Profiles of Effective College and University Teachers

Suzanne Young; Dale G. Shaw

The Journal of Higher Education is currently published by Ohio State University Press.

⨎ Suzanne Young
Dale G. Shaw

Profiles of Effective College and University Teachers

Although considerable research has been conducted in the arena of teacher effectiveness, important questions continue to persist: What is effective teaching? How may it be defined? How may it be measured? To date, educators and researchers have failed to reach agreement about clear-cut answers to these questions; indeed, consensus may not be possible. The answers undoubtedly are affected by a number of things, including the recognition that teaching effectiveness comprises multiple perspectives (Abrami, d'Apollania, & Rosenfield, 1997; Marsh & Dunkin, 1997) as well as such characteristics as the type of course, class size, student abilities, and grading practices (Abrami, d'Apollonia, & Cohen, 1990; Greenwald & Gillmore, 1997). Additionally, methodology may have a differential effect on findings. For example, according to d'Apollonia and Abrami (1997), reviews of multisection validity studies often reach different conclusions due, in part, to reviewers' biases.

The primary issues addressed in this study deal with how teaching effectiveness might be defined or, more accurately, whether it might be defined in multiple ways, and if so, how it might be defined. For example, does the multidimensionality of the construct dictate multiple definitions, or might a single definition be able to capture it? We have addressed these matters in this study from the standpoint of student-rated teacher effectiveness and with methodology that, though somewhat unconventional, was selected in an effort to maximize interpretability of results by magnifying relationships and at the same time minimizing

Suzanne Young is assistant professor in educational research, University of Wyoming, Laramie, and Dale G. Shaw is professor and department chair in applied statistics and research methods, University of Northern Colorado, Greeley.

measurement error. Our approach was to ask students to rate instructors from whom they had taken a college or university course in the recent past, using high-inference items garnered from the teaching effectiveness literature that have shown to be strong correlates of teaching effectiveness. Prior to conducting their ratings, the students were provided with a brief training and question/answer session that dealt with avoidance of rating errors and clarification of items, procedures, the referent, and the like. The data they provided us were analyzed to see if multiple definitions of teacher effectiveness were indicated.

A large segment of the teacher effectiveness literature describes investigations that seek identifying characteristics, factors, traits, classroom behaviors, and so on, of effective teachers by rating instruction; student ratings, in particular, have received a great deal of attention. The validity of student ratings has been thoroughly analyzed and generally supported in the literature during the past 25 years (Centra, 1994; Cohen, 1981, 1987; Feldman, 1989; Marsh, 1987; Marsh & Bailey, 1993). Indeed, Greenwald (1997) suggests that reviews of research conducted since about 1980 indicate overwhelming evidence supporting the construct, convergent, discriminant, and consequential validity of student ratings.

Student ratings of instruction have been found to correlate highly with instructor personality traits (Feldman, 1986; Murray, Rushton, & Paunonen, 1990; Renaud & Murray, 1996). The Dr. Fox experiments of the 1970s (Marsh, 1987; Naftulin, Ware, & Donnelly, 1973) illustrated that students rated charismatic and expressive instructors as highly effective, regardless of the substantive content of a lecture. Murray et al. (1990) correlated peer ratings of personality traits with student ratings and found that personality traits differed among course types. Renaud and Murray (1996) also investigated relationships between instructor personality and student ratings but did not differentiate between course types. Their patterns of correlations were stronger than those found in the Murray et al. study, but could be due to restricted ranges in the earlier study.

Student achievement and student ratings have been found to be related (Cohen, 1981, 1987; Greenwald & Gillmore, 1997; Marsh, 1987). Koon and Murray (1995) correlated student ratings with final examination scores for 36 instructors of an introductory psychology course and found a zero-order correlation of 0.41. In a meta-analysis of 41 independent studies, Cohen (1981) found an average correlation of 0.43 between student ratings of instructors and student achievement. Cohen also found that the average correlation between student ratings of the course and student achievement was 0.47. Marsh (1987) addressed the confounding issue of grades and student evaluations. He argued that it is difficult to explain the relationship between grades and evaluations.

When grades and exam scores are significantly correlated, then higher eval-
uations by students may be due to (a) more effective teaching that produces
greater learning and higher evaluations by students; (b) increased student
satisfaction with higher grades which causes them to 'reward' the instructor
with higher ratings independent of more effective teaching or greater learn-
ing; or (c) initial differences in student characteristics that affect both teach-
ing effectiveness and performance. (p. 290)

Marsh and Roche (1997) summarized research on grading leniency by
stating that a bias does indeed exist but that its effects are inconsequential.

Greenwald and Gillmore (1997) discussed the relationship between
student ratings and course grades and concluded that the strong relation-
ship between the two is indication of a causal relationship; grading le-
niency by instructors leads to more positive student evaluations. Overall,
the research on teacher effectiveness indicates a moderately positive rela-
tionship between student ratings of instruction and student achievement.

Researchers have validated student evaluation of instruction by corre-
lating student ratings of specific dimensions with overall measures, with
ratings by other observers, with self-evaluations, and even with ratings
by former students (for example, Arubayi, 1986; Cohen, 1981; Feldman,
1976, 1988; Marsh, 1987; Marsh & Overall, 1979; Miller, 1988). Over-
all and Marsh (1980) found a strong positive correlation between ratings
of students with the same students for the same courses several years
later, indicating that student ratings are stable over time. According to
McKeachie (1983), "Students are in class almost every day and they
know what's going on. They are the ones we are trying to affect, and
they have some sense of whether they are learning" (p. 38). And again,
in 1997, McKeachie was consistent in his views when he stated "Stu-
dents will continue to be those most affected by teaching. Therefore, stu-
dent ratings will continue to be useful" (p. 1224).

Teacher effectiveness is a complex, multidimensional construct (Chau
& Hocevar, 1994; Marsh, 1987; Marsh & Roche, 1997) that may be cap-
tured by a variety of models. However, our methods of modeling such
constructs are somewhat primitive. Correlational approaches are useful
but their limitations are well known. For example, an item such as
"course is well organized" will usually correlate highly with an item like
"overall teacher effectiveness," and will often emerge as a significant
contributor among any set of items in explaining or accounting for
teacher effectiveness. Yet not all well-organized courses are presented
by effective teachers and not all effective teachers present well-orga-
nized courses. Should we then include in a definition of teacher effec-
tiveness that courses be well organized? It is neither a necessary nor suf-
ficient trait, yet it bears a high relationship to effectiveness. How should

this, or may this, be modeled? A related matter is the inherent additive nature of linear models in regression and discriminant analyses that serves to compound the correlation problem and leads to additional issues regarding appropriate or "best" weightings for items. Feldman (1988), in a review of studies examining the effectiveness of faculty from both student and faculty perspectives, offers an interesting and similar discussion of this issue. He suggests that correlates of good teaching, such as subject matter knowledge, might be seen as highly important to students but, in fact, may not discriminate between good teachers and poor teachers.

Methods

This study sought to address the issue of effective teaching by attempting to identify alternative models to be compared with these traditional approaches, to take into account the possibility that there may be multiple definitions of teacher effectiveness, and to do this in the context of the university setting from the students' point of view. Using a 25-item instrument, students ($n = 912$) were asked to rate a teacher of their choice from whom they had taken a college or university course in their recent past. These data were submitted to a variety of analyses, including clustering procedures, to derive various profiles of effective teachers. A detailed description of these procedures follows.

Subjects

The study was conducted in a medium-sized western university with an enrollment of about 11,000 students. The university consists of five colleges, which offer nearly 3,000 courses, and embraces a wide range of teachers with respect to age, sex, experience, subject matter expertise, commitment to teaching, communication style, and the like. The courses themselves represent important curriculum variables and vary widely in terms of content, student task time, learning environment, workload, and difficulty.

Initially, 52 university faculty members from a wide variety of disciplines throughout the university were contacted to acquire permission for us to enter their classrooms to collect data. Twenty-nine faculty members granted us access to one of their classes; two granted access to 2 of their classes. Although an occasional student declined to participate when apprised of the nature and purpose of the study, virtually all students ($n = 912$) in the 31 classes elected to participate. These students were asked to select a teacher of a university course that they had taken in the recent past; a teacher of their choice, who, in their judgment,

could be highly effective, highly ineffective, or anywhere in between; and someone whom they thought they could recall in sufficient detail so that they could respond accurately to specific rating items about that teacher and that teacher's instructional effectiveness.

Of the participating students, 54% ($n = 530$) of the sample were graduate students, and 46% ($n = 382$) were undergraduates. The students ranged in age from 18 to 60 years old, with a mean age of 29 ($SD = 10$). They represented 71 different major areas. Two-thirds of the students were female and one-third were male. Students rated courses they were enrolled in as freshmen (20%), sophomores (14%), juniors (14%), seniors (12%), graduate level master's students (23%), and graduate level doctoral students (17%).

The students rated university teachers who taught courses in 152 different subject areas. The classes ranged in size from 1 student to 830 students, with a median class size of 26. Fifty-nine percent of the teachers rated were male and forty-one percent were female.

Items

A large body of literature on teacher evaluation supports the notion that the content of evaluation instruments should be multidimensional. Based on student evaluations and faculty self-evaluations, Marsh (1987) reported consistent identification of nine factors on the Students' Evaluations of Educational Quality (SEEQ): learning/value, enthusiasm, organization, breadth of coverage, group interaction, individual rapport, examinations/grading, assignments, and workload difficulty. In an extensive synthesis of the literature on college students' views toward effective teaching, Feldman (1976) found that stimulation of interest and clarity of presentation were the two most highly related dimensions of good teaching. He also found that the more effective teachers generally were seen as very knowledgeable about subject matter, were organized and prepared for class, and demonstrated enthusiasm. Other less important characteristics, according to Feldman, were related to classroom management. These included items such as course difficulty, workload, and clarity of objectives. Additionally, he found that interpersonal traits such as friendliness, helpfulness, and openness to opinions of others' were considered by students to be important traits of good teachers but not as important as other characteristics. Feldman (1988) compared student and faculty views of effective teaching on the twenty-two instructional characteristics; nineteen were identified in his 1976 review and three new ones were added as a result of this review (motivation, encouragement of self-initiated activities, and research productivity). Other research (for example, Chau & Hocevar, 1994; Koon & Murray,

1995; Marsh & Bailey, 1993; Marsh & Dunkin, 1997) has supported this view of the multidimensionality of effective teaching.

The instrument for this study was developed using items from this extensive literature on student evaluation and effective teaching. From a pool of 56 items selected from this research literature, duplicates and near-duplicates were eliminated, a uniform style of presentation was achieved, and a resultant pool of 25 items remained. These items are presented in Table 1 along with an overall, or global, measure of teacher effectiveness. All items, including the global measure, were rated on a scale from 1 to 9, where 1 was "not at all descriptive," and 9 was "very descriptive."

Administration

University teachers are usually evaluated by students in their classes at the end of every course. Students rate their instructors on the course

TABLE 1
Item Descriptions

Item Number	Item Description
1	The instructor was knowledgeable about subject matter.
2	The instructor communicated effectively.
3	The instructor was enthusiastic about teaching.
4	The instructor was well prepared for each class.
5	The instructor created a comfortable learning atmosphere.
6	The instructor adapted to student needs.
7	The instructor was tolerant of others' ideas and views.
8	The instructor was genuinely respectful of students.
9	The instructor was warm and friendly.
10	The instructor had a good sense of humor.
11	The instructor motivated students to do their best.
12	The instructor was self-confident.
13	The instructor genuinely enjoyed teaching.
14	The instructor was concerned about student learning.
15	The instructor was able to explain material clearly.
16	The instructor identified important ideas.
17	The instructor used good examples to explain concepts.
18	The instructor was accessible outside of class.
19	The assignments were appropriate in amount and level.
20	The evaluation methods were appropriate.
21	The course increased my interest in the subject matter.
22	The course was well organized.
23	The course materials (text, readings, etc.) were worthwhile.
24	The course improved my understanding of concepts in the field.
25	The course was valuable to me.
Global item	Compared with other college or university instructors I have had, I would rate this instructor as extremely effective.

content, the course delivery, and the personal attributes of the instructor. Often students are asked to identify expected course grades, possibly affecting the evaluations. In this study, students were not asked to consider grades, with the expectation that possible confounding effects would be minimized by using courses from their past.

Students were asked to respond to items about college and university teachers of their choice from their recent past. A training session was held for each group of students participating in the study, in an effort to reduce rater errors so that the reliability and accuracy of their ratings would be enhanced. The training consisted of a brief presentation as well as a written summary describing the biases that can affect student ratings (leniency and halo effect were emphasized), and the necessity of providing the ratings relative to a understandable referent. The students then participated in a brief discussion about responses to items of the variety that they would be using. Students were cautioned not to succumb to these rating errors and to rate teachers carefully, honestly, and accurately. Also, interpretation of the scale was discussed, so that students would understand they would be rating teachers in comparison to other university instructors they had encountered. The global item, "Compared with other college or university instructors I have had, I would rate this instructor as extremely effective," was discussed with regard to the referent the student was to use in making the ratings. Because each student rated only one teacher, the student was to use past encounters with other teachers as a reference for making judgments as to this teacher's effectiveness. It may be noteworthy that, in the course of briefing and training the students prior to conducting their rating of a teacher, judging from the nature and depth of the questions they asked, it was obvious the students came to understand that this study was more about what teacher effectiveness is than about simply evaluating teachers. It became quite obvious that they were interested in characterizing the teacher they rated very accurately.

Analyses and Results

A multiple regression procedure and a discriminant analysis were conducted to reduce the 25 items to a more workable number for further analysis. In the regression analysis, the global measure of teacher effectiveness was regressed onto the 25 predictor items. A discriminant analysis was used to find a subset of predictor items that differentiated effective and ineffective teachers. Results from both the regression analysis and the discriminant analysis were used in the cluster analysis to describe effective teachers.

Student ratings ranged from 1 to 9 on all but one of the 25 items and on the global teacher effectiveness item. Responses on the subject matter knowledge item ranged from 2 to 9. Item means and standard deviations are presented in Table 2. Examination of these shows that, on average, the university teachers that students recalled were rated very high on subject matter knowledge ($M = 8.06$) and above 6 on every item.

In the regression analysis, a maximum multiple correlation procedure was used to identify a set of items that, as a group, produced the highest multiple correlation with the global measure. Six items, value of the course, motivating students to do their best, comfortable learning atmosphere, course organization, effective communication, and concern for student learning accounted for 87% of the variability in the criterion of teacher effectiveness (see Table 3). Adding the remaining 19 items to the model increased the explained variance by only 1%; in other words,

TABLE 2

Item Means and Standard Deviations

Item Number	Item Description	Mean	Standard Deviation
1	subject matter knowledge	8.03	1.38
2	effective communication	6.89	2.09
3	enthusiasm for teaching	7.17	1.99
4	preparation for class	7.21	1.99
5	comfortable learning atmosphere	6.53	2.37
6	adapting to student needs	6.17	2.39
7	tolerant of others' ideas	6.59	2.31
8	genuine respect for students	6.96	2.23
9	warm and friendly	6.89	2.32
10	good sense of humor	6.83	2.30
11	motivating students to do their best	6.38	2.37
12	self-confident	7.66	1.66
13	enjoy teaching	7.24	1.94
14	concern for student learning	6.87	2.11
15	clear explanations	6.65	2.30
16	identify important ideas	7.02	2.07
17	use of good examples	6.93	2.09
18	accessible outside of class	6.61	2.11
19	appropriate assignments	6.70	2.13
20	appropriate evaluation methods	6.67	2.27
21	increased interest in subject matter	6.59	2.61
22	course organization	6.69	2.37
23	worthwhile materials	6.49	2.28
24	improved understanding of concepts	6.99	2.56
25	value of the course	6.83	2.46
Global	effective teaching	6.58	2.46

NOTE: $n = 912$ for all items.

TABLE 3
Summary of the Maximum Multiple Correlation Regression

Item Number	Item Description	Total R^2
25	value of the course	0.690
11	motivating students to do their best	0.796
5	comfortable learning atmosphere	0.852
22	course organization	0.861
2	effective communication	0.863
14	concern for student learning	0.869

NOTE: $n = 912$.

nearly all of the variance in teacher effectiveness was captured by the
6 items, and adding the other items increased the explanation very little.

A discriminant analysis was conducted on two groups of teachers so
that the variables that best differentiate between effective and ineffective
teachers could be found. The teachers scoring high on the global mea-
sure of teacher effectiveness were collapsed into one group and were
characterized as the effective teachers. These were 589 teachers whom
students rated as 7, 8, or 9 on the global item. Teachers who scored low
on teacher effectiveness were identified as the ineffective teachers. This
group included 149 teachers, who received student ratings of 1, 2, or 3
on the global item. Five items were found to differentiate significantly
between the two groups of teachers, accounting for 82% of the variance
in the dichotomous dependent variable: value of the course, motivating
students to do their best, effective communication, course organization,
and genuine respect for the student. Table 4 shows the means and stan-
dard deviations for each of the 5 items. The item that was found to be
most important in characterizing the effective teachers was value of the
course. Ineffective teachers were viewed as those who did not motivate
their students to learn. The discriminant function generated using these
items predicted group membership 97% correctly for the ineffective
teachers and 99% for the effective teachers.

In all, 7 unique items were found to account for teacher effectiveness in
the regression and discriminant analyses. Four items, value of the course,
motivating students, course organization, and effective communication,
were identified in both the regression and discriminant analyses. Three ad-
ditional items, comfortable learning atmosphere and concern for student
learning, from the regression analysis, and genuine respect for students,
from the discriminant analysis, were also selected. These 7 items were
submitted to clustering procedures in an attempt to find groups of effec-
tive teachers that differed in their profiles on these descriptor variables.

TABLE 4
Means and Standard Deviations for Items in the Discriminant Function by Group

		Effective Teachers		Ineffective Teachers	
Item Number	Item Description	Mean	SD	Mean	SD
25	value of the course	8.17	1.14	2.93	1.96
11	motivating students	7.65	1.37	2.77	1.59
2	effective communication	7.99	1.10	3.98	1.93
22	course organization	7.85	1.33	3.47	2.36
8	genuine respect for students	7.94	1.37	4.16	2.43

NOTE: $n = 589$ for the effective teachers and $n = 149$ for the ineffective teachers.

Only the teachers rated 9[1] on the global item ($n = 246$) were used in the cluster analysis. These teachers were rated as the most effective, and the profiles of the groups of variables describing them were the primary focus for this study. Scores on the items were not standardized because items were equally scaled. Due to the large number of similarities required in this analysis, both divisive and hierarchical clustering methods were used.

Divisive clustering reduced the 246 teachers into 35 clusters, placing each teacher into only one cluster. These 35 clusters of the 246 highly effective teachers were further clustered in a hierarchical fashion.

The hierarchical clustering procedure combined the two closest clusters and continued the process iteratively until all clusters were reduced to an interpretable set of clusters. Similarity between data points was found using Euclidean distance. Average linkage was used as the clustering method because it is not distorted by outliers and generally maintains the original structure of the data. Five clusters of effective teachers were found, and their profiles will be described beginning with the largest and most dense group.

Profile 1

The largest cluster of teachers ($n = 107$) may be described as receiving high ratings on all items considered in the analysis. Figure 1 illustrates that each item has a mean score above 8.84 for the teachers in this cluster.

Profile 2

This group of teachers ($n = 26$) was very similar to Profile 1. They had very high mean ratings on all items except for item 2, effective communication. The mean for this item was 7.50, compared to means of 8.35 and above for all other items. Figure 2 displays the relationships.
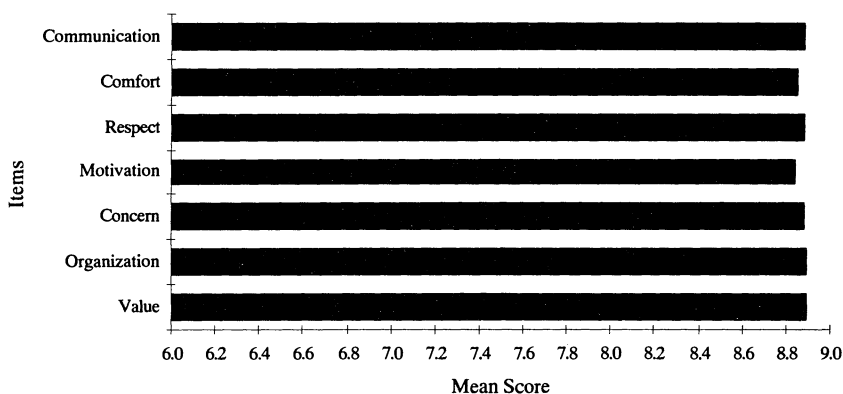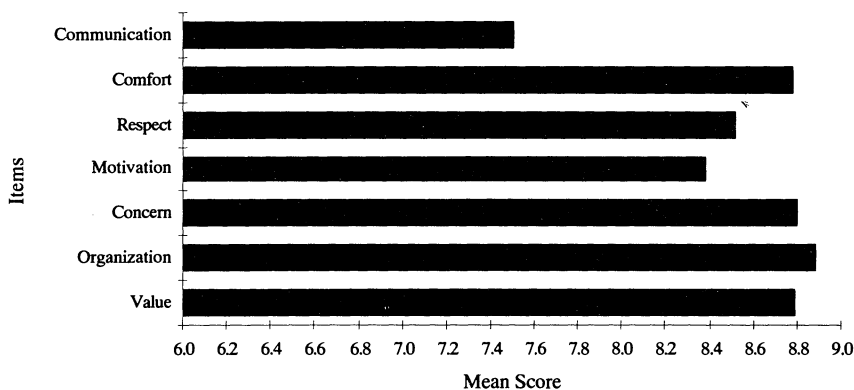
FIG. 1. Profile 1

FIG. 2. Profile 2

Profile 3

This third cluster of teachers was characterized as having only a moderately high mean rating, 7.04, on item 22, course organization, as compared to means of the other items. All other item means for these 26 teachers were greater than 8.31. Means for each item are shown in Figure 3.

Profile 4

The Fourth Cluster included teachers who were rated high on all items but had only a moderately high mean rating, 6.58, on item 11, motivating students to do their best. These teachers ($n = 19$) had mean scores above 8.26 on all items except item 11 (see Figure 4).
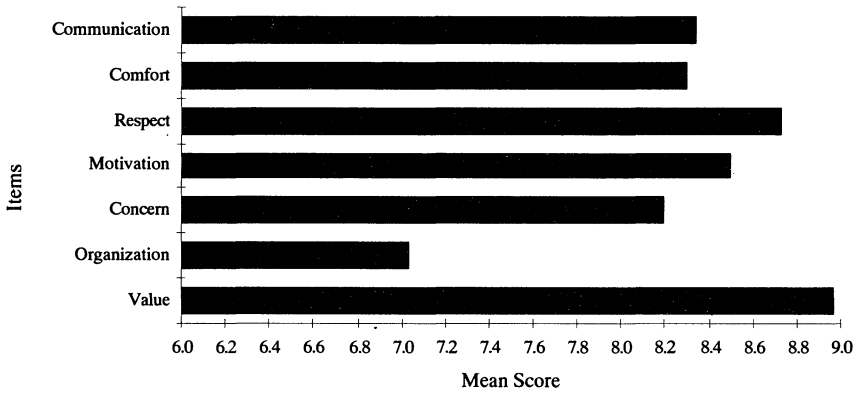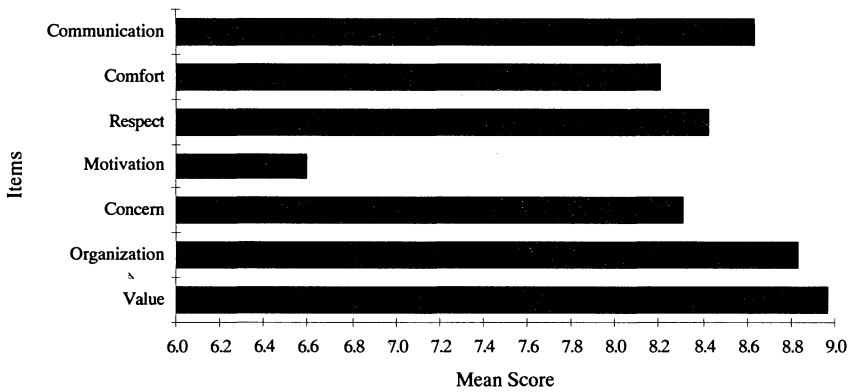
FIG. 3. Profile 3

FIG. 4. Profile 4

Profile 5

Seventeen teachers were profiled as having a moderately high mean rating on item 5, creating a comfortable learning atmosphere. The mean score for this item was 6.82, while the other items all had means above 8.24. The means for all 7 items are shown in Figure 5.

The cluster analysis identified five profiles of effective teachers that were very much alike and yet different in specific ways. All 246 teachers were rated high on most of the seven variables. Each profile, with the exception of profile 1, included 1 item on which the teachers were rated substantially lower but still above average, as compared to the other items in the analysis. However, the means for most items were very high.
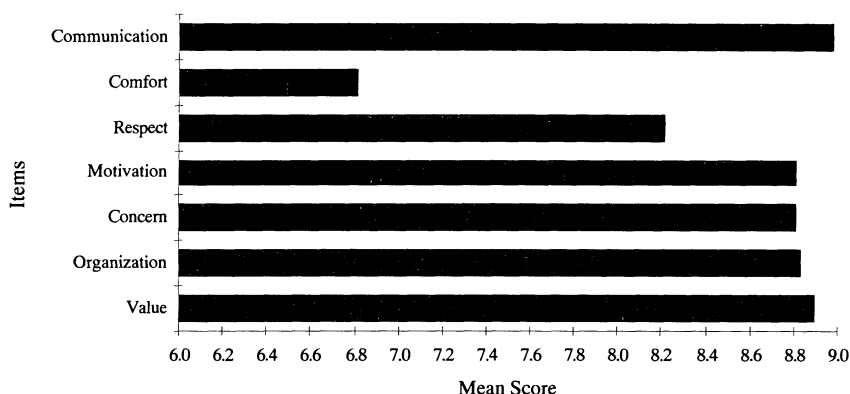
FIG. 5. Profile 5

Fifty-one teachers in this high group were not included in any of the five clusters because their profiles were unique. Although these teachers might be considered outliers, it is possible that instead they represent other clusters that were poorly sampled. However, these 51 teachers could be characterized as being rated *very* low, even as low as 1 in some cases, on up to 3 of the 7 items, yet the teachers were rated 9 overall on the global item. For all of these 51 teachers, it should also be noted that they were rated 9 on at least 2 other items.

Summary and Discussion

Is an effective teacher one who encourages students to think for themselves or one who provides structure, organization, and varied examples of concepts? Is a teacher who is organized but lacks enthusiasm perceived as being as effective as one who is enthusiastic yet provides little structure? The research literature on teacher effectiveness has clearly identified many of the important ingredients of good teaching (Perry & Smart, 1997). The findings of this study may serve to strengthen theoretical models and reveal some potentially important and intriguing possibilities.

Regression techniques produced a model of teacher effectiveness that had a high degree of predictive power with a small number of items. It was not a surprise that effective communication, a comfortable learning atmosphere, concern for student learning, student motivation, and course organization were found to be highly related, as a group, to the criterion measure of teacher effectiveness. This finding supports past and current research in this field. However, it was not expected that the value of the course would emerge so strongly as a predictor of teacher effectiveness in the analysis. This item was highly correlated with the

global measure for graduate students separately, undergraduates separately, and the groups combined. It entered the regression model first and was not removed at any step of the analysis. It also entered the discriminant model first in differentiating between effective and ineffective teachers. It was rated very high in each of the identified profiles of effective teachers. In this study, the worth of a course for the university students was the most important predictor of teacher effectiveness and this may be one of the most significant findings to emerge from this investigation. It is possible that course value was so important because, as time passed, students became much more aware of its usefulness. Recognizing that students believe that the relevancy and value of a course are so important may provide college and university teachers with a valuable insight into understanding effective teaching.

The cluster analysis clearly profiled the highly effective teachers ($n = 246$) into five natural groups. The items used in the cluster analysis were a subset of 7 of the original 25 course and teacher characteristics. A very large group of teachers ($n = 107$) was rated high, 8.84 and above, on all 7 items. Four smaller groups of teachers, ranging in size from 26 to 17, were each rated moderately high on 1 item and very high on all the others. Three items were not rated lower in any of the clusters; genuine respect for students, concern for student learning, and value of the course were consistently rated very high. Teachers who were recognized as highly effective in this study, then, were rated very high on at least 6 of the 7 items but were rated, on average, as low as 6.6 on the 7th item, and still considered very effective by students.

Results from the clustering procedures will likely provide the greatest impetus for further research in this field. To be rated as effective, a teacher had to be rated very high on genuine respect for students, concern for student learning, and value of the course. Very effective teachers are not necessarily rated high on every other important variable; they can have some deficiencies. McKeachie (1997), in his summary commentary on several articles addressing the validity of student ratings, suggested that "effective teachers come in all shapes and sizes" (p. 1218). A definition of teacher effectiveness must take this into account, and measurement of the construct may not be captured well by an additive model. This analysis illustrated that effective teachers were generally regarded by students as high on all variables included in this analysis but not equally high on all important characteristics. It appears that effective teachers can compensate for deficiencies in one or two areas by demonstrating outstanding skills in other areas. For example, one teacher in cluster 3 was rated only a 4 on organization with very high ratings on the other 6 items and was perceived to be a highly effective

teacher. Even though organization was generally seen as an important correlate of effective teaching, organization is not a necessary characteristic if a teacher is excellent in other aspects. According to the students in this study, teachers were considered effective despite a single low rating on a characteristic that was seen as very important.

Finally, the inclusion of the course value item in this study and its prominence in the analyses deserves additional comment. Asking students to rate teachers after a course has been completed allows students to reflect on the effectiveness of the teacher to deliver a valuable course. To our way of thinking, regardless of how difficult or mundane the course content is, the truly effective teacher of that course is assuming some responsibility for making it a valuable course for the student. It may well be that ratings collected *as courses end* show a substantial separation between overall course value and overall teacher effectiveness (Marsh, 1984), but ratings collected *after courses are completed,* show that student raters do not disentangle these two items to the same degree.

Notes

[1]In the discriminant analysis conducted prior to the clustering procedure, we used cut points that combined teachers rated overall 1–3 and 7–9 into the groups labeled "ineffective" and "effective," respectively. We did this because the focus in this analysis was on the identification of *items* that discriminate between persons at opposite ends of the effectiveness continuum. The choice of cut points is not overly critical here; our rationale was that 7, 8 , and 9 represent degrees of effective teachers, with 7 representing "well above average" to 9 representing something akin to "outstanding," all being effective to some degree. In the clustering procedure, however, we used only the teachers rated by their students as 9s, because we were focussing on definitional matters that are likely anchored best by teachers at the very highest point on the continuum of effectiveness.

References

Abrami, P. C., d'Apollonia, S., & Cohen, P. A. (1990). Validity of student ratings of instruction: What we know and what we do not. *Journal of Educational Psychology, 82,* 219–231.

Abrami, P. C., d'Apollonia, S., & Rosenfield, S. (1997). The dimensionality of student ratings of instruction: What we know and what we do not. In R. P. Perry & J. C. Smart (Eds.), *Effective teaching in higher education* (pp. 321–367). New York: Agathon.

Arubayi, E. A. (1986). Students' evaluations of instruction in higher education: A review. *Assessment and Evaluation in Higher Education, 11,* 1–10.

Centra, J. A. (1994). The use of the teaching portfolio and student evaluations for summative evaluations. *Journal of Higher Education, 65,* 555–570.

Chau, H., & Hocevar, D. (1994, April). *Higher-order factor analysis of multidimensional students' evaluations of teaching effectiveness.* Paper presented at the annual conference of the American Educational Research Association, New Orleans, LA.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research, 51,* 281–309.

Cohen, P. A. (1987, April). *A critical analysis and reanalysis of the multisection validity meta-analysis.* Paper presented at the 1987 annual meeting of the American Educational Research Association. Washington, DC (ERIC Document Reproduction Service No. ED 283 876).

d'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist, 52,* 1198–1208.

Feldman, K. A. (1976). The superior college teacher from the student's view. *Research in Higher Education, 5,* 243–288.

Feldman, K. A. (1986). The perceived instructional effectiveness of college teachers as related to evaluations they receive from students. *Research in Higher Education, 18,* 3–124.

Feldman, K. A. (1988). Effective college teaching from the students' and facultys' view: Matched or mismatched priorities. *Research in Higher Education, 28,* 291–344.

Feldman, K. A. (1989). Instructional effectiveness of college teachers as judged by teachers themselves, current and former students, colleagues, administrators, and external (neutral) observers. *Research in Higher Education, 30,* 113–135.

Greenwald, A. G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist, 52,* 1182–1186.

Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist, 52,* 1209–1217.

Koon, J., & Murray, H. G. (1995). Using multiple outcomes to validate student ratings of overall teacher effectiveness. *Journal of Higher Education, 66,* 61–81.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology, 76,* 707–754.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for further research. *Journal of Educational Research, 11,* 253–388.

Marsh, H. W., & Bailey, M. (1993). Multidimensional students' evaluations of teaching effectiveness. *Journal of Higher Education, 64,* 1–18.

Marsh, H. W., & Dunkin, M. J. (1997). Students' evaluations of university teaching: A multidimensional perspective. In R. P. Perry & J. C. Smart (Eds.) *Effective teaching in higher education* (pp. 241–320). New York: Agathon.

Marsh, H. W., & Overall, J. U. (1979). Long-term stability of students' evaluations of teaching effectiveness: A note on Feldman's "Consistency and variability among college students in rating their teachers and courses." *Research in Higher Education, 10,* 139–147.

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective. *American Psychologist, 52,* 1187–1197.

McKeachie, W. J. (1983). The role of faculty evaluation in enhancing college teaching. *National Forum, 63*(1) 37–39.

McKeachie, W. J. (1997). Student ratings. *American Psychologist, 52,* 1218–1225.

Miller, A. H. (1988). Student assessment of teaching in higher education. *Higher Education, 17,* 3–15.

Murray, H. G., Rushton, J. P., & Paunonen, S. V. (1990). Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology, 82,* 250–261.

Naftulin, D. H., Ware, J. E., Jr., & Donnelly, F. A. (1973). The Dr. Fox lecture: A paradigm of educational seduction. *Journal of Medical Evaluation, 48,* 630–635.

Overall, J. U., & Marsh, H. W. (1980). Students' evaluations of instruction: A longitudinal study of their stability. *Journal of Educational Psychology, 72,* 321–325.

Perry, R. P., & Smart J. C. (Eds.). (1997). *Effective teaching in higher education.* New York: Agathon.

Renaud, R. D., & Murray, H. G. (1996). Aging, personality, and teaching effectiveness in academic psychologists. *Research in Higher Education, 37,* 323–340.