

Robert J. Wilson

and

Rhonda L. Martinussen

Queen's University

## Factors Affecting the Assessment of Student Achievement

*To what extent do expectations, disconfirming information, and degree of parental involvement in schooling affect teachers' judgments about a student's growth and achievement? This study manipulated these variables with 147 preservice teacher candidates as they assessed the progress of a student named Chris in language arts over a 10-week period. As predicted by social cognition findings, these results showed that early expectations and differential growth patterns were substantial contributors to differences in this student's reported final grade. For example, these assessors were impressed by their student's socioeconomic background and other contextual data, and this early information affected the grade awarded him or her weeks later. In addition, the pattern of achievement exhibited by Chris was also significantly related to the report card grade. If Chris seemed to improve, the grade improved, but if he or she remained steady or even fell behind, the grade was unaffected. The implications of these findings for assessment practice are discussed briefly in the conclusion.*

*Dans quelle mesure les attentes, l'information invalidante, et le degré d'implication parentale dans l'éducation affectent-ils l'évaluation que font les enseignants du progrès et des réalisations des élèves? Voilà les variables que la présente étude a manipulé avec 147 stagiaires pendant que ceux-ci évaluaient le progrès sur dix semaines qu'a réalisé un élève nommé Chris dans un cours des arts du langage. Tels que prédit par les connaissances sur la cognition sociale, les résultats ont indiqué que les attentes du début du stage et les schémas de croissance différentiels ont grandement contribué à l'écart dans la note finale que l'on accordait à l'élève. Par exemple, les évaluateurs ont été influencés par le milieu socio-économique ainsi que par d'autres données contextuelles qu'on leur avait présentés au début de l'étude. Ces renseignements ont affecté la note qu'on accordait à Chris plusieurs semaines plus tard. De plus, il existait une relation significative entre le modèle de réussites effectué par Chris et la note qu'on lui donnait sur son bulletin. Quand Chris semblait faire des progrès, sa note s'améliorait; par contre, quand il restait au même niveau ou qu'il réussissait moins bien, sa note restait inchangée. On discute rapidement des implications de ces résultats pour la pratique d'évaluation.*

Teachers report that the most important evidence they collect for assessing students' growth and achievement comes from observations. Observations may be made informally and directly (Nicholson & Anderson, 1993) or formally and indirectly through the scoring and analysis of student products (Bachor & Anderson, 1994). Concurrently these teachers' schools often require assess-

---

Robert Wilson is a professor in the Faculty of Education. He specializes in assessment and evaluation.

Rhonda Martinussen is a doctoral candidate conducting her research at the Hospital for Sick Children under the supervision of Rosemary Tannock in the Department of Brain and Behaviour Research.

ments from teachers using more structured, objective procedures, typically for grading and reporting. But even with these policies in place, evidence exists that the formal systems are modified by observations and their interpretations (Wilson, 1990). Not much is understood about the processes teachers use to make their judgments about student achievement. A major goal of the present study was to begin the process of identifying some of the factors that bear on these decisions.

Observations involve data gathering and interpretations conducted in dynamic, interactive situations. Although not studied extensively in school-based environments, social judgments involving observation have been studied in the laboratory. Some work has been done with children (Babad, Bernieri, & Rosenthal, 1991), but more often investigations have been directed using college students in short-term, structured situations (Lord, Lepper, & Preston, 1984; Snyder & White, 1981). Application of these findings to the relatively unstructured and sustained interactions of teachers with younger individuals remains to be tested.

Some of these laboratory-based findings, however, do seem to have analogues with reports of teachers' observations of students. For example, teachers have recorded a large measure of trust in their ability to make early and valid judgments about students' abilities and potential performance (Wilson, 1990). The social cognition literature has found similar confidence displayed by participants, with early judgments being both easy to create and difficult to change (Edwards, 1998; Higgins, Rholes, & Jones, 1977; Lord, Ross, & Lepper, 1979; Srull & Wyer, 1980). This confidence may arise from another result found in the literature: Individuals routinely believe that their own views are the majority views, that those who believe similarly display appropriate character strength, whereas those who believe differently have faulty dispositions (Higgins & Bryant, 1982; Ross, 1977). Traits generally tend to be overestimated as factors affecting behavior in the laboratory, whereas the impact of the situation in which the behavior is displayed is often underestimated (Erber & Fiske, 1984; O'Sullivan & Durso, 1984; Stapel & Schwarz, 1998). Because classrooms are often self-contained and are also places where teachers are required to ascribe reasons for behaviors, such a tendency may also be evident there.

How individuals handle disconfirming information for a judgment may also be important in contributing to valid assessment of student work. In the laboratory such information is noticed, and then (typically) explained away rather than confronted (O'Sullivan & Durso, 1984). Report writing has also been found to affect what is remembered about an event, especially when the report itself may have been composed to fit the expectations of the receiver and shaped by a context provided by the other participants (Allison, Mackie, & Messick, 1996; Higgins & Lurie, 1983; Higgins & Rholes, 1978; Snyder & Cantor, 1979). Informal and formal reports are, of course, a ubiquitous feature of school life.

All these attributes if generalizable to teachers' behaviors carry with them the potential for invalid conclusions about students' achievements. Some experimental work using teachers as participants (Chase, 1986) has demonstrated

that the interaction of both student and teacher characteristics can influence the assessment of student work.

The main purpose of this aspect of the overall study was to determine whether teachers' expectations and observations influenced their assessment of achievement. A second purpose included investigating whether contextual factors such as the degree to which parents were interested in their child's progress might also affect the assessment.

### *Method*

#### *Participants*

The sample consisted of 147 teacher candidates (110 females and 37 males) attending an eastern Canadian university's Faculty of Education. The investigation was carried out in five sections of a required course in teaching skills taught by three different instructors. All participants were training to be teachers of students enrolled in grades 4-10.

#### *Materials*

A collection of assessment materials in language arts was prepared specifically for this study. Some materials were adapted from the work of real students, teachers, and schools, whereas others were either published worksheet materials from commercial sources or specially designed assignments that appeared to be teacher-prepared. All these materials were used to track the progress of an imaginary grade 8 student named Chris. Additional materials contained information about Chris's parents and siblings, number of schools attended, as well as recent standardized achievement test scores and whether Chris had been supported through special education. In addition, a brief essay entitled "Meet Me" and purporting to be written by Chris was included in which Chris described aspects of his or her life. All of these materials were provided the participants in the first week of the study.

In subsequent weeks the participants were given information about Chris's class and "regular" teacher (i.e., the layout of the room, requests for parent-teacher conferences, library readings, scores on group tasks, and class achievement results in all subjects) as well as further examples of Chris's work in language arts (i.e., writing samples, objectively and subjectively scorable worksheets, a tape of Chris's reading and a miscue analysis of it, and a final examination for language arts). Finally, all participants were provided with a report card on which they were asked to record a term percentage and letter grade in language arts for Chris along with a general comment.

#### *Design*

These materials were distributed to each participating section on seven occasions over a 10-week calendar period. Each participant kept a folder of Chris's work for the entire period of the study. Participants were requested to read all the materials that were provided to them about Chris and his or her work in school, and mark all the tasks that Chris had to do as part of his or her regular work in language arts. The scenario included the information that Chris's regular teacher would use these judgments in her own assessment of Chris's progress. When the participants were provided with updates of the

Table 1  
Variable Labels, Levels, and Definitions

<i>Variables</i>	<i>Levels and Definitions</i>
<i>Manipulated Independent Variables</i>	
Expectations	High, Middle, and Low Socioeconomic Status
Growth	Improving, Steady, and Falling Behind Previous Performance
Parental Involvement	Parents Did or Did Not Respond to School's Invitations
Gender	Male or Female Chris as Revealed in Oral Reading Tape
<i>Work Samples</i>	
Trip to the Mall	Directed Writing Exercise
Salmon for Simon	Embedded Multiple Choice
Elephant	Short-Answer Exercise
School Dance	Directed Writing Exercise
New Kid on the Block	Short Answer
Ghost Ship	Completion
Mending Wall	Directed Writing Exercise
Final Examination	Combination of Objective and Subjective Items
Report Card Grade	Percentage Linked to Letter Grade Categories

class's achievement, Chris's scores were blank so that the participants' marks could then be incorporated into the class record.

This aspect of the overall study used a fully randomized factorial design incorporating three levels of *expectations*, three levels of *growth*, two levels of *parental involvement*, and two levels of *gender* (see Table 1 for the variable labels, levels, and definitions).

Expectations were varied by providing information on the portfolio itself and through Chris's own statements in the "Meet Me" essay. High expectations for Chris were created by having information on the portfolio that Chris had an executive for a father, a mother who stayed home, and that Chris showed superior achievement on a standardized test series. In addition, the "Meet Me" essay indicated Chris had expensive hobbies and a self-contained room for studying and sleeping. Middle-level expectations were created for other Chrises by having information that indicated that both parents worked outside the home and that Chris had achieved near the median on standardized tests. Low expectations were created by having both parents unemployed, with Chris having moved from school to school and a sense in the "Meet Me" essay that Chris was alienated from adult and family influences.

The growth variable was created throughout the study in two ways. First, each participant was provided with achievement reports in the form of class marks that showed Chris *improving*, *remaining steady*, or *falling behind* in academic production across a range of subjects and work. Second, real samples of Chris's work demonstrated the same pattern. A Chris in the *improving* category, for example, produced work in the first few weeks that was generally poor in quality but that improved as the term went on. A *steady* Chris maintained a mid-range performance throughout, and the *falling behind* Chris began well, but demonstrated a deteriorating pattern of performance throughout the

term. All writing samples and most short answer exercises, therefore, had three levels of achievement embedded in them in order to implement this aspect of the design.

Through messages from the principal and other data that showed which parents turned up at parent-teacher conferences, two levels of *parental involvement* were implied, one indicating high interest on the part of the parents in Chris's school and the other little interest. The gender designation could only be obtained by listening to a tape-recording of Chris's reading that was provided each participant as part of Chris's reading miscue exercise. By checking the gender of each teacher participant, it was also possible to examine the interactions of teacher and student gender. Not all participants were able to distinguish Chris's gender accurately from the tape although all participants ascribed a gender to their Chris as shown by their comments on Chris's work and/or report card. All other information was gender-neutral. As a result, the gender variable was dropped from the inferential analyses.

A final examination was developed that was common across participants. All participants received exactly the same student-completed examination to mark. The first section assessed grammar concepts and comprised five subsections worth 20, 15, 25, 10, and 15 marks in total respectively. The second section assessed passage comprehension through a combination of completion, true-false, and multiple-choice items worth 25 marks. The final section was a combination reading-writing task asking students to provide the main idea for a nonfiction reading passage and then write a summary of the article. This section was valued at 20 marks.

Both quantitative and qualitative data were collected during the course of the study. The quantitative data consisted of scores on writing samples, exercises, the final examination, and a report card mark. In addition, the manipulated contextual variables of expectations, growth, and parental involvement were coded for analysis. Two checks on the manipulation of expectations and growth were also coded and analyzed. There were 18 missing data points (out of a possible 2,628) on the teacher-scored variables, and these were replaced with scores generated by regressing the missing values onto the *expectations*, *parental involvement*, *growth*, and *gender* variables. Pearson correlations and analysis of variance were used to explore the numerical data in this aspect of the study.

All members of the participating sections were involved in the study. As a consequence, four complete iterations of the design were implemented plus some partial replication. The three oversampled portfolios resulted in an overrepresentation of three boys, *interested* parents, *low* expectations, and *falling behind* students for a total of 147.

The experimental hypotheses were that variations in expectations, growth, and parental involvement would each independently affect the reported grade given to Chris for the term. Specifically, Chrises whose parents came from a high socioeconomic background would score significantly higher than Chrises whose parents came from a middle-class background and that both groups would score significantly higher than Chrises from a lower-class background. Second, Chrises whose production pattern showed significant improvement over the term (although their average achievement across all instruments

would not necessarily differ from the other two groups) would be rewarded with significantly higher grades than those students whose patterns were *steady* or who showed a *falling behind* pattern. It was further hypothesized that the *falling behind* group would not differ significantly from the *steady* group. These predictions are based on the social cognition findings that suggest that disconfirming information may be explained away. If these novice teachers believed they were helping Chris, then information suggesting Chris was not improving under their care would probably be ignored or explained away. Finally, it was hypothesized that those Chrises whose parents were involved in schooling, who appeared interested in their progress, would have a significantly higher reported grade than those Chrises whose parents exhibited little or no interest in their schooling. No interactions were hypothesized.

### Results

Descriptive statistics for all variables are reported in Table 2. These data indicate that the teachers' scoring distributions were fairly consistent across the writing samples and the short-answer exercises. Most distributions were negatively skewed with medians slightly higher than the mid-point of the range, and in all but one case higher than the mean. No answer keys were provided the participants for any of the assignments. Thus the marks that were produced reflected each participant's views of what would constitute correct and incorrect responses from Chris, or in the case of the writing samples, degrees of acceptability. The "School Dance" essay was common across participants, and this is reflected in the smaller variability.

The final task given the participants was to create a report card mark for Chris in percentage format. According to the report form's definitions of letter grades, the lower limit of an A grade was 80%; that of a B grade 70% and a C 60%. The other two possibilities were D and E. None of the participants awarded their Chris either of these two grades, and only five of them awarded their Chris a grade as low as C. The median award was 77% with a range of 65%

Table 2  
Descriptive Statistics for all Variables

Variables	Minimum	Maximum	Mean	SD
Expectations	1	3	1.98	.82
Growth	1	3	1.96	.83
Parental Involvement	1	2	1.49	.50
Gender	1	2	1.51	.50
Trip to the Mall	13	25	18.82	2.72
Salmon for Simon	1	10	6.04	2.16
Elephant	8	15	12.41	1.42
School Dance	15	24	20.17	2.06
New Kid on the Block	7	18	14.33	2.53
Ghost Ship of Mahone Bay	2	9	5.43	2.11
Mending Wall	6	25	19.68	3.34
Final Examination	79	122	100.26	6.70
Report Card Grade	65	90	77.24	4.75

Table 3  
Analysis of Variance for Expectation, Growth, and Parental Involvement on Report Card Grade

Source	Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	405.90 <sup>a</sup>	5	81.18	3.962	.002
Intercept	875,195.16	1	875,195.16	42,717.86	<.001
Expectations	194.50	2	97.25	4.75	.01
Growth	203.90	2	101.95	4.98	.008
Parental Involvement	.17	1	.17	.01	.93
Error	2,888.78	141	20.49		
Total	880,410.50	147			
Corrected Total	3,294.68	146			

<sup>a</sup> $R^2 = .123$ .

to 90%. The distribution of marks for the report card was virtually symmetrical, as was the distribution for the final examination. The medians for the final examination and the report card were both 77%.

To test the experimental hypotheses, a multi-way ANOVA was conducted (see Table 3). In the full model, expectations ( $F=4.75$ ,  $p=.010$ ) and growth ( $F=4.98$ ,  $p=.009$ ) were found to be significantly related to the report card mark. Parental involvement was not significantly related. The corrected model in total accounted for 12.3% of the variability in the report card marks ( $F=3.962$ ,  $p=.002$ ).

Figure 1 provides a summary of the key relationships. In terms of the specific comparisons hypothesized, it was found that the high level of expectation was significantly different from the other two levels (contrast of high vs. low= $-2.6225$ ,  $p<.05$ ; mid vs. high= $-2.333$ ,  $p<.05$ ) but not the difference between mid and low (mid vs. low = $.2892$ ,  $p>.05$ ). The improving level of growth was also found to be significantly different from the falling behind and steady levels combined (contrast of improving vs. falling behind and steady= $2.481$ ,  $p<.05$ ). The contrast between the falling behind group and the steady group, however, was not significant (contrast of falling behind and steady= $-1.757$ ,  $p>.05$ ). All tests were conducted using the Bonferroni procedure.

To begin to explore the relationships further, Pearson correlations were calculated between all pairs of variables in the study. These results are provided in Table 4. As might be expected, the report card percentage was significantly correlated with almost all the elements the teachers were asked to mark. The only exception was the short answer assignment called "A Salmon for Simon." This particular assignment contained multiple-choice questions embedded in the story where no one answer was correct. Creating a key may well have challenged the participants unduly. In addition, some of the answers made by Chris were obscured by poor handwriting, as in an example where a lower case *c* could well have been a lower case *a* and so on. The difficulty caused by these deliberate ambiguities probably accounts for the fact that this story fails to correlate with most of the other exercises, including the final

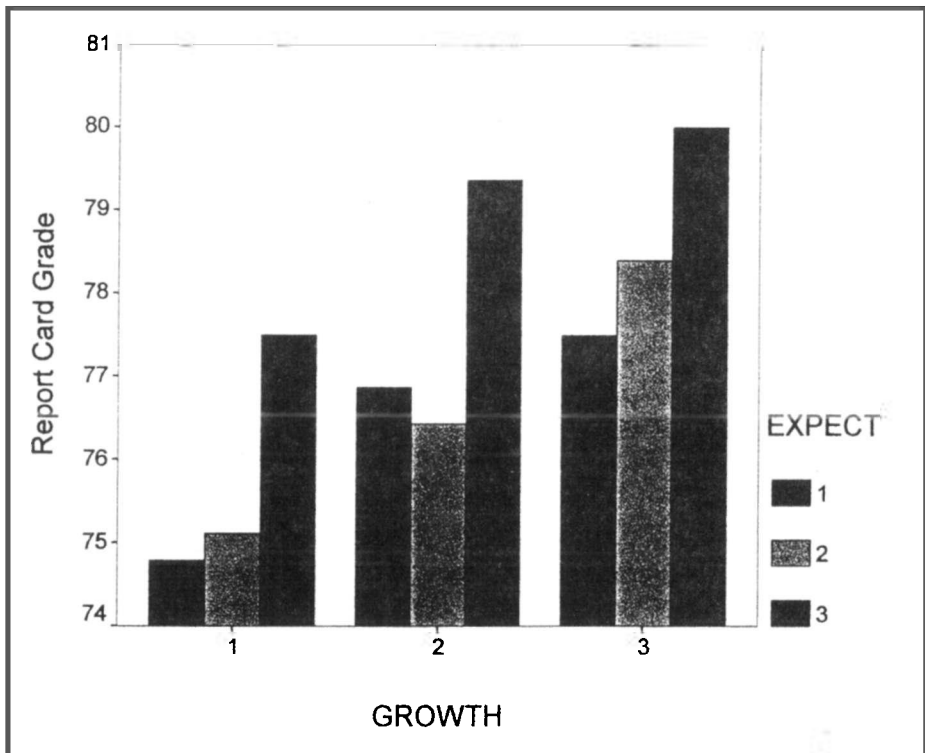


Figure 1. Report card grade by growth and expectations.

examination and the report card grades, and where it does correlate it does so negatively in all but one instance.

The final examination's total of 130 marks was equivalent to the total of the other exercises combined. The report card grade, then, might be expected to correlate highly with the final examination, and it does ( $r=.49, p<.01$ ). This correlation, however, is virtually equalled or exceeded by two of the writing samples' correlations with the report card grade indicating that these teachers put a great deal of emphasis on Chris's writing ability and less emphasis on those exercises that did not demand it. There are at least three plausible reasons for this emphasis. The first is that the participants felt that these types of responses gave a more valid picture of Chris's abilities than did the other exercises. The second is that in marking these assignments the teachers provided responses to Chris personally and became more committed, therefore, to their validity than they would have when they were merely grading exercises and recording marks. Finally, these teacher candidates may have felt reasonably confident responding subjectively to a writing sample but less confident of their ability to score more objective items adequately.

This final interpretation is supported indirectly by how the common final examination was scored. As with the other tasks, no scoring key was given the teachers, only the total number of marks to be awarded each section. The absence of a key may have contributed in part to the finding that the *objective* items produced a wider range of scores than did the *subjective* items. The objective items required teachers to produce a scoring approach, and qualita-



Table 4  
Pearson Correlations for all Variables

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Expectations		.02	.01	.02	.16	-.02	.22	.32	.02	.08	.06	-.03	.23
2. Growth	.02		.03	-.02	.31	-.74	-.15	-.18	.00	.86	.61	-.03	.25
3. Parental Involvement	.00	.03		.03	.01	.00	-.02	.07	.00	.04	-.04	-.03	.02
4. Gender	.03	-.02	.03		-.02	.06	-.09	.05	-.01	.01	.00	.11	-.04
5. Trip to the Mall	.16	.31	.01	-.02		-.31	.08	.32	.20	.23	.31	.30	.51
6. Salmon for Simon	-.02	-.74	.00	.06	-.31		.29	.03	.03	-.58	-.39	.02	-.06
7. Elephant	.22	-.15	-.02	-.09	.08	.29		.14	.10	-.03	.14	.04	.27
8. School Dance	.32	-.18	.07	.05	.32	.03	.14		.22	-.13	.01	.25	.38
9. New Kid on the Block	.02	-.01	.00	-.01	.20	.03	.10	.22		.06	.14	.15	.41
10. Ghost Ship	.08	.86	.04	.01	.23	-.58	-.03	-.13	.06		.59	-.03	.36
11. Mending Wall	.06	.61	-.04	.00	.31	-.39	.14	.01	.14	.59		.11	.47
12. Final Examination	-.03	-.03	-.03	.11	.30	.02	.04	.25	.15	-.03	.11		.49
13. Report Card Grade	.23	.25	.02	-.04	.51	-.06	.27	.38	.41	.36	.47	.49	

Note.  $r > .16$ , significant  $p < .05$ .

$r > .21$ , significant  $p < .01$ .

tive analysis of their decisions indicates they varied widely in their criteria. In some cases the criteria for awarding of marks by a participant were specific and followed; in others a mark was given with no justification or criteria evident. In addition, the weights within question groups varied considerably. For example, Chris could receive 5 out of 7 or 10 out of 12 for the same item depending on the internal weights decided on by the individual teacher.

To pursue the analysis further, the background variables (expectations, growth, gender, and parental involvement) were regressed on each of the subtest scores: objective and subjective. Neither of these analyses was significant. Using Chris's total marks before the examination as a predictor, however, did result in a significant prediction of the examination total score, but not highly so ( $R^2 = .11$ ,  $p < .05$ ).

Checks were made on the expectation manipulation intended in the study. Participants were asked in Week 6 to recall whether certain statements had been made in the "Meet Me" essay provided to them in Week 1. Four statements were provided. Two were taken directly from the essay, another was not, but was plausible given the essay's contents, and a fourth was neither plausible nor stated. Between 70% and 80% of the participants were able to indicate correctly which statements had appeared in the essay and which had not, whether plausible or not.

#### Discussion

The main finding from the data collected in this study is that objective evidence provided about a student's performance did not in and of itself determine that student's grade. These novice teachers allowed their expectations about how Chris might do to affect their judgments about performance. For some of them too, if Chris was showing improvement over the term, this would be rewarded with higher grades.

Early judgments did seem to be made about Chris because of what Chris said and did and seemed to be in terms of interests and background. Of particular interest were the grading behaviors of those novice teachers whose Chrises declined in performance over the term. For these participants, Chris's end of term performance was significantly poorer both in comparison with others in the class and with Chris's own earlier performance. For most of this group the report card grade is indeed significantly below that of the improving group. An exception occurred in the subgroup of Chrises who were *falling behind* but who came from the high socioeconomic expectation level. These performances exceeded those of the Chrises in the other two socioeconomic levels not only in the *falling behind* category, but in the *steady* level as well. Consistent with the social cognition literature, participants may have treated disconfirming information differently for those students for whom they had higher expectations and who were *falling behind* as the term progressed. Information about slowing performance for this subgroup was either ignored or explained away.

Even though the school policy required these teachers to produce an objective grade based on weighting classroom work and tests, this policy was typically not followed. A plausible explanation could be that these teachers, once having summed up the scores and computed an average mark, "adjusted" it to fit with their more overall qualitative judgments about Chris and what they themselves were attempting to reward as teachers. This judgment usually resulted in an inflated grade as well. For 126 of the 147 grades the reported score was higher by more than rounding than what would have been reported had the school's policies been implemented accurately.

To some extent the circumstances of the present study should have made it easier to provide objective judgments about Chris's performance. None of the participants met Chris; all they had to rate his or her performance was work purporting to be produced by a single person. These young teachers, however, although having some experience in schools, are by no means expert in the area of assessing learning and may not reflect what their more experienced colleagues would do in a real classroom with a real Chris before them. Nonetheless, they did seem to infer many of Chris's personal qualities from what they were given and used those inferred characteristics to shape their judgments about what Chris was doing and how adequate this was in terms of achievement and growth.

If subsequent work in schools finds similar patterns, it would call into question attempts by school policies and policy-makers to pretend that judgments about student achievement are arrived at by formal, objective means. The notion of equity implied by many of these policies—that everyone is being treated fairly because everyone is being treated alike—may have to be examined and perhaps replaced with a view of equity that treats individuals according to their needs and abilities. Teachers who work with students daily, and who come to know them as individuals, find it difficult to replace an interactive, ongoing assessment of the person with objective models that do not fit the life of the classroom as they and their students know it. The task for teacher educators, researchers, and administrators then becomes one of finding out how to do this type of assessment in reliable and valid ways.

References

- Allison, S.T., Mackie, D.M., & Messick, D.M. (1996). Outcome biases in social perception: Implications for dispositional inference, attitude change, stereotyping, and social behavior. In M.P. Zanna (Ed.), *Advances in experimental social psychology*, 28 (pp. 56-57). Toronto, ON: Academic Press.
- Babad, E., Bernieri, F., & Rosenthal, R. (1991). Students as judges of teachers' verbal and nonverbal behavior. *American Educational Research Journal*, 28, 211-234.
- Bachor, D.G., & Anderson, J.O. (1994). Perspectives on assessment practices in the elementary classroom in British Columbia, Canada. *Assessment in Education: Principles, Policy and Practices*, 1, 65-95.
- Chase, C.K. (1986). Essay test scoring: Interaction of relevant variables. *Journal of Educational Measurement*, 23(1), 33-41.
- Edwards, J.A. (1998). Effects of causal uncertainty on the dispositional attribution process. *Journal of Experimental Social Psychology*, 34, 109-135.
- Erber, R., & Fiske, S.T. (1984). Outcome dependency and attention to inconsistent information. *Journal of Personality and Social Psychology*, 47(4), 709-728.
- Higgins, E.R., & Bryant, S.L. (1982). Consensus information and the fundamental attribution error: The role of development and in-group versus out-group knowledge. *Journal of Personality and Social Psychology*, 43, 889-900.
- Higgins, E.R., & Lurie, L. (1983). Context, categorization, and recall: The "change of standard" effect. *Cognitive Psychology*, 15, 525-547.
- Higgins, E.R., & Rholes, W.S. (1978). "Saying is believing": Effects of message modification on memory and liking for the person described. *Journal of Experimental Social Psychology*, 14, 363-378.
- Higgins, E.R., Rholes, W.S., & Jones, C.R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, 13, 141-154.
- Lord, C.G., Lepper, M.R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, 47, 1231-1243.
- Lord, C.G., Ross, L., & Lepper, M.R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37, 2098-2109.
- Nicholson, D.J., & Anderson, J.O. (1993). A time and place for observations: Talking with primary teachers about classroom assessment. *Alberta Journal of Educational Research*, 39, 3, 363-374.
- O'Sullivan, C.S., & Durso, F.T. (1984). Effect of schema-incongruent information on memory for stereotypical attributes. *Journal of Personality and Social Psychology*, 47, 55-70.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in Experimental and Social Psychology*, 10, 173-220.
- Snyder, M., & Cantor, N. (1979). Testing hypotheses about other people: The use of historical knowledge. *Journal of Experimental Social Psychology*, 15, 330-342.
- Snyder, M., & White, P. (1981). Testing hypotheses about other people: Strategies of verification and falsification. *Personality and Social Psychological Bulletin*, 7, 39-43.
- Srull, R.K., & Wyer, Jr., R.S. (1980). Category accessibility and social perception: Some implications for the study of person memory and interpersonal judgments. *Journal of Personality and Social Psychology*, 38, 841-856.
- Stapel, D.A., & Schwarz, N. (1998). Similarities and differences between the impact of traits and expectancies: What matters is whether the target stimulus is ambiguous or mixed. *Journal of Experimental Social Psychology*, 34, 227-245.
- Wilson, R.J. (1990). Classroom practices in evaluating student achievement. *Alberta Journal of Educational Research*, 36, 4-17.