Preservice Success and Professional Performance: A Comment on Greene and Dravland*

William J. Hunter
Mount Saint Vincent University

The issue presented in the title as the focus of Greene and Dravland's (1981) recent article is not the same as the issue presented in the abstract. The title suggests a focus on the prediction of individual career performance from knowledge of performance in preservice training, while the abstract implies an evaluation of the effectiveness of a teacher education program: "In the final analysis, a teacher education program must be evaluated in terms of the success of its graduates" (p. 5). The abstract goes on to describe a study designed to address the issue presented in the title.

This confusion of purpose persists into the body of the paper as the writers argue in the first sentence that successful completion of a teacher-education program ought to be interpretable as indicative of success in teaching. They then project (in the second sentence) that "it follows that employers should be able to expect those who obtained the greatest success in the education program to be the most successful teachers" (p. 5). While the two statements are certainly not incompatible, there is no logical necessity that the second must be true given the veracity of the first. It is entirely possible that faculties of education could succeed in discriminating between those who can perform acceptably as teachers and those who cannot without being capable of accurately predicting the level of success that will be achieved. (In fact, this conclusion seems to be consistent with both the research cited by Greene and Dravland and their own results.) That is, it is possible to separate wheat from chaff without being able to judge the quality of the kernels. The widespread use of pass/fail marks in student teaching may reflect the extent to which education faculty subscribe to this notion.

The major purpose of the study is stated as "to develop a collection of variables which appear to be the most valid and reliable for evaluating a teacher education program" (p. 7), yet the four specific objectives pertain to measures of individual teaching success and the interrelationships among them. While the kinds of measurements proposed might

* This comment is made on a previously published article by Myrna L. Greene and Vern Dravland, "Relationships Between Success in an Education Program and Success in the Teaching Profession," *Canadian Journal of Education*, 1981, *6*(1), 5-17.

conceivably be used in evaluating a teacher education program, the primary thrust of Greene and Dravland's literature review was to provide evidence and opinion to the effect that this would be a misuse of the measures. Recognizing this inconsistency, the authors say: "In spite of the weaknesses in the utilization of principal and self-evaluations, for this study it was decided to use these assessments as the initial measures of teaching success" (p. 7). A "more sophisticated" system was promised for future research.

The problem continues into the method section, which informs the reader that the study is part of a program that "has as its ultimate goal the development of a model for the evaluation of teacher education programs" (p. 7).

Surely one does not require advanced training in research methods or logic to see that there are a great many variables other than the quality of teacher training received that contribute to the amount of success achieved in practice. Likewise, one need not be a profound thinker to realize that many of the individual differences measured by the ratings used in this study may well have existed independent of (or even prior to) the teacher training program. To use these as measures of the success of the preservice training is analogous to using the IQ scores of students as data for evaluation of teaching performance.

It might be argued that the consistency between student-teaching evaluations and on-the-job evaluations is a source of evidence that might be used to evaluate teacher-education programs. Better programs, then, would be those in which consistency was high. This argument seems to underlie the Greene and Dravland study, but it has not been made explicit. Nevertheless, to use a measure of consistency as program evaluation data would seem to imply that the judgments made by principals or supervisors, serving as the criterion, are the more valid indicators of success. Again, the literature reviewed by Greene and Dravland would call this conclusion to question. So would a good many practising teachers who know the nature of (some) principals' classroom observations.

This confusion of purpose does, however, suggest a fundamental question about the goals of teacher preparation programs. Greene and Dravland seem to be operating on a premise that individual differences in achievement in preservice training may serve as a measure of the quality of the teacher preparation program. The extent to which these differences predict differences in teaching performance is apparently regarded by the authors to be an indication of quality in the teacher preparation program. If this were so, then faculties of education should strive to increase individual differences — to magnify the achievements of the successful — so that selection of those most likely to succeed will be enhanced and the program will be seen as effective. It seems to this writer that one might defensibly argue the converse — that a successful program is one that assures that all graduates achieve some minimal stan-

dard of performance — and that this might serve to minimize individual differences in teaching skills and consequently reduce the correlation with measures of on-the-job success (due to restriction of range).

If one accepts the Greene and Dravland paper for what the title claims it to be and ignores the confusion of purpose, one must ask: Why should this study be done? The literature clearly indicates that there is little, if any, correlation between the measures being studied. It would have been far better to pursue a study based on the "more sophisticated evaluation system" that is apparently being developed.

Not surprisingly, Greene and Dravland failed to find correlations between teachers' self ratings and principals' ratings; between practicum marks and principals' ratings; between practicum marks and self ratings; among "overall assessments by professors, teachers, or principals"; or among 39 individual evaluation items completed by principals and professors. Greene and Dravland also said they did not find this last result surprising, but for the wrong reasons. Given 39 variables, it would be reasonable to expect some correlations to exceed the .05 significance level even if the data were random digits. The reason these results are not surprising is because they have been so consistently found before. Even more surprising was the conclusion in the abstract that "there appeared to be a definite relationship between success in the program and success in teaching" (p. 5). This simply ignores the data.

Although it has not been explicitly stated, the multiple regressions reported appear to be stepwise regressions. The tendency of such procedures to capitalize on chance relationships leads most methodologists to require large n ($n \geqslant 200$) and/or replications on an independent sample. Greene and Dravland warn that the statistical results should be "interpreted with caution." Perhaps this explains their willingness to draw conclusions that contradict the results. In the case of the multiple regressions presented, I seriously question whether anything beyond the first variable stepped in (i.e., the zero-order correlations) ought to be interpreted at all.

In their penultimate section, Greene and Dravland present a "descriptive analysis" which must be the real basis of the conclusion discussed above. This analysis would have been more informative had the authors reported chi-square analyses of their predictions. It is not possible to reconstruct the contingency tables from the data given, but the data given are not obviously different from chance expectations. Perhaps the authors concluded that cell sizes were too small to allow such analyses, but they have not so indicated. Particularly puzzling about the descriptive analysis is the fact that it is based on $n = 30$ (11 *A* students, 8 *B* students, and 11 *C* students), although 46 subjects were identified initially. In fact, the number of subjects varies with each analysis, with little or no explanation.

If I have been unduly harsh in this criticism, it is for two reasons: (a) the abstract and statement of purpose raised hopes which the article did not fulfil, and (b) the article stands in sharp contrast to the other articles in what was an otherwise enjoyable issue. Most disturbing of all is that the authors come at length to a conclusion they could have reached on the basis of their literature review: the need for a study that uses evaluation by trained observers as a major criterion of successful teaching. Let us hope that that study makes more rewarding reading.

REFERENCE

Greene, M. L., & Dravland, V. Relationships between success in an education program and success in the teaching profession. *Canadian Journal of Education*, 1981, *6*(1), 5-17.

William J. Hunter is an Assistant Professor and Co-ordinator of Graduate Studies in Educational and School Psychology at Mount Saint Vincent University, Halifax, Nova Scotia B3M 2J6.