

## Assessment of Criteria of Good Lecturing

Adrian Fisher

John Alder

Mark Avasalu

Department of Psychology  
Victoria University of Technology  
St Albans Campus

Paper presented at the AARE/NZARE Joint Conference, Deakin  
University,  
Geelong, November 1992.

This research was supported by a grant from the Western Institute  
Council  
Teaching-Learning Fund. The findings and interpretations are those of the  
authors and do  
not necessarily represent the views of the Western Institute or the  
Victoria University of  
Technology.

Abstract

Employing previous literature, focus groups and individual interviews, 21 criteria were identified as being potential contributors to an effective lecture. These criteria were assessed using a rating scale survey of 88 academic staff and 320 students at one University campus. Students placed more importance on pace of the lecture and the public speaking skills. Academic staff placed more importance on a series of criteria related to critical thinking and pedagogy: non-sexist language, non-discriminatory examples, challenging students' world view, independent learning, curiosity, and building on previous knowledge.

When factor analyses were used, the academic pattern of results showed a social equity factor, and others which related to expanding the ways in which students thought and worked independently. Student factors related much more to the structure and instrumental nature of lectures -- being taught the information in easily accessible ways.

As student appraisals represent one form of staff evaluation, it is crucial to recognise that staff and students understand the function of the lecture and the role of the lecture in very different ways. The implications of these differences are discussed in relation to the development of academic lecturing skills.

### Assessment of Criteria of Good Lecturing

There is currently a movement within Australian higher education to introduce a formal system of academic staff appraisal. Dawkins (1988), the Federal Minister for the Department for Employment, Education and Training (DEET), set out in his policy statement on higher education a requirement of tertiary institution to initiate assessment procedures. Assessment of performance would be made as a part of the so called quality control of tertiary education.

The rationale for a summative approach was that systematic procedures for staff appraisal would facilitate the rewarding of excellence, assist decision making about tenure and promotion, and ensure accountability of academic staff. According to the principles of staff appraisal promoted by DEET (Lonsdale, Dennis, Openshaw, & Mullins, 1988), one of the specific purposes of staff appraisal procedures is to provide the basis for staff development.

Four basic sampling methods have been employed to collect information for teaching appraisal in higher education in Australia. Information has been solicited from: (a) students, (b) colleagues, (c) expert/trained raters, and (d) self-reports by academics (Marsh, 1986; Moses, 1986; Thomson, Deer, Fitzgerald, Kensell, Low, & Porter, 1990). By far the most

widely used method in higher education has been student appraisal (Cruse, 1987), but self-appraisal of teaching performance is perhaps the most political neutral and politically easiest evaluation method to implement. All approaches, however, have been shown to suffer from measurement flaws, and a variety of studies have found serious inconsistencies between different types of raters.

There are a number of methodological problems associated with self-appraisal that centre on the accuracy of ratings (Meyer, 1980; Thornton, 1980). Self-ratings have been found to suffer from inflation -- in comparison with others' judgements -- and a tendency for the raters to exhibit socially desirable response patterns (Howard, Conway, & Maxwell, 1986; Moses, 1986). One particular distortion of self-ratings is the self-serving bias: self-appraisers are likely to take personal responsibility for good performance, but to attribute poor performance to external causes (Campbell & Lee, 1988). The presence of any of any of these errors is considered to have the potential to adversely affect the value of self-ratings when employed to assess past performance. Self-appraisal is considered to be less problematic when used by individuals to predict their future performance for developmental purposes (Campbell & Lee, 1988, Thompson et al., 1990).

Colleague and expert appraisals have been proposed as means to overcome some of the limitations of self appraisal. However, colleague and expert appraisals also pose a number of unique problems. For practical reasons, colleague and expert raters are not likely to be as familiar with an appraisee's teaching as either their students or the appraisee themselves. Consequently, sampling bias is considered to be a potential problem for colleague and expert appraisals (Cohen & McKeachie, 1981; Doyle, 1975; Scriven, 1987). The validity of appraisals based on limited observations of teaching

performance can be questioned on the basis that such observations cannot be assured of being representative of a lecturer's performance over the length of a course. Colleague ratings have also been shown to be influenced by halo effects (Nafturlin, Ware & Donnelly, 1973), and colleague and expert ratings are dependent on the appraiser's knowledge of the subject matter being taught (Sorcinelli, 1984; Scriven, 1987). Despite these effects, colleague and expert ratings are considered by some to be more reliable than student ratings, even though some evidence indicates student ratings have better validity than colleague, expert, or self-ratings of teaching effectiveness (Costin, Greenough & Menges, 1971; Howard, Conway & Maxwell, 1985; Marsh, 1982; McKeachie, 1986).

Student ratings, have been found liable to be affected by a wide variety of factors that have been categorised by Kulik

and Kulik (1974) under two broad headings: 1) course setting -- class size, elective versus compulsory subjects, subject level, workload/difficulty, and the academic discipline (e.g., Arts students tend to give higher ratings than Applied Science students); 2) teacher characteristics -- personality, verbal fluency, and lecturer's rank. Other reviewers have identified physical attractiveness, prior interest, expected grades, and timing of rating as additional factors that may affect student ratings (Marsh 1987; Braskamp, 1984).

A number of studies (e.g., Rodin & Rodin, 1972), have shown student ratings failed to correlate with purportedly valid indicators of teaching effectiveness and were, therefore, deemed to exhibit poor construct validity.

However, subsequent studies have refuted these findings, and contend student ratings, in contrast to colleague, expert and self-ratings, display reasonable levels of internal consistency and stability -- reliability (Arubayi, 1987) -- and provide valid measures of teaching effectiveness (Howard et.al., 1985).

Lecturing is still the most widely used teaching method in higher education, and importantly from a formative stand point, it is essentially a skill that can be learnt (McKeachie, 1986). It is, however, widely accepted that an effective lecture is a multi-faceted construct which is difficult to measure directly (Marsh, 1986; McKeachie, 1986) as only the indicators of its presence can be tested for. To this end, Marsh (1987) proposed criteria that form a number of dimensions of effective teaching, 5 of which relate specifically to lecturing: organisation/ clarity, learning/value, breadth of coverage, enthusiasm, and group interaction. Murray (1980) and Marsh (1987) both contend that there is agreement between academics and students as to what constitutes good university teaching.

Marsh (1984; 1986; 1987) used Multi-Trait Multi-Method (MTMM) procedures to test the validity of student and self-ratings by academics of criteria of effective teaching. Marsh reported a high correlation between student and self-ratings across the studies, and contended convergent and divergent validity was demonstrated for both the criteria used and the methods of measurement. Thus, Marsh concluded both the

criteria and the types of measures used (i.e., student and self-rating) displayed construct validity. Other studies, that have employed MTMM procedures, however, do not fully support Marsh's finding. Howard, Conway, and Maxwell (1986), used MTMM procedure to assess the validity of student, colleague, expert, and self-ratings of effective teaching, but found colleague, expert and self-ratings to have poor

reliability. Only student rating were found to be reliable and to display construct validity.

One source of the discrepancy between the two studies may be related to the issue of the inferential accuracy of raters (Nathan & Alexander, 1985). That is to say, whilst agreement may be reached between academics and students as to the characteristics/criteria that are indicative of effective teaching, academics' and students' cognitions as to the implicit properties of such criteria may differ. Such cognitive differences may result in the two groups making dissimilar judgements about a lecturer's performance despite the same measures being used, and lead to the conclusion that student or academic ratings are unreliable.

This study is based in the value stance that the establishment of criteria of performance and their application are to be used for the development of staff competencies, and that the development will be reflected in improved education for our students. The identification and acceptance of specific aspects of lecturing forms only one part of the duties and responsibilities of academics, and only one of many modes of delivery of education in higher education. Simple acceptance of these as "the way" to assess staff or for staff to assess themselves would be invalid and dangerous.

#### Method

##### Subjects:

**Academics:** With the exception of sessional lecturers, all academics from across the 4 higher education Faculties (Arts, Health Science, Applied Science, and Business) were approached to participate in the survey. A total 89 academic returned surveys, of which 88 were useable for statistical analyses. The number of respondents represented approximately one third of the academic staff. There were approximately equal numbers of males and females in the sample.

**Students:** A representative sample of undergraduate students from across the four Faculties were approached to participate in the survey. A total 320 student surveys were returned, all of which were included in the study. The student sample represented approximately equal number of students from the four faculties. The age of student in the sample varied from less than 21 year old to more than 25 years old, with vast majority being less than 21 years old. Seventy percent of the students were female.

##### Procedure:

Academics and students from each of the Faculties at St Albans Campus of Victoria University of Technology, were asked to participate in separate small informal discussion groups. Group members were asked to relate and discuss what they thought were the characteristics of an effective lecture. The

Table 1  
Twenty-one Lecturing Criteria Used in Study

Criterion	Abbreviation
Provide clear explanations	CLEAR
Present material in an interesting way	INTEREST
Stimulate students' interest	STIMUL
Pace lecture to allow note taking	PACE
Arouse students' curiosity	CURIOS
Use non-sexist language	NOSEXIST
Ensure lectures have defined structure	STRUCT
Use examples relevant to students	RELEVANT
Display mastery of subject matter	MASTERY
Provide up to date research	UPTODATE
Stimulate independent learning	INDEPEND
Use inclusive examples and expression	DISCRIM
Display high level of verbal fluency	FLUENCY
Interact with students	INTERACT
Possess good public speaking qualities	PUBSPEAK
Challenge students' world views	CHALLENG
Build on students' previous knowledge	BUILD
Pause to allow memory consolidation	MEMORY
Provide periodic summaries during lecture	SUMMARY
Act as academic role model	ROLEMOD
Project enthusiasm for the subject matter	ENTHUSI

discussions were recorded on audio tape and systematically analysed using theme analysis techniques.

The criteria derived from the nominal groups were combined with criteria that appeared in the literature. A total of 21 criteria were identified by these processes and are shown Table 1. A survey instrument was then designed incorporating the 21 proposed criteria in order to measure academics' and students' attitudes as to the importance of each of the criteria. The survey instrument consisted of 21 statements, each statement addressing only one criterion. Respondents were asked to indicate their level of agreement or disagreement with each of the statement on a 9 point interval

scale. Fourteen of the statements were framed as positive statements, and seven in the negative. The scale ranged from strongly disagree (1) too strongly agree (9).

## Results

The data for the 7 negatively framed statement were recoded so that analyses could be conducted with all items scored in the same direction (low values -- low importance, high values -- high importance). Separate group means were then calculated for the academic and student ratings of each of the 21, these are shown in Figure 1. From Figure 1 it can be seen that the scores for each group were similar, but that the academics placed greater value on 19 of the 21 criteria. The only two criteria that students rated as more important were the pace of the lecture for note taking, and the public speaking skills of the lecturer.

ANOVAs. In order to determine the significance of the observed differences in ratings of the criteria, one-way analyses of variance (ANOVAs) between the academic and student group scores were conducted. (The results are presented in Table 2.) From Table 2 it can be seen that students rated only one criterion as significantly more important than did the academics -- pace of the lecture for note taking ( $F = 26.88$ ,  $p < .001$ ). The academics, on the other hand, rated a number of criteria as more important than did the students.

For academics, challenging the students' world view ( $F = 60.67$ ,  $p < .001$ ) and stimulating independent learning ( $F = 60.50$ ,  $p < .001$ ) were far more important than for the

Figure 1

Table 2  
Comparisons of Academic and Student Ratings of 21 Criteria

Criterion	Academic		Student		F
	Mean	SD	Mean	SD	
Explanation	8.33	0.81	8.26	0.99	0.34
Interesting	8.09	0.87	7.75	1.26	5.49*
Stimulate	8.03	1.30	7.63	1.36	6.27*
Curiosity	7.90	0.98	6.97	1.46	31.46***
Non-sexist language	7.71	1.55	6.48	2.02	27.72***
Structure	7.66	1.48	7.61	1.31	0.08
Enthusiasm	7.60	1.57	6.67	2.02	15.72***

Up to date	7.59	1.65	7.50	1.41	0.21
Mastery	7.58	1.34	7.37	1.41	1.62
Independent Learning	7.57	1.59	5.70	2.07	
60.50***					
Relevant Examples	7.54	1.47	7.06	1.52	7.11**
Inclusive presentation	7.46	1.57	6.36	1.88	24.26***
Verbal fluency	7.38	1.08	7.30	1.49	0.22
Interact	7.34	1.85	6.98	1.98	2.23
Public Speaking	7.12	1.92	7.27	1.83	0.42
Challenge	7.04	2.10	5.23	1.89	
60.67***					
Pace	6.94	1.50	7.80	1.31	
26.88***					
Build	6.67	2.09	5.93	2.24	7.84**
Pause	6.52	1.40	6.26	1.81	1.59
Summaries	6.44	1.68	6.19	2.02	1.01
Role model	5.86	2.07	5.51	1.86	2.23

—  
\*  $p < .05$  \*\*  $p < .01$  \*\*\*  $p < .001$

students. These seem to indicate that the academics placed a high value on going well beyond the simple delivery of material in assessing what an effective lecture means. This would be supported by their higher values placed on stimulating students' curiosity ( $F = 31.47$ ,  $p < .001$ ), stimulating students' interest in the material ( $F = 6.27$ ,  $p < .05$ ), and demonstrating enthusiasm for the material ( $F = 15.72$ ,  $p < .01$ ).

The academics, however, also placed more value on some of the structural elements of lecturing to make them effective.

In particular, they saw using non-sexist language ( $F = 27.72$ ,  $p < .001$ ), using examples that did not discriminate or alienate certain groups of students ( $F = 24.26$ ,  $p < .001$ ), and examples that were specifically relevant to the students' experiences ( $F = 7.11$ ,  $p < .05$ ) as significantly more important than did students.

**Factor Analysis.** In order to understand how the individual criteria related to each other, principal component factor analyses were conducted. Because of the differences predicted and found between the student and academic scores, factor analyses were conducted separately for each group. The solution for each group was forced to seven factors, with the criteria for inclusion of a factor loading of at least .5, with no higher loading on any other factor. (The resultant factor structure matrix for academic staff appears in Tables 3, with inter-factor correlations in Table 4. For students,

the factor matrix is in Table 5 and the inter-factor correlations in Table 6.)

For academics, the criteria that loaded highest on the seven factors reflected reasonably distinct and meaningful dimensions. Criteria loading highly on Factor 1 were interpreted as reflecting social equity/content emphasis -- a general dimension concerned with avoiding discrimination in the content of lectures. While the criterion Up to Date Research does not appear to match conceptually the other two criteria, it was apparent from the qualitative data collected from academics that they interpreted, up to date research, not necessarily in terms of its recency, but its concurrence with

Table 3  
Factor Structure for Academic Staff

Factor/Variables	1	2	3	4	5	6	7
<b>Social Equity</b>							
NONSEXIST	.79						
DISCRIM	.77						
UPTODATE	.72						
<b>Presentation Skills</b>							
MASTERY	.73						
PUBSPEAK	.52						
<b>Motivation</b>							
CURIOS		.79					
STIMUL		.69					
INTEREST		.68					
<b>Modelling</b>							
ENTHUSI		.81					
ROLEMOD			.73				
<b>Lecturing Mode</b>							
PACER			.70				
SUMMARY			.69				
RELEVANT			.60				
<b>Critical Thinking</b>							

INDEPEN .65  
CHALLENG .65

Cognitive Processes

STRUCT -.79  
BUILD -.77  
MEMORY -.58

Table 4  
Factor Correlation Matrix for Academic Staff

Factor	1	2	3	4	5	6
Factor 1	1.00					
Factor 2	.06	1.00				
Factor 3	.20*	-.08	1.00			
Factor 4	.11	.06	.04	1.00		
Factor 5	.20*	-.01	.13	.10	1.00	
Factor 6	.15	.06	.00	.17*	.03	1.00
Factor 7	-.25*	-.12	-.21*	-.14	-.13	-.10

\*  $p < .05$

the dominant educational milieu; social equity. Criteria loading highest on the remaining six factors, posed few problems for interpretation. Factor 2: presentation skills; knowledgeable, communicate proficiently, and confidence in dealing with students. Factor 3: motivation; intellectual stimulation of students. Factor 4: modelling; behaviour exemplary of positive academic attributes. Factor 5: lecturing mode; relating to the mechanics of disseminating information. Factor 6: critical thinking. Factor 7: cognitive processes; use of mnemonic and cognitive principles to assist students information processing.

The correlations between the academics' factors tended to support the interpretations and the appropriateness of the factor solution, by demonstrating convergence between factors that were conceptually related, and divergences between conceptual distinct factors. Factor 1, Social Equity, which was interpreted as a general dimension was found to correlate

highly with five of the other six factors and, thereby, tended

to support the factor's generality. Conversely, Factor 2, Presentation Skills, which was interpreted as forming a distinct factor, correlated in ways consistent with its uniqueness, by correlating poorly with the six other factors. Factors 4, Modelling, and Factor 6, Critical Thinking, were conceptually related as both were concerned with higher education values. The resultant high correlations between these two factors, combined with their equally low correlations with Factor 3, Motivation, and Factor 7, Cognitive Processes demonstrate the validity of the interpretation of Factors 4 and 6 and Factors 3 and 7.

Factor 5, Lecturing Mode, bore some conceptual relationship to Factors 3 and 7, due its focus on the mechanics of lecturing. Conversely, Factor 5, bore little relationship with Factors 2, 4, or 6, due to their concern with more intellectual aspects of instruction. The resultant high correlation between Factor 5, and Factors 3 and 7, combined with its lower correlation with Factors 2, 4, and 6, suggests Factor 5 also forms a reasonably coherent sub-construct of an effective lecture.

For students, the first factor seemed in part to be a synthesis of criteria that matched two of Marsh's dimensions (1987); Organisation/ Clarity, and Breadth of Coverage. In this sense, Factor 1 seemed to be concerned primarily with the quality of a lecture's content, and represented Content Merit. Criteria loading highest on Factor 2, suggested Social Skills. This differed conceptually from the academics' presentation skills factor in that the configuration of criteria only

Table 5

Factor Structure for Students

Factor/variables	1	2	3	4	5	6	7
<b>Content Merit</b>							
CLEAR	.70						
UPTODATE	.61						
MASTERY	.61						
INTEREST	.61						
STRUCT	.50						
<b>Social Skills</b>							

PUBSPEAK .77  
 FLUENCY .55  
 INTERACT .52

Logical Progression

SUMMARY .73  
 BUILD .60

Social Equity

DISCRIM .83  
 NOSEXIST .81

Flow of Lecture

PACE .78  
 MEMORY .60  
 STIMUL .60

Instructional Environment

ROLEMOD .74  
 CURIOS .60

Dynamism

CHALLENGE -.79  
 ENTHUSIASM -.58

Table 6

Factor Correlation Matrix for Student

Factor	1	2	3	4	5	6
Factor 1	1.00					
Factor 2	.09	1.00				
Factor 3	.02	.08	1.00			
Factor 4	.13*	.06	-.05	1.00		
Factor 5	.19*	.05	.12*	.07	1.00	
Factor 6	.18*	.01	.07	.11	.20*	1.00
Factor 7	-.04	-.20*	-.16	-.07	-.08	-.05

\*  $p < .05$

addressed lecturers' competence in communicating with students; whereas the academics' presentation factor relates, some what, more to projecting a competent image. There was little ambiguity in the criteria loading highest on Factor 4, they clearly represent Social Equity/Content Bias.

The configuration of criteria loading highest on Factors 3, 5, 6 and 7, all reflect some aspect of lecturer style. Criteria loading highest on Factors 3 and 5, suggest procedural aspects of lecture presentation: Factor 3 Logical Progression; Factor 5 the Flow of a Lecture. Criteria loading highest on Factors 6 and 7 seemed to relate to certain traits of a lecturer that enhance their presentation -- Factor 6, Instructional Environment. In this sense, students seemed to have applied the criterion Act as an Academic Role Model, differentially from academics, by perceiving it in terms of what is a good presenter. Factor 7, Dynamism, encompassed a sense of entertainment value in a lecture's lecturing style.

The resultant correlations between the student factors

tended to demonstrate the validity of the factor's interpretation. Factor 1, Content Merit, correlated highly with Factors 4, 5, and 6, all of which were concerned to some degree with lecture content. Conversely, Factor 1, correlated poorly with Factors 2 and 7, both of which correlated highly with each other, and related to a conceptual distinct dimension concerned with social skills. Factors 3 and 5, both of which were conceptually related to the mechanics of lecture presentation, and correlated highly, discriminated from Factor 2 (a conceptually distant factor), by both correlating poorly with it. Factor 6, which was concerned with students' perceived instructional needs, correlated highly with Factors 1 and 5, (both of which related to the content of a lecture, and reflected the dominant theme of the student factor structure). Conversely, Factor 6 correlated lower with Factors 2 and 7, both of which were related and were not directly related to the content of a lecture. Factor 6, therefore, was also shown to have some degree of convergent and discriminant validity.

From Tables 3 and 5 and the previous discussion, it can be seen that there are strong differences in the factor structures for the students and academics. The academic structure has a stronger focus on the social equity issues that are beginning to be much more important in higher education -- and which match the official ethos of the institution. For students, the most important factor relates

much more to the specific presentation aspects of the lecture -- clarity, interesting and well structured.  
Discussion

Analysis of the rating scales indicated that all 21 criteria were rated by academics and students as having a bearing on the effectiveness of a lecturer's performance. All 21 criteria were rated towards the "agree" end of the scales by students and academics. This provides some face validation of the criteria as representing recognisable characteristics which apply to good lecturing. The results tended to support the contention drawn by Murray (1980) and Marsh (1987) that there is agreement between academics and students as to what constitutes good university teaching. However, there was strong evidence in this study that the relative importance placed on the criteria varied greatly between staff and students -- and may represent quite different interpretations of what a good lecture should be.

The analyses of variance demonstrated that the academic staff placed greater emphasis on a number of the criteria. These were aspects of lecturing that are focussed on challenging the status quo for students. Among the criteria rated more important were the use of nonsexist language, stimulating the students' curiosity, emphasis on independent learning, etc. The only two areas in which students placed a greater emphasis on the components of lecturing were the speed of the lecture and the speaking voice of the lecturer. Each of these represents a fairly instrumental component of the lecture for students. They need to be able to hear the material clearly and be able to write it down. For the

students overall, the notion that lectures are meant to go beyond the presentation and recording of facts was not evidenced.

These comparisons of the academics' and students' ratings of the criteria, indicated significant differences existed between them as to the relative importance each group attributed to the criteria. Furthermore, differences were also evident between students and academics in the cognitions of the underlying dimensions that the criteria related to. In short, a surface level of analysis showed a consensus between academics and students that masked deeper ontological differences between the groups.

To go beyond the comparison of single criteria, factor analyses were conducted separately on the staff and student data. The aim of the factor analyses were to determine the ways in which criteria combined to provide meaningful clusters of criteria that could be used to explain in more depth the ways in which the groups perceived a good lecture. Factor

analysis of the academic and student data resulted in different factor structure for the two groups. The simple fact that the two factor structures are dissimilar, is of little analytic value, unless the factors can be meaningfully interpreted. One means of determine the appropriateness of a factor structure relates to the convergence and divergence between factors. That is to say, the validity of the factors' interpretations may be demonstrated by showing that factors correlate higher with factors they are conceptually related to, than with factors they are not conceptually related to.

For academics, the criteria that loaded highest on the seven factors reflected reasonably distinct and meaningful dimensions. Criteria loading highly on Factor 1 were interpreted as reflecting social equity/content bias; a general dimension concerned with avoiding passive discrimination in the content of lectures. Criteria loading highest on the remaining six factors, posed few problem for interpretation for the resulting factors (see Table 3): Factor 2: presentation skills; Factor 3: motivation; Factor 4: modelling; Factor 5: lecturing mode; Factor 6: critical thinking; Factor 7: cognitive processes. The correlations between the academic factors tended to support their interpretations and the appropriateness of the factor solution, by demonstrating convergence between factors that were conceptually related, and divergences between conceptual distinct factors.

For students, the first factor seemed in part to be a synthesis of criteria that matched two of Marsh's dimensions: organisation/clarity, and breadth of coverage. In this senses, the criteria seemed to be concerned primarily with the quality of a lecture's content, and represented content merit. The remaining factors could be described as Factor 2: social skills; Factor 3: logical progression; Factor 4: social equity; Factor 5: flow of lecture; Factor 6: instructional environment; and Factor 7: dynamism. The configuration of criteria loading highest on Factors 3, 5, 6 & 7, all reflect some aspect of lecturer style.

Analysis of the inter-factor correlation matrices (see

Tables 4 and 6), therefore, was able to demonstrate both the academic, and student, factor structures were reasonably appropriate solutions for each groups data. The divergence between the academic and student factor structure, and the dimensions they related, tends to suggest, academics and students conceptualise an effective lecture differently. While, both groups contended that the 21 criteria were relevant, the context in which academics and students perceived the criteria to be relevant appeared to be sharply delineated between the two groups.

One possible source of the divergence between the academic and student factor structures, may relate to the intrinsically different roles academics and students possess within the higher education system. Whilst, both academic and student data were collected using the same survey instrument, academics' and students' differentially experience of higher education, may have resulted in academics and students responding in two contextually different ways to the survey questions. The role of a student is in many respects that of a consumer. Students, therefore, may have responded from the perspective of: what do I want from a lecturer? So, student data may reflect students' more immediate or instrumental concerns about instruction. Conversely, the role of academics is that of providers of education. In which case, academics may have responded from the perspective of: what do students require from a lecturer? Academic data may reflect academics' identification with the more abstract goal of higher education, that correspond with their dual identity within higher education of researcher and educationalist.

To some extent, academics appeared to be reiterating institutional values of equity in, and access to, higher education, which are proclaimed to be basic tenets of the institution in question. Academics' affirmation of these values was quite evident in the importance academics attributed to those criteria related to social justice. The academic data, therefore, may in part reflect, either, their fundamental concern for educational equity issues, or a degree of socially desirable response (Howard et al., 1986; Moses, 1986).

In contrast to academics, students appeared to projected a more instrumental orientation, and seemed to have made experiential judgements about what is an effective lecture, based on their perceived needs, and their socio-cultural expectations of higher education. The instrumental orientation of students, is to a large degree evident in the dominant theme of the student factor structure which centres on the informational content of lectures. Further support for the instrumentality of students in the present study, relates to broader development in higher education, in particular, the recent expansion of higher education in Australia, which mirror similar trends over a decade earlier, in North America.

Commenting on the North American experience, McKeachie (1986) noted that, whereas in the past only the academic [and social] elite students entered university, students populations now exhibit a broader range of academic skills.

Consequently, there may be a greater demand from students participating in higher education for academics to teach, were

as in the past students were (or were expected to be) more academically self-reliant.

The reliable agreement between academics and students as to the criteria for effective teaching found by Marsh (1987) and Murray (1980) was not evident in the present study. The results of this study clearly indicate some caution is warranted in assuming the generality of criteria for effective teaching. Changes in student populations over time may result in criteria for effective teaching that previously was shown to be valid and reliable no longer exhibiting these characteristics. Similarly, the findings question the transferability of criteria validated in one educational setting to another. Or for that matter, criteria validated by academics to be used by students to evaluate lecturer's performance. Thus, the assessment of academics by students will represent a very different form of measurement than either self or peer-review as the construct being utilised will be totally different. Such difference provide explanations for the findings that there appeared in the past to be discrepancies between student and peer-appraisals (Arubayi, 1987; Howard, et al., 1985; Marsh, 1987).

In sum, the academic and student data in the present study may be thought of as representing samples from two different populations due to differences in the role academics and students possess within the higher education setting, their psychological experience of higher education, and their perceptions of the purpose and goals of higher education. The factor analyses demonstrated support for this premise, by identifying what appears to be marked conceptual differences between the two groups, as to how the proposed criteria relate to the construct of an effective lecture. Convergent validity could not be demonstrated, due to the dissimilarity in the factor structures between academics and students. Nor, was discriminant validity demonstrated, due to the criteria that combined to form conceptual distinct factors for the academic data, combined differently to form conceptually different and distinct factors for the student data.

#### References

- Arubayi, E. (1987). Improvement of instruction and teacher effectiveness: Are student rating reliable and valid? *Higher Education*, 16, 267-278.
- Braskamp, L. A. (1984). *Evaluating teacher effectiveness*. Beverly Hills, CA: Sage.
- Campbell, D. J., & Lee, C. (1988). Self-appraisal in performance evaluation: Developmental versus evaluation. *Academy of Management Review*, 13, 302-314.

Cohen, P. A., & McKeachie, W. J. (1981). The role of colleagues in the evaluation of college teaching. *Improving College and University Teaching*, 28, 147-154.

Collier, M. J., & Powell, R. (1990). Ethnicity, instructional communication, and classroom systems. *Communication Quarterly*, 34, 334-349.

Costin, F., Greenough, W. T., & Menge, R. J. (1971). Student ratings of college teaching: Reliability, validity, and usefulness. *Review of Educational Research*, 41, 511-535.

Cruse, D. B. (1987). Student evaluation and the university professor: Caveat professor. *Higher Education*, 16, 723-737.

Dawkins, J. S. (1988). *Higher education: A policy statement*. Canberra: Australian Government Printing Service.

Doyle, K. (1975). *Student evaluation of instruction*. Lexington, MA: Lexington Books.

Howard, G. S., Conway, C. G., & Maxwell, S. E. (1985). Construct validity of measures of college teaching effectiveness. *Journal of Educational Psychology*, 77, 187-196.

Kulik, J. A., & Kulik, C. C. (1974). Student ratings of instruction. *Teaching of Psychology*, 1, 51-57.

Lonsdale, A., Dennis, N., Openshaw, D., & Mullins, G. (1988). *Academic staff appraisal in Australian higher education, Part 1: Principles and guidelines*. Canberra: A.G.P.S.

Marsh, W. H. (1982). Validity of students' evaluation of college teaching: A multitrait-multimethod analysis. *Journal of Educational Psychology*, 67, 833-839.

Marsh, W. H. (1984). Student evaluation of university teaching: Dimensionality, reliability, validity, potential bias, and utility. *Journal of Educational Psychology*, 70, 707-754.

Marsh, W. H. (1986). Applicability paradigm: Students' evaluation of teaching effectiveness in different countries. *Journal of Education*, 78, 465-473.

- Marsh, W. H. (1987). Students' evaluation of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11, 253-288.
- McKeachie, W. J. (1986). *Teaching tips*, (8th Ed.). New York: Heath.
- Meyer, H. (1980). Self-appraisal of job performance. *Personnel Psychology*, 33, 291-295.
- Moses, I. (1986). Self and student evaluation of academic staff. *Assessment and Evaluation in Higher Education*, 11, 76-86.
- Murray, H. G. (1980). *A comprehensive plan for the evaluation of teaching at the University of Queensland*. St. Lucia, QLD: Tertiary Education Institute.
- Nafturlin, D. H., Ware, J. E., & Donnelly, F. A. (1973). The Doctor Fox lecture: A paradigm for educational seduction. *Journal of Medical Education*, 48, 630-635.
- Nathan, B. R., & Alexander, R. A. (1985). The role of inferential accuracy in performance ratings. *Academy of Management Review*, 10, 109-115.
- Rodin, M., & Rodin, B. (1972). Student evaluation of teaching. *Science*, 177, 1164-1166.
- Runco, M. A., & Thurston, B. J. (1987). Students' ratings of college teaching: A social validation. *Teaching Psychology*, 14, 89-91.
- Scriven, M. (1987). *The validity of student ratings*. Nedlands, W.A.: The Department of Education, University of Western Australia.
- Sorcinelli, M. D. (1984). An approach to colleague evaluation of classroom instruction. *Journal of Instructional Development*, 7, 11-17.
- Thompson, H. R., Deer, C. E., Fitzgerald, J. A., Kensell, H. G., Low, B. C., & Porter, R. A. (1990). Staff self-appraisal at the departmental level: A case study. *Higher Education Research and Development*, 9, 39-48.
- Thornton, G. (1980). Psychometric properties of self-appraisal and job performance. *Personnel Psychology*, 33, 263-271.



Paper Presented at the Joint AARE/  
NZARE Conference, Geelong, 1992