Development of a Questionnaire to Measure the Level of Reflective Thinking

David Kember , Doris Y. P. Leung , Alice Jones , Alice Yuen Loke , Jan McKay , Kit Sinclair , Harrison Tse , Celia Webb , Frances Kam Yuet Wong , Marian Wong & Ella Yeung

Development of a Questionnaire to Measure the Level of Reflective Thinking

DAVID KEMBER & DORIS Y. P. LEUNG WITH ALICE JONES, ALICE YUEN LOKE, JAN MCKAY, KIT SINCLAIR, HARRISON TSE, CELIA WEBB, FRANCES KAM YUET WONG, MARIAN WONG & ELLA YEUNG, *Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*

ABSTRACT   *Many courses aim to promote reflective thinking or reflection upon practice, but there is a scarcity of readily usable instruments to determine whether students engage in reflective thinking and, if so, to what extent. This paper reports the development and testing of such an instrument. To ensure validity, the constructs measured were derived from the extensive literature on reflective thinking, particularly the writing of Mezirow. A combination of the literature review and initial testing led to the development of a four-scale instrument measuring four constructs; habitual action, understanding, reflection and critical reflection. The final version of the instrument was tested with a sample of 303 students from eight classes of a health science faculty. The reliability of the scales was established by acceptable Cronbach alpha values. Confirmatory factor analysis showed a good fit to the proposed four-factor structure. Comparison of mean scores between the eight classes showed predicted significant differences on each of the four scales between undergraduate and postgraduate students.*

Introduction

It is now widely recognised that most of the work of professionals deals with issues or problems which have been variously described as ill-defined, wicked, messy, indeterminate or occupying the swampy lowland. The problems are not clearly identified, have multiple facets and do not have ideal solutions.

Schön (1987) observed that many professional education courses had not recognised the nature of professional practice so used a technical-rational approach which taught procedures for solving well-defined problems with unique solutions. He argued that a more appropriate model for professional education was equipping students to become reflective practitioners in order to deal with the multi-faceted problems.

The message has obviously been widely received as innumerable professional courses in many disciplines and countries now claim to be based upon a reflective practitioner approach. Alongside this growth of reflective curricula, there has developed a voluminous literature on reflection, reflective thinking, curricula and teaching for reflective practice and related topics.

What is surprising, given the number of courses which aim to produce reflective practitioners, is how little attention has been paid to methods for assessing whether students do engage in reflective thinking and if so to what extent. There are many papers that describe courses, which aim to promote reflection upon practice, but few which explain how to determine whether the aim has been met. We can only assume that in many cases the aim is not assessed and the curricula are not evaluated to see whether they are meeting their goal of developing reflective thinking.

As we have been involved in a major project to synthesise conclusions about curriculum design for promoting reflective thinking (Kember *et al*., 1996a, 1996b), we experienced a need for methods to determine whether students were being prompted to reflect upon their practice in the courses under study. Initial data was qualitative in nature, from reflective journals, student interviews and classroom observation. This provided valuable insights into the effects of various curriculum initiatives upon the levels of reflective thinking, but the data gathering and analysis required time commitments beyond that normally available for routine curriculum evaluation.

More useful to normal teaching situations were protocols we developed for assessing the level of reflective thinking in journal writing (Kember *et al*., 1999; Wong *et al*., 1995a). These provided procedures for teachers to assess the level of reflection students displaying in journal writing. The coding or assessment scheme in the former paper was based upon types and levels of reflective thing described in Mezirow's writing (1991). The method can be used to assess the level of reflective thinking achieved by individual students in their reflective writing. By aggregating results across a class it could also be used as a course evaluation procedure to determine whether the curriculum is promoting reflective practice.

A limitation of this approach is that it can only be utilised in courses that require reflective journal writing. It could not be used in a pre- and post-design to see whether there was a change in the level of reflective thinking. It requires judgement of level by someone who had become familiar with the category definitions based upon Mezirow's work.

To complement the procedure for assessing the level of reflection in journal writing, we thought it would be valuable to develop a questionnaire to measure the level of reflective thinking. To be most useful this would need to be reasonably short so that class time was not taken away for its completion. It should be easy to administer and analyse so that teachers could easily use it. The results should be readily interpretable and not require expert researchers to conduct tests and analyse results.

In view of the number of courses based on the reflective practitioner premise it seemed surprising that there did not appear to be any widely accepted questionnaire for determining whether reflective thinking takes place or assessing the level of reflective thinking. The closest were instruments that measured critical thinking. King and Kitchener (1994, p. 12) reviewed literature on two of the most popular of these instruments, the Cornell Critical Thinking Test (CCTT) and the Watson-Glaser Critical Thinking Appraisal (WGCTA). They concluded that the former examined the ability to solve well-structured problems and the latter a mixture of well- and ill-structured problems. Further, other instruments tended also to concentrate upon well-structured

problems so were of little utility for programmes, which adopted a reflective practice model because of recognition, that professional practice consists of dealing with ill-structured problems.

The Literature Base on Reflective Thinking

Having established a rationale for proceeding to develop an instrument to measure the level of reflective thinking, it is then necessary to establish a framework for the constructs to be incorporated. For a field with an extensive literature, this is clearly the first place to look.

In developing a protocol for assessing the level of reflection in journal writing, we found the work of Jack Mezirow provided a comprehensive, logical and, most importantly, workable framework for developing a method to assess reflective thinking. Mezirow has written extensively on the subject of reflective thinking as an essential component of his model of transformative learning for adults. The protocol we developed (Kember *et al.*, 1999) was 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.

Mezirow separates reflective action from non-reflective action. He identified three types of non-reflective actions: habitual action, thoughtful action and introspection. There were two levels of reflective action of which the lower or less critical level was sub-divided into content and process reflection. This terminology is taken from Dewey (1933) who used the term critical reflection to refer to deeper, more thoughtful and more profound reflection. Mezirow labels the more critical form of reflection premise reflection.

Explanations will be provided of the types of reflective and non-reflective thinking assessed in the instrument. However, first we must anticipate the outcomes of development work on the questionnaire to explain why the number of constructs in the eventual questionnaire was narrowed down to four. An initial version of the questionnaire contained six scales, but had poor psychometric properties. Items devised for particular scales loaded onto other scales because of the similarity or overlaps between the constructs.

To produce a more usable instrument, the sub-division between content and process reflection was dropped and the two seen as components of one reflective thinking scale. The introspection scale was excluded, partly on psychometric grounds and partly because it refers to the affective domain. We recognised that there was an affective dimension to developing reflective thinking (Boud & Walker, 1993; Wong *et al.*, 1995b) but felt that the instrument should concentrate upon assessing outcomes in terms of the level of reflective thinking displayed. Inclusion of a scale on introspection contributes little to this aim so can be left out on the grounds of parsimony. This left us with four constructs or scales that will now be described by drawing from Mezirow's work and substantiating definitions with work from other prominent writers.

Habitual Action

Habitual action is that which has been learnt before and through frequent use becomes an activity that is performed automatically or with little conscious thought. Common examples are using a keyboard or riding a bicycle. The work of experienced professionals dealing with normal cases or issues can become quite habitual. When they have

experienced a particular type of problem many times, their way of dealing with similar cases becomes quite routine. Schön (1983) called this type of behaviour *knowing-in-action*.

Understanding

Mezirow describes a type of thinking or learning that he calls thoughtful action. This 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. Much of the 'book learning' which takes place in universities is best classified as thoughtful action.

The Bloom's taxonomy (1979) categories of knowledge, comprehension, application, analysis and synthesis would normally be placed in Mezirow's thoughtful action category. Incorporating this breadth of categories into a questionnaire scale soon proved to be unworkable. A scale encompassing thoughtful action with broad attributes, in an initial trial version of a questionnaire, had very poor psychometric properties.

For a scale to be reliable, it needed to concentrate upon a narrower construct. For it to possess discrimination, the type of thinking needed to be distinguishable from the others incorporated in the questionnaire. For this reason the scale was narrowed down to focus upon understanding or comprehension, hence the title for the scale is 'understanding'. Bloom's (1979) definition of comprehension as "understanding without relating to other situations" captured the distinction we wished to make between an academic type of learning in which the student might reach an understanding of a concept without reflecting upon its significance in personal or practical situations.

Reflection

Dewey is normally considered to be the originator of the concept of reflective thinking as an aspect of learning and education. His definition (1933, p. 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. (p. 9)

Mezirow interprets Dewey's definition as implying that "reflection means validity testing" (Mezirow, 1991, p. 101).

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, p. 105)

Two further definitions which are consistent with Dewey's are given by Boud *et al.*, (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 *et al.*, 1985, p. 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 & Fales, 1983, p. 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.

Critical Reflection

Several writers also recognise a higher level of reflective thinking through which we can transform our meaning framework. Mezirow uses the term *premise reflection*.

Premise reflection involves us becoming aware of *why* we perceive, think, feel or act as we do. (1991, p. 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 that 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.

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 recognises the difficulty of perspective transformation.

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

We, therefore, decided that premise reflection was unlikely to be observed frequently, as understanding or reflective thinking since, it needed a significant change of perspective. This would be particularly true of topics that are central to principal 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.

Dewy (1933) also recognised a deeper level of reflection by distinguishing between *critical reflection* and less considered reflection. He argued that a person who was not sufficiently critical could reach a hasty conclusion without examining all the possibilities. The term *critical reflection* has more commonly been used for this more profound level of reflection so we have adopted it in naming the scale of our instrument.

Development and Testing

Having determined the identities of scales through the literature, the next step was to draw up items for the scales. The items were based upon two sources. The first was the literature review discussed above. The second was the research project on reflective

teaching. In this we analysed an extensive qualitative database of reflective journals and student interviews (Kember *et al.*, 1996a, 1996b).

This first trial version of the questionnaire was then tested with 350 students from the health science faculty of a university in Hong Kong. The properties of the questionnaire were examined by computing Cronbach alpha values for each scale to determine its reliability. Factor analysis was also performed to see whether items contributed to intended scales.

The results of the initial trial have been anticipated in the preceding section to provide the rationale for a questionnaire incorporating four scales. Following the initial trial, the questionnaire was revised by modifying items and incorporating new ones which seemed to more closely fit the emerging identity of the scales.

This process of trial and revision was repeated for three further cycles. At each stage scales were examined by reliability tests and confirmatory factor analysis. Items that did not contribute to a scale were either removed or modified. They were replaced by new items more in line with the emerging identity of the scales, reflected in the retained items.

It is normal for questionnaires to need a development process incorporating several trial and evaluation steps. In this case the major difficulty was that the scales were measuring qualities that could complement each other or be evident in the same student, or even used in the same task. We did not, therefore, benefit from the neat discrimination that appears in questionnaires incorporating dichotomous qualities.

Sample

Results from the final test of the questionnaire are given in detail. The final version of the questionnaire contains four scales. They are:

(1) Habitual action (HA),
(2) Understanding (U),
(3) Reflection (R),
(4) Critical reflection (CR).

Each of the four scales is measured by four items. The questionnaire is included as Appendix A. The version of the questionnaire completed by the students did not include the scale headings nor separate the items into the scale groupings. Items appeared in the order given by the item numbers.

The final version of the questionnaire was completed by 303 students from eight classes of the health science faculty of a university in Hong Kong. In each case the questionnaire was handed out in class and the students were asked to complete the questionnaire before leaving the class. The return rate depends on the number of students attending classes, and is generally at a comparable and acceptable rate. The returns and course information for the sample is given in Table 1.

Reliability

To determine the reliability of the scales in the instrument, Cronbach alpha values for each scale were computed and are given in Table 2. The values all reach acceptable levels indicating that the scales can be interpreted as reliable.

TABLE 1. Return rates of the questionnaires from each class

Discipline	Undergraduate or postgraduate	Year of study	Sample N	Return rate %
Occupational therapy	Undergraduate	2	47	94
Occupational therapy	Undergraduate	3	42	89
Physiotherapy	Undergraduate	2	47	47
Radiography	Undergraduate	1	60	85
Radiography	Undergraduate	2	40	80
Nursing	Undergraduate	2	29	83
Nursing	Postgraduate Diploma	1	14	88
Nursing	Masters	mixed	24	80

TABLE 2. Cronbach alpha values for the four scales

Scale	Cronbach alpha
Habitual action (HA)	0.621
Understanding (U)	0.757
Reflection (R)	0.631
Critical reflection (CR)	0.675

Structural Analysis

The next step was to show that the four items for each scale were measuring that scale and not contributing to others. The fit of the items to the intended scales was tested using confirmatory factor analysis. The scales were constructed with a four-factor model in mind so it was appropriate to test the fit to the hypothesised model, rather than use exploratory factor analysis, which is meant for data with no prior structure in mind. A single factor model was also tested to check if there was only one dimension for the items used (Schmitt, 1996).

The goodness of fits of the hypothesised four-factor and single-factor models were tested with confirmatory factor analysis using the EQS program (Bentler, 1995). Table 3 details the corresponding covariance matrix used in the analysis. The extent to which the model was a good fit to the data was measured by the model chi-squares statistic χ^2 with associated degree of freedom (*df*) and Bentler's comparative fit index (CFI). Models with small chi-squares value and CFI values greater than 0.9 are normally considered to indicate an acceptable fit (Bentler, 1990). The single-factor model fits the data poorly as indexed by a large χ^2 ($= 432.4$), $df = 104$, and a small CFI ($= 0.542$). On the other hand, the values obtained for the four-factor model were $\chi^2 = 179.3$, $df = 100$, and CFI $= 0.903$, so the questionnaire scales were judged to be acceptable indicators of the intended four constructs.

The path coefficient (link from a scale to an item) can be interpreted as a measure to describe how strongly the item is affected by its corresponding scale that is considered as a latent factor. The arrow between two latent variables denotes the correlation between these two variables. The correlation should have a value between -1 and 1. An out-of-range correlation implies a theoretical problem.

The standardised solution for the model tested is shown in Figure 1. All hypothesised paths are statistically significant at 5% level. Each item is a statistical significant

TABLE 3. The variance-covariance matrix used in the analysis

Items	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1.3															
2	0.04	0.85														
3	0.06	0.12	0.62													
4	0.06	0.19	0.11	1.1												
5	0.54	−0.00	−0.07	0.15	1.1											
6	−0.08	0.45	0.01	0.18	0.00	1.0										
7	−0.11	0.08	0.20	0.13	−0.03	0.10	0.67									
8	0.01	0.02	0.07	0.33	0.02	0.16	0.12	0.93								
9	0.24	−0.09	0.02	0.06	0.26	−0.12	−0.03	−0.09	1.3							
10	−0.01	0.35	0.00	0.20	0.04	0.43	0.09	0.21	−0.06	0.97						
11	−0.08	0.10	0.18	0.17	0.00	0.07	0.25	0.12	−0.12	0.10	0.73					
12	0.05	0.09	0.09	0.42	0.06	0.08	0.07	0.25	−0.04	0.03	0.03	0.94				
13	0.29	−0.00	−0.02	0.05	0.30	0.07	−0.03	0.03	0.35	0.08	−0.12	0.09	0.92			
14	−0.06	0.32	0.03	0.22	0.05	0.39	0.08	0.22	−0.05	0.45	0.13	0.19	0.02	0.78		
15	0.01	0.04	0.15	0.12	−0.02	0.05	0.20	−0.02	0.11	0.02	0.20	0.07	0.08	0.08	0.57	
16	0.11	0.23	0.09	0.31	0.03	0.11	0.06	0.38	0.11	0.17	0.18	0.33	0.09	0.16	0.08	0.96

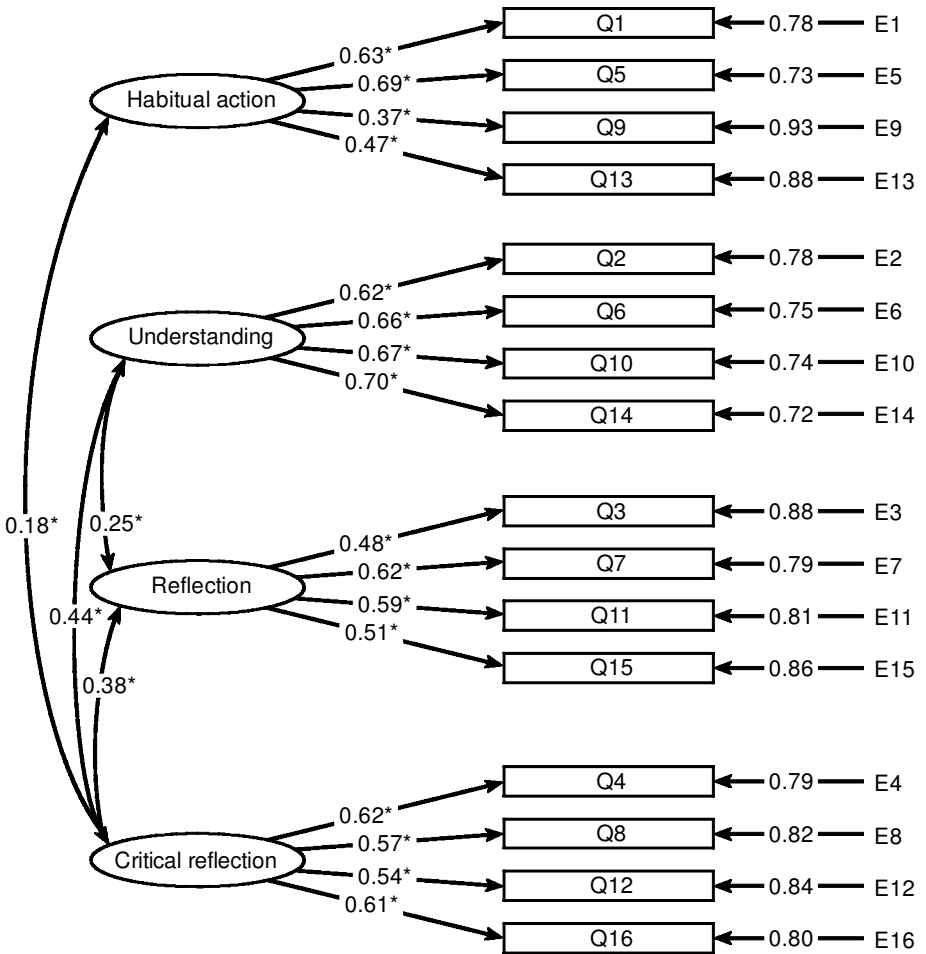*Note*: Only lower triangular non-duplicated elements are shown.

FIG. 1. Standardised parameter estimates of the structural model of reflective thinking.

Note: Variables in circles are latent constructs and variables in squares are observable measures. Paths with * are statistically significant at 5% level.

indicator for its hypothesised latent factor or scale. No item has a statistical significant loading on any other scale.

The model with the best fit showed inter-correlation between the scales or latent factors. This was predicted conceptually. Those who engage in critical reflection are also likely to have reflected upon their practice. Students who engage in either form of reflection may also have a tendency to study for understanding, particularly in more theoretical parts of a course, which have less obvious relationships to practice. The correlation between critical reflection and habitual action may be explained by the common mode of professional practice observed by Schön (1983). Experienced profes- sionals tend to deal with common cases in a routine or habitual way, often without thinking very deeply about them. Unusual problems or cases, though, do give rise to critical reflection as the practitioner considers the implications of the case for future practice.

TABLE 4. Sample means and standard deviations for the four scales

Discipline	Undergraduate or postgraduate	Year of study	HA Mean (SD)	U Mean (SD)	R Mean (SD)	CR Mean (SD)
Occupational therapy	Undergraduate	2	10.62 (2.46)	16.68 (2.41)	14.57 (2.34)	12.57 (2.60)
Occupational therapy	Undergraduate	3	9.55 (2.79)	13.14 (3.32)	15.76 (2.34)	11.14 (3.33)
Physiotherapy	Undergraduate	2	11.55 (2.89)	16.02 (2.50)	14.77 (2.14)	12.74 (2.14)
Radiography	Undergraduate	1	10.78 (2.29)	16.72 (2.03)	14.57 (1.89)	13.10 (2.69)
Radiography	Undergraduate	2	10.80 (3.02)	16.05 (2.30)	15.10 (1.71)	12.63 (2.57)
Nursing	Undergraduate	2	12.17 (3.26)	14.86 (3.68)	15.72 (1.91)	12.79 (2.96)
Nursing	Postgraduate					
Nursing	Diploma	1	8.93 (3.10)	17.00 (2.08)	16.64 (1.86)	14.57 (2.98)
Nursing	Masters	mix	8.58 (2.83)	17.00 (2.77)	17.21 (1.93)	13.50 (3.01)
OVERALL			10.58 (2.91)	15.88 (2.90)	15.25 (2.21)	12.70 (2.82)

Note: HA = habitual action; U = understanding; R = reflection; CR = critical reflection.

TABLE 5. Pair-wise comparisons for mean score
differences of the four scales

Pair	Mean difference	p-value
HA-U	− 5.30	0.000
HA-R	− 4.67	0.000
CR-U	− 3.19	0.000
CR-R	− 2.55	0.000

Note: HA = habitual action; U = under-
standing; R = reflection; CR = critical
reflection.

The testing process has shown that, for each of the four scales, the four items together
form a reliable scale. The items make significant contributions to no scale other than the
intended ones. The inter-correlations between scales are acceptable, considering the
constructs do have relationships.

Group Comparison

The next step was to examine practical applications of the instrument by computing
individual student scores on each scale and from the class means. A student's score on
each scale is computed simply by adding the response score for each of the four items.
Strongly agree was scored as 5, through to strongly disagree as 1. Hence, the scores for
the four scales could range from 4 (strongly disagree) to 20 (strongly agree). A mean
score was then computed for each class group for each of the four scales. The resulting
mean scores are shown in Table 4 with sample standard deviations in parentheses.

As expected the mean scores for habitual action and critical reflection are lower than
those for understanding and reflection. We accept that the values may not be directly
comparable but for the questionnaire to be accepted as valid these should be an
indication that the students in the sample were less inclined to employ habitual action
and critical reflection than understanding and reflection. Critical reflection requires a
major change of perspective and alteration to deep-seated beliefs which is a difficult,
lengthy and often painful process (e.g. Champagne *et al.*, 1985; Strike & Posner, 1985).
Habitual action would also not be common in university courses as there is insufficient
time for curricula to require students to repeatedly perform particular actions.

We performed a test to see whether the differences between the mean scale scores
were statistically significant. We used the pair-wise comparisons in repeated measure
technique to compute the differences between values of scores of each pair of two scales
for each student and test whether the average differs from zero. The results of the
corresponding mean differences and the associated *p*-values for the four pairs considered
in the paper are shown in Table 5. The results show that the lower scores for the habitual
action and critical reflection scales are statistically significant at 5% level.

Undergraduate and Postgraduate Comparison

Of particular interest in the table is the comparison between undergraduate and
postgraduate students. The undergraduate students in the sample were a fairly homo-
genous group in the sense that most entered university either directly upon completion

TABLE 6. Comparison between undergraduate and postgraduate

Undergraduate or postgraduate	Group size	HA Mean (SD)	U Mean (SD)	R Mean (SD)	CR Mean (SD)
Undergraduate	265	10.8 (2.82)	15.7 (2.92)	15.0 (2.14)	12.5 (2.76)
Postgraduate	38	8.7 (2.89)	17.0 (2.50)	17.0 (1.90)	13.89 (3.00)
t-statistic for mean difference (*p*-value)		4.36 (0.000)	−2.58 (0.010)	−5.45 (0.000)	−2.82 (0.011)

Note: HA = habitual action; U = understanding; R = reflection; CR = critical reflection.

of high school or within a year or two of doing so. Their professional experiences, therefore, would have been limited to that obtained on periods of professional practice incorporated within their degree programmes.

The two classes of postgraduate students were by contrast studying part-time while still practising their professional role as nurses. They would all have several years of professional experience before enrolling for their current course. As the questionnaires were completed towards the end of an academic year those enrolled in the postgraduate courses would have experienced nearly one year of the postgraduate diploma or two or more years of the Masters degree. Both courses encourage the nurses to reflect upon their professional practice as one of their main themes. The combined group, therefore, would have been expected to have higher scores on the reflection and critical reflection scales than the undergraduate students.

Table 6 compares the overall mean scores for undergraduate and postgraduate students for each of the four scales. In each cell the mean score is given, with the standard deviation in parentheses.

The *t*-tests show that in each case the difference in mean scores is statistically significant at 5% level. The postgraduate students are significantly less likely to engage in habitual action and significantly more likely to seek understanding or engage in reflection or critical reflection than the undergraduates.

Conclusion

The aim of this study was to develop a simple instrument that would examine the extent to which students engage in reflective thinking in professional preparation courses. The outcome has been an instrument with four scales or factors, each of which has four contributing items. The four scales or constructs measured are habitual action, understanding, reflection and critical reflection.

The psychometric properties of the instrument have been established by the use of confirmatory factor analysis. This showed that the 16 questionnaire items showed a good fit to the intended factor structure. The reliability of each scale was shown to be satisfactory by the use of Cronbach alpha. It is argued that the scales should be valid because they were derived from a well-established literature on the nature of reflective thinking. Further evidence for the reliability (and also the discrimination and utility) of the instrument comes from the way that the mean of each scale significantly distinguished between postgraduate and undergraduate students in a trial sample. The relative values of mean scores for the four scales were also in line with predictions about the likelihood of that type of thinking being present in the sample.

The instrument has been developed and tested on students from four disciplines, all in the health sciences. There seems to be no reason why it should not be suitable for other disciplines. The literature, from which the framework was derived, referred to reflective thinking as a generic construct rather than specific to particular disciplines. The wording of items excludes any terminology specific to particular disciplines or professions. The questionnaire is designed for use in academic programmes. It would require some modification if the intention was to measure the level of reflective thinking by professionals engaged in their professional practice. However, it is suitable for students enrolled on courses that include a professional practice component.

We believe the instrument will have value as a diagnostic tool in courses that aim to promote reflective thinking. A principal use is as a tool to investigate the effects of the teaching and learning environment on reflective thinking. To determine the effect of a course on reflection, the instrument may be used in a repeated measure design. Students are asked to complete the instrument shortly after the start of a course, reporting their normal study patterns prior to the course. Near the end of the course, and possibly at intermediate points for longer courses, the students are again asked to complete the questionnaire, this time reporting modes of thinking and learning used for the course. Any changes to reflective thinking can then be reasonably attributed to the course and its teaching and learning environment.

The instrument may also be usable for diagnosing study patterns of individual students. If a course aims to promote reflective thinking, it could be useful to see which students are meeting the goal. Study counselling could be given to those who are not.

Another use for the instrument would be examining the inter-relationships between scores on the scales with scores on scales measuring other constructs related to the teaching and learning environment. The rationale for this use is that reflective thinking is seen as, to some extent at least, responsive to that environment.

The questionnaire would also be suitable for comparing groups of students subjected to different conditions or treatments. An example of this type of use is given in this paper for the comparison of undergraduate and postgraduate students. If the questionnaire becomes widely used it may be useful to compare scores for particular cohorts to norms for those enrolled in similar courses. This approach should be used with caution, though, as many university classes now have heterogeneous enrolments. Establishing a norm for a particular type of course is considerably complicated by the presence of mature students, international students or those studying part-time.

Acknowledgements

Notes on Contributors

This paper is an outcome of a collaborative action research project that synthesised conclusions from five courses. David Kember is the Coordinator of the Action Learning Project, which is a collaborative project operating across the eight universities in Hong Kong. Doris Leung is a Research Fellow in the Educational Development Unit.

The remaining authors are academics within disciplines in the health sciences. Alice Jones is an Associate Professor in Physiotherapy. Alice Yuen Loke is an Associate Professor in Nursing. Jan McKay is an Associate Professor in Radiography. Kit Sinclairis an Assistant Professor in Occupational Therapy. Harrison Tse is a former Research Fellow for the project. Celia Webb is an Associate Professor in Physiotherapy. Frances Kam Yuet Wong is an Associate Professor in Nursing. Marion W. L. Wong is an Assistant Professor in Nursing and Ella Yeung is an Assistant Professor in Physiotherapy. *Correspondence:* Dr Doris Y. P. Leung, Educational Development Unit, Hong Kong Polytechnic University, Hunghom, Hong Kong. Tel: (852) 2766 6288. Fax: (852) 2334 1569. E-mail: etyleung@polyu.edu.hk

REFERENCES

BENTLER, P. M. (1990) Comparative fit indexes in structural models, *Psychological Bulletin,* 107(2), pp. 238–246.

BENTLER, P. M. (1995) *EQS Structural Equations Program Manual* (Encino, CA, Multivariate Software).

BLOOM, B. S. (1979) *Taxonomy of Educational Objectives, Book I: Cognitive domain* (London, Longman).

BOUD, D., KEOGH, R. & WALKER, D. (1985) *Reflection: turning experience into learning* (London, Kogan Page).

BOUD, D. & WALKER, D. (1993) Barriers to reflection on experience, in: D. BOUD, R. COHEN, & D. WALKER (Eds) *Using Experience for Learning* (Bristol, Open University Press).

BOYD, E. M. & FALES, A. W. (1983) Reflective learning: key to learning from experience, *Journal of Humanistic Psychology,* 23(2), pp. 99–117.

CHAMPAGNE, A. B., GUNSTONE, R. F. & KLOPFER, L. E. (1985) Effecting changes in cognitive structures among physics students, in: L. H. T. WEST & A. L. PINES (Eds) *Cognitive Structure and Conceptual Change* (New York, Academic Press).

DEWEY, J. (1933) *How We Think: a restatement of the relation of reflective thinking to the educative process* (Boston, D.C. Heath).

HABERMAS, J. (1970) Towards a theory of communicative competence, *Inquiry*, 13, whole issue.

HABERMAS, J. (1972) *Knowledge and Human Interests*, trans. J.J. SHAPIRO (London, Heinemann).

HABERMAS, J. (1974) *Theory and Practice*, trans. J. VIERTEL (London, Heinemann).

KEMBER, D., JONES, A., LOKE, A., MCKAY, J., SINCLAIR, K., TSE, H., WEBB, C., WONG, F., WONG, M., YAN, P. W. & YEUNG, E. (1996a) Developing curricula to encourage students to write reflective journals, *Educational Action Research*, 4(3), pp. 329–348.

KEMBER, D., JONES, A., LOKE, A., MCKAY, J., SINCLAIR, K., TSE, H., WEBB, C., WONG, F., WONG, M., YAN, P. W. & YEUNG, E. (1996b) Encouraging critical reflection through small group discussion of journal writing, *Innovations in Education and Training International*, 33(4), pp. 203–212.

KEMBER, D., JONES, A., LOKE, A., MCKAY, J., SINCLAIR, K., TSE, H., WEBB, C., WONG, F., WONG, M. & YEUNG, E. (1999) Determining the level of reflective thinking from students' written journals using a coding scheme based on the work of Mezirow, *International Journal of Lifelong Education*, 18(1), pp. 18–30**.**

KING, P. M. & KITCHENER, K. S. (1994) *Developing Reflective Judgement: understanding and promoting intellectual growth and critical thinking in adolescents and adults* (San Francisco, Jossey-Brass).

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. (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. (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.

SCHMITT, N. (1996) Uses and abuses of coefficient alpha, *Psychological Assessment*, 8(4), pp. 350–353.

SCHÖN, D. A. (1983) *The reflective practitioner: How professionals think in action* (New York, Basic Books).

SCHÖN, D. A. (1987) *Educating The Reflective Practitioner* (San Francisco, Jossey-Bass).

STRIKE, K. A. & POSNER, G. J. (1985) A conceptual change view of learning and understanding, in: L. H. T. WEST & A. L. PINES (Eds) *Cognitive Structure and Conceptual Change* (New York, Academic Press).

WONG, F. K. Y., KEMBER, D., CHUNG, L. Y. F. & YAN, L. (1995a) Assessing the level of reflection from reflective journals, *Journal of Advanced Nursing,* 22, pp. 48–57

WONG, M. W., WONG, F. K. Y. & LOKE, A. Y. (1995b) The affective dimension in reflection: the lived experience of reflective learners, *Asia-Pacific Human Science Research Conference Proceedings,* pp. 306–313 (Monash University, Australia).

Appendix A

Reflection Questionnaire

Please fill in the appropriate circle to indicate your level of agreement with statements about your actions and thinking in this course.

A—definitely agree
B—agree with reservation
C—only to be used if a definite answer is not possible
D—disagree with reservation
E—definitely disagree

Habitual Action

1. When I am working on some activities, I can do them without thinking about what I am doing.
5. In this course we do things so many times that I started doing them without thinking about it.
9. As long as I can remember handout material for examinations, I do not have to think too much.
13. If I follow what the lecturer says, I do not have to think too much on this course.

Understanding

2. This course requires us to understand concepts taught by the lecturer.
6. To pass this course you need to understand the content.
10. I need to understand the material taught by the teacher in order to perform practical tasks.
14. In this course you have to continually think about the material you are being taught.

Reflection

3. I sometimes question the way others do something and try to think of a better way.
7. I like to think over what I have been doing and consider alternative ways of doing it.
11. I often reflect on my actions to see whether I could have improved on what I did.
15. I often re-appraise my experience so I can learn from it and improve for my next performance.

Critical Reflection

4. As a result of this course I have changed the way I look at myself.
8. This course has challenged some of my firmly held ideas.
12. As a result of this course I have changed my normal way of doing things.
16. During this course I discovered faults in what I had previously believed to be right.