A survey-based component for programme assessment in undergraduate pre-service teacher education

Curriculum and programme review are an important part of the improvement of university instructional programmes. The purpose of this article is to describe our efforts to develop and implement a survey-based component for programme assessment that provides a framework for on-going programme evaluation and improvement. In addition we will report on results from our first cycle of data collection, discuss how these results were used in evaluating our undergraduate instructional programmes, and propose a theoretical model of the dimensions of pre-service teacher education revealed by factor analytic techniques.

The purpose of the survey instrument we sought to use was to help us learn how recent graduates of our pre-service teacher education programme perceived the quality of their undergraduate preparation. In particular, we wished to look at our pre-service graduates' perceptions of their general education and professional education preparation as distinct elements in our assessment process.

In the United States, institutions that offer undergraduate and graduate teacher education programmes intended to prepare students for licensure must have those programmes approved by their state department of education. Individual states accredit these institutions in order to assure that teacher education programmes prepare pre-service teachers in accordance with state-mandated guidelines and requirements.

In addition to being accredited by their state department of teacher education, some institutions also seek accreditation through the National Association for Colleges of Teacher Education (NCATE). This external, independent accrediting agency is the only national professional accrediting body for teacher education institutions in the United States. Both the state education agencies and NCATE often require schools and colleges of education to show evidence of systematic programme evaluation through follow-up studies of their graduates as criteria for
professional accreditation. While standards are set both by the state and by NCATE and programmes are designed to meet those standards, it was the perceived effectiveness of the programmes that we sought to determine.

With this problem in mind we turned to the teacher education programme assessment literature, hoping to find models for the development of the instrument we would ultimately use in our study. Initially the literature review suggested a less than optimistic outlook. Galluzzo and Craig (1990) have, for example, characterised programme evaluation as the ‘orphan’ of teacher education. According to their review of the literature, efforts to establish evaluation models for teacher education programmes have typically been ‘site-specific, idiosyncratic models designed for particular applications at individual institutions’ (p. 602). Although it can be reasonably argued that the proposal to develop an assessment instrument for evaluating our pre-service teacher education programme was, in fact, site-specific, there were two reasons why we wished to adopt a broader view.

One reason for developing an assessment instrument that went beyond immediate ‘in-house’ needs is that, while such an instrument might provide some useful information about pre-service graduates’ perceptions of their programmes, it would not provide meaningful comparisons with similar studies conducted at other institutions. A second reason for adopting a broader view was that one of the most common criticisms of programme evaluation studies in the teacher education literature is that they tend to be one-shot studies that are theoretically ill grounded and generally fail to play a role in a larger on-going evaluation scheme.

We found in our review, however, that there were at least two different approaches to evaluation that appeared relevant to our needs. One approach involved adapting an instrument developed for use in an objectives-based teacher assessment programme and the other involved the use of a ‘standard’ assessment instrument specially developed to allow comparisons across institutions.

The concept behind objectives-based evaluation is that teaching competence is reflected in the extent to which teachers demonstrate specific objectives in their teaching. Presumably the objectives upon which the instrument is based are related to teaching effectiveness. On that assumption, it seems reasonable to think of the objectives as one standard by which teacher preparation programmes can be evaluated. In one study (Murray, 1991) similar to that described here, a random sample of 1985–90 pre-service graduates of the Berry College teacher education programme were mailed surveys asking them to report how adequately they felt their education had prepared them for completing the competence-based assessment programme for beginning teachers that had been in place in Georgia since 1980. In effect the Georgia Teacher Performance Assessment Instrument (TPAI) provided the conceptual framework within which graduates were asked to evaluate their professional training. One important limitation of the Berry College study was that only recent pre-service graduates were surveyed, yet results reported by Adams (1987) suggest that beginning teachers’ perception of the adequacy of their
undergraduate training appears to be unstable until about the fifth year of teaching. Given Adams’s findings, we decided to include a ‘years of teaching experience’ item in the demographic data we collected.

In addition to the TPAI, a variety of other competence-based assessment instruments have been developed for use in the United States. Some of these instruments (like the TPAI) are characterised as high-inference measures that provide numerical scores intended to represent level of competence. Other measures, like the Classroom Observations Keyed for Effectiveness Research (COKER), developed at Toledo University, are characterised as low-inference, intended simply to discriminate between competent and non-competent performance in each of the areas assessed. Factor analytic studies (Wiersma et al., 1983) suggest that although the TPAI and the COKER appear to account for the same degree of variance, the two instruments appear to have very different factor structures, the TPAI resulting in two factors whereas the COKER produced five consistently strong factors. Since the goal of our programme assessment is to identify specific programme strengths and weaknesses, the factor structure of the COKER would appear to meet our needs more adequately but in this case, since practising teachers are the intended source of data, the COKER is subject to the criticism that it will not pick up on important differences, since, presumably, non-competent individuals are both less likely to have obtained a teaching position and more likely to have left the profession. The TPAI, by providing a high-inference measure, presumably would allow us to assess degrees of programme effectiveness, but its factor structure does not appear to support distinct curricular areas.

A second approach to programme evaluation employed a general-purpose questionnaire specifically developed to allow comparisons between pre-service teacher education programmes at different universities. In the study that employed this approach (Loadman, 1989), programme evaluators from ten institutions came together at the National Center for Research on Teacher Education at Michigan State University in order to develop a common questionnaire that could be used by a wide variety of institutions so that meaningful comparisons could be made. Preliminary results reported by Loadman and Gustafson (1990) suggested that the questionnaire might help us both conceptualise our assessment programme and allow us to make comparisons with pre-service programmes at other universities.

Our review of the literature suggested that, while we might serve our immediate short-term goals by designing an instrument tailored specially for our curricular areas and students, we would as a result fail to achieve other important long-term goals related to the ways we conceptualise and implement our programmes in pre-service teacher education and how our programme to prepare beginning teachers compares with other institutions across the country. We felt it was in the best interest of our long-term development as a programme to adopt a ‘generalisable’ assessment model and as result adapted the short form of the National Database survey for our use.
ADAPTATION OF THE SURVEY AND RESEARCH METHODOLOGY

Adaptation of the National Database survey for our needs involved three modifications. Since our data were collected by telephone, items were rephrased, if necessary, to be consistent with the telephone format. Secondly, we adapted the response scale employed in the survey. The original short-form National Database survey employed two different response scales for attitudinal items, a three-point scale for professional skills and a seven-point scale for all other scaled items. We decided to employ a single seven-point Likert scale for all scaled items. We did this both because a single scale would make interpretation and comparisons easier and because we felt respondents might be confused by switching the response scale during the course of the telephone interview. Our third modification was to add a number of items to our survey from the long form of the National Database survey. Items 64, 65, 72, 76 and 81 from the long-form survey were added (becoming items 23–7 in our survey), since they addressed professional preparation issues of special interest to us. Items 46–58 from the long form were also added (becoming items 31–43 on our survey) in order to provide a measure of broader subject area knowledge both in professional and in general education.

The adapted survey consisted of a total of forty-six items. The first ten and the last three of these items requested demographic information, information about the subjects’ current professional status and some general attitudinal measures reflecting subjects’ general philosophy of learning and teaching. Seventeen items (items 11–27) were designated Skill items since they focused on teaching skills such as selecting, preparing and using media in instruction (item 23), planning stimulating lessons (item 11), and other specifically professional skills. Three items (items 28–30) were designated Quality items, since they asked respondents to rate the overall quality of their student teaching/internship experiences (28), liberal arts/general education courses (29) and courses in their professional education programme of study (30). Thirteen items (items 31–43) were designated Subject items, since they focused on pre-service teacher education graduates’ perceived level of knowledge in broader subject areas that included both general and professional education.

Subjects
A list of approximately 400 pre-service Teacher Education 1988–92 programme graduates and their telephone numbers was generated from our files. This list consisted of the total number of persons graduating from our pre-service teacher education programmes during that period. An attempt was made to call each individual on the list and administer the survey as a telephone interview, with subjects’ responses coded direct into electronic files. If, after three telephone call attempts over the course of one week, individuals were not successfully contacted (i.e. no one answered the telephone, or an answering machine message was played)
they were dropped from the subject pool. A total of 201 interviews were completed with those who were reached by telephone. All of those contacted were willing to respond to all items on the survey. The 50 per cent response rate, therefore, was the total number of pre-service graduates over the years in the study who were reached by telephone. This total of 201 respondents included 131 Elementary Education, 54 Secondary Education and 16 Special Education pre-service teacher education programme graduates.

Data analysis
Since we had opted to make use of the National Database survey form, our first analyses consisted of comparing our data with those reported by Loadman and Gustafson (1990) and Freeman (1993). Moreover, since means and standard deviations were reported, it was possible to carry out a number of population t tests (treating the database as the population) in order to assess the likelihood that differences could be attributed to random variation.

In addition to comparisons with the National Database, we were interested in looking at student responses across the five years of pre-service graduates included in the survey. If, for instance, there was a declining trend in the respondents’ attitudes toward their programmes we reasoned that could be a cause for concern, whereas an improving trend would be viewed more positively. Analysis of data across time was carried out on three levels.

The first level of analysis across time involved calculating a mean rating across all thirty-three Skill, Quality and Subject items, resulting in a single overall score that we interpreted as reflecting a global evaluation of our students’ educational experience. These global means were then analysed in a one-way ANOVA (alpha = 0.05) to assess whether any changes had occurred across the years of interest (1988–92).

In the second level of data analysis across time we calculated mean scores for each subject within the Skill, Quality and Subject sets of items. We interpreted these scores as reflecting sub-scale measures for our instrument. Each of these sub-scale means was then analysed in a one-way ANOVA (alpha = 0.05) to assess whether any changes had occurred across the years of interest (1988–92).

In the third level of data analysis across time each of the thirty-three Skill, Quality and Subject items was individually analysed. Analysis of individual items involved a series of one-way ANOVAs to determine whether there were statistically significant differences in individual item means across years. The purpose of the analysis was to determine what, if any, changes had occurred across time in the evaluation of specific areas in our programme by our graduates. Presumably, if specific aspects of our programme had improved (or declined) in quality, our pre-service graduates’ evaluations would be influenced. In addition, if major changes in our programme could be related to specific changes in the way pre-service graduates responded to
our survey we conjectured that we might be in a better position to evaluate the impact of those changes.

Since, in the analysis of individual items, we planned numerous statistical tests (one ANOVA for each of the thirty-three items of interest), it became necessary to consider both familywise and per comparison alpha values where the familywise alpha set an upper limit for type I error across all comparisons and the per comparison alpha set an upper limit for type I error on each comparison. Setting the familywise alpha at a maximum of 0.05 resulted in a per comparison alpha of approximately 0.001, where the relationship between the error rates is given by $\alpha_{FW} = 1 - (1 - \alpha_{PC})^c$ and $c$ represents the number of planned comparisons (in this case, thirty-three) (Keppel, 1982).

We were also interested to know whether subsequent teaching experience influenced the way our pre-service graduates evaluated their prior educational experience. Since the data had been collected as a part of our demographics, we simply sorted students into groups according to whether or not they had ever taught and reanalysed our data, using average values across all five years in the study.

Finally, we attempted to validate the conceptual structure presumed by our division of items into three categories (Skill, Quality, Subject). Our effort at conceptual validation relied on factor analytic techniques used to sort items into categories based on the way our pre-service graduates had responded on the survey.

RESULTS

Comparison with the National Database

Results of the analysis comparing our data with the National Database were quite positive, with our pre-service graduates reporting significantly higher ratings across Skill ($t(200) = 9.95564$, $p < 0.0001$), Quality ($t(200) = 5.92466$, $p < 0.0001$), and Subject ($t(200) = 5.13016$, $p < 0.0001$) sub-scales. Since use of the National Database involved comparisons back to the original three-point Likert scale on the Skill items, means and standard deviations reported by Loadman and Gustafson (1990) for this scale were adjusted to range across a seven-point scale (adjusted mean = 4 + ((mean – 2) × 3), adjusted sd = sd × 3). Long-form Skill items that were added to the survey (items 23–7) were not included in comparisons with means reported by Loadman and Gustafson on the short-form survey. The Subject scale mean and standard deviation required the same adjustments noted above and were drawn from the long form of the National Database survey, since all of these items were added from that form. The set of Quality items and the scale employed did not differ from the original short-form survey and thus did not require any adjustments for comparisons with the reported short-form data. In addition, comparison with the data reported by Freeman (1993) revealed significantly more positive ratings (with per comparison alpha = 0.003 in each case) on all individual
Skill items except item 11 (Planning stimulating lessons) and item 25 (Monitoring students’ progress).

**Comparisons across time**

Results of analyses across time indicated increasingly positive ratings across the years 1988 to 1992. A one-way ANOVA indicated significant differences across time for our global measure (AVE) averaging all thirty-three items ($F(4,196) = 4.0293, p = 0.0037$). A post hoc Scheffe analysis revealed a significant group difference between the 1989 and the 1992 graduates ($p = 0.0196$). Results of the analysis of the Skill, Quality and Subject sub-scale scores revealed a significant difference across years for the Skill sub-scale only ($F(4,196) = 5.5158, p = 0.0003$). Scheffe post hoc analyses of the Skill sub-scale data revealed significant differences between 1989 graduates and graduates from 1991 ($p = 0.0143$) and 1992 ($p = 0.0023$). Analysis of the Quality and Subject sub-scores did not indicate significant differences across years.

Results of the individual item data analysis revealed three Skill items with significant increases of mean scores across time, including item 13 (Teaching problem-solving, $F(4,196) = 4.651, p < 0.001$), item 26 (Using co-operative learning strategies, $F(4,196) = 6.353, p < 0.001$), and item 27 (Reflecting on teaching performance, $F(4,196) = 5.401, p < 0.001$). In addition, the means of Skill items 11 (Planning stimulating lessons, $F(4,196) = 4.0684, p = 0.0034$) and 25 (Monitoring students’ progress, $F(4,196) = 4.5657, p = 0.0033$) showed marginal increases across time.

**Comparisons by whether or not respondents had subsequently taught**

Results of analyses according to whether or not respondents had ever taught also revealed differences. Significantly more positive ratings by non-teachers resulted for both the overall average score ($t(199) = 3.25, p = 0.001$) and the Skill sub-scale score ($t(199) = 3.84, p < 0.001$). There were no apparent differences between non-teachers and teachers for the Quality and Subject sub-scales.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Analyses of overall and sub-scale scores by whether respondents had taught</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sub-scale</strong></td>
<td><strong>Mean score (sd)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>have taught</strong></td>
</tr>
<tr>
<td>Average (All items)</td>
<td>5.0773 (0.682)</td>
</tr>
<tr>
<td>Skill</td>
<td>4.9367 (0.841)</td>
</tr>
<tr>
<td>Quality</td>
<td>5.3736 (0.853)</td>
</tr>
<tr>
<td>Subject</td>
<td>5.1927 (0.713)</td>
</tr>
</tbody>
</table>
Factor analysis

Following preliminary analysis of the data within the conceptual framework that had been presumed, we carried out a factor analysis of all thirty-three items. A principal components analysis resulted in nine factors with eigen values greater than one, accounting for 66 per cent of the total variance. These nine factors were then subjected to an oblique rotation, since we did not expect that the factors of interest would be independent. Of the nine factors identified only the first six were clearly interpretable. On the basis of item loadings we labelled these six interpretable factors (1) Professional skills, (2) General professional knowledge, (3) Specialised professional knowledge, (4) Humanities, (5) Psycho-educational foundations, and (6) Mathematics/science.

Numerous items from the Skills sub-scale loaded heavily on factor 1 (Professional skills). Nearly all of the items loading heavily on factor 2 (General professional knowledge) tapped general knowledge having to do with historical, legal, and multicultural issues in education, content typically addressed in social foundation survey courses in our programme. Items having to do with individualised learning, special education and communicating with parents loaded heavily on factor 3, which was labelled Specialised professional knowledge. Since the survey was intended to assess general education outcomes (as well as professional education) we were not surprised to find that content typically delivered in courses our students take in the Division of Liberal Arts and Sciences fell out in a separate factor (factor 4), which we labelled the Humanities factor. Items loading most heavily on factor 5 were those that dealt with psychologically oriented theory (child development, learning theory), content typically addressed in our educational psychology foundation courses. Finally, items 31 and 32, dealing with mathematics and science, fell out in factor 6.

Since the rotation employed was oblique, the analysis also resulted in a set of inter-factor correlations (see Table 2) that appears to support our interpretation of

<table>
<thead>
<tr>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
<th>Factor 5</th>
<th>Factor 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1</td>
<td>100a</td>
<td>43a</td>
<td>42a</td>
<td>28</td>
<td>30</td>
</tr>
<tr>
<td>Factor 2</td>
<td>43a</td>
<td>100a</td>
<td>34</td>
<td>27</td>
<td>16</td>
</tr>
<tr>
<td>Factor 3</td>
<td>42a</td>
<td>34</td>
<td>100a</td>
<td>25</td>
<td>11</td>
</tr>
<tr>
<td>Factor 4</td>
<td>28</td>
<td>27</td>
<td>25</td>
<td>100a</td>
<td>12</td>
</tr>
<tr>
<td>Factor 5</td>
<td>30</td>
<td>16</td>
<td>11</td>
<td>12</td>
<td>100a</td>
</tr>
<tr>
<td>Factor 6</td>
<td>24</td>
<td>25</td>
<td>13</td>
<td>16</td>
<td>4</td>
</tr>
</tbody>
</table>

Notes: Printed values are multiplied by 100 and rounded to the nearest integer.

* Values greater than 0.387986.
the factors. Factor 1 correlates fairly highly with factors 2 and 3 but none of the other factors correlate highly with one another.

Following this factor analysis we reanalysed our data, using the conceptual structure that had emerged. New sub-scale scores for each of the six interpretable factors were calculated for each student. New analyses by year of graduation and whether or not the respondent had taught were carried out. Results of these reanalyses (see Table 3) reinforced our earlier findings but also suggested more subtle effects we had not seen before. Factor 1 (Professional skills) showed significant differences across time ($F(4,196) = 6.59614$, $p < 0.0001$), with more recent pre-service graduates reporting more positive evaluations. The Specialised professional knowledge factor showed a marginal difference ($F(4,196) = 2.53208$, $p = 0.417$) across time. Analysis according to whether respondents had taught revealed significant differences for the Professional skills factor ($F(1,199) = 16.4084$, $p < 0.0001$), the Humanities factor ($F(1,199) = 9.7403$, $p = 0.0021$) and the General professional knowledge factor ($F(1,199) = 4.85$, $p = 0.0288$). The Specialised professional knowledge factor again showed marginal changes across groups ($F(1,199) = 8.5095$, $p = 0.0411$).

Table 3 Results of analyses of variance for factors 1–6 across years of graduation and whether or not the respondent had ever taught

<table>
<thead>
<tr>
<th>Factor</th>
<th>Changes across year of graduation</th>
<th>Teaching/Non-teaching differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Professional skills</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>$F(4,196) = 6.59$, $p &lt; 0.0001^a$</td>
<td>$F(1,199) = 16.41$, $p &lt; 0.0001^b$</td>
</tr>
<tr>
<td>2 General professional knowledge</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>$F(4,196) = 1.35$, $p = 0.2522$</td>
<td>$F(1,199) = 4.85$, $p = 0.0288^b$</td>
</tr>
<tr>
<td>3 Specialised professional knowledge</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>$F(4,196) = 2.53$, $p = 0.0417^a$</td>
<td>$F(1,199) = 4.23$, $p = 0.0411^b$</td>
</tr>
<tr>
<td>4 Humanities</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>$F(4,196) = 1.35$, $p = 0.2506$</td>
<td>$F(1,199) = 9.74$, $p = 0.0021^b$</td>
</tr>
<tr>
<td>5 Psycho-educational foundations</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>$F(4,196) = 1.71$, $p = 0.1475$</td>
<td>$F(1,199) = 0.64$, $p = 0.4254$</td>
</tr>
<tr>
<td>6 Maths/science</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>$F(4,196) = 0.13$, $p = 0.9677$</td>
<td>$F(1,199) = 0.28$, $p = 0.5949$</td>
</tr>
</tbody>
</table>

Notes
a  More recent graduates reporting more positive evaluations.
b  Non-teaching graduates reporting more positive evaluation.
DISCUSSION

Two results emerge from our analysis that are of greatest interest to us. One result is that, consistent with a number of other previous studies (Gaede, 1978; Ligana, 1970; Hummel and Strom, 1987), more recent pre-service graduates appear to have a more favourable evaluation of their educational experience than do graduates with more teaching experience. A careful analysis of the data suggests, however, that most of the variation across years can be attributed to changes in the Skill sub-scale. Moreover, item-by-item analysis of the complete set of sixteen Skill items suggests that much of the variation across time can be attributed to a relatively small number of items on the Skill sub-scale. Specific items for which significant differences were found include:

- Teaching problem-solving and higher-order thinking skills (item 13),
- Reflecting and improving your teaching performance (item 27), and
- Using co-operative learning techniques (item 26).

Two additional items showed marginally increased means across time:

- Monitoring students’ progress and adjusting instruction (item 25).
- Planning stimulating lessons (item 11).

Several explanations of the results are possible. One is that the more positive response of more recent graduates is simply a halo effect that tarnishes with time, an effect Ligana (1970) has termed the ‘curve of disenchantment’. It may be that pre-service graduates leave the programme feeling quite positive about their educational experience but that, with exposure to the difficulties of teaching in real classrooms, they begin to feel less positively. Arguing against this interpretation, however, is the fact that only three of thirty-three items in the survey showed significant changes across time and only two more showed marginal changes. In addition, all of those changes occurred in the Skills sub-scale. Presumably, if changes across time were due to more general disenchantment we would expect to see more consistent changes across our items across all of our sub-scales rather than only a few in a single sub-scale.

The suggestion that survey responses actually reflect programme characteristics is further supported by changes instituted in our programme over the years investigated in the study. One of the programme changes that occurred at about the mid-point of our survey sample was increased emphasis on co-operative learning techniques in our courses. An in-service graduate credit workshop specially devoted to this topic was started, and undergraduate pre-service students began to receive more instruction in this area in their general and specific methods courses. Inspection of year-by-year means for this item reveals a substantial jump at about the mid-point of the survey span. Another important influence on our programme during the years studied was our effort to clarify the roles of decision-making and reflective evaluation as components of our teacher education model. The increased emphasis on these areas was also reflected in the responses of our graduates. It
therefore appears that undergraduate pre-service programme changes are related to our survey results in ways we would expect, thus strengthening the argument that this kind of data can provide insights into programme characteristics and the effects of changes.

Our second major finding is that, although our initial categorisation of Skill, Subject and Quality items helped us organise our assessment effort, the data we collected suggest a more informative conceptual framework composed of at least six factors that seem to correspond to curricular blocks of courses. This was especially important for us, since it was programme assessment that drove the study from the beginning. The factor structure that resulted thus appears well suited to our interest in identifying strengths and weaknesses in our undergraduate teacher education pre-service programme.

Although survey responses seem to be related to programme changes implemented over the years of the study it is almost certain that other influences also have been at work. Our instructional programmes function within, and are influenced by, broader social, cultural and professional ideas and events that are also felt in the classrooms where our pre-service graduates teach. In addition, surveys like the one employed in this study are subject to ‘noise’ as a consequence of both the self-report format (Linn and Gronlund, 1995, p. 284) and the nature of judgements about the past. These limitations notwithstanding, it is quite clear that information of this sort is widely viewed as useful, since both state departments of education and national accrediting bodies recommend or require this kind of data to be collected in the course of evaluating instructional programmes.

It is also important to remember that the theoretical factor structure identified on the basis of the data collected in this study is specific to the instrument employed. As in any other factor analytic study, modification of the original variables can result in different factor structures. The ratio of cases (201) to variables (thirty-three) in this study, however, substantially exceeds the generally recommended minimum of 4:1 (Rummel, 1970; Cattell, 1952) and thus it is reasonable to suppose that the factor structure will generalise in a fairly robust manner in the absence of significant changes to the instrument. Obviously, as we continue to collect data using this survey form we intend further to test and refine the theoretical structure that has emerged in this first round of data collection.

SUMMARY AND CONCLUSIONS

This article describes the development, implementation and results of a survey-based component to support programme assessment in pre-service teacher education. Primary data in the research consist of a telephone interview survey of pre-service teacher education programme graduates to assess their perceptions of the adequacy of their training both in general education and in their professional preparation.
programmes. Data analysis involved comparisons of our results with those reported in a national survey, analyses to reveal changes across time and demographic dimensions, and a factor analysis of our survey items in an attempt to refine the conceptual framework of our instrument. Our results are similar to those of previous studies indicating more positive evaluations from more recent pre-service graduates and from those pre-service graduates who have never taught. In addition, although our intuitive assessment of the items used suggested three item categories, the data we collected suggest a more informative conceptual framework composed of at least six factors that seem to correspond to curricular blocks of courses. The factor structure that resulted thus appears to be well suited to identifying relative strengths and weaknesses in undergraduate pre-service teacher education programmes.

What is perhaps most encouraging about our survey results at the conclusion of this first cycle of data collection is the finding that documented programme changes appear to be reflected in the responses of our pre-service graduates. Specific programme modifications that occurred during the years of our survey span appear to have resulted in more positive responses in those areas, thus supporting the intended use of this kind of survey data. It has also been encouraging to find that the factor analysis of our data suggests that students perceive their relative strengths and weaknesses in terms of meaningful curricular areas, since this makes the interpretation of survey results and the application of those results to programme improvement much more straightforward.

Our intention is to continue to look for relationships between survey results and specific programme characteristics or changes that we can document from other sources. We also intend to rank all of the scaled items in the survey in order to identify those on which respondents were most positive and most negative. We expect to continue our programme assessment on the basis of these data by asking why our pre-service graduates rated the items the way they did. Can we find any explanations for their responses in what we know about our programme? In other words, is there any other evidence to support these areas as relative strengths and weaknesses? Finally, as has been suggested, it is our intent to make this kind of survey a regular part of an on-going assessment programme which will allow us to understand better what, if any, consideration must be given to the stability of response factor noted by Adams (1987) and Ligana’s (1970) ‘curve of disenchantment’ described above, both of which are related to recency of graduation and thus could have implications for interpreting our results.

REFERENCES


A survey-based component for programme assessment


Address for correspondence
Dr John E. McEnaney, Indiana University South Bend, Division of Education, Greenlawn Hall, 1700 Mishawaka Avenue, P.O. box 7111, South Bend, IN 46634–7111, USA.