

## **Criterion-referenced Assessment: The Students' Views**

Cathy S.P. Wong  
Department of English  
The Hong Kong Polytechnic University  
Hung Hom  
Hong Kong  
egcathyw@polyu.edu.hk

&

Anna S.C. Cheung  
Department of Linguistics  
The University of Hong Kong  
Pokfulam Road  
Hong Kong  
anna.cheung@hku.hk

### **ABSTRACT**

*This paper reports on the findings of students' reaction to the implementation of criterion referenced assessment (CRA) in a university in Hong Kong. The university started to adopt the CRA model in 2005. Wong (2006) has reported that teachers in general are positive of adopting the SOLO Taxonomy as a CRA model. A follow-up study was conducted to investigate students' reaction to the CRA approach. Studies on CRA have shown the importance of students' involvement in a successful implementation (Cuffe & Jackson, 2006; O'Donovan, Price & Rust, 2000; Rust, Price & O'Donovan, 2003). Therefore, a study was conducted to gather students' feedback on the implementation of the CRA model.*

*18 First Year language majors volunteered to participate in focus group discussion sessions to share their views on the assessment methods. Students commented that CRA was a fairer method and that the given grades could reflect the level of knowledge they had acquired and the standard they had achieved of a certain subject. However, the students also pointed out a number of problems of CRA. First, there was inconsistency among subjects. Some subjects followed CRA while some did not. Second, students were not very clear about the assessment criteria or the grading mechanism. Third, they were concerned about the disparity among teachers who did use CRA. When asked to suggest ways to improve, most students suggested more transparency: for example, more detailed assessment criteria should be given to students beforehand; guidelines should be provided to help students cope with assignments; teachers should use the WebCT as a public forum for disclosing assessment criteria or posting sample assignments for students to have easy access.*

## I. INTRODUCTION

The criterion-referenced assessment model was officially adopted by the university in the fall semester of the academic year 2005/06. Wong (2006) has reported on the findings of the study which examines the teachers' opinions on the change from a norm-referenced assessment model to a criterion-referenced assessment model within the Department of English. As a follow-up study of Wong (2006), a similar study was conducted one year afterwards. One of the major aims of the follow-up 2007 study is to investigate whether students are receptive of the criterion-referenced assessment method and what the issues are of concern are from the students' stance.

Based on the results from the in-depth focus group discussion, this paper presents the results and discusses findings with regard to the adoption and implementation of the criterion-referenced assessment model in the Department of English.

## II. WHY "CRA" AND WHY "SOLO"?

In selecting a model of criterion-referenced assessment (CRA), the Department of English has adopted the Structured of Observed Learning Outcomes (SOLO) Taxonomy (Biggs 1995, 1999a, 1999b, Biggs and Collis 1982) for most of the subjects offered by the department. For one thing, the SOLO Taxonomy has been applied to a number of different subject areas (see review in Wong 2006). The other advantage is that the five stages of development are explicit and theoretically sound. Wong (2006) has reported that most teachers welcomed the introduction of CRA and found that the SOLO Taxonomy was useful in helping them to formulate assessment tasks and it was also meaningful to have a common platform for discussion of assessment tasks and assessment criteria across subjects.

The findings reported in Hodges and Harvey (2003) have also shown that the SOLO Taxonomy was found to be useful in identifying students' difficulties or confusion in learning in a chemistry faculty in the US. Two classes of around 20 to 30 students were being studied. The teachers used the SOLO Taxonomy to classify students' answers in an examination held after the first semester. They then identified the areas of difficulties or misunderstanding among students. Then in the second semester, more focus is given to these areas in teaching and subsequent assessments. As a result, students' performance improved. They concluded that the SOLO Taxonomy is very useful in help charting students' progress in learning.

Another study conducted in a Sweden university (Lundberg 2004) investigated the changes brought about by the implementation of the SOLO Taxonomy to a group

of Civil and Environmental Engineering students. The implementation process was very time-consuming but it led to some changes in the learning process. The students felt that they had learned more deep learning but the breadth of knowledge had been sacrificed. They also found that the SOLO Taxonomy was useful in designing the assessment tasks, a point which is echoed in the results reported in the study reported in Wong (2006).

While CRA seems to be welcomed by some colleagues in Hong Kong, the US, and Sweden, the CRA project reported in Jewels, Ford and Jones (2007) has presented some negative feedback on the implementation of CRA in an Australian university. Jewels et al (2007) reported on the procedure and details of how CRA was implemented and how an elaborate marking grid was developed. The feedback they received from the teachers was mostly negative. Teachers complained that it was very time-consuming to develop the procedure although the teachers agreed that more consistency and quality of marking has been achieved. In addition, on the one hand, they appreciated working together to develop the elaborate marking grid, the marking grid also called for complaints on the other. Some commented that the marking criteria confused the students. According to their study, students' feedback towards CRA was also negative because they thought the marking grid caused confusion. It also "stifled" creativity. Nor did it enhance learning.

### III. WHY STUDENTS' VIEWS?

Several studies (Cuffe & Jackson, 2006; O'Donovan, Price & Rust, 2000; Rust, Price & O'Donovan, 2003) on the implementation of CRA have shown students' involvement was very important in a successful implementation.

O'Donovan, Price and Rust (2000) surveyed students' views on their experience in CRA. The results clearly show that students often found that the criteria tended to be too subjective and they preferred more explicit assessment criteria and a more transparent process. In response to such a need, Rust, Price and O'Donovan (2003) conducted a follow-up study to develop students' understanding in the assessment criteria through a more transparent process. They found that students who had received training about the CRA process performed better than those who did not attend the training. However, they also argue that while explicitness, clarity and transparency of the assessment criteria and assessment process would contribute to better performance, a "socialization" process in which feedback and discussion between teachers and students is what leads to success in CRA.

Cuffe and Jackson (2006) discuss the pros and cons of CRA implemented in Queensland University of Technology in 2004/05 for all first year undergraduate

students. The study involved students pursuing a law degree in the Law School. According to Cuffe and Jackson, the implementation of CRA involves aligning assessment with teaching and learning so a CRA sheet was developed and used. Afterwards, a survey was conducted on 288 students, around 60% of all Year 1 law students. In spite of the fact that the overall feedback was quite positive, individual students commented on the inadequacies of the implementation of CRA. For example, some mentioned that more explanation was needed on the criteria; mock exercises using the assessment criteria; more personal feedback was needed, etc. As a result of these comments, students were then engaged more with the CRA sheets in the following academic year 2005/06. When the second survey was conducted, much more positive feedback was received. Therefore, Cuffe and Jackson (2006) concluded that students should be more involved in CRA.

Based on the findings summarized above, it is evident that students play an indispensable role in the successful implementation of CRA. Therefore, it is the aim of this study to explore the students' views on the implementation of CRA in the case of the Department of English as reported in Wong (2006).

#### IV. METHODOLOGY

The study was carried out between November 2006 and May 2007. A group of 18 Year 1 undergraduate students who majored in languages were invited to five sessions of focus group discussion. Each session was facilitated by one of the co-authors of this paper following a set of discussion questions developed by the two authors (see Appendix <i>). Four sessions were conducted in the students' preferred language, Cantonese. One of the sessions (Focus group 2) was conducted in English because one of the participants was a non-local student who did not speak Cantonese. Each session was audio-taped and each session was translated / transcribed in English immediately afterwards. The two authors listened to the audio tapes again to check the content of the translation / transcription for reliability.

After the content of the transcripts were double-checked, the students' responses were categorized according to the issues being investigated. The results and findings are presented and discussed in the next section.

#### V. FINDINGS AND DISCUSSION

##### (1) Preference for CRA

The major aim of the study is to find out whether students are receptive to CRA. Therefore, students' responses to the question "*Do you prefer the assessment method*

*in PolyU (your assignment is compared to a set of explicit assessment criteria and not with other students) or the one that you had experienced in the public examination (every student is compared with others and the results are spread over a bell curve)?” (Q.7) are crucial.*

14 out of the 18 students gave positive comments to this question. They all said that they preferred the assessment method in the university rather than that of the public exams (the norm-referenced model).

The following are some of the reasons why students thought that CRA was a better way of assessment:

- *The level of the university is different from that of the secondary school. If the assessment methods are the same, then there will be no difference. The students have to achieve a certain level in order to get a certain grade, and so the requirements have to be higher (1D)<sup>1</sup>; the purpose of studying at the university is not for getting good grades (2C), but for future use (3B); students should look at what they have learned and not the grade they get. (3D).*
- *Fairer: the grades are given according the assessment criteria met (1A, 3B, 3D, 5B, 5C); you get a certain grade no matter how many people have got it. (2C, 3B).*
- *(CRA) can reflect a student’s knowledge and standard (3B, 3C, 4B).*
- *For every assignment, students know the criteria for each grade. All they need to do is try to meet the. (1D).*
- *The ability of the students should be assessed according to their performance rather than comparing with other students (2A).*
- *Students are evaluated as individuals and not compared with each other (2D).*
- *Feel like comparing with oneself rather than with the others, thus helping the students get improved (5A, 5C).*
- *The subjects are not based on numerical scores (5B).*
- *Should get the grade that the student deserves (4B).*
- *The grades given are quite reasonable (4B).*

The following are some of the criticism against the norm-referenced assessment model which is being used in most public examinations in Hong Kong:

- *The bell curve distribution method is unfair; the difference between two grades is just 1-2 marks (1A, 3B); only a certain amount of people can get top grades while a lot of people get the middle ones (2C); there is a quota*

---

<sup>1</sup> This indicates the identity of the student—“1D” refers to Student D in the 1<sup>st</sup> Focus Group. If a view is shared by more than one student, this is reflected by the entries indicated after each comment.

*for each grade (3B, 3C, 3D); it is possible for everyone to have the same ability or the same knowledge (2A); the actual knowledge level of the students can not be revealed (3B, 3C) and make them lose confidence in getting a better result. (3C) Examples: in a Chinese subject, a C+ is given to a student who has got nearly 80 marks (3C); students might get a low grade even if they get high scores (5B).*

- *Disadvantages of the bell curve distribution method: (students) do not know how to achieve the target grade (1B); the meaning of grades will not be clear if the bell curve distribution is used (3B); it makes no sense to differentiate students by this method (5A).*

## (2) Understanding of Assessment Criteria

The assessment criteria are crucial in the success of CRA. Therefore, the students were asked whether they were clear about the assessment criteria of their assessment tasks (Q.5) *“Do you think your understanding of the assessment criteria is the same as your teachers? If not, can you suggest some ways to improve the situation?”*

The feedback from the students reflected a glitch in the implementation.

First, the majority of the 18 students commented that different teachers had different practices. Here are their responses:

- *Depends on the subject or the teacher (1A, 1B, 1C, 1D, 2A, 2B, 2D, 3A, 3B, 3C, 3E, 4A, 5A, 5B, 5C).*
  - *Some teachers give very clear and detailed criteria, while others do not (3B).*
  - *Though EPC was taught by 5 teachers, the criteria were very clear. Perhaps the teachers have discussed them in advance. The students can see the difference between an A grade piece of work and a C grade one (5A).*
  - *Different teachers have different opinions, especially for term papers (3A).*
  - *Though teachers of different seminar groups have discussed the criteria in advance, their understanding might not be exactly the same though this does not often happen (3E).*
  - *Some teachers would not make the criteria clear (1C) and some even would not mention at all (1C, 4A). For example: the difference between a good and a poor piece of work. Therefore, students do not know (sometimes) how to produce a good piece of work (4A).*
  - *Some teachers gave brief explanations about the criteria even when the*

- students asked them (1C).*
- *Some subjects require the students to do a lot of assignments, but the teacher did not inform them clearly about the weightings and also the subsequent changes (1A).*
  - *Some teachers did not provide an additional guideline for doing the tasks (2A)*
  - *Some teachers did not even tell the students about the scope of the tests and made them feel anxious (3B).*
  - *Some teachers seem to use the bell curve distribution method even though they said it was not the case because the students are not sure about the criteria of the grades. The difference between two grades was too little, only 1-2 marks.*
  - *Essays are marked according to the subjectivity of the teachers. They misled the students into doing those topics that they are familiar with, thereby reducing students' freedom to choose the topics they like. (1B, 1D).*
  - *The assessment criteria of a term paper on translation are subjective (to a certain extent) (3A, 5C) as it really depends on personal judgement (5C).*
  - *The criteria of the English subjects are clearer than the Chinese ones. For example: there are no clear instructions for doing the task of comparing the Chinese variant characters (2A, 2B); students have no idea about the assessment method in a Chinese subject (2B, 2D), and thus feel annoyed (2B) and frightened (2D).*
  - *Chinese subjects use tests as the means of assessment a lot more than the other methods (5A).*

Second, many students said that the assessment criteria were vague.

- *Descriptions of the criteria are ambiguous / vague / unclear (1A, 1C, 1D, 2D, 3A, 3B, 3D, 5A, 5B, 5C).*
  - *Examples: the meaning of 'a good understating of this subject' should be clarified (1A). A+ means very good. But it is difficult to define the meaning of 'very good' in a translation task (5A); never know whether they have met the criteria set by the teacher or not (5C).*
  - *Some teachers would not make the criteria clear and some even would not mention them at all. Only brief answers were given when asked (1C).*
  - *Though it is difficult for this programme to have a clear set of criteria, there should be at least model answers for tests (3D).*



- *Teachers are not used to tell students about the criteria for tests. Therefore, students do not know what they think (3A, 3B, 5A), especially when they do not know him/her well (3A).*
- *Students have no clear guidelines for writing papers (3B).*
- *The criteria of a Chinese subject are written in English (5A).*
- *Perhaps the criteria are set according to the performance of the students (5B).*

It is interesting but unsettling to learn that one of the students made the following comment:

- *The criteria have to be unclear. Besides, teachers should not tell the students about their criteria. Otherwise, everyone will get an A. A general picture of what the students should do is enough (2D).*

This reflects that some students did not really want everyone to perform well, which indicates that they have been “brainwashed” by the norm-referenced assessment method in which only a small fraction of the student population should be awarded high grades while the majority should not.

With reference to the understanding of the assessment criteria, the students were also asked to discuss whether they thought their understanding of the assessment criteria was the same as their teachers’. 9 students responded by saying “No.” while only 4 students said “Yes, basically the same”. Below are comments from students who believed that they did not share the same understanding as their teachers’:

- *The criteria are not clear (3E, 5A) and thus can not get the meaning of the teacher (3E).*
- *One of the assessment criteria was on the number of words used which increases with each higher grade. It gave the students an impression that their grade depends on this criterion. After the first assignment, the teacher told them that s/he had lowered the number of words required. But the students still wrote a few thousand words more than the minimum requirement when they submitted the second assignment. They regarded the number of words used as the first criteria because it was put before the other criteria (3E). The teacher should tell the students the word range of different grades (3B, 3E); changes are needed if the teacher really uses this mechanism to grade the students (3C).*
- *The number of words allowed to use in a term paper has restricted the ideas that the students can elaborate (3B).*
- *Students would believe all the hearsays about the teacher (1C).*
- *Sometimes, a student thinks that his/her presentation is quite good, but for some reason, s/he has to present at a later time, and the teacher might not*

*be interested in the topic that was already presented by the other groups. This would affect the grade (1B).*

- *Teachers are subjective in marking essays (1D); the assessment criteria of a term paper on translation are subjective (to a certain extent) (3A, 5C) as it really depends on personal judgement (5C).*

### (3) Suggestions for Improvement

Although the overall feedback from the students is quite positive, students have also pointed a number of problematic areas which need improvement. Below is a summary of the problems identified by students' and their suggestions for improvement. The problems that students have identified include: parity, assessment criteria, and communication between teachers and students.

#### Parity

The major concern for many students is parity among different groups. For example, they commented:

- *(The assessment is) not fair sometimes. For example: a subject is taught by different teachers / tutors. These people have very different assessment criteria even though they have discussed them in advance. The students in the lecturer's group had slightly lower grades than the tutors' groups (3E). Suggestion: to state the criteria clearly to the students because they do not know the grading differences between their teachers (3E); ask the teachers to discuss the criteria a few more times (3C).*
- *Teachers should stick to their criteria when giving grades. For example: tutors cannot give higher grades because they are lenient (3B); the lecturer should look at the criteria of the tutors if there are any grading differences / check whether the other groups are reasonably graded (3B, 3E). Suggestion: let other teachers have a look at their graded scripts so as to minimise subjectivity (3B); students should be allowed to sit in and compare with other groups (3E); both teachers and students should do something to make the assessment criteria fairer (3E). Suggestion: have more than one teacher teaching the same subject so that they can exchange some of the student scripts for grading. Students would thus feel safer as there will be a second teacher looking at their work, even though this would also increase the workload of the teachers (1C).*

#### Assessment criteria

In terms of assessment criteria, students made suggestions in three areas:

- (a) the right to know what the assessment criteria are
- *Students have the right to know the rules of the game (i.e. the criteria) (1C). Teachers should follow the rules of the game and allow students to have their say instead of unfairly forcing them to accept everything (1A).*
  - *Guidelines are essential for producing a good piece of work within a short period of time (5B).*
  - *Teachers should not tell the students to write anything they like because they do not understand what s/he wants. Guidelines should be given for free topics, like what the teacher wants the students to demonstrate in the task (3B).*
  - *For tasks that do not have a clear set of criteria, teachers should tell the students their perception of the grades. For example: students should not write in primary school level English (5A).*
  - *If the teacher has a set of criteria in mind, then it is already a guideline for the assignment or an outline for the essay (2B).*
  - *Teachers should post their criteria on WebCT or show them in the lecture rather than announcing them in the seminars, because they might give slightly different criteria to different groups (3D). For example: one of the teachers had mentioned different test scopes to different classes, which was not fair, because some students had studied something that was not tested. This would make the student hate the teacher as well as the subject (3C).*
- (b) clarity of the assessment criteria
- *More detailed criteria (1A, 1C, 1D) / clear guidelines (3D, 5A, 5B, 5C) / guidance (3E).*
  - *Clear guidelines or assessment criteria are essential for having nearly the same understanding as the teachers (3B, 3E).*
  - *Students will enjoy doing the task if a clear guideline is given (3B).*
  - *Clear criteria and guidelines should be given for writing papers so that students would have a goal to achieve or know the areas that they need to improve. (3B)*
  - *A clear revision guideline and assessment criteria are essential for tests because students are not clear about the scope for revision. For example: the content tested is more than that mentioned by the teacher (3C).*
  - *The criteria / guidance should be given at the beginning of the semester (1C, 3E). A clear guideline to write freely can be given near the end (3E).*
- (c) feedback on students' work
- *Teachers should encourage students to pay more attention to quality rather than quantity (3B, 3E). For example: Though the teacher explained that*

*the number of words used is not that important, but the student doubted whether the teacher would give her an A if the number of words she used just met the requirement for a C. Nothing has been learned by following this criterion (3E).*

- *Students have to do the same kind of tasks repeatedly. If the teacher can make the weighting of the second one heavier than the first, then they will know what to do after getting the feedback on the first one (4B). As a result, the students will know more about the criteria (4A, 4B).*
- *Feedback is quite important (4B).*

### Communication between Teachers and Students

Some students thought that

- *Students can also do something to ensure that they are on the right track and thus avoid getting low grades. For example: ask the teacher to give some comments on their work during the lecture break or even outside class (3C).*
- *Students have the responsibility to tell the teacher about the problems they have faced as soon as possible, like system errors or clarifications about the vague criteria. They should not complain about unfairness after the assessment results have been finalised (3E).*
- *The best way for students to learn is to take the initiative to ask their teachers about the vague criteria / their papers or areas that they need to improve (3B, 3C, 3D).*
- *If the teacher knows that the students do not quite understand a certain theory, s/he can explain more on this. It would be much better if s/he could explain what s/he wants them to do and allow them to write freely without following a certain pattern (4A).*

Our results show that although students are quite receptive to CRA, they have a few concerns which may hamper the successful implementation of CRA in higher education in Hong Kong. For example, students are very concerned about their right to know what the assessment criteria are. They would like to engage themselves more in the assessment process, which may explain why in Cuffe and Jackson (2006) found that more involvement of the students in the assessment process has improved the students' perception of CRA. Another major problem that these students have identified is the issue of parity. This coincides with their teachers' concern as reported in Wong (2006). The university should find a solution to this problem which has been repeatedly identified by both teachers and students, indicating that this is a serious problem for CRA.

## VI. CONCLUSION

This paper has presented the views that were collected from students in five different focus group discussion sessions. Students were asked to compare CRA implemented in the university and the norm-referenced assessment model that they were used to in public examinations. Most of them preferred CRA because they felt that they were no longer being compared against each other but rather, they were assessed with reference to what they have learned. This, they believed, is the true value of education. At the same time, they were quite vocal in their views towards the weaknesses of the CRA being implemented in its current mode. They criticized how the assessment criteria were presented to students and the how assessment was being handled by some of the teachers. As some previous studies have shown, engaging students in CRA is a crucial element in making it a success. This paper has provided further evidence to support that view. It is a must that students should be more involved and that they should have more say in CRA.

## REFERENCES:

- Biggs, J. 1995. Assumptions Underlying New Approaches to Educational Assessment: Implications for Hong Kong. *Curriculum Forum* 4, 1-22.
- Biggs, J. 1999a. *Teaching for quality learning in university*. Society for Research into Higher Education and Open University Press.
- Biggs, J. 1999b. What the Student Does: teaching for enhanced learning. *Higher Education Research & Development*, 18(1), 57-75.
- Biggs, J. and Collis, K. 1982. *Evaluating the Quality of Learning*, Academic Press.
- Cuffe, N. and S. Jackson. 2006. Engaging students in the implementation of criterion referenced assessment in first year law. Engaging Students: 9<sup>th</sup> Pacific Rim First Year in Higher Education Conference.  
<[http://www.fyhe.qut.edu.au/past\\_papers/2006/Papers/Cuffe.pdf](http://www.fyhe.qut.edu.au/past_papers/2006/Papers/Cuffe.pdf)> at 15 February 2008.
- Hodges, L.C. and L.C. Harvey. 2003. Evaluation of Student Learning in Organic Chemistry Using the SOLO Taxonomy. *Journal of Chemical Education*, Vol.80(7), 785-787.
- Jewels, T., M., Ford and W. Jones. 2007. What Exactly Do You Want Me To Do? Analysis of a Criterion Referenced Assessment Project. *Journal of Information Technology Education*, Vol.6, 311-325.
- Lundberg, A. 2004. Students and teacher experience of assessing different levels of understanding. *Assessment & Evaluation in Higher Education*, Vol.29(3), 323-333.
- O'Donovan, B., M. Price and C. Rust. 2000. The Student Experience of Criterion-Referenced Assessment (Through the Introduction of a Common Criteria Assessment Grid). *Innovation in Education & Teaching International*, Vol.38(1), 74-85.
- Rust, C., M. Price and B. O'Donovan. 2003. Improving Students' Learning by Developing their Understanding of Assessment Criteria and Processes. *Assessment & Evaluation in Higher Education*, Vol.28(2), 147-164.
- Wong, C.S.P. 2006. Views on the Adoption and Implementation of the SOLO Taxonomy. In Steve Frankland (Ed.), *Proceedings of the First International Conference on Enhancing Teaching and Learning through Assessment*, pp.4-15.

Appendix <i>

**DISCUSSION QUESTIONS**  
**FOR STUDENTS IN THE FOCUS GROUP SESSIONS**

1. How do you feel about the assessment method / approach in PolyU?
2. Please give two examples of how your course work was evaluated in the last semester. (i.e. 1 good assessment method / task and 1 poor assessment method / task.)
3. Is this assessment method different from the one that you had experienced in the secondary school? If so, in what ways are they different?
4. Do you think the assessment tasks have helped you to learn the subject?
5. Are you clear about the assessment criteria of these tasks?
6. Do you think your understanding of the assessment criteria is the same as your teachers? If not, can you suggest some ways to improve the situation?
7. Do you prefer the assessment method in PolyU (your assignment is compared to a set of explicit assessment criteria and not with other students) or the one that you had experienced in the public examination (every student is compared with others and the results are spread over a bell curve)?
8. Other comments (in general)?