

Half a Minute: Predicting Teacher Evaluations From Thin Slices of Nonverbal Behavior and Physical Attractiveness

Nalini Ambady and Robert Rosenthal

The accuracy of strangers' consensual judgments of personality based on "thin slices" of targets' nonverbal behavior were examined in relation to an ecologically valid criterion variable. In the 1st study, consensual judgments of college teachers' molar nonverbal behavior based on very brief (under 30 s) silent video clips significantly predicted global end-of-semester student evaluations of teachers. In the 2nd study, similar judgments predicted a principal's ratings of high school teachers. In the 3rd study, ratings of even thinner slices (6-s and 15-s clips) were strongly related to the criterion variables. Ratings of specific micrononverbal behaviors and ratings of teachers' physical attractiveness were not as strongly related to the criterion variable. These findings have important implications for the areas of personality judgment, impression formation, and nonverbal behavior.

The ability to form impressions of others is a critical human skill. "This remarkable capacity we possess to understand something of the character of another person, to form a conception of him as a human being . . . with particular characteristics forming a distinct individuality is a precondition of social life" (Asch, 1946, p. 258). In the present article, we show that this capacity is even more remarkable than Asch suggested: Our consensual impressions of others, even when based on very brief observations of nonverbal behavior, can sometimes be unexpectedly accurate.

Kruglanski (1989) outlined the following definitions of the construct of accuracy in personality and social psychological research: (a) the degree of correspondence between a judgment and a criterion (the most popular definition in psychological research), (b) interpersonal consensus, and (c) a construct possessing pragmatic utility. Much of the recent research on impression formation and personality judgment has been focused on the second definition regarding interpersonal consensus among individuals in their judgments regarding the personality traits of others. This research has revealed three quite surprising findings regarding strangers' judgments of others. First, people can concur remarkably in some of their judgments of complete strangers, thus exhibiting high consensual accuracy (Albright, Kenny, & Malloy, 1988; Kenny, Horner, Kashy, & Chu, 1992; Paunonen, 1991). Second, these judgments can be

unexpectedly accurate in predicting targets' self-reported traits, even when they are based on very superficial interactions (Albright et al., 1988; Passini & Norman, 1966; Watson, 1989). Third, these impressions can be accurate, as defined both by high consensus and by the prediction of a criterion (self-ratings), even when there is absolutely no interaction between the targets and raters—for example, when impressions are based on 5-min videoclips of the targets (Colvin & Funder, 1991; Funder & Colvin, 1988; Watson, 1989).

In this article, we attempt to extend the research on the accuracy of strangers' judgments in four ways. First, we assess the accuracy of consensual judgments based on exposures to extremely "thin slices" of nonverbal behavior: videoclips 10 s or less in length. Second, we assess how well such consensual judgments predict a novel criterion variable. Third, we discuss how prediction of this criterion is of pragmatic utility because the criterion variable is ecologically valid and has "real life" consequences. Fourth, we examine the influence on judgmental accuracy of two possible mediating variables: nonverbal behavior and physical attractiveness.

Thin Slices of Behavior

In previous studies on the accuracy of strangers' ratings in the absence of any interaction, strangers rated targets after viewing a 5-min videotaped clip of them (Funder & Colvin, 1988). Our goal was to examine the accuracy of ratings based on much briefer noninteractive observations. Both Erving Goffman and Gordon Allport suggested that people are able to form accurate impressions of others from mere glimpses of their behavior (Allport, 1937; Goffman, 1979), and there is meta-analytic evidence demonstrating that people are unexpectedly accurate in their consensual judgments of others from observations of thin slices of targets' expressive behavior (Ambady & Rosenthal, 1992). Furthermore, ecological approaches to person perception and impression formation suggest that individuals communicate certain stimulus information or affordances that enable others to quickly form valid impressions of them (Baron & Boudreau, 1987; McArthur & Baron, 1983).

Part of this article is based on Nalini Ambady's doctoral dissertation conducted at Harvard University.

Preparation of this article was supported in part by the Spencer Foundation. We thank the Derek Bok Center for Teaching and Learning for their cooperation and assistance, without which this research would not have been possible. We also thank Ross Buck, David Funder, Raj Marphatia, and two anonymous reviewers for their very helpful comments and suggestions.

The content of this article is solely our responsibility.

Correspondence concerning this article should be addressed to Nalini Ambady or Robert Rosenthal, Department of Psychology, William James Hall, Harvard University, Cambridge, Massachusetts 02138.

With the goal of extending previous research, we examined the accuracy of strangers' judgments regarding personality attributes from very minimal noninteractive information, that is, very brief video clips of nonverbal behavior. We also examined the effects of decreasing exposure lengths on judgmental accuracy. Thus, we obtained ratings of video clips 10 s, 5 s, and 2 s long.

Ecologically Valid Criterion Variable

Although the findings mentioned above regarding the accuracy of strangers' judgments are striking, the criterion variable used to assess accuracy in virtually all the studies was targets' self-ratings. An examination of criteria besides self-reports would add considerably to our knowledge concerning the accuracy of interpersonal perception, especially because self-report data are likely to be biased (Cheek, 1982; Kagan, 1988; Wiggins, 1973). Of particular value would be any evidence regarding the relationship between strangers' judgments of targets observed in a naturalistic situation and a criterion variable characterized by both pragmatic utility and ecological validity—one that is used in everyday decisions about people. Any evidence that people can make accurate judgments regarding others related to some external outcome of high ecological validity and utility would provide even more striking evidence for the accuracy of person perception.

To this end, we selected a criterion variable that seemed to be both ecologically valid and of considerable consequence: teacher effectiveness usually measured by any one or some combination of student ratings, peer ratings, and supervisor ratings. Such evaluations are important because of their practical significance: They influence salary, promotion, and tenure decisions.

An additional practical objective of this study was to identify the nonverbal correlates of good teaching. Nonverbal behavior relates to the "affective" aspect of teaching (Rosenthal, 1989) and often communicates unintended yet significant information to students (Cooper & Good, 1983; Dusek, 1985; Harris & Rosenthal, 1985). Knowledge of the nonverbal correlates of effective teaching would be of value not only in increasing our basic understanding of the importance of affective behavior in teaching and learning processes, but might also be of practical importance in guiding the selection and training of future teachers.

Nonverbal Behavior, Physical Attractiveness, and Judgmental Accuracy

In most of the research on interpersonal judgmental accuracy, judgments were based on exposures to both verbal and nonverbal channels of targets' behavior. Some authors, however, have suggested that the degree of consensus among and accuracy of strangers' judgments might be related to the presence of observable cues, particularly nonverbal behavior and physical appearance cues (Albright et al., 1988; Kenny et al., 1992; Paunonen, 1991) and that it might be worthwhile to provide strangers only with such cues and then examine the accuracy of their judgments (Albright et al., 1988; Watson, 1989). In a meta-analysis of the accuracy of judgments based on thin slices of behavior, we found that people were quite accurate in their judgments based on expressive behavior, but we were not

able to disentangle the effects of the possible mediating variables of specific nonverbal behaviors and physical attractiveness, yet previous research has indicated that people who are nonverbally more spontaneously expressive are also judged to be more attractive, and both variables influence the accuracy of nonverbal encoding (Sabatelli & Rubin, 1986).

Clearly, nonverbal cues do influence judgments about some personality attributes (DePaulo, 1992). For example, Kenny and his associates found that certain behavioral cues such as smiling and rapid body movements coded from 20-s clips were related to observers' ratings of targets' extraversion (Kenny et al., 1992), although they did not examine the accuracy of such judgments.

Physical appearance cues, particularly physical attractiveness, also influence judgments regarding a variety of characteristics (Berscheid & Walster, 1974; Dion, 1986), particularly those characteristics related to sociability and social competence (Albright et al., 1988; Eagly, Ashmore, Makhijani, & Longo, 1991; Kenny et al., 1992). Judgments of strangers might be influenced by targets' attractiveness and might actually be accurate in assessing targets' self-ratings. One explanation for this accuracy could be that attractive targets might actually rate themselves more positively, because of the influence of self-fulfilling prophecies (Snyder, Tanke, & Berscheid, 1977).

To examine the mediating effect of nonverbal behavior and physical appearance on the accuracy of personality judgments, we controlled the information available to raters so that targets were rated solely on the basis of their nonverbal behavior. We also obtained separate judgments of the physical attractiveness of targets to examine the relationship between physical appearance and the accuracy of personality judgments.

Study 1: College Teachers

In the first study, we had judges rate 13 teachers on several personality dimensions on the basis of three 10-s silent video-clips of each teacher. The sample of teachers was drawn from a university setting.

The criterion was end-of-the-semester student evaluations. Although student achievement (adjusted for student ability) might be the best possible criterion of effective teaching, it is very difficult to obtain such data. Teacher effectiveness in the real world is often evaluated solely on the basis of ratings of supervisors and students. Therefore, we used end-of-the-semester student ratings of teachers as a measure of teacher effectiveness. Considerable evidence supports the validity of student evaluations: Student ratings are consistent over time and across raters; correlate positively with expert, colleague, and administrator ratings; are independent of extraneous characteristics or characteristics of the students themselves; correlate significantly with how much students actually learn; and, last, do not change appreciably with greater age of the student rater and reflection by the student (Abrami, d'Apollonia, & Cohen, 1990; Centra, 1979; Cohen, 1981; Feldman, 1989a, 1989b; Howard, Conway, & Maxwell, 1985; Kulik & Kulik, 1974; Leventhal, Perry, & Abrami, 1977; Marsh, 1984; McKeachie, 1979; Trent & Cohen, 1973). Thus, student evaluations seem to be a valid means of evaluating teacher effectiveness.

Method

Sample and Criterion

Permission was obtained to use videotapes of classes taught by 13 teachers at the Teaching Laboratory of a private university. The teachers (6 women and 7 men) were all graduate student Teaching Fellows who were videotaped while teaching sections for undergraduate courses at the Teaching Laboratory to receive feedback about their teaching from specialists at the center. Each section lasted approximately 1 hr. The sections covered diverse areas of the curriculum: eight of the sections were in the area of the humanities and social sciences, three in the natural sciences, and two in languages. The number of students in the sections varied from 8 to over 20. The sample of teachers was selected by members of the Teaching Center from their library of videotapes. The criterion used for selection was that the sample should represent a fairly wide range in their ratings on the teaching criterion measure. The experimenters were blind to the evaluations of the teachers until the data analysis stage.

As mentioned earlier, we used end-of-the-semester student ratings of teachers as the criterion measure. We used the mean end-of-the-semester ratings of the teacher by the students in the section for which the teacher was videotaped. Ratings, compiled by the Committee on Undergraduate Education, were obtained by averaging the responses to the following items: "Rate the quality of the section overall" and "Rate section leader's performance overall." We were provided with the average of the two ratings for each teacher in our sample but not with the two separate ratings. Because some teachers were rated on a 5-point scale and others on a 7-point scale, these ratings were converted into percentages for this analysis, which varied from 41.67% to 91.67%. Only after all the data had been collected did we receive the ratings for the sample of teachers in this study.

Stimulus Material

One master videotape was derived from the 13 individual videotapes using the following method: For each teacher, 10 s from the first 10 min of the class video that was focused on the teacher alone (without any of the students), 10 s from the middle of the class video with the teacher alone, and 10 s from the last 10 min of the class with the teacher alone were assembled and rerecorded onto one videotape. The order of the 3 clips showing the teacher alone from each of the 13 tapes was randomized in a design like a Latin square. The final stimulus tape contained 39 clips: 3 clips each of 13 teachers.

Judges and Ratings

Ratings of molar nonverbal behaviors. Nine female undergraduates were paid to rate the 39 clips of the teachers individually. Female judges were chosen because of the evidence that women are better decoders of nonverbal behavior (Hall, 1984). They were told that they would see short segments of teachers teaching a class and would be asked to rate the nonverbal behavior of the teacher in each segment on 15 dimensions on a scale running from 1 (*not at all*) to 9 (*very*). Besides these instructions, the judges received no training. Previous research has demonstrated that naive subjects are able to reliably make such ratings without training (Rosenthal, 1987; Rosenthal, Blanck, & Vannicelli, 1984). The dimensions rated were accepting, active, anxious (this dimension was reverse scored), attentive, competent, confident, dominant, empathic, enthusiastic, honest, likable, optimistic, professional, supportive, and warm. Each segment was played once with the audio turned down completely, and the judges were given enough time to complete their ratings before the next segment was shown.

Coding of molecular nonverbal behaviors. The frequencies per clip of specific molecular nonverbal behaviors were coded by two female undergraduates. For each clip the raters counted the number of head

nods, head shakes, smiles, laughs, yawns, frowns, biting of the lips, downward gazes, self-touches, fidgets by shaking the hands or legs or by fiddling with an object, emphatic gestures (pointing, clapping, etc.), and weak gestures (using hands while talking); noted the position of the hands (symmetrical or asymmetrical, folded or open), position of legs (crossed or open), and position of the torso (leaning forward or backward); and also coded whether the teacher was sitting or standing.

Ratings of physical attractiveness. As mentioned earlier, both the criterion variable (student evaluations) and the predictor variable (judges' ratings of the video clips) could be influenced by the physical appearance of the teachers. In that case any correlation between the criterion and predictor variables would be attributable to physical appearance rather than expressive behavior. To evaluate the influence of physical appearance on the criterion variable, we asked two female judges to rate the physical attractiveness of each teacher on a 5-point scale on the basis of a single still video clip.

Results

Reliability of Judges' Ratings

Molar nonverbal behaviors. The reliabilities of the judges' ratings of the molar nonverbal behaviors were computed by means of intraclass correlations for all judges combined as well as for individual judges (Rosenthal, 1987). The effective reliabilities of the mean of nine judges' ratings ranged from .60 to .89, with a mean of .72. These reliabilities are displayed in Table 1 and were based on video clips just 30 s long.

The mean of the judges' ratings of each of the 15 molar dimensions was computed across the 39 clips. These means were intercorrelated in a 15 × 15 correlation matrix, which was subjected to a principal-components analysis. The principal-components analysis before rotation indicated the presence of a single factor accounting for 71% of the variance. On the basis of this finding, a composite variable was created, composed of 15 scales for the video channel. The anxiety variable was reverse scored as not anxious because it was negatively related to all the other molar variables. Thus, one new composite variable was

Table 1
Reliabilities of Judges' Ratings of College
Teachers' Molar Nonverbal Behaviors

Molar behaviors	M of 9 judges	1 judge
Accepting	.60	.15
Active	.89	.47
Attentive	.60	.15
Competent	.70	.21
Confident	.87	.42
Dominant	.75	.25
Empathic	.71	.23
Enthusiastic	.89	.46
Honest	.60	.15
Likable	.70	.21
(Not) anxious	.67	.19
Optimistic	.87	.42
Professional	.64	.17
Supportive	.70	.21
Warm	.63	.16
M	.72	.24
Mdn	.70	.21
Global variable	.85	.36

created by summing the mean rating for the video channel of 15 of the nonverbal variables with the sign of the variable taken into account. The standard deviations of the variables were similar enough, so that transformations into standard scores were not needed. The reliability of the mean of the nine judges' ratings for this new variable was .85. This reliability, an Armor's theta of .97 (an index of the internal consistency of this composite; Armor, 1974; Rosenthal, 1987), and a mean intercorrelation among the 15 variables of $r = .67$, supported the decision to use the sum of ratings of the 15 dimensions as an additional global variable. A one-way analysis of variance (ANOVA) of this composite variable revealed no significant main effect for the order of the 3 clips for each teacher.

Molecular nonverbal behaviors. Some of the categories for the molecular ratings of nonverbal behavior were not observed by one or both of the two judges. These categories (biting of lips, emphatic and weak gestures, arms folded or open, and legs crossed or open) were dropped from further analysis. The effective reliabilities of the remaining categories ranged from .51 to .98 (see Table 2). These ratings were subjected to a principal-components analysis, but no interpretable factors emerged. Furthermore, the mean intercorrelation among the molecular variables ($r = .38$) was considerably lower than the mean intercorrelation among the molar variables ($r = .67$). Behaviors were considered individually in later analyses.

Molar Nonverbal Behavior and Teacher Evaluations

Ten of the 15 molar nonverbal dimensions were significantly predictive of the criterion variable of teacher effectiveness, as can be seen from Table 3. On the whole, teachers with better evaluations were judged more favorably on the average of 15 dimensions of nonverbal behavior, $r(11) = .76, p < .005$. Specifically, teachers who were rated higher by their students were judged to be significantly more optimistic, confident, dominant, active, enthusiastic, likable, warm, competent, and supportive on the basis of their nonverbal behavior. Note also that dimensions that did not reach significance correlated substantially and in the predicted direction with the criterion. It is interesting that these results are consistent with other findings

Table 2
Reliabilities of Judges' Ratings of 10-s Clips of College Teachers' Specific Molecular Nonverbal Behaviors

Variable	<i>M</i> of 2 judges	1 judge
Sit	.98	.95
Frown	.90	.81
Touch head	.86	.75
Gaze down	.85	.74
Lean forward	.85	.74
Fidget with hands	.84	.73
Laugh	.84	.73
Smile	.80	.68
Arms symmetrical	.77	.62
Nod head	.70	.55
Shake head	.60	.44
Fidget with object	.51	.35
<i>M</i>	.81	.70
<i>Mdn</i>	.84	.73

Table 3

Correlations of Molar Nonverbal Behaviors With College Teacher Effectiveness Ratings (Student Ratings)

Variable	<i>r</i>
Accepting	.50
Active	.77**
Attentive	.48
Competent	.56*
Confident	.82***
Dominant	.79**
Empathic	.45
Enthusiastic	.76**
Honest	.32
Likable	.73**
(Not) anxious	.26
Optimistic	.84***
Professional	.53
Supportive	.55*
Warm	.67*
Global variable	.76**

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

based on far longer observations in suggesting that achievement-oriented positive behaviors (confidence and dominance) and interpersonally oriented behaviors (warmth and supportiveness) are both related to students' ratings of teachers (Brophy & Good, 1974; Erdle, Murray, & Rushton, 1985; Rosenshine, 1971; Rosenshine & Furst, 1973; Ryans, 1960).

Molecular Nonverbal Behaviors and Teacher Evaluations

Each molecular behavior was correlated with the global overall impression variable in an attempt to understand the specific behaviors influencing molar impressions. As can be seen from the first data column of Table 4, there was a significant negative relationship between teachers' frowning and judges' ratings of them on the overall molar variable, $r(11) = -.61, p < .05$. Simi-

Table 4

Correlations of Molecular Nonverbal Behaviors With Global Molar Behavior Rating and College Teachers' Effectiveness Ratings (Student Ratings)

Molecular nonverbal behaviors	<i>r</i> with molar global rating	<i>r</i> with teacher effectiveness rating
Arms symmetrical	.21	-.18
Gaze down	-.11	-.34
Frown	-.61*	-.46
Fidget with hands	-.39	-.70**
Fidget with object	-.50	-.56*
Nod head	.37	.38
Shake head	-.25	.02
Laugh	.20	.34
Lean forward	.02	-.49
Sit	-.21	-.34
Smile	.23	.20
Touch head	-.06	.12

* $p < .05$. ** $p < .01$.

larly, fidgeting (with an object or with hands) was negatively correlated with global molar impressions.

Each molecular behavior was also correlated with the criterion variable. Fidgeting was negatively correlated with the criterion variable: Teachers who fidgeted more with their hands or fiddled with an object (such as chalk or a pen) received significantly lower ratings from their students (see the second data column of Table 4). Teachers with lower ratings also tended to gaze down, frown, and lean forward more. This result seems to contradict previous research (Harris & Rosenthal, 1985; Mehrabian, 1969), in which forward lean has been associated with involvement, warmth, and so forth. However, the coding of forward leans might have been confounded with the position of the teacher, because teachers who were seated were judged as less effective. The negative correlation between leaning forward and standing, $r(11) = -.36$, *ns*, indicated this might have been likely, although the correlation between leaning forward and teaching effectiveness dropped only from $-.49$ to $-.42$ when sitting position was partialled out from both the variables and to $-.39$ when sitting position was partialled out from forward lean. Another possibility is that a teacher leaning forward might have been perceived as threatening by the students, but this needs to be explored further. Teacher effectiveness was not found to be significantly related to such behaviors as smiling, shaking the head, laughing, or touching the head, although the results suggest that nodding and laughing are positively related to evaluations of teachers (see Table 4).

Physical Attractiveness and Teacher Evaluations

The reliability of the mean of the two judges' ratings of physical attractiveness was $R = .80$. The correlation between ratings of teachers' physical attractiveness and the criterion variable was $r(11) = .32$, *ns*, suggesting that students' ratings of teachers were somewhat influenced by the physical appearance of the teachers.

Partial correlation coefficients were computed to examine the relationship between physical attractiveness and the criterion variable, controlling for nonverbal behavior, and to examine the relationship between nonverbal behavior and the criterion variable, controlling for physical attractiveness. Results indicated that controlling for physical attractiveness very slightly reduced the relationship between the global impression variable and the criterion, from $r(11) = .76$ to $r(11) = .74$, $p < .01$. Furthermore, controlling for nonverbal behavior decreased more noticeably the relationship between physical attractiveness and the criterion, from $r(11) = .32$ to $r(11) = .14$.

The overall result of this study was quite remarkable: On the basis of observations of video clips just half a minute in length, complete strangers were able to predict quite accurately the ratings of teachers by students who had interacted with them over the course of a whole semester! Furthermore, these predictions retained their accuracy after we adjusted for physical appearance of the teachers, indicating that the judges were picking up very subtle nonverbal cues.

Study 2: High School Teachers

Because of the surprising results of the previous study and the small sample size, Study 2 was a replication of the first study

with a different population of teachers and a somewhat different criterion variable. For this study we used a sample of high school teachers.

Method

Sample and Criterion

Thirteen high school teachers (8 female and 5 male) were videotaped in the classroom at a private high school in a large city. Each class lasted 50 min. Teachers were told about the study and they volunteered to be videotaped. A camcorder was used to videotape the classes. The criterion measure used to assess the effectiveness of the teachers was the evaluation by the principal of the high school on a 5-point scale, in response to the request, "Rate the overall effectiveness of _____ as a teacher." However, all the teachers received ratings of 4 or 5, possibly because the school was very selective and did not renew the contracts of teachers thought to be unsatisfactory.

Stimulus Material

One master videotape was derived from the 13 videotapes using the same method as in the previous study. The final stimulus tape was composed of 39 clips: 3 clips each of 13 teachers.

Judges and Ratings

Ratings of molar nonverbal behaviors. Eight female undergraduates were paid to rate the 39 video clips of the teachers individually on the molar dimensions described in Study 1.

Coding of molecular nonverbal behaviors. The frequencies per clip of specific molecular nonverbal behaviors described in Study 1 were coded by four undergraduates.

Ratings of physical attractiveness. As in the previous study, to evaluate the influence of physical appearance on the criterion variable, we asked two female judges to rate the physical attractiveness of each teacher on a 5-point scale on the basis of a single still video clip.

Results

Reliability of Judges' Ratings

Molar nonverbal behaviors. The reliabilities of the judges' ratings of the molar nonverbal behaviors were computed by using intraclass correlations (Rosenthal, 1987). The effective reliabilities of the mean of the judges' ratings for the video channel ranged from $.39$ to $.84$, with a mean reliability of $.69$. These reliabilities are displayed in Table 5.

As in the previous study, a composite variable was created by summing the mean rating of 14 of the nonverbal variables. The variable of anxiety was dropped because of its low reliability. Again, because the standard deviations of the variables were similar enough, transformations into standard scores were not needed. The reliability of the mean of the eight judges' ratings for this new variable was $.80$. This reliability, a mean intercorrelation among the 15 variables of $.49$ and an Armor's theta of $.96$ (an index of the internal consistency of this composite, Armor, 1974; Rosenthal, 1987), supported the decision to use the sum of the same dimensions as in Study 1 as an additional global variable. A one-way ANOVA on this composite variable revealed no significant main effect for the order of the three clips for each teacher.

Molecular nonverbal behaviors. The reliabilities of the mo-

Table 5
Reliabilities of Judges' Ratings of High School Teachers' Molar Nonverbal Behaviors

Molar behaviors	<i>M</i> of 8 judges	1 judge
Accepting	.64	.18
Active	.75	.27
Attentive	.71	.25
Competent	.69	.24
Confident	.59	.15
Dominant	.62	.16
Empathic	.78	.30
Enthusiastic	.79	.31
Honest	.60	.16
Likable	.80	.34
(Not) anxious	.39	.06
Optimistic	.84	.40
Professional	.49	.11
Supportive	.70	.23
Warm	.81	.35
<i>M</i>	.69	.24
<i>Mdn</i>	.79	.31
Global variable	.80	.34

lecular behaviors ranged from .31 to .94 (see Table 6). The mean reliability for the 13 coded behaviors was .74.

Molar Nonverbal Behavior and Teacher Evaluations

Seven of the 15 molar nonverbal dimensions significantly predicted the criterion variable, as can be seen from Table 7. On the whole, teachers rated more highly by the principal were judged more favorably on the average of all 14 dimensions of video nonverbal behavior, $r(11) = .68$, $p < .05$. Specifically, teachers who were rated higher were judged to be significantly more supportive, likable, accepting, attentive, enthusiastic, warm, and optimistic on the basis of their nonverbal behavior. As in the previous study, dimensions that failed to reach signifi-

Table 6
Reliability of Judges' Ratings of 10-s Clips of High School Teachers' Molecular Nonverbal Behaviors

Variable	<i>M</i> of 2 judges	1 judge
Arms symmetrical	.65	.47
Frown	.31	.19
Nod head	.59	.43
Shake head	.73	.58
Point	.77	.62
Sit	.96	.94
Smile	.86	.75
Stand	.89	.80
Strong gesture	.63	.45
Touch head	.94	.89
Touch upper torso	.84	.73
Walk	.78	.64
Weak gesture	.66	.49
<i>M</i>	.74	.58
<i>Mdn</i>	.77	.62

Table 7
Correlations of Molar Behaviors With High School Teachers' Effectiveness Ratings (Principal's Rating)

Variable	<i>r</i>
Accepting	.64*
Active	.41
Attentive	.64*
Competent	.47
Confident	.35
Dominant	.27
Empathic	.53
Enthusiastic	.62*
Honest	.42
Likable	.64*
Optimistic	.58*
Professional	.35
Supportive	.74**
Warm	.63*
Global variable ^a	.68**
Not anxious	-.12

^a The not anxious variable was dropped from the global variable because of low reliability.

* $p < .05$. ** $p < .01$.

cance were of substantial magnitude and were related to the criterion in the predicted direction.

Molecular Nonverbal Behaviors and Teacher Evaluations

These variables were considered individually because a principal-components analysis revealed no interpretable factors. Moreover, the mean intercorrelation among the variables was quite low ($r = .11$). Correlations of the global molar variable with the individual molecular variables are shown in the first column of Table 8. Teachers who smiled more were rated significantly higher on the composite global variable, $r(11) = .71$, $p < .05$.

Teachers with higher ratings tended to be more nonverbally

Table 8
Correlations of Molecular Nonverbal Behaviors With Global Molar Rating and High School Teachers' Effectiveness Rating (Principal's Rating)

Molecular nonverbal behavior	<i>r</i> with global molar rating	<i>r</i> with principal's rating
Arms symmetrical	.26	.25
Frown	-.06	-.14
Nod head	.21	.25
Shake head	.09	-.11
Point	-.04	.21
Sit	.07	-.24
Smile	.71*	.33
Strong gesture	.23	.05
Touch head	-.21	-.49
Touch upper torso	.52	.53
Walk	.25	.47
Weak gesture	.27	.20

* $p < .01$.

active and expressive. They were more likely to walk around, touch their upper torsos, and smile (see Table 8). Less effective teachers were more likely to sit, touch their heads, and shake rather than nod their heads. Touching the head was significantly related to sitting, $r(11) = .62, p < .05$, and negatively related to standing, $r(11) = -.67, p < .05$. Sitting was also significantly associated with shaking the head, $r(11) = .73, p < .01$. These results suggest that teachers with higher ratings showed more nonverbal expressiveness and involvement than less effective teachers.

Physical Attractiveness and Teacher Evaluations

The reliability of the mean of two judges' ratings of physical attractiveness was $R = .75$; the reliability of a single judge was $.60$. The correlation between ratings of teachers' physical attractiveness and the criterion variable was $r(11) = -.18$, indicating that the principal was not significantly influenced by the physical appearance of teachers in making her ratings. Any correlation obtained between the judges' ratings and the criterion variable in this study could, therefore, be attributed to factors other than the appearance of the teachers.

As in the previous study, partial correlation coefficients were computed to examine the relationship between physical attractiveness and the criterion variable, controlling for nonverbal behavior, and to examine the relationship between nonverbal behavior and the criterion variable, controlling for physical attractiveness. Controlling for physical attractiveness actually increased the relationship between the global impression variable and the criterion from $r(11) = .68$ to $r(11) = .73, p < .01$. Furthermore, controlling for nonverbal behavior decreased the relationship between physical attractiveness and the criterion from $r(11) = -.18$ to $r(11) = -.40$.

Thus, in this study we were able to replicate the results of the previous study with a different sample of teachers and a different criterion variable. Judges' ratings of teachers' personality based on thin slices of behavior significantly predicted the principal's evaluation of the teacher.

Study 3: "Thinning" the Slices

In the third study, we investigated whether strangers' ratings of teachers would predict the two different criterion variables from the previous studies if we thinned the slices even more. Accordingly, we had judges rate two sets of shorter video clips on the same dimensions.

Method

The original 10-s clips from the previous two studies were reduced to 5-s and 2-s clips. This was done by a research assistant unfamiliar with the ratings of the teachers or the hypotheses and results of the previous two studies. She randomly selected portions from the longer clips to comprise the four sets of the shorter clips. Thus, four new videotapes were created: two for each of the two sets of teachers, one with three 5-s clips for each teacher and the other with three 2-s clips for each teacher.

Thirty-two female undergraduates were paid to rate the four video-

tapes; eight judges rated each tape and no judge saw more than one tape. The method used was the same as in the previous two studies.

Results

The reliabilities of the eight judges and the typical single judge for each videotape are presented in Table 9. As expected, the mean reliabilities for the 2-s clips were slightly lower than the reliabilities of the 5-s clips. The means of the ratings for each dimension were summed as in the previous two studies to create a composite variable. The reliabilities of the means of the judges' ratings for the new composite variables were satisfactory (.90 and .73 for the college teachers, and .89 and .77 for the school teachers), as were the mean intercorrelations ($r_s = .85$ and .61 for the 5-s and 2-s clips of the college teachers and $r_s = .66$ and .63 for the 5-s and 2-s clips of the college teachers), and Armor's theta was also high (.95 and .94 for the college teachers and .95 for both ratings of the school teachers).

Correlations of the 15 variables as well as the composite variables for the college and high school teachers' 5-s and 2-s clips are presented in Table 10. As can be seen from Table 10, the correlation of the composite based on the 5-s clips with the criterion variable was slightly higher than the correlation of the composite based on the 2-s clips for the high school teachers. For the college sample, however, there was an unexpected reversal: The composite based on the 2-s clips correlated unexpectedly highly with the criterion variable. However, as indicated below, the effect sizes within as well as between the two samples did not differ significantly.

To summarize the results of all the studies, we combined the effect sizes for ratings on the global variable (composed of the individual dimensions) using meta-analytic techniques (Rosenthal, 1991). Table 11 displays the effect sizes of the composite molar variables for each clip length for both sets of teachers. The combined overall meta-analytic effect size from all six sub-studies was .59, associated with a significant Z ($Z = 5.06, p < .00001$). We tested the effect sizes for clip length from each sample for heterogeneity by comparing three effect sizes for each sample, using a method for comparing correlated coefficients (Meng, Rosenthal, & Rubin, 1992). For the college teacher sample the chi-square value, $\chi^2(2) = 1.52, ns$, indicated that the effect sizes were homogeneous. For the schoolteacher sample the chi-square value, $\chi^2(2) = 2.06, ns$, indicated that these effect sizes were also homogeneous. We also performed two contrasts on both sets of effect sizes. Contrasts testing for a linear trend using weights of 1, 0, and -1, for the 10-s, 5-s, and 2-s clips, respectively, were computed and were not significant for the college teacher sample ($Z = .24, p = .40$) or the schoolteacher sample ($Z = 1.14, p = .13$). Contrasts comparing the 10-s clips with the 5-s and 2-s clips using contrast weights of 2, -1, and -1, respectively, were also not significant ($Z = .82, p = .21$, for the college teachers and $Z = 1.07, p = .14$, for the schoolteachers). We also compared effect sizes for the two samples for each clip length and found no significant differences between the two samples of teachers (for 10-s clips, $Z = .37, p = .35$; for 5-s clips, $Z = .46, p = .47$; and for 2-s clips, $Z = 1.27, p = .10$).

These analyses show that there were no significant differences in the accuracy of judgments based on video clips 10 s, 5 s, and 2 s in length. In addition, there were no significant differ-

Table 9
*Reliabilities of Judges' (Mean of Eight Judges' and One Judge's)
 Ratings of Molar Nonverbal Behaviors*

Molar behavior	College teachers				School teachers			
	5 s		2 s		5 s		2 s	
	8 Js	1 J	8 Js	1 J	8 Js	1 J	8 Js	1 J
Accepting	.62	.16	.60	.16	.88	.49	.77	.29
Active	.89	.50	.83	.39	.87	.45	.71	.25
Attentive	.78	.30	.58	.14	.84	.40	.69	.24
Competent	.75	.27	.49	.11	.79	.31	.46	.10
Confident	.78	.30	.67	.20	.84	.40	.75	.27
Dominant	.26	.03	.58	.14	.77	.30	.52	.13
Empathic	.81	.35	.56	.14	.77	.30	.83	.39
Enthusiastic	.92	.60	.83	.39	.89	.50	.79	.31
Honest	.78	.30	.30	.05	.78	.30	.41	.09
Likable	.88	.49	.71	.25	.85	.41	.81	.35
(Not) anxious	.21	.03	.38	.06	.67	.20	.11	.02
Optimistic	.90	.54	.78	.30	.87	.45	.86	.44
Professional	.88	.49	.48	.11	.75	.27	.48	.11
Supportive	.86	.44	.67	.20	.88	.49	.76	.29
Warm	.88	.49	.73	.25	.90	.54	.84	.40
<i>M</i>	.75	.27	.61	.16	.82	.37	.65	.19
<i>Mdn</i>	.81	.35	.60	.16	.84	.40	.71	.25
Global variable	.90	.54	.73	.25	.89	.50	.77	.30

Note. J = judge.

ences in the accuracy of judgments for the two samples of teachers.

General Discussion

These results were striking. First, we found that the ratings of complete strangers based on very thin slices of teachers' non-

verbal behavior (video clips from 2 s to 10 s long) predicted with surprising accuracy the ratings of the same teachers by people who had substantial interactions with those teachers (students and supervisors, for example). Moreover, judgments based on 30-s exposures (three 10-s clips of each teacher) were not significantly more accurate than judgments based on 6-s exposures (three 2-s clips of each teacher). We found that judges were accurate, as indicated by the high degree of consensus in their judgments, the criterion relied on by most studies on the accuracy of personality judgment (Colvin & Funder, 1991). More remarkably, we found that judges were accurate on the basis of the high correlations between their judgments and a real-life criterion variable. These findings demonstrate the wealth of information conveyed in thin slices of behavior and the unexpected accuracy of judgments based on these slices. Second, our results demonstrated the value of using ecologically valid criteria and everyday behavior in assessing personality and affective variables. Third, our results highlighted the invaluable

Table 10
*Correlations of Teacher Effectiveness Ratings With Mean
 Judges' Ratings of 5-s and 2-s Video Clips*

Molar behavior	College teachers		School teachers	
	5 s	2 s	5 s	2 s
Accepting	.38	.37	.44	.31
Active	.55*	.69**	.28	.05
Attentive	.25	.58*	.50	.38
Competent	.52	.57*	.35	.28
Confident	.58*	.62*	.48	.17
Dominant	.40	.38	.10	.14
Empathic	.35	.59*	.50	.29
Enthusiastic	.45	.61*	.36	.18
Honest	.28	.63*	.43	.26
Likable	.34	.72**	.44	.39
Optimistic	.35	.64*	.41	.26
Professional	.36	.37	.52	.42
Supportive	.36	.66*	.40	.33
Warm	.50	.54	.37	.23
Global variable ^a	.44	.71**	.47	.31
(Not) anxious	.51	.14	.27	.29

^a The not anxious variable was dropped from the global variable because of its low reliability.

* $p < .05$. ** $p < .01$.

Table 11
Combined Effect Sizes for Clip Length and Teacher Samples

Clip length	High school <i>r</i>	College <i>r</i>	Combined <i>r</i>	Combined <i>Z</i> ^a
10 s	.68**	.76**	.72****	3.98
5 s	.47	.44	.46*	2.14
2 s	.31	.71**	.54***	2.64
Combined <i>r</i>	.50****	.66****	.59****	5.06
Combined <i>Z</i>	2.99	4.16		

^a *Z* is the standard normal deviate corresponding to the *p* value.

* $p < .05$. ** $p < .01$. *** $p < .005$. **** $p < .001$.

information provided by nonverbal cues in making judgments about others.

To explain these results, we examined the role of two possible mediating variables on the accuracy of judgments from thin slices. The first variable was physical attractiveness. Contrary to the evidence regarding the biasing effects of physical appearance, we found that judges were not strongly influenced by the physical attractiveness of the teacher. This result could be attributed to methodological differences: Studies using photographs generally report strong effects for physical attractiveness, but studies providing additional information such as expressive behavior cues do not always find such effects (Albright et al., 1988; Bull & Rumsey, 1988). Thus, when raters actually interact with targets or when they watch video clips of targets, the effects of physical attractiveness become diluted by the other information available (Barnes & Rosenthal, 1985). For example, Zuckerman, Miyake, & Hodgins (1991) found that unattractive voices dilute the effects of attractive faces and vice versa. Another explanation for the relatively low influence of physical appearance cues is suggested by the work of Eagly and her colleagues, who in a meta-analysis examined the influence of physical attractiveness on beliefs about social competence, intellectual competence, concern for others, integrity, adjustment, and potency (Eagly et al., 1991). They concluded that physical attractiveness was strongly related to judgments of social competence; somewhat related to judgments of potency, adjustment, and intellectual competence; and unrelated to judgments regarding integrity and concern for others. The prototype of the ideal teacher used by subjects in making their judgments probably involved more than just social competence. Perhaps judges rated teachers on a combination of intellectual competence (intelligence and high task motivation), concern for others (nurturance and social sensitivity), adjustment (self-esteem, mental health, and maturity), as well as social competence (sociability and likability). At an intuitive level judges may have known that "teaching can proceed quite successfully no matter how unfavored the teacher is by nature" (Allport, 1953, p. 875), and they may have focused on cues other than physical attractiveness in forming their impressions.

This hypothesis is supported by the significant increase in accuracy from ratings based solely on appearance (conceived of as a 0-s exposure) to ratings based on 2-s clips. This increase in accuracy is probably attributable to the presence of nonverbal movement cues, as suggested by earlier results showing that judgments based on clips with even micromomentary movement cues (42 ms) substantially increase the accuracy in decoding behavior over still frames (McLeod & Rosenthal, 1983). In addition, people who provide more spontaneous nonverbal information are generally better encoders of information and are also judged to be more attractive (Sabatelli & Rubin, 1986).

Accordingly, the second mediating variable we investigated was nonverbal behavior. We found that judgmental accuracy based on thin slices of behavior was not strongly linked to specific nonverbal behaviors of the targets. Judgmental accuracy was, however, strongly linked to gestalt, molar impressions based on nonverbal behavior. This result supports previous research showing that judgments of molar impressions, although vaguer and fuzzier, generally yield more useful information than the coding of specific behaviors. This makes intuitive sense: Although we might think that a smiling teacher should

be judged as more warm and more effective, judges interpret the smile in the overall context (Ekman, 1992; DePaulo, 1992). A smile could be interpreted as positive (genuinely warm, nurturant, and interested), or as negative (condescending or threatening), depending on accompanying behaviors and the context in which the behavior occurs. But judges can identify whether a person is being warm or not, although they might not be able to isolate the specific cues driving these perceptions. An interesting example of the accuracy of the more fuzzy molar dimensions comes from the work of Holt and Luborsky (1958), who attempted to predict the success of psychiatric residents at the Menninger Clinic. They found that judges' ratings of how much they liked the resident predicted psychiatric competence better than residents' scores on a variety of projective and non-projective assessment measures.

A possible explanation for the strong influence of nonverbal behavior on impression formation from thin slices of behavior is suggested by ecological perspectives on person perception (Baron & Boudreau, 1987; McArthur & Baron, 1983). This theory posits that individuals communicate certain "affordances" (stimulus information) that are adaptive for perceivers to detect, and perceivers detect this information if they are attuned to detecting it. For example, perceiving characteristics associated with infancy has adaptive consequences; thus, baby-faced individuals afford or elicit certain specific reactions from people (Zebrowitz, 1990). Moreover, people typically tend to use affordance-related information when they make predictions regarding circumscribed contexts, as in the present study (Baron & Boudreau, 1987). Perhaps the affordances related to teaching behavior are communicated through nonverbal behavior, and judges are attuned to pick these up. The present results support the ecological theory of interpersonal perception in suggesting that such affordances are communicated primarily through expressive behavior and have real life consequences (Zebrowitz, 1990).

Can the results of the present study be generalized to the accuracy of impressions in everyday life? The answer is not clear. Certain settings are particularly conducive to the expression and easy observation of traits; these settings include academic and business settings (Kenrick, McCreath, Govern, King, & Bordin, 1990). In the present study, we were examining *circumscribed*, or situation-specific, accuracy (Swann, 1984): Judges were making predictions regarding a criterion variable relevant to the situation in which they were observing the targets; that is, they were predicting teacher effectiveness on the basis of observations of actual teaching behavior. This setting may have contributed to increased judgmental accuracy because it permitted the clear expression of relevant traits and consequently the more accurate judgment of such traits. Nevertheless, these results indicate that the consensual judgments of strangers in real life can be quite accurate if they are based on observations, even very brief ones, of behaviors that are relevant to the criterion situation (see also Colvin & Funder, 1991).

It is also worth noting that the explicit goal of judges in this study was to make judgments about others from minimal information. This goal could have increased both the attention they paid to the stimuli as well as their motivation to be accurate (Swann, 1984). In everyday situations, people might not always be as accurate in their judgments under ordinary circum-

stances. Our results, however, indicate that if people are aware that they must make a rapid judgment, these judgments are likely to be quite accurate.

These results have important implications for education. They highlight the considerable influence of very subtle affective nonverbal behaviors on the teaching process and reveal that these subtle influences might be identifiable from thin slices of behavior. Teachers with higher ratings were judged more positively on affective dimensions of their personality from their nonverbal behavior, confirming the importance of teacher affect (Abrami, Leventhal, & Perry, 1982; Babad, Bernieri, & Rosenthal, 1989; Harris & Rosenthal, 1985; Woolfolk & Woolfolk, 1977) and suggesting that better teachers might be higher in nonverbal sending accuracy, that is, in their ability to spontaneously communicate nonverbal information (Buck, 1979). Indeed, the agreement between ratings of students who had known the teacher for a whole semester or the ratings of a supervisor who had probably known the teacher for even longer and the ratings based on thin slices suggests that some stable quality of the teacher was being communicated; impressions were consistent between different sets of raters, although the variables being judged differed. This agreement supports ecological perspectives on social perception indicating that raters might be tuning in to the same affordances.

These findings also suggest that teachers should be made aware of the possible impact of their nonverbal behavior and perhaps even trained in nonverbal skills (Smith, 1979; Wolfgang, 1979; Woolfolk & Brooks, 1983). But there is little unequivocal evidence that teaching effectiveness can be improved by training in nonverbal skills. In generalizing the results of this article, certain caveats should be kept in mind. First, judgments based on thin slices are probably most accurate for interpersonal or affective variables (Ambady & Rosenthal, 1992). Second, such judgments should be based on actual observations of the person in the situation to which the criterion is applicable. Third, aggregate judgments should be used rather than judgments by single individuals.

Some limitations of this research should be kept in mind. First, multiple criteria such as peer evaluations, student evaluations, and supervisor ratings should be used to evaluate teacher effectiveness. In particular, student achievement and student learning should be used as a criterion in future research (Abrami et al., 1982; Feldman, 1989b), although student ratings are positively related to student achievement (Cohen, 1981; Feldman, 1989b). Second, we examined the accuracy of predictions regarding personality dimensions from judgments based on video clips. Future studies might examine other channels of behavior to investigate whether these results generalize to other behavioral channels such as the audio channel. Furthermore, future studies should examine whether these results generalize to other ecologically valid criterion variables.

Nevertheless, these findings are surprising and provocative. They suggest, first, that our consensual intuitive judgments might be unexpectedly accurate, and second, that we communicate—unwittingly—a great deal of information about ourselves. Not only do we possess the remarkable ability to form impressions of others, as Asch (1946) suggested, but, perhaps more remarkably, the impressions that we form can be quite accurate!

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., Leventhal, L., & Perry, R. P. (1982). Educational seduction. *Review of Educational Research*, 52, 446–464.
- Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology*, 55, 387–395.
- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt.
- Allport, G. W. (1953). The teaching–learning situation. *Public Health Reports*, 68, 875–879.
- Ambady, N., & Rosenthal, R. (1992). Thin slices of behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 112, 256–274.
- Armor, D. J. (1974). Theta reliability and factor scaling. In H. L. Costner (Ed.), *Sociological methodology, 1973–1976* (pp. 17–50). San Francisco: Jossey-Bass.
- Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, 41, 258–290.
- Babad, E., Bernieri, F., & Rosenthal, R. (1989). Nonverbal communication and leakage in the behavior of biased and unbiased teachers. *Journal of Personality and Social Psychology*, 56, 89–94.
- Barnes, M., & Rosenthal, R. (1985). Interpersonal effects of experimenter attractiveness, attire, and gender. *Journal of Personality and Social Psychology*, 48, 435–446.
- Baron, R. M., & Boudreau, L. A. (1987). An ecological perspective on integrating personality and social psychology. *Journal of Personality and Social Psychology*, 53, 1222–1228.
- Berscheid, E., & Walster, E. (1974). Physical attractiveness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 158–216). San Diego, CA: Academic Press.
- Brophy, J. E., & Good, T. L. (1974). *Teacher–student relationships: Causes and consequences*. New York: Holt, Rinehart & Wilson.
- Buck, R. (1979). Measuring individual differences in the nonverbal communication of affect. *Human Communication Research*, 6, 47–57.
- Bull, R., & Rumsey, N. (1988). *The social psychology of facial appearance*. New York: Springer-Verlag.
- Centra, J. A. (1979). *Determining faculty effectiveness*. San Francisco: Jossey-Bass.
- Cheek, J. M. (1982). Aggregation, moderator variables, and the validity of personality tests: A peer-rating study. *Journal of Personality and Social Psychology*, 43, 1254–1269.
- Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity results. *Review of Educational Research*, 51, 511–517.
- Colvin, C. R., & Funder, D. C. (1991). Predicting personality and behavior: A boundary on the acquaintanceship effect. *Journal of Personality and Social Psychology*, 60, 884–894.
- Cooper, H., & Good, T. (1983). *Pygmalion grows up: Studies in the expectation communication process*. New York: Longman.
- DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin*, 111, 203–243.
- Dion, K. K. (1986). Stereotyping based on physical attractiveness: Issues and conceptual perspectives. In C. P. Herman, M. P. Zanna, & E. T. Higgins (Eds.), *Physical appearance, stigma, and social behavior: The Ontario Symposium* (Vol. 3, pp. 7–21). Hillsdale, NJ: Erlbaum.
- Dusek, J. B. (Ed.). (1985). *Teacher expectancies*. Hillsdale, NJ: Erlbaum.
- Eagly, A. H., Ashmore, R. D., Makhijani, M. D., & Longo, L. C. (1991). What is beautiful is good, but . . . : A meta-analytic review of re-

- search on the physical attractiveness stereotype. *Psychological Bulletin*, 110, 109–128.
- Ekman, P. (1992). *Telling lies* (expanded ed.). New York: Norton.
- Erdle, S., Murray, H. G., & Rushton, J. P. (1985). Personality, classroom behavior, and student ratings of college teaching effectiveness: A path analysis. *Journal of Educational Psychology*, 77, 394–407.
- Feldman, K. A. (1989a). 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, 137–194.
- Feldman, K. A. (1989b). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education*, 30, 583–645.
- Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology*, 55, 149–158.
- Goffman, E. (1979). *Gender advertisements*. New York: Harper & Row.
- Hall, J. (1984). *Nonverbal sex differences*. Baltimore: Johns Hopkins University Press.
- Harris, M. J., & Rosenthal, R. R. (1985). The mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin*, 97, 363–386.
- Holt, R. R., & Luborsky, L. (1958). *Personality patterns of psychiatrists* (Vol. 1). New York: Basic 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.
- Kagan, J. (1988). The meaning of personality predicates. *American Psychologist*, 43, 614–620.
- Kenny, D. A., Horner, C., Kashy, D. A., & Chu, L. (1992). Consensus at zero acquaintance: Replication, behavioral cues, and stability. *Journal of Personality and Social Psychology*, 62, 88–97.
- Kenrick, D. T., McCreath, H. E., Govern, J., King, R., & Bordin, J. (1990). Person-environment intersections: Everyday settings and common trait dimensions. *Journal of Personality and Social Psychology*, 58, 685–698.
- Kruglanski, A. W. (1989). The psychology of being “right”: The problem of accuracy in social perception and cognition. *Psychological Bulletin*, 106, 395–409.
- Kulik, J. A., & Kulik, C. C. (1974). Student ratings of instruction. *Teaching of Psychology*, 1, 51–57.
- Leventhal, L., Perry, R. P., & Abrami, P. C. (1977). Effects of lecturer quality and student perception of lecturer’s experience on teacher ratings and student achievement. *Journal of Educational Psychology*, 69, 360–374.
- Marsh, H. W. (1984). Student’s evaluation of university teaching: Dimensionality, reliability, validity, potential biases and utility. *Journal of Educational Psychology*, 76, 707–754.
- McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, 90, 215–238.
- McKeachie, W. J. (1979). Student ratings of faculty: A reprise. *Academe*, 65, 384–397.
- McLeod, P. L., & Rosenthal, R. (1983). Micromomentary movement and the decoding of face and body cues. *Journal of Nonverbal Behavior*, 8, 83–90.
- Mehrabian, A. (1969). Significance of posture and position in the communication of attitude and status relationships. *Psychological Bulletin*, 71, 359–372.
- Meng, X., Rosenthal, R., & Rubin, D. R. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, 111, 172–175.
- Passini, F. T., & Norman, W. T. (1966). A universal conception of personality structure? *Journal of Personality and Social Psychology*, 4, 44–49.
- Paunonen, S. V. (1991). On the accuracy of ratings of personality by strangers. *Journal of Personality and Social Psychology*, 61, 471–477.
- Rosenshine, B. (1971). *Teaching behaviors and student achievement*. London: National Foundation for Educational Research.
- Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In R. Travers (Ed.), *Second handbook of research on teaching*. Chicago: Rand McNally.
- Rosenthal, R. (1987). *Judgment studies*. New York: Cambridge University Press.
- Rosenthal, R. (1989, August). *Experimenter expectancy, covert communication, and meta-analytic methods*. Donald T. Campbell award address presented at the 97th Annual Convention of the American Psychological Association, New Orleans, LA.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Newbury Park, CA: Sage.
- Rosenthal, R., Blanck, P. D., & Vannicelli, M. (1984). Speaking to and about patients: Predicting therapists’ tone of voice. *Journal of Consulting and Clinical Psychology*, 52, 679–686.
- Ryans, D. G. (1960). *Characteristics of teachers*. Washington, DC: American Council on Education.
- Sabatelli, R. M., & Rubin, M. (1986). Nonverbal expressiveness and physical attractiveness as mediators of interpersonal perceptions. *Journal of Nonverbal Behavior*, 10, 120–133.
- Snyder, M., Tanke, E., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology*, 44, 656–666.
- Smith, H. A. (1979). Nonverbal communication in teaching. *Review of Educational Research*, 49, 631–672.
- Swann, W. B., Jr. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review*, 91, 457–477.
- Trent, J. W., & Cohen, A. M. (1973). Research on teaching in higher education. In R. Travers (Ed.), *Second handbook of research on teaching* (pp. 997–1071). Chicago: Rand McNally.
- Watson, D. (1989). Strangers’ ratings of five robust personality factors: Evidence of a surprising convergence with self-report. *Journal of Personality and Social Psychology*, 57, 120–128.
- Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley.
- Wolfgang, A. (1979). The teacher and nonverbal behavior in the multicultural classroom. In A. Wolfgang (Ed.), *Nonverbal behavior: Applications and cultural implications* (pp. 159–174). San Diego, CA: Academic Press.
- Woolfolk, A. E., & Brooks, D. M. (1983). Nonverbal communication in teaching. In E. Gordon (Ed.), *Review of research in education* (Vol. 10, pp. 103–150). Washington, DC: American Educational Research Association.
- Woolfolk, R. L., & Woolfolk, A. E. (1977). Effects of teacher verbal and nonverbal behavior on student perceptions and attitudes. *American Educational Research Journal*, 11, 297–303.
- Zebrowitz, L. A. (1990). *Social perception*. Pacific Grove, CA: Brooks/Cole.
- Zuckerman, M., Miyake, K., & Hodgins, H. S. (1991). Cross-channel effects of vocal and physical attractiveness and their implications for interpersonal perception. *Journal of Personality and Social Psychology*, 60, 545–554.

Received April 6, 1992

Revision received August 5, 1992

Accepted September 26, 1992 ■