



Soden, Rebecca and Maclellan, Euphemia (2005) Helping education undergraduates to use appropriate criteria for evaluating accounts of motivation. *Studies in Higher Education*, 30 (4). pp. 445-458. ISSN 0307-5079

<http://strathprints.strath.ac.uk/3359/>

This is an author-produced version of a paper published in *Studies in Higher Education*, 30 (4). pp. 445-458. ISSN 0307-5079. This version has been peer-reviewed, but does not include the final publisher proof corrections, published layout, or pagination.

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (<http://eprints.cdlr.strath.ac.uk>) and the content of this paper for research or study, educational, or not-for-profit purposes without prior permission or charge. You may freely distribute the url (<http://eprints.cdlr.strath.ac.uk>) of the Strathprints website.

Any correspondence concerning this service should be sent to The Strathprints Administrator: eprints@cis.strath.ac.uk

**Helping education undergraduates to use appropriate criteria for evaluating
accounts of motivation**

Rebecca Soden and Effie Maclellan

Department of Educational Studies

Faculty of Education,

University Of Strathclyde,

Jordanhill Campus,

Southbrae Drive,

Glasgow G13 1 PP

e-mail r.soden@strath.ac.uk

Tel 0141 950 3155

Abstract. The aim of the study was to compare students in a control group (n = 45) with those in a treatment group (n = 29) with respect to evaluative comments on psychological accounts of motivation. The treatment group systematically scrutinized the nature and interpretation of evidence that supported different accounts, and the assumptions, logic, coherence and clarity of accounts. Content analysis of 74 scripts (using three categories) showed that the control group students made more assertions than either evidential or evaluative points, whereas the treatment group used evaluative statements as often as they used assertion. The findings provide support for privileging activities that develop understanding of how knowledge might be contested, and suggest a need for further research on pedagogies to serve this end. The idea is considered that such understanding has a pivotal role in the development of critical thinking (Kuhn, 1999, King & Kitchener, 1994).

Introduction

Current advocacy of the idea that constructivist accounts of learning (Piaget 1977, Vygotsky 1978) provide powerful theoretical frameworks for instruction (Salomon & Perkins, 1998, McGuinness 1999, Shayer & Adey 2002), implies that it is important to understand how students might learn to evaluate knowledge they study. A central tenet of these constructivist accounts is that learning is an active restructuring of knowledge by learners that is provoked when they interpret experiences as dissonant with their current understandings. This tenet implies that students need to learn to construct and deploy criteria, such as available evidence, in making judgements about whatever they study. These processes often are included in the term critical thinking. Kuhn's (1991) description of critical thinking includes being able to: (a) separate beliefs from evidence; (b) imagine beliefs alternative to one's own and to know what evidence would support these; (c) provide evidence which supports one's own beliefs while rebutting the alternatives; (d) to weigh up reasons for believing what is alleged to be known. Thus, Kuhn implies that students need to learn to restructure knowledge in the light of criteria.

That such thinking seems to be poorly achieved has been attributed in part to its poor delineation in interventions (Kuhn, 1999). Livingston et al (2003) reported that, often, the term critical thinking is used to describe whatever thinking processes the researchers wanted students to learn, such as analysing and articulating concepts, sorting through ideas, connecting thoughts, constructing and reflecting on ideas, constructing courses of action from analyses and hypothetical-deductive thinking. Often referred to as cognitive skills, these processes appear in the many theoretical frameworks that purport to describe thinking (Moseley et al. in press), and were the focus of the vast majority of interventions reviewed by Livingston et al. As Moseley et al point out, some frameworks include almost every known cognitive process, on the basis of

evidence that it is difficult to think critically about any task without deploying processes that often have been considered more basic, such as memory skills for organising information meaningfully.

It has been argued that cognitive skills training is unlikely to lead to significant and sustainable progress with thinking unless it is accompanied by measures that help students to challenge their assumptions about the nature of knowledge and how it is gained (King & Kitchener 1994, Kuhn, 1999, King, 2000, Palmer & Marra, 2003). These assumptions and their sequence of development are described in King & Kitchener's (1994) Reflective Judgement Model, which describes three broad levels of epistemic development that reflect different beliefs about knowledge and how it is constructed. People who reason at a pre-reflective level believe that knowledge is certain, that knowledge from good authorities is right, that justification is about showing that one's own beliefs are right and that this is achieved by simply stating them, whereas those who attain either of the two higher levels (quasi-reflective and reflective reasoning) understand that knowledge is contestable rather than absolute, and that rules of inquiry for the context can be invoked for establishing the veracity of knowledge claims.

Although critical thinking probably depends on the execution of many cognitive processes (Halpern 1997), some writers argue that cognitive skills training cannot be expected to bring about critical thinking unless the training invokes criteria for evaluating knowledge (Lipman 1991, Kuhn 1999). The central idea in Lipman's (1991, p.116) definition of critical thinking, that it is thinking 'that can be assessed by appeal to criteria' illuminates what students need to learn if they are to progress beyond the restrictive pre-reflective level of epistemological reasoning described in King & Kitchener's (1994) Reflective Judgement Model. King (2000) reported that, although very few college students attain fully reflective levels of thinking, some

students can use rules of enquiry to examine claims, a central feature of quasi-reflective thinking. However, King reports that those in courses with a vocational orientation, such as the teacher education course that is the focus of the present study, are less likely to achieve quasi-reflective thinking. Since both King (2000) and Kuhn (1999) argue that improvements in thinking are likely to be insignificant until students change their epistemological assumptions, the present study seeks to influence such change by helping students to use constructs that are helpful in forming rules of enquiry.

Concern that thinking aims are not being achieved has prompted spates of educational interventions designed to identify instructional practices that encourage such thinking (for a recent evaluation of interventions see Livingston et al. 2003). As Livingston et al. (2003) noted, there have been few rigorously evaluated interventions that have sought to develop abilities connected with the epistemological reasoning described by King & Kitchener (1994) and by Kuhn (1991, 1999). A notable exception was Jones & Merrit's (1999) TALESSI (Teaching and Learning at the Environment-Science-Society Interface) Project, which indicated that undergraduates could be moved towards a critical epistemology. Jones & Merrit's (1999) approach focused on teaching the means to question and reveal the contestable character of knowledge claims about environmental questions, and this was directed toward constructing knowledge bases, on the grounds that important differences between experts and novices lie in the quality of their knowledge (e.g. Hoffman 1998). Activities to promote these ends were built into the redesign of curricula. However, the evaluation of the project focused on teachers' views of the redesigned curricula rather than on quantifying the incidence of targeted aspects of thinking in students' writing, as does the present study.

Lawson et al's (2000) findings from a rigorous evaluation of a programme that was designed to promote the development of the ability to raise and answer a set of questions before drawing a firm conclusion about the relative truth or falsity of any particular causal claim suggest that there are benefits in this precise operationalizing of a critical epistemology. Students learned to handle questions about what alternative causes were possible, in addition to the proposed cause, about how each possibility could be tested, and about how the evidence, once gathered, matched the expectations. Lawson et al's (2000) findings support their claim that students were able not only to use this form of thinking in the classes in which it was taught but were able also to transfer the form of thinking within the domain of science. They claim that the approach promoted concept construction and thereby helped students to construct a knowledge base in science, and that students used the approach competently on the course examinations covering a wide range of theoretical topics. Given the scarcity of evidence that what is learned in one context transfers to others (Haskell 2001), these findings provide powerful support for privileging activities that develop understanding of how knowledge might be contested.

Taken together, the findings outlined above provide grounds for the present modest intervention, which focuses on learning to use appropriate constructs to carry out a rudimentary evaluation of psychological perspectives. By avoiding looser definition of thinking used in many interventions reviewed by Livingston et al, it was hoped to maintain an achievable and measurable student goal. While generalisable results are the ultimate aim, it is important to emphasise that the study reported here did not have such lofty aims, partly because of the sample frame used and partly because of the authors' desire to infuse the aims into existing curricula. While the authors acknowledge the strength of arguments against the inclusion of some kinds of theory in professional courses, and the complexity of relationships between theory and practice

(Rowlands 1999; Thomas 1997), they would want to emphasise that these are not the central concerns of this paper. Rather, their concern is the more modest one that students should not accept uncritically the ideas in content prescribed for their course. Therefore, the aim was to evaluate an intervention designed to decrease the incidence of assertion and increase the incidence of evaluative comment in undergraduate writing about three psychological perspectives. Specifically, the question to be answered is:

To what extent are patterns of evaluating, asserting, and evidencing in an undergraduate assessment task in a teacher education course influenced by brief participation in an approach informed by critical thinking literature?

The intervention that was evaluated

The learning activities

The authors tried out an approach to developing critical thinking (designated Approach A), based on systematic practice in using constructs included in many taxonomies of critical thinking (e.g. see Moseley et al. 2004) to question theoretical accounts of motivation. This approach was tried with twenty-nine students during four slots, each lasting ninety minutes, in a second year module in a Bachelor's degree in Primary Education that focused on three theoretical frameworks (Applied Behavioural Analysis, Cognitive Behaviourism and Humanism), with an emphasis on the contribution of these frameworks to understanding classroom behaviour. A non-treatment control group from the same cohort spent the same time in tutorial activities described below that had been used for several years, designated Approach B. Thus, Approach B can be described as

an unchanged pedagogy condition. Approach A and Approach B had a common aim: to enable students to critique the psychological ideas they were studying.

In Approach A, the students about the following constructs and their use in forming questions about the three perspectives they studied:

- assumptions
- coherence and clarity
- nature of evidence
- interpretation of evidence
- generalizations

Following Bensley (1998), the students practised using the constructs (assumptions etc) in relation to the module content and to possible responses to children's behaviour. In general, Approach A incorporated pedagogical principles discussed in Salomon and Perkins, (1998).

In Approach B, instead of building in systematic practice in using the above constructs, the tutorials were loosely structured around questions about what might be incomplete or problematic in the account provided by each perspective, and tutors facilitated students' attempts to comment critically on the module content. This can be contrasted with Approach A, where the tutors explicitly and systematically asked students to examine assumptions, coherence, clarity, evidence and its interpretation, and justifiable generalisations, all in relation to the module content.

Both Approach A and Approach B tutorials included discussion of the same assessment criteria, both the control and treatment group attended two lectures, and before each tutorial

meeting all the students completed prescribed reading (from Porter 2000), which included a chapter that introduced the targeted constructs. In both approaches the students engaged in discussion in small groups and tutors chaired plenary sessions.

The following is a typical activity in Approach A:

Students were asked to discuss their responses to the following questions about an article they had read:

- What was the author's purpose in writing the article?
(for example to report on an empirical study, to examine an issue critically or what?)
- What are the author's main findings or conclusions?
- What evidence does the author offer in support of these findings?
- Might there be better evidence for the findings? If so what would that be like?
- Might there be other ways of explaining the findings?
- What, if anything, does the author seem to assume about the nature of children, the nature of discipline and the role of the teacher. To what extent do you agree with these assumptions?
- How far might the authors' conclusions be generalized?

The assessment task

All students enrolled on the module (n = 74) wrote a response to the following question, but only responses to Part c of the question were used as research data:

- a. Provide a brief account of Applied Behavioural Analysis (ABA).

- b. Comment on how ABA helps you to understand the behaviour of children.
- c. Consider in what ways this account is incomplete.

Method

Research Design

A quasi-experimental design was employed on an opportunity sample of one cohort of students. The study compared post-treatment patterns of evaluating, asserting and evidencing by a treatment group and an untreated control group. Evidence of the effectiveness of such interventions (Bereiter, 2002) is slender enough to suppose that students in a control group would not be disadvantaged by non-participation in the intervention.

Participants

A cohort of seventy-four (74) students (64 female and 10 male) who were registered for a Bachelor's degree in Primary Education (a professional qualification with a license to practice), and who were enrolled on the targeted module, were randomly allocated to the treatment group (n = 29) and a control group (n = 45). The pedagogical treatment consisted of Approach A, described above. The control group students were not exposed to any changes in pedagogy. This no-change pedagogy is described above as Approach B.

Data/Data analysis

The data consisted of responses written under examination conditions to Part c of the assessment task described above. Since there was little variation in the responses to Parts a and b of the assessment task, which did not invite a substantially analytical response, they were not coded. However, given posited connections between knowledge and thinking in many taxonomies of

thinking (Moseley et al.), the authors looked at responses to Parts a and b to check that evaluation had not been impeded by inability to recall module content.

Following Tesch (1990), the unit of analysis was a segment of text that is comprehensible by itself and contains one idea, episode or piece of information. The development of the categories was guided by both literature on critical thinking (outlined earlier in this paper) and the data itself. The procedure used in developing the system involved iteration between literature and data, following a process described by Anderson et al. (2001), and which involved a research assistant as well as the authors. The category system was revised and developed through this process of iteration until agreement was reached on a set of codes. The research assistant used the final category system to code all assignments. Cronbach's alpha was used to calculate the reliability of the coding scheme. The results indicated that the coding could be carried out reliably over a two-week period. (Cronbach's alpha = 0.70).

The final set of codes was as follows. Further examples of statements allocated to codes are presented in the discussion section of the paper.

1. 'Evaluates' assertions or evidence

To be counted as evaluative, statements must match any one of the following indicators:

- Discusses assumptions in views cited.
- Comments on logic, coherence, clarity of views cited.
- Considers the nature of the evidence, e.g. reliability, validity, generalisability, different possible interpretations of evidence, what agenda might be served by evidence/the extent to which the evidence is from a reliable, disinterested source, discusses the extent to which significant evidence is absent.

- Cites an alternative explanations of phenomena discussed, considers limitations of explanations, and what any account explains well.
- Considers adequacy of inferences from evidence.

Sample statement in this category:

Thinking is missing from ABA accounts, which makes them incomplete because other accounts of learning, such as attribution, provide evidence of that there is a link between how children think about reasons for their success and failure and their achievement.

2. ‘Asserts’ – an opinion that is not supported by any kind of evidence, or by any reference to the indicators set out above for ‘Evaluates’ statements, or by reasoned argument:

Sample statement:

ABA is an incomplete account because all children are different.

3. ‘Evidences’ – something offered to support assertions

Any point of view or empirical finding which is attributed either to a publicly available source such as an acknowledged authority, a journal article, book, or newspaper, or is said to be based on placement experience; it is presented as a reason for believing that an adjacent assertion might be valid. The difference between *Evaluates* and *Evidences* is that *Evidence* statements simply recite evidence without commentary, whereas *Evaluates* statements comment on the evidence in a way that matches one of the indicators set out above.

Sample evidence statement:

... during my third placement in school (or alternatively, according to McGrath) there were improvements in behaviour in children who were on an ABA programme.

Results

Tables 1 and 2 below show that there are some differences in the pattern of responses for the control and treatment groups. Differences between categories are significant at least to the level of $p < 0.05$, but some are significant at the more stringent level of $p < 0.001$. The control group scores are presented in Table 1 below:

Please insert Table 1 about here

The significant differences that emerged between categories ($F_{2,88} = 84.74, p < 0.001$) for the control group data (Table 1) were followed-up by Bonferroni t-tests, which indicated that there were significantly more statements coded as *Asserts* (Category 2) than statements coded as either *Evaluates* (Category 1) or *Evidences* (Category 3), but there was no significant difference between the amount of statements coded as *Evaluates* and *Evidences* (Categories 1 and 3). Table 2 below shows the post-treatment test scores.

Please insert Table 2 about here

The significant difference that emerged between categories ($F_{2,56} = 6.75, p < 0.01$) for the treatment group data (Table 2) was followed-up by Bonferroni t-tests, which indicated that there were no significant differences between the number of statements coded as *Evaluates* (Category 1) and statements coded as *Asserts* (Category 2). Tables 1 and 2 show that there were no significant differences between the amount of evaluation and evidence statements written by students in either the treatment or control group, although the means for the treatment group exceed those of the control group for both these categories.

Discussion

Overview

All that the authors claim is that, on a part of a task that asked specifically for evaluation, the control group students asserted (mean = 10.18) four times as often as they appealed to criteria (mean = 2.53), whereas the treatment group did not assert (mean = 5.69) significantly more often than they evaluated knowledge (mean = 4.00). This main finding can be related to Kuhn's (1999) discussion of Lipman's (1991, p.116) point that critical thinking is thinking that can be assessed by appeal to criteria. There are grounds for connecting the results with the pedagogy, in that the criteria most commonly used by the treatment group were ones that had been formed from the constructs that had been taught in Approach A. Since the only student statements that were coded were the ones written in response to a request to say 'what is incomplete' about a particular perspective, it might be inferred that the more explicit and systematic approach in the treatment group tutorials had greater impact on students' grasp of criteria for addressing this question. During tutorials, students in both the control and treatment groups discussed the need to weigh up evidence for their views, and were informed in writing that evaluative statements

would be rewarded through higher grades. Therefore, the findings suggest the different assertion/evaluation patterns in the two groups do not simply reflect ignorance of what was valued by the assessors. but a firmer grasp by the treatment groups students of appropriate means of enquiry to contest the ideas they were studying.

Although modest, the improvement in the treatment group is not trivial if it is considered in the light of themes that pervades the literature: the difficulty in bringing about *any* improvement of thinking (Lawson et al. (2000) and the scarcity of evidence of transfer of what has been learned from one context to another, such as the transfer from tutorial work to the examination in the present study (Haskell, 2001). Moreover, these differences arose at the end of an intervention that was much shorter than is usually reported and which can be easily embedded into higher education teaching. However, it is important to highlight some limitations of this type of study.

First, there are problematic assumptions in any categorization system that is used to quantify differences in the incidence of types of statements that students write. Definitions of categories developed from a different body of research for the three categories (*evaluates*, *asserts* and *evidences*) might have produced a different pattern of results. In the present study the definitions used are ones that describe forms of thinking that are thought to be important in the discipline (psychology) that the students studied (Bensley 1998), as well as in many scientific disciplines (Lawson et al. 2000). In future research, there is a case for taking into account the distinction that Keefer (1996) made between practical arguments and theoretical arguments, and his point that different standards are appropriate for judging practical arguments.

Second, many unknown variables may have influenced performance in both groups. Although Approach A included much more systematic practice on aspects of critical thinking, as is the case in similar studies, it is difficult to isolate the effect of such practice from other variables, particularly since there was no dialogue analysis of interaction during tutorial sessions. It is possible too that students in the treatment group shared their understandings with their peers in the control group. While it would have been unethical to discourage such sharing, it is possible that it did attenuate differences between the groups. Like most studies of its type, it is open to a fundamental objection that differences in post-tutorial reading and reflection contributed to the different pattern of statements in the two groups.

A closer look at the main finding

For the most part students invoked the adequacy of assumptions, the strengths and weaknesses in evidence and alternative interpretations of evidence. In addressing the question of ‘what is incomplete’ the control group mainly asserted what they ‘felt’ about a perspective. Their responses were characterised by an absence of criteria. The use of constructs learned in the treatment condition to form criteria can be illustrated by reference to the indicators listed in the Data/Data Analysis section; a match with any one indicator allowed a statement to be coded as *evaluates*. An example of an attempt to use as a criterion the validity of assumptions appears in the following statements:

ABA accounts assume that teachers have a right and responsibility to be in charge, but Humanists point out an alternative assumption is that students and teachers have equal rights to have their needs met but occupy different roles. Thus, not all research shares the assumption in behaviour modification that it is ethically acceptable for adults to

decide what are desirable behaviours for children. We need to judge it against our own and others' ideas about ethics.

The statements coded as *evaluates* reflect thinking that has some resemblance to a critical epistemology (Perkins 1989) in that these students seem to be trying to consider possible objections to ideas they have encountered in their module, whereas ones coded as *Asserts* are closer to Perkins' description of a makes-sense epistemology, in that they assert, but do not consider if their assertion stands up to scrutiny. Rather they go straight on to consider practices that flow from their assertion: students who asserted (with no evidence) that 'all children are different' went on to advocate classroom practices that reflected this belief. They did not question what the differences might be, or consider why they should believe that these differences existed. Nor did they refer to the practical possibility of implementing the practices they advocated to deal with the supposed differences. Like the adults in Kuhn's (1991) sample, they seemed to be thinking with a personal theory but not about this theory, since there were no adjacent statements that suggested enquiry into its veracity. This can be contrasted with the treatment group's propensity to raise possible objections as often as asserting views. For example, they raised objections in statements such as:

Behaviourists argue that it's not worth considering thinking processes but my placement experience suggests that if a teacher does not ask children to talk about how they thought their way through a problem she does not know enough about the children's difficulties to help them.

I tried to apply ideas from Humanism in my last placement but it was hard to say how they should be interpreted in the classroom because they're so woolly and slippery.

What emerges from closer inspection of the control group data is that many statements begin with an apparently evaluative intention, such as ‘Cognitive Behaviourism is better than Humanism’. However, the statement is completed by simply stating an opinion, unaccompanied by evidence or reasoning, implying that students believed that opinion alone was a sufficient justification. In King & Kitchener’s (1994) Model of Reflective Judgement, such responses are characterised as pre-reflective thinking because they do not invoke rules of inquiry for the context. While the control group students asserted much more often than they evaluated or evidenced, attempts to evaluate psychological knowledge by weighing up evidence and querying assumptions were not missing altogether from the control group data.

However, while the pattern of results in the treatment group can be interpreted as consistent with quasi-reflective reasoning in King & Kitchener’s (1994) Model of Reflective Judgement, these students sometimes used pre-reflective reasoning, such as simply stating their beliefs, as well as querying their veracity in appropriate ways. Within each of King & Kitchener’s (1994) levels are stages. They suggest that the reasoning of a less sophisticated stage continues as the reasoning of the next stage develops and manifests itself. According to King & Kitchener (1994), at each stage the cognitive seeds of the next level are developing. Inspection of the treatment group data suggests that the pedagogy influenced some students more than others in this group to use rules of enquiry. It is possible that those who evaluated less often might have benefited from a longer period of systematic instruction about appropriate enquiry constructs.

Incidence of evaluation in expert and undergraduate writing

What might be a reasonable level of evaluative commentary in second year students' assessment tasks in a programme that included activities to promote such thinking is difficult to judge because few other studies have defined and measured critical thinking in a similar way and in a similar population. Anderson et al. (2001) used a similar category system to capture critical thinking in an intervention and reported means in three pieces of writing that ranged from 1.07 to 1.7 for evidence statements but found no evaluative statements in their data. In a study that allocated PGCE students' writing to categories that included critical reflection, Maclellan (1999) allocated only two out of her twenty-five samples of writing to this category, with the others falling into descriptive categories. Maclellan's indicators of critical reflection were similar to those used in the present study to allow statements to be coded as *evaluates*. Overall, the treatment group students in the present study constructed more evaluative statements than those in either of the other two studies.

What might be regarded as typical expert levels of competence in writing evaluative statements was assessed by coding a critique of Cognitive Behaviourism from the text used by the students (Porter 2000), using the same category system. Almost all the statements were allocated to the *evaluates* and *evidences* codes. As would be predicted by expert/novice research (e.g. Hoffman 1998), the students in both groups fell far short of this level. It might be helpful for students to practise identifying how the author of a critique questions assumptions and weighs up evidence. However, the paucity of students' knowledge also hampered their attempts at evaluation. For example, many students erroneously imagined that Applied Behaviour Analysis could not work because 'it does not allow for involving parents', whereas others had read enough to understand that consistency in the use of this approach 'inside and outside

school' is important. Many writers argue that knowledge and thinking are intimately connected (Bonnett 1995; Gardner & Johnson 1996). Reasons for believing that appropriate forms of thinking arise from sound knowledge of a domain, such as psychology, have been well rehearsed by Hirschfeld & Gelman (1994). The literature provides sound reasons for supposing that, although the students' responses to the other parts of the assessment task suggested that they had basic knowledge of the perspectives studied, the impact of Approach A might have been attenuated by insufficient student reading.

Implications

The findings are consistent with Bensley's (1998) claim that sustained, systematic practice in weighing up underlying assumptions, coherence and quality of supporting and conflicting evidence in prescribed reading is a promising way of helping students to think critically about what they are studying (Bensley 1998). Plainly, substantial assessment credit has to be given for engaging in such scrutiny. The approach seems to have potential for disturbing students' beliefs that knowledge is absolute and entirely a matter of opinion, and for providing some conceptual tools for acting on this understanding. Nevertheless, the modest differences imply that a mindset might exist, such as that described by Perkins' (1989) makes-sense epistemology, that will take more to change than the measures adopted in this study. Livingston et al's (2003) conclusions support a belief that more substantial progress might have been achieved if students had practised similar aspects of thinking across the curriculum.

The fact that students in the control group predominantly reasoned in ways that are consistent with the most limited 'pre-reflective' level in King & Kitchener's (1994) Model of Reflective Judgement suggests that much more systematic measures are required across the

curriculum to ensure that students constantly bring to bear on their course content the sort of enquiry tools that were practised in the present study. According to King (2000) pre-reflective thinking is a barrier to reflective judgement, and disadvantages teachers and other professionals who have to deal with ill-structured problems such as curriculum development and classroom management. A critical epistemology (Perkins 1989) encapsulates thinking that seems to have professional utility in that it implies that people understand the pitfalls of justification and try to test their arguments against possible objections. It might be supposed that those who think about their practice in this way are likely to be thorough in their evaluation of innovations proposed by themselves or others.

According to King and Kitchener (1994), to be deemed capable of fully reflective thinking (rather than quasi-reflective) students must provide evidence of interrelating considerations such as the explanatory value of a perspective, the strength of evidence that supports it and the practical possibility of working with the perspective. This level of thinking seems likely to be promoted by studying rigorously a range of disciplinary perspectives, since it is difficult to entertain such considerations in the absence of some knowledge of alternative explanations, and the arguments that underpin them. The ability to interrelate psychological and sociological explanations of the educational difficulties that children in disadvantaged communities sometimes experience would be a step towards fully reflective thinking. How such progress might be achieved is well illustrated in the Jones & Merrit intervention, but requires institutional changes that were beyond the scope of the present study. Although King & Kitchener's (1994) model leaves questions to be resolved about how individuals progress through the stages and about the relationship between maturation, education and culture, the findings in the present study support the view that introducing learners to constructs that are

useful in making more reasoned judgments might help them to advance their epistemological understanding.

A good deal of research remains to be done on identifying a range of conditions that support the development of aspects of thinking (Zohar et al 2001, Browne & Freeman 2000). These questions seem to be worth addressing if higher education is to enhance employability and citizenship, which seem to be well served by a capacity to arrive at reasoned judgments by weighing up evidence in the light of multiple and often conflicting criteria, and by a growing understanding that knowledge is neither entirely objective and certain, nor is it simply a matter of opinion.

References

- ANDERSON, A., HOWE, C., SODEN, R., HALLIDAY, J. & LOWE, J. (2001) Peer interaction and the learning of critical thinking skills in further education students, *Instructional Science*, 29, pp. 1-32.
- BENSLEY, D. A. (1998) *Critical Thinking in Psychology: A Unified Skills Approach* (Pacific Grove CA, USA, Brooks/Cole Publishing Company).
- BEREITER, C., (2002), *Education and Mind in the Knowledge Age*. NJ: Lawrence Erlbaum Associates.
- BONNETT, M. (1995) Teaching thinking and the sanctity of content, *Journal of Philosophy of Education*, 29 (3), pp. 295-309.
- BROWNE, M. N. & FREEMAN, K. (2000) Distinguishing Features of Critical Thinking Classrooms, *Teaching in Higher Education*, 5 (3) pp. 301-309.
- GARDNER, P. & JOHNSON, S. (1996) Thinking Critically about Critical thinking: an unskilled enquiry into Quinn and McPeck, *Journal of Philosophy of Education* 30 (3) pp. 441-456.

- HALPERN, D.F. (1997) *Critical Thinking Across the Curriculum: a brief edition of thought and knowledge* (Mahwah, N.J., Lawrence Erlbaum Associates).
- HASKELL, R.E. (2001) *Transfer of Learning: Cognition, Instruction and Reasoning* (London, Academic Press).
- HOFFMAN, R. R. (1998) How can expertise be defined? Implications of research from cognitive psychology. In R. Williams, W. Faulkner and J. Fleck (eds.), *Exploring expertise: issues and perspectives*, 81-100 (Basingstoke: MacMillan).
- HIRSCHFELD, L & GELMAN, S 1994 *Mapping the mind: domain specificity in cognition and culture* (Cambridge, Cambridge University Press). HYLAND, T. & JOHNSON, S. (1998) Of Cabbages and Key Skills: exploding the mythology of core transferable skills in post-school education, *Journal of Further and Higher Education*, 22 (2), pp. 163-172.
- JONES, P. C. & MERRIT, J. Q. (1999) The TALESSI Project: Promoting Active Learning for Interdisciplinarity, Values Awareness and Critical Thinking in Environmental Higher Education, *Journal of Geography in Higher Education*, 23 (3), pp. 335-348
- KEEFER, M. W. (1996) Distinguishing practical and theoretical reasoning: A critique of Deanna Kuhn's theory of informal argument, *Informal Logic*, 18(1), pp. 35-55.
- KING, P. (2000) Learning to make reflective judgements, *New directions for teaching and learning*, 82, pp. 15-26.
- KING, P. & KITCHENER, (1994) *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults* (San Francisco, Jossey-Bass).
- KUHN, D. (1991) *The skills of argument* (Cambridge: Cambridge University Press).

KUHN, D. (1999) A Developmental Model of Critical Thinking. *Educational Researcher*, 28, (2), pp. 16-25.

LAWSON, E., CLARK, B., CRAMER-MELDRUM, E., FALCONER, K., SEQUIST, J. & KWON, Y. (2000) Development of scientific reasoning in college biology: do two levels of general hypothesis testing skills exist? *Journal of Research in Science Teaching* 37 (1) pp. 81-101.

LIPMAN, M. (1991) *Thinking in Education* (2nd. edition) (Cambridge, Cambridge University Press).

LIVINGSTON, K., SODEN, R. & KIRKWOOD, M. (2003 in press) 'Post 16 Pedagogy and Thinking Skills: an evaluation' (London, Learning and Skills Development Agency).

McGUINNESS, C. (1999) *From Thinking Skills to Thinking Classrooms: a review and evaluation of approaches for developing pupils' thinking*. Research Report No. 115, Department for Education and Employment (Norwich, HMSO).

MACLELLAN, E. (1999) Reflective Commentaries: what do they say about learning? *Educational Action Research*, 7 (3) pp. 433-449.

MOSELEY, D.V., BAUMFIELD, V., M., ELLIOTT, J., GREGSON, HIGGINS, S.E., LIN, M., MILLER, J., NEWTON, D.P. AND ROBSON, S. (2004 in press) Thinking Skill Frameworks for post-16 Learners: an Evaluation (London: Learning and Skills Development Agency).

PALMER, B. & MARRA, R. (2003) College student epistemological perspectives across knowledge domains: A proposed grounded theory, *Higher Education*, pp. 1-25.

PERKINS, D. N. (1989) Reasoning as it is and as it could be: an empirical

- perspective. In Topping, D. M., Crowell, D. C. & V. N. Kobayashi (eds.) *Thinking across cultures: the third international conference on thinking* (Hillsdale, N.J., Erlbaum, 175-94).
- PIAGET, J. (1977) *The Principles of Genetic Epistemology* (London, Routledge and Kegan Paul).
- PORTER, L. (2000) *Behaviour in Schools*. (Buckingham, The Open University Press).
- ROWLAND, S. (1999) The role of theory in a pedagogical model for lecturers in higher education, *Studies in Higher Education*, 24,3, pp. 303-314.
- SALOMON, G. AND PERKINS, D. (1998) Individual and Social Aspects of Learning, *Review of Research in Education*, 23, pp. 1-24.
- SHAYER, M. AND ADEY, P. (2002) *Learning Intelligence*. (Buckingham, Open University Press).
- THOMAS, G. (1997) What's the use of theory? *Harvard Educational Review*, 67, 1, pp. 76-105.
- TESCH, R. (1990) *Qualitative research analysis types and software tools* (New York, Falmer Press).
- VYGOTSKY, L. S. (1978) *Mind in society: The development of higher psychological processes* (Cambridge, Mass: Harvard University Press).
- ZOHAR, A., DEGANI, A. & VAAKNIN, E. (2001) Teachers' beliefs about low-achieving students and higher order thinking. *Teaching and Teacher Education*, 17, pp. 469-485.

Table I Control group test scores: Mean number of statements in each of the three categories and standard deviations (SD)

Description of Category	Category Number	Mean (n=45)	S.D.
Evaluates	1	2.53 _a	2.19
Asserts	2	10.18 _b	5.25
Evidences	3	1.71 _a	2.06

Table II Post-treatment test scores: Mean number of statements in each of the three categories

Description of Category	Category Number	Mean (n=29)	S.D.
Evaluates	1	4.00 _{ab}	2.02
Asserts	2	5.69 _a	4.76
Evidences	3	2.66 _b	3.47

Total number of words 6260)