

Determining the level of reflective thinking from students' written journals using a coding scheme based on the work of Mezirow

DAVID KEMBER, ALICE JONES, ALICE LOKE,
JAN MCKAY, KIT SINCLAIR, HARRISON TSE,
CELIA WEBB, FRANCES WONG, MARIAN WONG
and ELLA YEUNG

Hong Kong Polytechnic University, Kowloon, Hong Kong

To determine whether students are engaged in reflective practice it is necessary to have some means of identifying reflective thought and a measure of the depth of reflective thinking. Several measures of reflectivity have been proposed but there appears to be no widely accepted and clearly formulated procedure for determining levels of reflective thinking from students' written reflective journals. In this study we propose a scheme for estimating the quality of reflective thinking in students' writing in reflective journals, using categories based on Mezirow's work on reflective thinking. In an initial test of the scheme, reasonable levels of agreement were obtained from eight judges. Disagreements over coding resulted from differing interpretations of the significance of what students had written rather than from a lack of precision in the guidelines for coding categories. A second test, using students' reflective papers, showed acceptable levels of reliability between four assessors. The method is recommended for both assessing students and evaluating courses in programs which aim to develop reflective thinking.

Assessing levels of reflective thinking

The aim of this project was to devise a method for assessing the kind of reflective thinking which students undertake in programs for professional education. The hope was to develop a method which would examine writing in reflective journals which are used to both develop reflective thinking and record the outcomes of such thinking in courses of this type. The rationale for this venture was that these programs frequently take as their main aim developing a reflective practitioner, but have no means of assessing whether students engage in reflective thinking in their courses or the periods of professional practice.

It is now widely accepted that successful professionals need to reflect upon their actions as most tasks they perform involve novel elements to which there are no defined solutions. This implies that courses educating professionals should aim to develop students' abilities to reflect upon their actions. The aim of encouraging student reflective thinking is now incorporated into many professional courses.

David Kember is the Coordinator of the Action Learning Project. *Alice Jones*, *Kit Sinclair*, *Celia Webb* and *Ella Yeung* are in the Department of Rehabilitation Sciences. *Alice Loke*, *Frances Wong* and *Marian Wong* are in the Department of Nursing and Health Sciences. *Jan McKay* is in the Department of Optometry and Radiography. *Harrison Tse* was the Research Fellow for the project.

Given the proliferation of courses which aim to develop students' abilities to reflect upon their practice, it is desirable to have a valid and reliable procedure for determining whether students achieve the aim. If the aim is to be directly assessed, the assessment should incorporate a process for determining whether students do reflect upon their practice and the depth of that reflective thinking. Such a measure would also be invaluable for course evaluation procedures. If a curriculum is designed with encouraging reflective thinking in mind, there is a need to determine whether students do indeed engage in reflective practice for the outcomes to be properly evaluated. What is perhaps surprising then is that there does not appear to be any widely accepted procedure for determining whether reflective thinking takes place or assessing the level of reflective thinking from written journals. It was possible to locate some attempts to determine levels of student reflective thinking, but we were also unable to find any instances of authors using procedures developed by others, which is a sign of a method achieving wider acceptance. None of the methods satisfied each of the criteria below we considered necessary.

The method should focus upon reflective thinking and should assess it directly. We specifically looked for a method which could be used with students' written reflective journals. We felt it desirable that any method of assessing the level of reflectivity should be based upon existing features of the course, both for authenticity and to avoid adding additional loads to students and faculty.

We looked for a naturalistic evaluation based on scenarios relevant to the requirements of the professional courses rather than unrelated additional testing procedures or instruments. The method should be specified in sufficient detail so that others could use it. There should be appropriate testing for reliability.

Of the schemes we examined, Sparks-Langer *et al.* (1990) report a seven level framework based upon the type of language used by students. The authors describe the levels in the coding scheme as a mirror of Gagné's (1968) hierarchy of thinking. The method was well documented but the levels equated more to the linguistic structure of discourse than to commonly accepted models of reflective thinking.

Powell (1989) presented two analyses of interview transcripts using a six-category scheme based upon Mezirow (1981) and a five-category scheme derived from Colaizzi (1973). However, the paper gave no details of the coding procedure and reported no attempt to determine the reliability or validity of the coding process. Hahnemann (1986) reported a method for assessing the content of journal entries. The focus was whether the answers to questions were correct, rather than whether there was evidence of reflective thinking.

A method for monitoring reflective judgement which did appear to be well developed and tested has been described by King and Kitchener (1994). The authors describe a method for determining which of seven stages of reflective judgement, based on a reflective judgement model, has been reached by a student. The method relies upon semi-structured interviews, by a trained interviewer, investigating students' processes in tackling four ill-structured problems such as how the Egyptian pyramids were built. The method appeared to be appropriate for the research purposes of its originators but hard to apply to regular use in taught professional courses. The resources and expertise required to conduct the tests and the time requirement for both students and faculty made the procedure unsuited to routine teaching assessment or course evaluation.

For various reasons, then, we felt that articles in the existing literature did not meet the criteria we specified for a method of assessing the level of reflective thinking. We, therefore, felt it was justifiable to proceed with the work of developing an appropriate

scheme which did meet the five criteria. The method built upon earlier work by Wong *et al.* (1995), but used a different theoretical framework for the coding categories.

Mezirow's writing on reflective thinking

Developing a coding scheme to assess the level of reflective thinking requires some framework for defining levels of reflective thinking. The outcomes of another attempt to develop a method of assessing reflection (Sumsion and Fleet 1996) suggests that the categories should be very well defined. The study achieved low levels of inter-coder reliability, but in the paper categories were defined only by labels such as 'highly reflective' or 'more-than-moderately reflective'. That such descriptors proved insufficient for reliable assessment led us to develop much more extensive descriptors which we include here.

In this study we chose to base our scheme upon the work of Jack Mezirow, who has written extensively on the subject of reflective thinking as an essential component of his model of transformative learning for adults. The coding scheme described in this paper is principally derived from Mezirow (1991), of which chapter 4 is most central to defining reflective thinking. Other works by Mezirow (1977, 1985, 1992) were also perused to clarify the meaning of important constructs.

The fact that Mezirow has written so extensively and in such detail makes his work well suited to the formation of a coding scheme. To use any classification system consistently, even an individual classifier needs a clear well-defined definition of each category which provides sufficient detail to encompass the range of material to be classified. Where there are multiple judges or assessors, the need for adequate definition becomes even more important. Individual coders can supplement the formal category definitions with their own working guidelines. Multiple coders, however, inevitably derive a variety of interpretations and idiosyncratic operational definitions when the provided category definitions lack precision and detail.

We noted an earlier Mezirow work (1981) which divides reflective thinking into no less than seven levels. However, we felt that this presented too fine a sub-division of levels to produce a usable and reliable coding scheme, despite it being used by Powell (1989). As Mezirow makes no reference to this categorization scheme in his later and more substantial work (1991), we assume that his own thinking has developed beyond the earlier model.

Non-reflective action

Mezirow separates reflective action from non-reflective action. While being more interested in reflective thinking, it is still important to clearly explain and define what is not reflection so that all content can be accurately assessed. Three types of non-reflective action are distinguished; habitual action, thoughtful action and introspection.

Habitual action

Habitual action is that which has been learnt before and through frequent use becomes an activity which is performed automatically or with little conscious thought. Examples are riding a bicycle or using a keyboard. Habitual actions are not recorded in journals,

or if they were they would cease to be so classified. This activity, then, is of little concern in categorizing journal writing.

Thoughtful action

Thoughtful action makes use of existing knowledge, without attempting to appraise that knowledge, so learning remains within pre-existing meaning schemes and perspectives. Thoughtful action can be described as a cognitive process. The Bloom's taxonomy (1979) categories of knowledge, comprehension, application, analysis and synthesis would normally be placed in Mezirow's thoughtful action category. Much of the 'book learning' which takes place in universities is best classified as thoughtful action.

Thoughtful action differs from habitual action in that the latter does not require us to think about the action while performing it. When we ride a bicycle we do not normally think about the technique or mechanics of riding.

For professionals in their everyday practice, much of their work can become fairly routine, in which case they tend not to reflect upon their actions. Their normal mode of operations becomes thoughtful actions in Mezirow's terminology or 'knowing-in-practice' according to Schön's work.

As practice becomes more repetitive and routine, and as knowing-in-practice becomes increasingly tacit and spontaneous, the practitioner may miss important opportunities to think about what he is doing. Through reflection, he can surface and criticize the tacit understandings that have grown up around the repetitive experiences of a specialized practice, and can make new sense of the situations of uncertainty or uniqueness which he may allow himself to experience. Practitioners do reflect on their knowing-in-practice. (Schön, 1983: 61–62)

It is only when cases occur which do not fit within the normally experienced framework that reflection becomes necessary.

Introspection

Unlike thoughtful action, which is concerned with cognition, introspection lies in the affective domain. It refers to feelings or thoughts about ourselves. The feelings can be personal, such as recognizing that we feel happy, upset or bored with something. It can involve the recognition that we have feelings towards others such as liking or disliking them. It does not, however, encompass us deciding how or why these feelings developed as that becomes reflective thinking. Introspection remains at the level of recognition or awareness of these feelings.

Mezirow (1991: 107) regarded introspection as not reflective because it involves no attempt to re-examine or test the validity of prior knowledge. It is worth noting that, as a result of this, our coding scheme, based on Mezirow's work, does not particularly address the affective domain of learning.

We noted the affective responses of the students during the process of reflective learning and, as part of our major study, we have analysed the affective dimension involved in reflective learning. Similar to the observations by Boud and Walker (1985), we found that triggering reflective learning usually involves a sense of discomfort. Boud and Walker explain that these emotional barriers may inhibit reflective learning if the frustrations are not acknowledged and addressed. The role of the supportive faculty is

important in facilitating students in making their paradigm shift, turning their emotional response into a positive learning experience. Recognition of the association of these emotional responses to reflective learning is necessary, but the responses do not normally have a direct relationship with the stage of reflection reached and of coding of the level. Our own findings on the affective dimensions of reflection are a study in their own right and will be reported elsewhere.

Reflective action

The most important aim of any coding scheme concerned with reflection would be that of distinguishing reflective thinking or writing from that which is non-reflective. Having considered some definitions of categories of non-reflective thinking, it is worthwhile, therefore, examining a range of definitions of reflection in general before moving to Mezirow's categories of reflective thinking.

What is perhaps surprising, in spite of the cult status of reflection and the volumes written about it, is that the concept is ill defined. Formal definitions are hard to find as has been observed by Atkins and Murphy (1993) and Sparks-Langer *et al.* (1990) among others. Many write about reflection with the apparent assumption that everyone knows what it is. However, the disparities in terminology, frames of reference, applications and usage make it clear that this assumption is not helpful. That the construct has now become so widely used means that it is applied within different philosophical frameworks and in diverse contexts. Even where attempts are made to define or explain the term, these may not correspond with use within another orientation.

Dewey is normally considered to be the originator of the concept of reflection as an aspect of learning and education. His definition (1933: 9) has been widely quoted.

active, persistent and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and the further conclusion to which it tends.

Mezirow interprets Dewey's definition as implying that 'reflection means validity testing' (Mezirow, 1991: 101). Two further definitions which are consistent with Dewey's are given by Boud and Walker (1985) and Boyd and Fales (1983).

Reflection in the context of learning is a generic term for those intellectual and affective activities in which individuals engage to explore their experiences in order to lead to new understandings and appreciations. (Boud and Walker 1985: 19)

Reflective learning is the process of internally examining and exploring an issue of concern, triggered by an experience, which creates and clarifies meaning in terms of self, and which results in a changed conceptual perspective (Boyd and Fales, 1983: 100)

The latter definition is of particular relevance to professional practice in that it views experience as the touchstone for reflection. This definition is moving more towards the framework of Schön (1983) and the reflective practitioner within the context of professional practice. When Mezirow himself considers reflection, the influence of critical theory upon his work becomes apparent. Mezirow defines reflection as:

Reflection involves the critique of assumptions about the content or process of problem solving.... The critique of premises or presuppositions pertains to

problem posing as distinct from problem solving. Problem posing involves making a taken-for-granted situation problematic, raising questions regarding its validity. (Mezirow 1991: 105)

Mezirow then proceeds to subdivide reflective thinking into three categories of content, process and premise reflection.

Content reflection

We interpreted content and process reflection as being equivalent in level. The two are distinguishable in terms of the subject matter of the reflection. Content reflection being concerned with what, while process examines how.

Mezirow defines content reflection as:

Reflection on what we perceive, think, feel or act upon. (Mezirow 1991: 107)

Process reflection

Process reflection is concerned more with our method or manner in which we think.

Examination of how one perform the functions of perceiving, thinking, feeling, or acting and an assessment of efficacy in performing them. (Mezirow 1991: 107–108)

Premise reflection

Premise reflection is seen as a higher level of reflective thinking as it is through premise reflection that we can transform our meaning framework as it opens the possibility of perspective transformation. In terms of definitions similar to those above, Mezirow views premise reflection as: Premise reflection involves us becoming aware of why we perceive, think, feel or act as we do. (1991: 108)

It is the category of premise reflection which borrows most from the foundation of Mezirow's work on critical theory (Mezirow 1981) and the writing of Habermas (e.g. 1970, 1972, 1974). To undergo a perspective transformation it is necessary to recognise that many of our actions are governed by a set of beliefs and values which have been almost unconsciously assimilated from the particular environment. Premise reflection then requires a critical review of presuppositions from conscious and unconscious prior learning and their consequences. In this discussion we have put aside the political connotations of perspective transformation which are present in Mezirow's and Habermas's work. These are not useful to defining our categories for use in an academic and professional education context.

Conventional wisdom and ingrained assumptions are hard to change, in part because they become so deeply embedded that we become unaware that they are assumptions or even that they exist. Mezirow clearly recognizes the difficulty of perspective transformation.

It must involve a hiatus in which a problem becomes redefined so that action may be redirected. (1991: 110)

We, therefore, decided that premise reflection was unlikely to be observed frequently within student journals. For writing to be coded as premise reflection, there needed to be evidence of a significant change of perspective. This would be particularly true of topics which are central to our main activities as these have the greatest store of, and the most deep-seated beliefs. Perspective transformation would be easier if the subject is more peripheral to the main interest and activity of the person.

Comparison

Mezirow takes a passage by Schön (1983: 50) and analyses it according to his three categories of content, process and premise reflection. The segments within the square brackets in the quotation below are Mezirow's (1991: 112–113) additions to Schön's original quotation.

They may ask themselves, for example, 'What features do I notice when I recognize this thing? [process reflection] What are the criteria by which I make this judgement? [premise reflection] What procedures am I enacting when I perform this skill? [process reflection] How am I framing the problem that I am trying to solve?' [premise reflection] Usually reflection on knowing-in-action goes together with reflection on the stuff at hand [content reflection].

This quotation serves to illustrate the differences between the three categories. It also implies that Mezirow feels that it is appropriate to interpret elements of written work by classifying them according to the type of reflective thinking.

Coding categories

The categories described above translate into the coding scheme in Figure 1. The level of reflective thinking increases from bottom to top. Categories shown on the same horizontal level are regarded as being equivalent in the level of reflective thinking. The shaded lower levels indicate non-reflective action while the upper unshaded ones are reflective.

A category for both content and process reflection was included as early trials of the coding scheme revealed journal entries displaying both types of reflective thinking. A further category (0) was used for elements of writing which did not fit into any of the defined categories.

In making use of this coding scheme, in the tests described below, the definitions were taken from the above interpretations of Mezirow's writing. The categories, and more particularly their descriptors, went through several stages of development. With each refinement, the work of Mezirow was re-examined for consistency and further illumination. Developmental stages were pursued until independent coders felt the guidelines provided sufficient detail and ambiguity was reduced to a minimum.

Initial test

Before proceeding to discuss the actual testing of the coding scheme, we would like to provide a brief account of the collection of journals from students for coding. The students were from four undergraduate degree programs in nursing, occupational

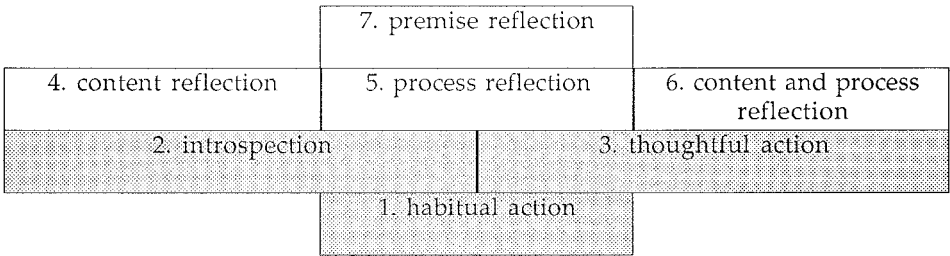


Figure 1. The coding categories for reflective thinking

therapy, physiotherapy and radiotherapy. The students were informed that the teaching team was trying out a new curriculum endeavour, which aimed to promote reflective learning among students in the health care disciplines. Alongside the curriculum innovation, the teaching team would undertake research to monitor and evaluate the effects of the new approach. The students were asked whether their journals could be used for analysis.

Initial journals were not graded for assessment. These journals were mainly used as documentation of learning progress, so that students themselves could review their own learning, and share the learning with peers and teachers. The final paper submitted at the end of the semester was graded. The grading criteria were based on the student’s progress, their ability to critically analyse the issues and their fulfilment of general academic expectations such as clear expression of thoughts and correct referencing of material.

The students were reassured that the grades they attained would not be associated with the results of coding. This association was in fact not possible since the grades were submitted at the end of the academic year, before the research team could arrive at the final decision on the coding. The students were also assured that confidentiality and anonymity of individual identity would be protected. Participation in the study was voluntary. Students had the right to ask not to be included in the study, and their journals would then not be included in the analysis. These students would not be in anyway be penalised. All students agreed to contribute their journals to the research.

In five initial rounds of the test exercise, which are not reported here, the discussion focused largely upon interpretation of the categories and their definition. After each test the definition was modified in the light of the discussion. For the final set of descriptors, described above, there was no disagreement or discussion among the coders of the category definitions. This was interpreted as implying that the definitions had been developed as far as possible. During these initial rounds the procedures for independent judging were also developed. The need to identify text blocks before the coding exercise was found to be important.

Once the coding scheme and procedures were finalized, the next step was to examine the reliability, ease and usefulness of the coding scheme reported above. Representative sections were taken from the journals of three first year students enrolled in an undergraduate health care course. The purpose of using journal writing was for students to reflect upon clinical visits and class activities in relation to the theory taught on the course and the students’ learning. The students were asked to make journal entries of about one page in length at relevant points of the course. Each selected journal entry was divided into a number of text segments corresponding to a coherent idea. The text segments were normally two to three paragraphs in length. Eight judges were then

asked to independently code each of the defined text segments in the three journal entries. The judges were academics from the Faculty in which the course was taught, though only one actually taught the course from which the journal entries were taken and was familiar with the students. The judges used the coding scheme in figure 1 together with the above text to define the coding categories. The coding sheets had a space for comments for each element to be coded. This facility was found to be useful as a means of developing clear and consistent guidelines in the various developmental stages.

After each judge had coded the journal sections independently the results were tabulated. The coders then set down as a group to discuss the coded results.

Reliability of coding

Cronbach alpha was computed as a measurement of the internal consistency of raters in the coding exercises. It is the most widely used and most general form of internal consistency estimate, representing the mean reliability coefficient one would obtain from all possible split-halves (Murphy and Davidshofer 1991). It can be used as an index of internal consistency without the need of splitting the assessment in half for scoring purposes. The higher the internal consistency, the greater confidence one may have on the scale of assessing reflective thinking. Cronbach alpha ranges in value from 0 to 1 (Norusis 1994: 147). When applied to the estimation of inter-rater consistency, the value of alpha is a function of the level of agreement among the raters. Murphy and Davidshofer (1991) suggested that an alpha value of 0.70 is a typical value in estimating the reliability of rating scales. For the initial coding using the seven levels in figure 1, the Cronbach alpha value was computed and found to be 0.65.

Differing interpretations of student writing

The discussion which took place after the coding exercise revealed that the differences arose from the coders' interpretation of the meaning or significance of what the students had written in their journals. It was not a function of ambiguity or lack of precision in the coding categories or their definitions.

The point is best illustrated by quoting two sections of one of the test journals and considering this against both the range of coding and the reasons given for the selected categories by judges in the subsequent discussion. The journal entries given below were made by a first year student after a clinical visit to a special school for multiple handicapped children:

Before the visit, I had already heard about this school because my elder sister is a social worker who had visited this school before. She said that these pupils are lovely and glad to have other people to visit them. So, I expected that they would all be glad to see us.

However, what I met was only part of the response that I expected. I heard that some pupils actually don't like others to visit them. One of the pupils said, 'That's good, you have finished seeing monkeys in the zoo!' 'Oh, no!' Actually, I think, none of my classmates think of them as a monkey!

But after I heard what he said, I feel guilty! Though this visit is a valuable chance for us to explore more real settings and situations, I still think, if we hadn't visited them, they would not get this unpleasant feeling!

From this experience, I found that although they have different kind of disabilities, they still retain a high self-esteem and self-image. They don't want others to watch them as if they were creatures! They know what was happening. They have their feelings!

For this text unit there was a division between judges who rated this as introspection and those who felt it should be recorded as content reflection. The former saw this entry as reporting an emotional response and that there was insufficient evidence that the student's position had changed to rate this as reflective writing. The latter group were more generous in their interpretation of what the visit meant to the student. Typical comments were:

The student learnt to view disabled people in a different light after the visit.

The experience enabled the writer to reach a deeper understanding/appreciation about the psychological needs of disabled persons.

A similar division occurred over the following passage in the student's journal:

Besides, during the fifteen minutes recess, I understood that pupils love each other. In a classroom, I saw a hand-injury pupil try to help his classmate to move his wheelchair and go to the toilet. This impressed me most, but I felt ashamed. Because in a normal school, people show even less concern and love for each other. Most of us are only concerned about our own studies and public examination, but neglect concern for others.

Again there was a faction which took a dubious view of the entry and coded it as introspection. One coder suggested that the student was merely 'explaining why s/he felt guilty'. Others were more generous in their interpretation of the student, so took the statement more at face value, and rated this as content reflection.

In an attempt to clarify the intentions of the student in this piece of journal writing, the student was interviewed by one of the research team a few weeks after the journal was written. The interview focused upon amplification and clarification of the written comments. It should be noted that this would not be seen as a normal accompaniment to the use of the reflective thinking category scheme in a practical application. In the initial development and testing of the process the interviews did turn out to be instructive. The most relevant sections of the interview are quoted below.

When you wrote this journal, did you have any special feeling? I mean when you were writing this paragraph.

Special feelings? I really think that if we did not visit the residential school, we may not know the truth that the child had such negative feeling. It may exist among other children who fear to express their feeling. I found that they did not like special ways of looking at them.

You used the word 'ashamed'. Why did you use such a word?

I think even these disabled children know how to take care of each other. However, I as a normal person, seldom have such experiences except from best friends. I used 'ashamed' because I think that our behaviour was worse than those

disabled children. ... the relationship among normal school children was worse than that of the disabled. They were not willing to help the others since there were no rewards for such behaviour. It included my classmates and I. I wrote this as a comparison between the children of the residential school and my past experience.

There is still an element of uncertainty, but the interview suggests that the writer was expressing an emotional reaction, which should be classified as introspection. There does not appear to be any strong evidence from the interview to substantiate the thoughts being classified as reflective thinking.

In practical applications of the scheme it may be possible that the assessor will have more evidence available to assist in the interpretation of the written work. Debriefing sessions in which students disclose and discuss their journal entries are normal features of courses designed to promote reflective thinking. It is also common to base tutorial discussions upon elements of written reflective journals.

Practical test

The next step was to move on to test the scheme in a direct practical application of how we envisage it being used in practice. Another course in the same Faculty required students to write a reflective paper derived from the reflective journals kept during a period of clinical placement. The paper was meant to be a final reflection upon important aspects of their clinical experience. The students were asked to focus upon communication and interaction with patients which was an important aspect of the course. The students wrote between three and six sides for their papers.

Four assessors then used the above coding scheme to assess the papers as they would if the assessment were to contribute to the students' grades. As the main aim of the course was to develop a reflective approach in the clinical practice setting and the paper was a written reflection upon the period of practice, this was considered to be an authentic method of assessment.

Each of the four assessors independently graded the written reflective papers of nine students using the seven category scheme described above. The assessors assessed the papers as a whole rather than dividing them into text segments as in the previous exercise. The papers were categorized according to the highest level of reflection shown by the student. This process is in line with use of the scheme as a normal assessment approach consistent with grading a student's achievement, or in this case level of reflection.

Of the four assessors, one was the course coordinator who knew the students well and had been involved in the development of the coding scheme. The second was from the same department, but not teaching on the particular course. She had not been involved in the scheme's development. The third was from another department in the same Faculty, who had been involved in the final stages of developing the scheme. The final assessor was the research fellow for this project.

Overall, then, the assessors would have had less contact with the students than would normally be expected of those grading student assignments. With one exception, their focus had to be exclusively upon the written papers with no additional insights in interpreting what had been written. In spite of this situation, the four assessors reaches a level of agreement indicated by a Cronbach alpha value of 0.74. This exceeds the alpha value of 0.70 suggested by Murphy and Davidshofer (1991) as acceptable in estimating the reliability of rating scales.

Conclusion

The level of agreement is considered reasonable in terms of the normal use of statistical tests for reliability. We consider that the initial, and more particularly, the practical test indicate a sufficient level of inter-judge agreement for the scheme to be considered a reliable method for determining whether reflective thinking takes place and the level of reflective thinking.

Those not convinced that the level of agreement is sufficient should take into account the fact that the levels of consistency between markers is often not as high as expected. Rowntree (1977) presents an extensive review of evidence of inconsistency between assignment markers dating back to Edgeworth in 1890. It is in more authentic forms of assessment, such as that attempted here, where it is usually most difficult to obtain good agreement.

What should also be taken into account is that, in both the tests, only one of the coders was familiar with the students and taught on the course. The remaining coders had limited, and in some cases negligible, appreciation of the student background and the context in which the journals were written. If the method were to be used for assessment or evaluation purposes it is more likely to be used by those teaching the target course so the assessors will be better placed to interpret the journal entries. It is our expectation that the level of reliability reached in these research exercises, with coders not familiar with the context, may well be exceeded if the method is used in normal teaching.

The coding schemes are based upon one of very few detailed discussions of the meaning and nature of reflective thinking (Mezirow 1991). This framework establishes the validity of the coding categories. The testing process shows that the category definitions are unambiguous and usable in practice for determining the levels of reflective thinking from journal writing.

We feel that the scheme can be recommended as a tool in assessing students enrolled in courses which have the aim of developing reflective practice and thinking. Many such courses will already require students to keep reflective journals so there would be no additional burden upon the students. The process is also of value in the evaluation of such courses. It provides a means of determining whether the various teaching and learning strategies for encouraging reflective thinking are succeeding. The scheme should be applicable to any professional development course which requires students to write reflective journals.

The justification for starting this work was that so many courses nowadays have the aim of developing reflective thinking. Without some method for assessing reflective thinking it is impossible to determine whether students have met the aims of the course. Without an evaluation tool for examining the levels of reflective thinking, it is hard to determine whether the course meets its objective in promoting reflective thinking.

We believe that we have developed an appropriate scheme for assessing the level of reflective thinking and carried out sufficient testing to report the details to others. Wider acceptance of the scheme obviously depends upon more extensive testing, preferably in other contexts and by those not involved in its development. This would show how readily interpretable the category descriptors are and how easy it is to apply them in authentic contexts. This further testing and development can only follow as a result of reporting our findings to date and inviting others to make use of the described process.

Acknowledgements

This research was supported by grants from the Action Learning Project, the Educational Development Fund of the Hong Kong Polytechnic University, the Research Committee of the Hong Kong Polytechnic University and the University Grants Committee of Hong Kong.

References

- ATKINS S. and MURPHY, K. (1993) Reflection: A review of the literature. *Journal of Advanced Nursing*, **18**, pp. 1188–1192.
- BLOOM, B. S. (ed.) (1979) *Taxonomy of Educational Objectives, Book I: Cognitive Domain*, (London: Longman).
- BOUD, D. and WALKER, D. (1985) Barriers to reflection on experience, in D. Boud, R. Keogh and D. Walker, *Reflection: Turning Experience into Learning*, (London: Kogan Page).
- BOYD, E. M. and FALES, A. W. (1983) Reflective learning: Key to learning from experience. *Journal of Humanistic Psychology*, **23**, (2,) pp.99–117.
- COLAIZZI, P. F. (1973) *Reflection and Research in Psychology: A Phenomenological Study of Learning*, (New York: Hunt Publications).
- DEWEY, J. (1933) *How we Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*, (Boston: D. C. Heath).
- EDGEWORTH, F. Y. (1890) The element of chance in competitive examinations. *Journal of the Royal Statistical Society*, **Vol. LIII**, pp. 400–475, 644–663.
- GAGNÉ, R. M. (1968) Learning hierarchies, in M. D. Merrill (ed.), *Instructional design: Readings* pp. 118–131, (Englewood Cliffs, NJ: Prentice-Hall).
- HABERMAS, J. (1970) Towards a theory of communicative competence. *Inquiry*, **13**.
- HABERMAS, J. (1972) *Knowledge and Human Interests*, tr. J. J. Shapiro. (London: Heinemann).
- HABERMAS, J. (1974) *Theory and Practice*, tr. J. Viertel. (London: Heinemann).
- HAHNEMANN, B. K. (1986) Journal writing: A key to promoting critical thinking in nursing students. *Journal of Nursing Education*, **25** (5), pp. 213–215.
- KING, P. M. and KITCHENER, K. S. (1994) *Developing Reflective Judgement: Understanding and Promoting Intellectual Growth and Critical Thinking in Adolescents and Adults*, (San Francisco: Jossey-Bass).
- MEZIROW, J. (1985) A critical theory of self-directed learning, in S. Brookfield (ed.), *Self-directed learning: From theory to practice*, pp.17–30, (San Francisco: Jossey-Bass).
- MEZIROW, J. (1977) Perspective transformation. *Studies in Adult Education*, **9** (2), pp. 153–164.
- MEZIROW, J. (1981) A critical theory of adult learning and education. *Adult Education*, **32** (1), pp. 3–24.
- MEZIROW, J. (1991) *Transformative Dimensions of Adult Learning*, (San Francisco: Jossey-Bass).
- MEZIROW, J. (1992) Transformation theory: Critique and confusion. *Adult Education Quarterly*, **42** (4), pp. 250–252.
- MURPHY, K. R. and DAVIDSHOFER, C. O. (1991) *Psychological Testing: Principle and Applications* (2nd ed.), (Englewood Cliffs, NJ: Prentice Hall).
- NORUSIS, M. J. (1994) *SPSS Advanced Statistics 6.1*, (Chicago, Illinois: SPSS).
- POWELL, J. H. (1989) The reflective practitioner in nursing. *Journal of Advanced Nursing*, **14**, pp. 824–832.
- ROWNTREE, D. (1977) *Assessing Students: How Shall we Know Them?* (London: Harper and Row).
- SCHÖN, D. A. (1983) *The Reflective Practitioner: How Professionals Think in Action*, (New York: Basic Books).
- SPARKS-LANGER, G. M., SIMMONS, J. M. and PASCH, M. (1990) Reflective pedagogical thinking: How can we promote it and measure it? *Journal of Teacher Education*, **41** (4), pp. 23–32.
- SUMSION, J. and FLEET, A. (1996) Reflection: can we assess it? Should we assess it? *Assessment and Evaluation in Higher Education*, **21** (2), pp.121–130.
- WONG, F. K. Y., KEMBER, D., CHUNG, L. Y. F. and YAN, L. (1995), Assessing the level of reflection from reflective journals. *Journal of Advanced Nursing*, **22** pp. 48–57