Validity concerns and usefulness
of student ratings of instruction

Anthony G. Greenwald, University of Washington

Greenwald served as editor for this collection of articles. In his introductory piece, he relates an experience he had during 1989-90. In 1989, he received the highest student ratings he ever received; a year later, teaching the same course with only slight modifications to the syllabus, he received his lowest ratings. The two sets of scores were 2.5 standard deviations apart, representing an 8 decile separation according to the university's norms.

The experience spurred him to read up on the literature about SETs (student evaluations of teaching) and to collect more data of his own. He begins here by summarizing historical trends in research on ratings, from the early 1970s to date. Electronic searches of PsychINFO and ERIC data bases indicate that activity in this area seems to have peaked with 71 publications in the 5-year period 1976-80, shrinking to a low of 8 publications in 1991-95. The 1976-80 period saw the largest proportion of publications critical of the validity of the SETs.

During the 1970s, a major source of concern was the possibility that grading practices were biasing the evaluations. Some experimental tests were done to show that manipulating grades upwards or downwards indeed influenced ratings. Although later authors criticized the methodology of these studies, Greenwald maintains that their conclusions have never been empirically refuted.

Publications in the 1980s focused on the "convergent validity" of student ratings; that is, the extent to which the ratings are correlated with other measures of teaching effectiveness. Among other things, this research pointed out that correlation between grades and ratings did not necessarily represent contamination of the ratings by grading practices, but might be attributable to third variables such as student motivation.

One might conclude from the publication record that earlier concerns about validity had been settled. But the four articles which follow each address different points of concern in this regard..

Discussion Questions

1. Do you think the overall distribution of teaching ratings at University follows a normal curve. Assuming that it did, what would the fact that a 2.5 standard deviation swing corresponded to 8 deciles tell you about the positions of these rankings on the curve?

2. What are some of the factors that could lead to such a dramatic swing?