

Relation Between Aging and Rated Teaching Effectiveness of Academic Psychologists

Karen L. Horner, Harry G. Murray, and J. Philippe Rushton
University of Western Ontario, London, Ontario, Canada

Within a combined cross-sectional and longitudinal design, several analyses showed that student-rated teaching effectiveness in university professors declined with age. For 106 full-time psychology teachers, aged 26 to 55, who were studied for time periods ranging from 2 to 15 years, an overall negative correlation of $r = -.33$ ($p < .05$) was found between age and general teaching effectiveness. A similar decrement with age occurred for items measuring specific teaching behaviors. Age accounted for 8% of the variance in general teaching effectiveness.

The relation between age and faculty performance has received considerable attention for more than 50 years. Currently, it may be particularly salient given the increasing average age of university faculty. The two indicators of faculty performance typically studied are teaching effectiveness and research productivity. In the case of research productivity, several investigators have found a curvilinear age function, with productivity beginning at a relatively low rate in the late 20s, peaking around 40, and declining thereafter (Cole, 1979; Dennis, 1956; Horner, Rushton, & Vernon, 1986; Lehman, 1953). With respect to teaching effectiveness, investigators have provided less definitive results. Recent reviewers have concluded that either no relation or a negative relation exists between age and quality of teaching (Blackburn & Lawrence, 1986; Feldman, 1983).

The usual indicator of teaching quality in prior research has been some form of student rating scale. A general review of what is known about the reliability and validity of student instructional ratings was recently provided by Murray (1984). In regard to year-to-year reliability for the same instructor and course, there is clear evidence for high stability of scores on scales such as that shown in Table 1. Retest reliabilities of .80 to .90 are typical and do not vary appreciably as a function of the type of course being taught or whether alternate rating scales are being used. Student ratings of classroom teaching have also been found to correlate between .50 and .90 with comparable ratings made by supervisors, colleagues, alumni, and paid classroom observers, indicating that student perceptions of good and poor teaching are similar to those of more expert, more mature, and more neutral observers (Murray, 1984).

Further evidence for the validity of student instructional ratings comes from studies reporting a moderate positive correlation between student ratings and objective measures of student achievement. As Murray (1984) discussed, most studies of this

type have been done in the context of a multiple-section course with a common, objectively scored final examination. This makes it possible to correlate mean student ratings with mean final examination scores across instructors or class sections. An average correlation of .43 between instructor ratings and student examination performance was reported by Cohen (1981), indicating that highly rated teachers do in fact tend to stimulate higher levels of student learning than less highly rated teachers.

Longitudinal studies of the relation between aging and student instructional ratings are curiously scarce. Most longitudinal research has collapsed across age groups and investigated changes in teaching effectiveness solely as a function of chronological time. An early finding by Heilman and Armentrout (1936), which remains representative, showed stability in teaching effectiveness ratings, collapsed across age, over a period ranging from 5 to 7 years.

Most studies of the relation between teacher effectiveness and age have used cross-sectional or between-subjects designs. For example, in a study of student ratings obtained for 70 college faculty in one semester, Cornwell (1974) showed that teacher's age was negatively related to teaching effectiveness, accounting for 6% of the criterion variance. Similarly, using years since PhD as an approximation of age in more than 4,000 faculty from a variety of disciplines at 16 colleges and universities, Linsky and Straus (1975) found teaching effectiveness to be negatively related to years since PhD when these variables were compared cross-sectionally.

Two problems are apparent with most studies relating age to teaching performance. First, age has generally been considered a between-subjects variable at one measurement time, rather than a within-subjects variable longitudinally. Although cross-sectional findings may be indicative of longitudinal performance, they may also be a function of cohort effects. That is, performance of different age groups may vary, not as a function of age but rather as the result of distinct generational influences that are variations in life experiences of groups of persons entering a social system at different points in time. Second, studies that have used longitudinal designs, intended to study stability, have collapsed across age groups, thus obscuring age effects. Most studies of this sort have reported stability across even relatively long periods of time. For example, Heilman and Armen-

Financial support was provided by an Ontario Ministry of Health Fellowship award to Karen L. Horner and by a grant from the Gerontology Research Council of Ontario to J. Philippe Rushton.

Correspondence concerning this article should be addressed to Karen L. Horner, Department of Psychology, University of Western Ontario, London, Ontario, Canada N6A 5C2.

Table 1
Teacher-Rating Form

1. The instructor is a good speaker.
2. The instructor is well prepared for classes.
3. The instructor presents material in a well-organized and coherent manner.
4. The instructor is able to explain difficult concepts in a clear and straightforward way.
5. The instructor makes effective use of examples and illustrations in explaining course materials.
6. Considering limitations due to class size, the instructor does a good job of answering questions that are asked in class sessions.
7. The instructor is enthusiastic about the subject matter.
8. Considering inherent limitations of the course content, the instructor is successful in presenting the subject matter in an interesting way.
9. The instructor is successful in encouraging students to think independently and do supplementary reading related to the subject matter of the course.
10. How would you rate your instructor in terms of general, overall effectiveness as a teacher?

Note. The rating scale is as follows: *strongly agree/outstanding* (5), *agree/very good* (4), *undecided/good* (3), *disagree/satisfactory* (2), *strongly disagree/poor* (1).

trout (1936) obtained stable instructor ratings across 7 years. In this type of design, an increase in effectiveness for younger teachers at one measurement time can be canceled by a decrease in effectiveness for older teachers at the same measurement time, thus masking any potential age differences in effectiveness and giving the illusion of greater stability over time than exists.

The present study was conducted to investigate more clearly and decisively the relation between teacher effectiveness and age over time. It was predicted that, as with research productivity, performance as a teacher would decline as the faculty member aged. This study is a combined cross-sectional and longitudinal design and goes beyond past research in that age groups were specifically defined. Moreover, rated effectiveness of various ages was assessed for up to 15 years, a time span considerably longer than that of previous research.

Method

Subjects

The sample consisted of 106 teachers who held full-time faculty appointments of at least 2 years duration within the period from autumn 1972 to spring 1987 in the Department of Psychology, University of Western Ontario. Age was obtained from the *APA Membership Register* (American Psychological Association, 1985) and from departmental records. Because of the small number of women, analyses were collapsed across sex.

Measure of Teacher Effectiveness

Beginning in the autumn of 1972, all of the teachers in the department were rated by students on the standardized questionnaire shown in Table 1. Ratings were obtained for each undergraduate course taught in each year. If more than one undergraduate course was taught in the same year by a given faculty member, a mean rating was used for that year. If fewer than six students contributed ratings, data for that course

were excluded from analysis. The average internal consistency reliability for the Teacher-Rating Form was .97.

Results

Within each of the 15 years from 1972 to 1986, Pearson product-moment correlations were computed between age and rated effectiveness on each of the 10 items. Within each year, age ranged from the late 20s to the late 50s. Sample size for each year ranged from 35 to 46. For the general item measuring overall effectiveness (Item 10), all of the correlations were negative, ranging from $-.06$ to $-.55$, with a mean Fisher z -transformed $r = -.33$ ($p < .05$). Cross-sectional correlations between age and rated overall effectiveness within each of the 15 years were also computed for eight teachers holding appointments for the complete 15-year period. This analysis yielded a mean correlation of $-.25$ (ns). Similar negative correlations were found for each of the nine specific teaching behaviors (Items 1-9) with z -transformed mean correlations ranging from $-.11$ (ns) to $-.36$ ($p < .05$).

Longitudinal analyses also examined age changes in teaching effectiveness. For the complete sample, and collapsing across assessment years, Figure 1 shows the mean ratings for 15 ages in 2-year periods from age 26-27 to age 54-55. Trend analyses were computed in which the sum of the products of each mean and its respective linear or quadratic coefficient was calculated as a proportion of the error variance. The amount of variance accounted for by the trends was also determined by eta-squared calculations. Significant linear trends were found for age for both the overall effectiveness rating (Item 10), accounting for 8.0% of the variance ($F = 29.87$, $p < .05$), as well as for each of Items 1-9 with the range of variance explained from 1.6% to 8.5% ($F = 5.316$ to $F = 32.29$, all $ps < .05$). In general, quadratic trends failed to reach significance.

Discussion

The results of this study show that across several types of analysis and several teaching behaviors, both specific and general, rated teaching effectiveness declines as teachers age, a finding that is consistent with past research (e.g., Cornwell,

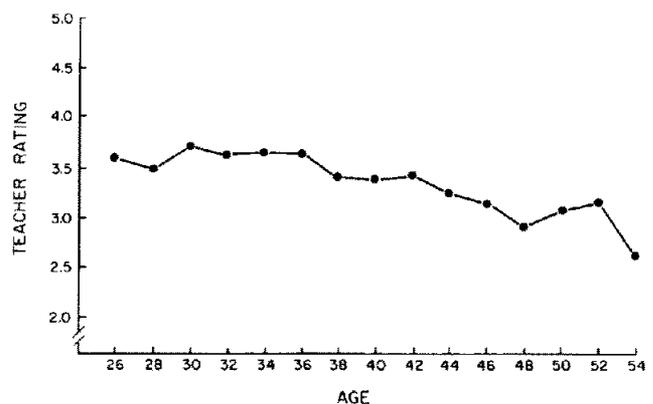


Figure 1. Mean effectiveness rating at 15 age intervals for 106 teachers rated for at least 2 years.

1974). Moreover, the negative relation may be more robust than has been shown here, inasmuch as younger teachers are often assigned large introductory courses (Baldwin & Blackburn, 1981), and large classes tend to give lower overall instructor effectiveness ratings than do small classes (Neumann & Neumann, 1985). In other words, the ratings of younger instructors may have been artificially suppressed in this study, thus weakening the overall negative relation between age and ratings. Although there has been much speculation that the possible curvilinear relation between rated teaching effectiveness and age has been obscured by sole reliance upon linear analyses (Blackburn & Lawrence, 1986; Feldman, 1983), the present study found only a linear relation between these two variables in an orthogonal polynomials analysis. Moreover, the general linear effect size of 8% for age is consistent with the 6% effect size found by Cornwell (1974).

One problem with the present study is the lack of available data over a long period of time on both cross-sectional and longitudinal bases. Teacher ratings were not systematically acquired until relatively recently, and therefore longitudinal studies have been restricted to short periods of time. Even the 15-year period used here, whereas substantially longer than previous research still remains a relatively short time span over which to assess stability and change. There is also the problem of longitudinal sample size, inasmuch as not everyone begins his or her appointment at the same age. For any group of teachers holding long-term appointments, the sample size at different ages will be extremely variable.

Lawrence and Blackburn (1988) argued that cohort effects may explain more of the variance in faculty performance than do age effects. Horner et al. (1986) found no support for this hypothesis in the case of research productivity, in which age effects outweighed cohort effects. The relative weighting could not be examined in the present study because of the problems noted previously—too short a time span and too small a sample to subdivide into smaller age groups.

One of the anticipated outcomes of instructor-rating forms is the improvement of teaching performance through feedback. In the short-term, Cohen (1980) showed through a meta-analysis of 22 studies that instruction does not improve substantially from ratings feedback alone. In addition, from the present study, it appears that teaching instruction does not necessarily improve on a long-term basis as a result of annual feedback alone.

On the assumption that a negative relation between age and teaching effectiveness continues to emerge in future research, several perhaps complementary influences could be posited as contributing to age-related decrements in teaching quality. For example, individual factors such as slowing of biological and intellectual processes (Shock, 1985) and/or changes in personality variables including motivation (Eysenck, in press) might be operative, which in turn will decrease perceived effectiveness as teachers. In addition, Blackburn and Lawrence (1986) argued that external rewards for good teaching are scarce. Thus, once internal variables begin to affect teaching quality, there are few external incentives to reverse the negative trend.

An alternative explanation for the decline in rated effectiveness with age is a negative halo effect on the students' perception of teaching behaviors. That is, as the teachers age, they become

members of a group generally believed to be less competent in a variety of physical and intellectual functions (Schaie, 1988). Although there is some evidence that age stereotyping may occur in job performance ratings, this appears to depend on the behavior being rated, the rank of the employee with respect to the rater, the occupation of the ratee, and whether the target ratee is seen as part of a group or as an individual (Braithwaite, 1986; Cleveland & Landy, 1981; Waldman & Avolio, 1986). Because the age effect was found both for the overall effectiveness item in Table 1 and for items assessing more specific teaching behaviors, no definitive conclusions can be reached regarding the potential influence of an age bias in the present study. Nevertheless, the fact that (a) the teachers are professionals, (b) they were rated as individuals, and (c) the age effect varied slightly for specific teaching behaviors may suggest that the students were rating behaviors as they appeared rather than according to an age bias (Braithwaite, 1986; Waldman & Avolio, 1986). Another alternative to actual age declines in teaching effectiveness is the *generation-gap* hypothesis, that is, the hypothesis that older faculty are perceived by younger students as less effective. This could be put to the test using older students who, if generation similarity were important, may even be expected to rate the younger faculty as less effective. Note, however, that in the review by Murray (1984), ratings made of teaching effectiveness tend to be stable across diverse raters.

Although there appears to be a small generalized decline with age in the teaching behaviors rated in the present study, considerable evidence suggests a strong influence of individual differences on the rate and pervasiveness of the aging process (Shock, 1985). This evidence extends to behaviors relevant to university professors (Horner et al., 1986). The perceived decline in teaching, therefore, may be expected to manifest itself at different ages interindividually and be greater for some persons than for others. It was not possible to test these effects in the present study because of the small sample size. Nevertheless, this remains an important issue. Indeed, because teaching and research abilities appear to be orthogonal to each other and to correlate differentially with established dimensions of personality (Rushton, Murray, & Paunonen, 1983), which in turn reliably alter with age (Eysenck, in press), these too could profitably be taken into account. For example, it might be predicted from this literature that extraverts will be more susceptible to age declines in research productivity, whereas introverts may suffer more rapid declines in teaching abilities.

Relevant to these last points, there is some suggestion that as faculty members age, they spend increasing time in managerial or administrative activities, serving on committees or as consultants. Thus, rather than having impact only through research or teaching, older faculty may also have a major role to play through administration. This third area of faculty performance remains a fascinating topic for future research (Rushton et al., 1983). For example, What are the characteristics necessary for success as a good administrator? How do these differ from the qualities necessary for success in other academic areas? How do these qualities change with age?

References

- American Psychological Association. (1985). *APA membership register*. Washington, DC: Author.

- Baldwin, R. G., & Blackburn, R. T. (1981). The academic career as a developmental process. *Journal of Higher Education*, 52, 598-614.
- Blackburn, R. T., & Lawrence, J. H. (1986). Aging and the quality of faculty job performance. *Review of Educational Research*, 56, 265-290.
- Braithwaite, V. A. (1986). Old age stereotypes: Reconciling contradictions. *Journal of Gerontology*, 41, 353-360.
- Cleveland, J. N., & Landy, F. J. (1981). The influence of rater and ratee age on two performance judgements. *Personnel Psychology*, 34, 19-29.
- Cohen, P. A. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. *Research in Higher Education*, 13, 321-341.
- 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.
- Cole, S. (1979). Age and scientific performance. *American Journal of Sociology*, 84, 958-977.
- Cornwell, C. D. (1974). Statistical treatment of data from student teaching evaluation questionnaires. *Journal of Chemical Education*, 51, 155-160.
- Dennis, W. (1956). Age and productivity among scientists. *Science*, 123, 724-725.
- Eysenck, H. J. (in press). Personality and aging: An exploratory analysis. *Journal of Social Behavior and Personality*.
- Feldman, K. A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. *Research in Higher Education*, 18, 3-124.
- Heilman, J. D., & Armentrout, W. D. (1936). The ratings of college teachers by their students. *The Journal of Educational Psychology*, 27, 197-216.
- Horner, K. L., Rushton, J. P., & Vernon, P. A. (1986). Relation between aging and research productivity of academic psychologists. *Psychology and Aging*, 1, 319-324.
- Lawrence, J. H., & Blackburn, R. T. (1988). Age as a predictor of faculty productivity: Three conceptual approaches. *Journal of Higher Education*, 59, 22-38.
- Lehman, H. (1953). *Age and achievement*. Princeton, NJ: Princeton University Press.
- Linsky, A. S., & Straus, M. A. (1975). Student evaluations, research productivity, and eminence of college faculty. *Journal of Higher Education*, 46, 89-102.
- Murray, H. G. (1984). The impact of formative and summative evaluation of teaching in North American universities. *Assessment and Evaluation in Higher Education*, 9, 117-132.
- Neumann, L., & Neumann, Y. (1985). Determinants of students' instructional evaluation: A comparison of four levels of academic areas. *Journal of Educational Research*, 78, 152-158.
- Rushton, J. P., Murray, H. G., & Paunonen, S. V. (1983). Personality, research creativity, and teaching effectiveness in university professors. *Scientometrics*, 5, 93-116.
- Schaie, K. W. (1988). Ageism in psychological research. *American Psychologist*, 43, 179-183.
- Shock, N. W. (1985). Longitudinal studies of aging in humans. In C. E. Finch & E. L. Schneider (Eds.), *Handbook of the biology of aging* (pp. 721-743). New York: Van Nostrand Reinhold.
- Waldman, D. A., & Avolio, B. J. (1986). A meta-analysis of age differences in job performance. *Journal of Applied Psychology*, 71, 33-38.

Received April 5, 1988

Revision received August 8, 1988

Accepted August 10, 1988 ■

Squire Appointed Editor of *Behavioral Neuroscience*, 1990-1995

The Publications and Communications Board of the American Psychological Association announces the appointment of Larry R. Squire, Veterans Administration Medical Center and University of California, San Diego, as editor of *Behavioral Neuroscience* for a 6-year term beginning in 1990. As of January 1, 1989, manuscripts should be directed to

Larry R. Squire
 Veterans Administration Medical Center (116 A)
 3350 La Jolla Village Drive
 San Diego, California 92161