities and their actual grades (Hyde et al., 2000). However, the questionnaires were anonymous and the researcher was not present during their completion, so it is likely that this type of responding was limited to some extent.

An alternative reason for the current findings, particularly the lack of an increase in students' grades from previous years, is that the expectations of the researcher marking the coursework may have influenced the assessment grades given. As it was not possible to blind the researcher to the aims and outcomes of the study, it is possible that their marking was more stringent, knowing that the students had received additional teaching and support with critical appraisal, that they would expect to be reflected in the coursework. However, this is speculation, and it is perhaps equally likely that the researcher might have given higher grades than in previous years, knowing that the students had performed well in the seminars when practicing critical appraisal. It is perhaps more likely that the lack of change, and significant decline in the case of the second piece of work might reflect the nature of the students being taught and their increased numbers. In previous years, the module had an intake of 68 and 102 students, but for the current year, the module numbered 140, reflecting the increased intake of the degree course and university overall. Naturally it is likely that some of these students would be poorer academic performers, given the widening of access to higher education and higher numbers in a study making it more reflective of the general population mean, and this would be reflected in the mean coursework grades. Some support for this explanation can be gleaned from Figure 2, where it is clear that there was a significant difference between the grades two years previously and the current year for the second piece of coursework (see Figure 2).

Another explanation for the current findings is that it is simply not possible for certain students to progress to this level of critical thinking. Piagetian theory suggests that some individuals do not ever fully progress through the stages of development to formal operations, which incorporates deductive logic and critical hypothetical thinking (Piaget, 1936). However, this is a rather negative interpretation of the current findings. It is possible that critical appraisal teaching may not be able to increase the grades of high ability students who are

already in possession of many of these skills, but is capable of improving the grades of those who usually perform poorly. Support for this contention can be found in the current results (see Table 2) where the originally more poorly performing Biosciences students improved their coursework grades over the year. Others have noted that certain types of critical appraisal teaching has been observed to significantly increase the mean score of lower performing students, make no change to the grades of those in the middle, and drop the grades of higher performing students (Magnussen et al., 2000). That the, originally better performing, Sport and Exercise Sciences students' grades dropped between the two pieces of coursework might also be explainable in this way. However, it is possible that the overall significant decrease in grades between the assessments simply reflects the more complex and challenging nature of the second piece of coursework. Alternatively, the improvement in the Biosciences students' grades might reflect increased efforts given the lower grades for the journal article review, and also practice effects, given that they are less accustomed to essay-style coursework.

An alternative argument might be that teaching critical appraisal can benefit students both in their practice of these skills and the objective assessment of it in their work, but that this type of improvement needs time before it can be observed. In the present study, it was not possible to continue to follow the students to see whether or not their grades improved over subsequent years, but perhaps this type of strategy should be employed in future to examine the possibility that application of critical appraisal skills to written coursework is a competency that develops gradually. It is also possible, that critical thinking skills taught in one module may not be easily transferable to other teaching throughout a degree course, making it more difficult to see overall improvements. However, it is thought that critical thinking is a generic set of skills that can be taught in a way that makes it transferable (Halpern, 1998). More recent research than that reviewed in the present article has led to the design of a new taxonomy of educational objectives (Marzano & Kendall, 2007). This taxonomy emphasises a framework for thinking skills and associated thinking processes which can be implemented in teaching objectives and assessments to help develop students towards independent effective thinking. Using this thinking framework, educators are not only trying to teach content material and critical and creative thinking, but also metacognition, in other words the awareness of ones' own thinking. Although the present study attempted to encourage this type of thinking in the seminar setting, it did not formally use this thinking skills framework in the design of the teaching sessions, which might have been useful both in terms of the implementation of teaching and evaluation of its success. This is because the thinking skills framework attempts to define the various stages of thinking in enough detail to be easily translatable into teaching practices and techniques (Marzano, 1984).

The present study, naturally, has several other limitations. First, the changes in the module across the years, in terms of student numbers and composition, means that comparison of coursework grades across the last few years may not be the most valid way to examine the influence of the additional critical appraisal teaching. Similarly, the incorporation of some of the aspects involved in this teaching in previous years also means that any changes in the current year are less likely to be detectable. However, in comparison to previous studies of teaching critical appraisal, the current study did attempt to include some aspect of objective evaluation of the effectiveness of the teaching in addition to students' self-reports, which is an advantage. Second, coursework grades were the only objective measurement of the success of the teaching, and it is debatable how well these might measure this. Presumably their validity as an actual marker of the impact of the teaching will only be as

good as the match between the learning outcomes expressed in the teaching and the aims of the coursework itself. However, strong attempts were made in this study to link the aims of the seminars teaching critical appraisal to the coursework aims and evaluation method, although, in retrospect, direct feedback to students on how well their coursework showed evidence of critical thinking would also have been useful. It remains possible that the critical appraisal teaching led to improvements in other ways which were not observed or measured in this project, through using coursework grades as the outcome measure. A better approach might be to include more specific outcome measures such as formal critical appraisal testing before and following the seminar programme, in order to directly measure improvements in critical thinking which might not be reflected in academic coursework. Alternative future developments might be the incorporation of some other type of follow-up such as objective assessment of the students' critical appraisal and argumentation in the seminars themselves. This method has been used by others (see, for example, Peters et al., 2002). Third, on reflection, it might have been useful to incorporate some form of linking students' questionnaire responses and objective performance by code number, in order to examine more thoroughly the association between ratings of improvements in critical appraisal and coursework grades. Fourth, as mentioned above, the researcher marking the assessments was not blind to the aim of the additional critical appraisal teaching due to being the module organiser, thus it might be important to use blind markers in any assessments during the seminars themselves. The limitations listed thus far suggest that designing a means of assessing the change in critical thinking, specific to a set of learning content is as challenging as designing the teaching techniques themselves. Finally, it is feasible that some of the critical appraisal teaching was not as beneficial as intended. The ratings for the usefulness of seminars three and four were around the middle of the rating scale, indicating that there was room for improvement. Indeed, some of the students commented that the hypothetical listing of the possible differences between papers in seminar three was not particularly helpful, and that seminar four was far more useful in terms of the coursework and critical appraisal generally. It is likely that some changes should be made to these later seminars to improve their effectiveness given the performance on the second piece coursework compared to the previous assessment.

The contribution of the present research to the psychology teaching literature is to demonstrate that both cognitive and behavioural approaches to critical thinking can be applied in a higher education setting, along with the many pedagogical tips and techniques associated with the definitions and operations of critical thinking. The present research particularly utilised the behavioural model of giving students practice with using a critical approach to real content material (Huitt, 1998). As such, these approaches certainly increased students' own confidence in and appraisal of their critical thinking skills. However, this article demonstrates that an understanding and application of these approaches is a separate issue to being able to adequately measure the effectiveness of such teaching approaches and observe positive changes in students' abilities. Indeed, one of the flaws identified in the critical thinking teaching literature, was that many studies defined and attempted such teaching but failed to assess its impact. This suggests that researchers need to go beyond the definitions of critical thinking and strategies to teaching it, towards psychological methods of assessing whether such strategies are, in fact, effective. A better set of measures for capturing the effectiveness of critical thinking teaching will illuminate the positive and negative aspects of such teaching, and thus enable further improvements in the teaching of this vital skill. Psychologists, with the skills of questionnaire design and interview techniques among a multitude of

research design and measurement skills, are well placed to lead the field forward toward the scientific and objective evaluation of teaching effectiveness.

Future directions from the present study would be to address some of the issues raised above such as that of making clearer links between learning outcomes and assessments and using blind markers for additional assessments. Additional assessments could be formative rather than summative, such as the inclusion of generic critical thinking tests, in order to help the students develop and practice their critical thinking skills and monitor their progress. Also, on the basis of the students' comments and peer discussions with the postgraduate seminar leaders, it would be useful to redesign seminars three and four to make the teaching more effective. Seminar three could be used to summarise the key differences between the actual two studies being used, with seminar four could then used solely to discuss the differences previously identified which most might have contributed to the different findings, and logical arguments of how this might be the case. Given that this aspect of logical and critical argument receives the majority of marks in the coursework, it intuitively makes sense to devote more time to it in the seminar teaching.

In conclusion, a series of seminars designed to gradually develop critical appraisal skills were used to increase undergraduate students' critical appraisal. Students' ratings of their critical appraisal skills showed improvements over time, although this was not reflected in their assessed coursework. This study provides the outline of a set of tools which can be used to engage students in critical thinking about scientific literature. However, it also highlights the difficulty in the assessment of critical thinking in learning, and suggests that the choice of assessment method merits as much attention as the teaching methods themselves with regard producing useful evaluations of teaching methods. It remains to be tested whether or not the use of more directly linked assessments, or longer follow-up over time might yield improvements in students' abilities in terms of this important skill.

Correspondence

Dr Anna Phillips, PhD MSc BSc(Hons)
Health Psychologist,
School of Sport & Exercise Sciences,
University of Birmingham,
Edgbaston,
Birmingham B15 2TT.

References

Anderson, L. & Krathwhol, D, (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Audet, N., Gagnon, R., Ladouceur, R. & Marcil, M. (1993). How effective is the teaching of critical analysis of scientific publications? Review of studies and their methodological quality. *Canadian Medical Association Journal, 148*, 945–952.

Bensley, D.A. & Haynes, C. (1995). The acquisitiong of general purpose strategic knowledge for argumentation. *Teaching of Psychology, 22*, 41–45.

Biggs, J. (2000). *Teaching for quality learning at university*. Buckingham: Society for Research into Higher Education & Open University Press.

Bloom, B.S. (1956). *Taxonomy of Educational Objectives: The classification of educational goals. Handbook 1: Cognitive domain by a committee of college and university examiners*. London: David McKay Co.

Bonk, C.J. & Smith, G.S. (1998). Alternative instructional strategies for creative and critical thinking in the accounting curriculum. *Journal of Accounting Education, 16*, 261–293.

California Academic Press. *The teaching for thinking student course evaluation form.* www.insightassessment.com/pdf_files/Eval%20Course%20Form%20CT.pdf.

Castle, A. (2006). Assessment of the critical thinking skills of student radiographers. *Radiography, 12*, 88–95.

Dexter, P., Applegate, M., Backer, J., Claytor, K., Keffer, J., Norton, B. & Ross, B. (1997). A proposed framework for teaching and evaluating critical thinking in nursing. *Journal of Professional Nursing, 13*, 160–167.

Duemler, D. & Mayer, R. (1988). Hidden costs of reflectiveness: Aspects of successful scientific reasoning. *Journal of Educational Psychology, 80*, 419–423.

Halpern, D.F., (1998). Teaching critical thinking for transfer across domains. *American Psychologist, 53*, 449–455.

Huitt, W. (1992). Problem solving and decision making: Consideration of individual differences using the Myers-Briggs Type Indicator. *Journal of Psychological Type, 24*, 33–44. Retrieved 18 November 2009, from: www.edpsycinteractive.org/files/prbsmbti.html

Huitt, W (1998). Critical thinking: An overview. *Educational Psychology Interactive*. Retrieved 17 November 2009, from: www.edpsycinteractive.org/topics/cogsys/critthnk.html.

Huitt, W., & Hummel, J. (2006). An overview of the behavioural perspective. *Educational Psychology Interactive*. Valdosta, GA: Valdosta State University. Retrieved 17 November 2009, from: www.edpsycinteractive.org/col/behsys/behsys.html

Hyde, C., Parkes, J., Deeks, J. & Milne, R., (2000). *Systematic review of the effectiveness of teaching critical appraisal*. Oxford: ICRF/NHS Centre for Statistics in Medicine.

Klimoviene, G., Urboniene, J. & Barzdžiukiene, R. (2006). Developing critical thinking through co-operative learning. *Studies About Languages, 9*, 77–84.

Kowalczyk, N. & Leggett, T.D. (2005). Teaching critical thinking skills through group-based learning. *Radiologic Technology, 77*, 24–31.

Magnussen, L., Ishida, D. & Itano, J. (2000). The impact of the use of inquiry-based learning as a teaching methodology on the development of critical thinking. *Journal of Nursing Education, 39*, 360–364.

Marzano, R. (1984). *The theoretical framework for an instructional model of higher order thinking skills*. Denver, CO: Mid-Continent Regional Education Lab.

Marzano, R. & Kendall, J. (2007). *The new taxonomy of educational objectives*. Thousand Oaks, CA: Corwin Press.

Mayer, R. & Goodchild, F. (1990). *The critical thinker*. New York: W.C. Brown.

Norman, G.R. & Shannon, S.I. (1998). Effectiveness of instruction in critical appraisal (evidence-based medicine) skills: A critical appraisal. *Canadian Medical Association Journal, 158*, 177–181.

Norris, S.P. (1985). Synthesis of research on critical thinking. *Educational Leadership, 42*, 40–45.

Peters, M.W., Smith, M.F. & Smith, G.W. (2002). Use of critical interactive thinking exercises in teaching reproductive physiology to undergraduate students. *Journal of Animal Sciences, 80*, 862–865.

Scheffer, B.K. & Rubenfeld, M.G. (2000). A consensus statement on critical thinking in nursing. *Journal of Nursing Education, 39*, 352–359.

Siegel, H. (1980). Critical thinking as an educational ideal. *Educational Forum, 45*, 7–23.

Taylor, R., Reeves, B., Ewings, P., Binns, S., Keast, J. & Mears, R. (2000). A systematic review of the effectiveness of critical appraisal skills training for clinicians. *Medical Education, 34*, 120–125.