A PARADIGM FOR PROGRAM EVALUATION

Ronald Byrd
Louisiana State University, Shreveport

Cay Evans
Louisiana State University, Shreveport

Abstract

The purpose of this study was to describe use of a paradigm that had been developed to examine strength of preparation specifically addressed on the Professional Knowledge section of the National Teachers' Examination (NTE). Components of the model included assessment, analysis, and adaptation. A total of 466 subjects (student teachers, graduates, and principals) responded to questionnaires on one of two occasions, with five years separating assessments. Steps taken following analysis of the first-year data for purposes of program improvement were described. The instrument consisted of 18 indicators in the six categories outlined in NTE Professional Knowledge feedback to universities. Results indicated general agreement that the weakest scores in the first evaluation were in classroom management and in social aspects (including cultural differences, racial, and gender bias) and that, despite steps taken to improve, no significant change had occurred after five years. Analysis of variance revealed several differences between groups, with clear indications that principals were most supportive and graduates most critical, with student teachers falling between the two in their assessments. It was concluded that: (a) the paradigm's use of multiple sources was successful in evaluating the status of the aspects of teacher education investigated, (b) the procedure yielded a high degree of confidence in conclusions drawn, and (c) consistency in results provided a strong foundation for further decision-making.

The National Council for Accreditation of Teacher Education (NCATE) stresses the importance of systematic data collection from a variety of groups in order to “...foster student achievement through the modification and improvement of the conceptual framework(s) and programs...” (1995, p. 16). NCATE urges that assessments be done on a regular basis and seeks evidence of application of data toward program improvement. While such preparation for NCATE accreditation might be viewed as being “research,” the process is more situation-specific and there is generally little interest or emphasis on generalizing to other institutions (Alkin & House, 1992). This study is a case in point, but the concepts underlying this paradigm could be of value to other institutions developing follow-up strategies. The primary purpose of this study was to describe application of a paradigm developed to
monitor strength of preparation specifically addressed on the Professional Knowledge section of the National Teachers’ Examination (NTE). A concomitant practical purpose was to address NCATE recommendations.

**Methods**

**The Paradigm**

The first element of the paradigm was *assessment*. Three separate groups of subjects were contacted for feedback regarding perceptions of selected aspects of professional preparation. The rationale for use of multiple sources was that viewing the program from different perspectives would maximize probability of identification of problems. Statistical *analysis* should be an integral part of any follow-up, but too often the analysis consists of a cursory viewing, perhaps including some loose qualitative analysis, prior to filing away. This paradigm included formal qualitative analysis in addition to use of correlations among the three types of subjects and analysis of variance for differences among groups and test items. The combination of qualitative and quantitative analytical procedures was an important characteristic of the paradigm. Finally, without *adaptation*, defined here as steps taken to shore up weaknesses or to solidify satisfactory aspects of the program, the process would have been an exercise in futility, a waste of time and effort. Thus, the paradigm depends on a repeating cycle of assessment, analysis, and adaptation; an application of this model is described herein.

**Subjects**

The entire population of student teachers in their final semester prior to graduation acted as subjects (initial year, n = 43; fifth year, n = 37). For the initial assessment, 158 graduates from all who had graduated the preceding five years took part; a smaller randomly selected sample was contacted for the fifth year follow-up, with 55 taking part. All principals from two local school districts who had hired program graduates were asked to take part; participants were 76 for the initial study and 97 for the latter. Questionnaire return rates ranged from 96% (student teachers, first assessment) to 69% (graduates, fifth-year).

**Instrument**

A questionnaire was developed to address the six subcategories on the NTE Professional Knowledge test: planning instruction, implementing instruction, managing student conduct, evaluating teaching/learning, recognizing extra-classroom influences on teaching/learning, and knowledge of professional behavior. Recognizing extra-classroom influences included cultural differences, racial and gender bias, and using community resources. These will be collectively referred to hereafter as social aspects. Validity was established by a panel of experts and reliability was calculated by test-retest of a group of student teachers (n = 24, \( r = .94 \)) and a group of graduates (teachers, n = 20, \( r = .97 \)). There were three indicators for each of the six categories, for a total of 18 questions. A 5-point scale was used for each question, yielding total category scores ranging from 3 to 15 (very poor, 3.0-4.4; inadequate, 4.5-7.4; adequate, 7.5-10.4; good, 10.5-13.4, and; excellent, 13.5-15.0).

**Data Collection**

Questionnaire data collection was carried out during two spring semesters, five years apart. Student teachers took part during the last month prior to graduation, while graduates and principals were contacted during this same period by mail, with self-addressed and stamped envelopes accompanying an introductory letter and the instrument.
Analysis

Pearson’s r and analysis of variance were used, with decisions at alpha = .05. The Tukey follow-up procedure was used for determining the location of significant differences.

Results and Discussion

Initial Assessment

Findings for the initial assessment revealed that the 18 means of scores of principals, graduates, and student teachers ranged from 10.1 to 12.8, all but two in the good category (Table 1). There were few significant differences in responses of the three groups of subjects (Table 3) and there was a clear implication of general satisfaction with these six aspects of professional preparation.

However, previous informal feedback had supported NTE scores that had consistently indicated relative weakness in the categories of managing student conduct and extraclassroom influences. For this reason and because these were the two categories with scores falling into the adequate range, special attention was given to these items immediately following the initial assessment.

It was not surprising that managing student conduct might be a problem since it is a nation-wide concern (Metropolitan Life Survey, 1996). To attack this problem, a psychology course in behavior modification oriented toward teacher education was added to the curriculum, hoping to improve this element of teacher education preparation.

Social aspects was the second relatively weak area addressed. Student teachers, principals, and graduates were all consistent in rating this among the lowest scores, with no significant differences among the groups detected by ANOVA. This weakness was attacked by developing a course in sociology to focus on cultural, racial, and gender issues.

Five Years Later

Reflecting results similar to those on the initial assessment, the 18 means ranged from 9.3 to 13.3, with three earning adequate ratings and the remaining 9 being classified as good (Table 2). As before, results indicated a general satisfaction by all three groups of subjects with professional preparation. However, managing student conduct and social aspects remaining generally lowest of the six indicators. It is clear that steps made to improve performance in these very important elements did not work.

Graduates, First to Fifth Year Comparisons

Though ANOVA revealed no significant differences (Table 3), means for the fifth year were lower, without exception, than for the initial data. A disappointing finding was that this particular group, most of whom had the benefit of curricular changes following the initial study, did not rate their preparation in managing student behavior and in social aspects (primarily culture, race, and gender) as improved. It is clear that the measures taken to strengthen these elements were ineffective. There was no significant change and means remained in the adequate category.

Student Teachers, First to Fifth Year Comparisons

There were no consistent changes from the first to the fifth year, nor were there any significant differences across time (Table 3). As with graduates, a disappointment was the failure of the perceived quality of preparation to manage student behavior or to master social elements to improve. In fact, means of these two categories exhibited small reductions from
### Table 1
Means and Standard Deviations for Subcategories of Profession Knowledge, Initial Assessment

<table>
<thead>
<tr>
<th>Group</th>
<th>Planning</th>
<th>Teaching</th>
<th>Managing</th>
<th>Evaluation</th>
<th>Social</th>
<th>Profession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graduates</td>
<td>12.4+.2</td>
<td>10.7+.2</td>
<td>10.1+.2</td>
<td>11.8+.2</td>
<td>10.2+.2</td>
<td>12.3+.2</td>
</tr>
<tr>
<td>Stu Tchrs</td>
<td>12.8+.3</td>
<td>11.2+.4</td>
<td>11.6+.4</td>
<td>11.3+.3</td>
<td>11.6+.4</td>
<td>12.4+.3</td>
</tr>
<tr>
<td>Principals</td>
<td>12.5+.2</td>
<td>11.5+.3</td>
<td>11.2+.3</td>
<td>11.3+.3</td>
<td>11.4+.3</td>
<td>12.4+.3</td>
</tr>
</tbody>
</table>

### Table 2
Means and Standard Deviations for Subcategories of Profession Knowledge, Fifth Year Assessment

<table>
<thead>
<tr>
<th>Group</th>
<th>Planning</th>
<th>Teaching</th>
<th>Managing</th>
<th>Evaluation</th>
<th>Social</th>
<th>Profession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graduates</td>
<td>12.0+.3</td>
<td>10.1+.3</td>
<td>9.3+.4</td>
<td>11.5+.3</td>
<td>9.8+.4</td>
<td>12.0+.3</td>
</tr>
<tr>
<td>Stu Tchrs</td>
<td>12.1+.3</td>
<td>11.6+.4</td>
<td>11.5+.4</td>
<td>11.7+.4</td>
<td>11.4+.4</td>
<td>13.0+.2</td>
</tr>
<tr>
<td>Principals</td>
<td>13.3+.2</td>
<td>12.7+.2</td>
<td>12.6+.3</td>
<td>12.5+.2</td>
<td>12.6+.3</td>
<td>12.8+.2</td>
</tr>
</tbody>
</table>

### Table 3
Results of Analysis of Variance

<table>
<thead>
<tr>
<th>Category</th>
<th>ANOVA</th>
<th>Significant differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning</td>
<td>$F = 3.91$, df = 5,462</td>
<td>6&gt;2, 6&gt;4, 6&gt;5</td>
</tr>
<tr>
<td>Teaching</td>
<td>$F = 11.13$, df = 5,462</td>
<td>6&gt;1, 6&gt;2, 6&gt;3, 6&gt;5</td>
</tr>
<tr>
<td>Managing</td>
<td>$F = 15.90$, df = 5,462</td>
<td>6&gt;2, 6&gt;3, 6&gt;5, 1&gt;2, 1&gt;5, 4&gt;5, 3&gt;2, 3&gt;5</td>
</tr>
<tr>
<td>Evaluation</td>
<td>$F = 2.76$, df = 5,462</td>
<td>6&gt;1, 6&gt;5</td>
</tr>
<tr>
<td>Social</td>
<td>$F = 12.66$, df = 5,462</td>
<td>6&gt;2, 6&gt;3, 6&gt;5, 5&lt;1, 5&lt;3, 5&lt;4, 5&lt;6</td>
</tr>
<tr>
<td>Professional</td>
<td>$F = 1.33$, df = 5,462</td>
<td>none</td>
</tr>
</tbody>
</table>

Notes:
Initial assessment: 1 = student teachers, 2 = graduates, 3 = principals
Fifth year assessment: 4 = student teachers, 5 = graduates, 6 = principals

the initial to the fifth year assessment. It is clear that these two relatively weak, but very important, areas of professional preparation merit continued attention.

**Principals, First to Fifth Year Comparisons**

Principals’ perceptions of teacher education preparation underwent improvement across time in all six categories, with three being statistically significant (Table 3). The three elements in which changes were significant were teaching, managing student conduct, and social aspects. Positive feelings were clear, with all six categories being rated good (means, 10.5-13.4) in both assessments. Interestingly, means for managing student behavior and social aspects increased significantly, but remained among the lowest ratings. This lends additional
support to feelings of both graduates and student teachers that these are among the weakest areas of professional knowledge preparation.

**Subgroups’ Overall Perceptions Compared**

Principals were clearly the most positive and graduates the most negative of the three groups of subjects. Marked differences in overall scoring by the three groups should be considered. There were 15 comparisons in which ANOVA showed means of principals to be significantly higher than means of either student teachers or graduates (Table 3). Five comparisons of means revealed student teacher scores significantly higher than those of graduates. In no instance was a graduate mean significantly higher than one of either of the other two groups. The pattern is clear; principals were most optimistic, student teachers were intermediate, and graduates the most critical. Which group’s ratings would probably be most valid? Logic dictates that the graduates are best suited to judge. Student teachers’ higher ranking precedes their contact with the real world of teaching and probably reflects less mature evaluation. Principals’ administrative positions demand that they assess that of which they often have little or no direct personal knowledge. That is not to denigrate their perceptions; while judgments as to teachers’ relative overall abilities and success might be valid, principals could not be expected to be as attuned to specifics covered by this instrument as would their teachers. The teacher in the classroom knows best, better than administrators, and certainly better than student teachers whose exposure is yet to come.

**Summary and Conclusions**

Assessment resulted in responses by the 466 subjects indicating an adequate job in most areas of preparation regarding professional knowledge, with the remainder seen as good. Statistical and qualitative analysis of the first-year data led to adjustments which included attempts at reform in managing student conduct and in social aspects (culture, race, gender). Assessment in the fifth year following the initial study led to analysis that showed a failure of scores by student teachers and graduates to improve, a disappointing finding. The lack of change following the initial assessment, analysis, and adjustment is neither alarming nor does it indicate a failure of the paradigm. The cyclical nature of the process is such that it is never-ending and failures of steps taken for improvement are always only temporary setbacks. It was concluded that this paradigm for evaluation worked well: (a) strengths and weaknesses were clearly identified, (b) input from multiple sources was in general agreement, (c) consistency in results provided a strong foundation for further adaptations, and (d) faculty will continue to look creatively for solutions to problems as they are revealed by on-going assessment and analysis.

**References**

