Student Self-Evaluation of
Teaching Abilities at Graduation from GC&SU
Teacher Education in the Initial Preparation Program
Validity Study April 2004

Mentor Leaders Participating:

Dr. Betty Block, Health & Physical Education
Dr. Paige Campbell, Early Childhood Education
Dr. Kevin Crabb, MAT Secondary Education
Dr. Heide Hlawaty, MAT Secondary Education at Macon Campus
Dr. Rosemary Jackson, Special Education Interrelated Disabilities
Dr. Patricia Klein, Middle Grades Education
Dr. Brian Mumma, Middle Grades Education at Macon Campus
Dr. Patricia Tolbert, Music Education

Data Analysis and Report by:
Dr. Sharene Smoot, Foundations & Secondary Education

Date of Report: August 23, 2004
Summary

- The participants were students in eight cohorts in the initial teacher preparation program.
- The 14 mentor leaders served as advisors for their group of student for both their junior and senior years in most cases. In the case of the two graduate level cohorts, the program was only for one year in secondary education. The mentor leaders also served as instructors for various (but not all) of the courses taken by their assigned students.
- The questionnaire had 13 questions related to the stated goals of the school of education in their conceptual framework as well as an overall readiness to begin teaching question. Both groups completed identical versions by circling the Likert scale elements. The forms were matched by student identifying numbers but were otherwise confidential.
- The data analysis used the Pearson $r$ to correlate the sets of answers. The results were as expected in that for the most part; the students rated themselves higher than their mentor leader did. Most of the students (67%) rated themselves as a 9 or a 10 on their overall readiness to begin teaching and on most of the different teaching skills. (This was from all 135 initial teaching candidates surveyed at exit from the program, not just the students in the cohort programs.)
- The mentor leaders also used a lot of 9’s or 10’s (58%) to rate their students on their overall readiness as beginning teachers. The difference between these two sets of ratings was statistically significant using a paired $t$-test at the $p < .001$). The proportion of their ratings of excellent ratings was much higher for the students than for the mentor leaders, which we expected to find.
- Then the percent of ratings of excellent were calculated for both the students and the mentor leaders for all of the 14 skills listed, actually 13 skills and one overall teaching readiness rating. When these numbers were correlated as a set, the value of the correlation was much higher ($r = .65$ n = 14, $p = .006$, one-tailed). This means that the agreement between students and mentor leaders of which of their teaching skills was considered excellent is very good.
- The data from this study supports the assertion that student self-ratings at exit from the program are valid in regard to comparison with those of their mentor leader (program advisor), although students tend to rate themselves more highly than their professors did.
- Both students and professors agreed on which specific aspects of the conceptual framework resulted in outcomes of perceived strengths and weaknesses in readiness to begin teaching.
- In regards to more accurate self-assessment of specific abilities, the mentor leaders could stress more self-evaluation or collect more data in regards to the lowest correlated skills or abilities as stated in the Conceptual Framework.
- A copy of the mentor leader version of the assessment form is at the end of this report. It shows the new scaling from 1 to 10, anchored by Poor and Excellent.
Student Self-Evaluation of Teaching Abilities 
at Graduation from GC&SU Teacher Education in the Initial Preparation Program

The purpose of this study was to investigate whether the self-ratings that the exiting seniors gave themselves on the 13 teaching skills/abilities emphasized in our conceptual framework would be comparable to the ratings given by them to their mentor leaders. It was expected that the seniors would probably rate themselves somewhat higher than their major professors would, but that there would be a positive correlation of these ratings.

Methods

Participants

There were a total of 135 students and seven professors participating in the 2004 study. There were 5 cohorts from the main campus at Milledgeville and 2 cohorts from the satellite campus at Macon State College. The Milledgeville cohorts were Early Childhood (n = 45), Health and Physical Education (n = 7), Middle Grades (n = 12), Special Education (n = 23), and Secondary Education (n = 36). The Macon cohorts were Middle Grades (n = 9) and Secondary Education (n = 13). The mentor leaders were all professors who served as advisors and course instructors in the major field. The cohorts were all preservice teachers working in their initial teacher preparation program.

Instrumentation

One-page questionnaires were used. See Appendix. The professors were instructed to rate each student in their group (cohort) at the end of his/her student teaching. “How would you rate his/her present abilities as a beginning teacher?” The students were asked “How would you rate your skills as a beginning teacher right now?” The 13 teaching skills were described and the rating scale was from one to ten, anchored by poor (1) and excellent (10). There was an additional question (which was actually question one) in which the person was asked to rate their ability to “begin teaching overall”. The last five digits of the students’ social security numbers were used to match the questionnaires for data analysis.

Procedures

The students were given the one page questionnaire during a session in which they all met with the SOE certification officer to complete their paperwork for applying for teacher certification during their last month at GC&SU. The mentor leaders were given a comparable version of the form at about the same time and asked to rate each of their graduating seniors using the same rating scale. The instructions on the forms stated that the participants were to base their ratings or self-ratings on their skills/abilities as a beginning teacher.

Data Analysis

This was a descriptive study and Pearson correlations were used. Data were analyzed as a whole and then cohort-by-cohort for the different programs.

Results

Most of the preservice teachers (67%) rated themselves with a nine or a ten on their overall readiness to begin teaching and on most of the different teaching skills. Their average was 9.0 ($SD = .91$). The mentor leaders’ ratings were lower; only 58% rated their students with a nine or a ten as on their overall readiness as beginning teachers. The range was from four to ten for the mentor leaders and from seven to ten for the preservice teachers. The average of the mentor leaders ratings was 8.6 ($SD = 1.4$). The difference between these two sets of ratings was statistically significant using a paired t-test [$t (125) = 3.64$, $p < .001$].
This is where the preservice teachers and their mentor leaders agreed most on those aspects of the conceptual framework that are easily observable and evaluated. Highest correlated were the ratings for planning and teaching lessons \((r = .42)\), asking questions \((r = .39)\), and managing student behavior \((r = .36)\) These ratings were statistically significant at the \(p < .001\) levels.

The actual value of the correlation coefficients were low to moderate for the whole set of participants even though all but the bottom three aspects on the chart above were statistically significant. However the correlations were twice as high this year, because we switched from the four point rating scale (Poor, OK, Good, Excellent) to a ten point scale anchored by Poor and Excellent.

Again the skills that correlated the highest are skills that are relatively easy to observe in performance or written products. The lowest correlated areas might be targets for improved assessment activities. However there is considerable variation by cohort when the results are disaggregated. Some mentor leaders have much higher correlations than others.
Then the percent of excellent (ratings of 9 or 10) were calculated for both the set of students’ self-ratings and the mentor leaders’ sets of evaluations for all of the 14 skills listed, actually 13 skills and one overall teaching readiness rating. When these numbers were correlated as a set, the value of this correlation was much higher \((r = .65 \ n = 14, \ p = .006, \ \text{one-tailed})\). This means that the agreement between students and mentor leaders of which of their teaching skills was considered excellent is very good. For example, the mentor leaders rated 81% of their students with a 9 or 10 in their knowledge of educational ethics and laws. However, 87% of the students rated themselves with a 9 or 10 for this same item. In this case, there is only a 6% discrepancy and the ratings match rather well.

**Table 1 –Percent of Ratings of Excellent (9 or 10) Item by Item**

<table>
<thead>
<tr>
<th>#</th>
<th>Question</th>
<th>Mentor Leaders</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Knowledge of Ethics &amp; Laws</td>
<td>81</td>
<td>87</td>
</tr>
<tr>
<td>10</td>
<td>Listen &amp; Respond</td>
<td>66</td>
<td>73</td>
</tr>
<tr>
<td>1</td>
<td>Begin Teaching Overall</td>
<td>58</td>
<td>67</td>
</tr>
<tr>
<td>14</td>
<td>Leadership Role</td>
<td>57</td>
<td>63</td>
</tr>
<tr>
<td>9</td>
<td>Ask Questions</td>
<td>55</td>
<td>55</td>
</tr>
<tr>
<td>8</td>
<td>Teach Diverse Students</td>
<td>54</td>
<td>62</td>
</tr>
<tr>
<td>11</td>
<td>Self-Evaluate</td>
<td>52</td>
<td>61</td>
</tr>
<tr>
<td>13</td>
<td>Evaluate Materials</td>
<td>49</td>
<td>71</td>
</tr>
<tr>
<td>5</td>
<td>Plan Lessons Day to Day</td>
<td>49</td>
<td>79</td>
</tr>
<tr>
<td>4</td>
<td>Manage Behavior</td>
<td>49</td>
<td>53</td>
</tr>
<tr>
<td>3</td>
<td>Use of Technology</td>
<td>45</td>
<td>49</td>
</tr>
<tr>
<td>7</td>
<td>Assess Learning</td>
<td>45</td>
<td>56</td>
</tr>
<tr>
<td>6</td>
<td>Design Units of Study</td>
<td>42</td>
<td>70</td>
</tr>
<tr>
<td>2</td>
<td>Knowledge of Content Areas</td>
<td>42</td>
<td>51</td>
</tr>
</tbody>
</table>

Scatter plots are included in this report, both for the dataset as a whole and individually by cohorts that illustrate the aspects of teaching where the agreements and disagreements are by quadrants. As expected the mentor leaders’ ratings are lower than the students’ self-ratings in most aspects, including their overall readiness to begin teacher.

Then the difference between the mentor leaders’ and students’ ratings of excellent in percents was calculated. A set of bar charts accompanies each scatterplot and shows the discrepancies.

**Discussion and Conclusions**

This study validated (for the most part) the self-ratings of the teacher education candidates and the mentor leaders. The JHL-SOE produces mostly excellent or good beginning teachers. This was also the finding in the entire set of employer surveys conducted. Our program has a unique conceptual framework for teacher education with a large field-based component and many other additional requirements designed to enhance the ability of these students to work cooperatively with others and develop their liberal arts skills; especially skills in communication, problem solving, leadership, and self-evaluation.
Figure 4 - Differences in the Ratings of Excellent (9 or 10)

**Students’ Self-Ratings are Higher than Mentor Leaders’ Ratings on almost Everything**

The biggest discrepancy was that 79% of the students rated themselves excellent (a 9 or 10) in planning lessons, but actually only 49% were rated excellent by their mentor leaders. However when the ratings of excellent for the students and the mentor leaders are correlated for each aspect listed above, the correlation between these ratings is high ($r = .65$, $n = 14$, $p = .006$ one-tailed).

See scatterplot on the next page showing the close alignment of mentor leaders’ ratings to preservice teachers’ self-ratings on most aspects of the conceptual framework.
In this large group of 135 students when the percent of ratings of excellent (9 or 10) for the students and the mentor leader are correlated for the 14 items rated, the correlation between these ratings is high ($r = .65, n = 14, p = .006$ one-tailed). See Scatterplot below which illustrates the agreement between these two sets of ratings on the 14 items illustrating the adherence to the conceptual framework.

Points closest to the regression line are the most highly in agreement with each other. There is a close alignment of mentor leader ratings to preservice teacher ratings on most aspects of the conceptual framework. This is evidence of the close adherence of the SOE instructors to the conceptual framework in the initial preparation of teachers.

The most discrepancies are in the upper left quadrant. The upper left quadrant has aspects where the students’ self-ratings were much higher than the mentor leaders’ ratings. For example, the ability to plan lessons day to day was rated lower by the mentor leaders than by the preservice teachers.
Early Childhood Cohort from Milledgeville 2004 (n = 46)

In this large group when the percent of ratings of excellent (9 or 10) for the students and the mentor leader are correlated for the 14 items rated, the correlation between these ratings is not significantly different from chance \( (r = .26, n = 14, p = .18 \text{ one-tailed}) \). See Scatterplot below.

Mentor Leaders' Ratings

The more the distance from the regression line, the more the discrepancy, see chart below. However, item by item Pearson correlations reveal areas where the self-ratings and mentor ratings agree fairly well such as leadership ability and use of technology.
Middle Grades Cohort from Macon 2004 (n = 9)

When the percent of ratings of excellent (9 or 10) for the students and the mentor leader are correlated, the correlation between these ratings is very high ($r = .75$, $n = 14$, $p = .002$). This high correlation was achieved by a different type of administration of the questionnaire. See bottom of page

These students rated themselves much higher in all aspects of the conceptual framework than their mentor leader did. See discrepancy chart below. However, the correlations are quite high.

These students and Dr. Mumma spent several classes over a week doing reflective activities to evaluate their progress towards their goal of becoming a teacher just before graduation. Then Dr. Mumma had them self-rate in his class and he independently rated them at this time also. See reliability study 2004.
Middle Grades Cohort from Milledgeville 2004 (n = 12)

However when the percent of ratings of excellent (9 or 10) for the students and the mentor leader are correlated, the correlation between these ratings is high and statistically significant (r = .65, n = 14, p = .03, one-tailed). See Scatterplot below.

The discrepancies farthest from the regression line, also see below for discrepancy chart. The Pearson correlations between the mentor leader and students are fairly high.
However when the percent of ratings of excellent for the students and the mentor leader are correlated, the correlation between these ratings is not significantly different from chance ($r = .26, n = 14, p = .37$). See Scatterplot below. Pearson correlations item by item were no better.

These students and their mentor leader perceived their abilities somewhat differently, also see below for discrepancy chart. The mentor leader was much lower than the cohort in rating these aspects. The actual item by item correlations were all nonsignificant (see below).
However when the percent of ratings of excellent (9 or 10) for the students and the mentor leader are correlated, the correlation between these ratings is too low to be statistically significant ($r = .44$, $n = 14$, $p = .11$, one-tailed). See Scatterplot below.

![Scatterplot showing correlation between student and mentor leader ratings]

This group of students rated themselves much higher than their GC&SU Mentor Leader on their overall readiness to begin teaching. See discrepancy chart below. The item by item correlations are mostly very high, the mentor leader rated them even higher than they did themselves on many aspects.

![Discrepancy chart showing differences in ratings and Pearson correlations]
**Special Education Cohort 2004 (n = 23)**

When the percent of ratings of excellent (9 or 10) for the students and the mentor leader are correlated, the correlation between these ratings is high and statistically significant. \((r = .66, n = 14, p = .01)\). See Scatterplot below.

![Scatterplot](image-url)

**Mentor Leader's Ratings**

There is only one major discrepancy and the item by item correlations are quite high; see charts below.

![Pearson Correlations Chart](image-url)

This mentor leader and students have high ratings indicating that the objectives in the conceptual framework have been met in this initial preparation cohort.
Health & Physical Education Cohort 2004 (n = 7)

When the percent of ratings of excellent for the students and the mentor leader are correlated, the correlation is not meaningful or statistically significant (r = .23, n = 14, p = .43). See scatterplot below.

This mentor leader rated all students as excellent in many aspects, the students were not so sure of their abilities. They disagreed most on their ability to teach diverse students. See charts below, the item by item correlations were extremely varied in range. This mentor leader uses a different conceptual framework in the School of Health Sciences. It is the Department of Health Sciences.
John H. Lounsbury School of Education  
Teacher Preparation Program Evaluation Validity Study  
Spring 2004

Please rate each student in your cohort at the end of her/his student teaching.

Student SSN__________________

How would you rate his/her present abilities as a beginning teacher?  
(Please circle your answer)

(Please circle your answer on a scale of one to ten where 1 is Poor and 10 is Excellent)

<table>
<thead>
<tr>
<th>Poor</th>
<th>Excellent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
</tbody>
</table>

Thank you, these ratings will be used for a variety of program assessment uses. These fourteen aspects of teaching are from our conceptual framework. Your ratings will be compared to the students’ self-ratings, to their Praxis II scores, to their portfolio scores, etc. This is a part of the NCATE requirement in program assessment.