

## Section 3 Task Development

### *Pilot Testing*

The first phase of pilot testing conducted in June of 2003, consisted of sending the survey to a sample of eight teachers from the University of Hartford Magnet School. These teachers were asked to complete the survey as if they were taking it, and to provide the required justification for their responses. Six of the teachers provided responses, and the numerical responses are provided in the tables below.

### *Professionalism*

Item Number	Teacher Number						% Agree	Mean	SD
	1	2	3	4	5	6			
1	1	1	1	2	1	1	100%	1.17	0.37
2	1	1	2	1	1	2	100%	1.33	0.47
3	5	1	5	5	5	6	83%	4.50	1.61
4	3	3	2	2	2	4	83%	2.67	0.75
5	5	6	5	5	5	4	100%	5.00	0.58
6	6	1	5	5	4	6	83%	4.50	1.71
7	2	1	2	4	2	4	67%	2.50	1.12
8	1	1	1	2	2	1	100%	1.33	0.47
9	5	1	6	4	2	4	67%	3.67	1.70
10	5	1	2	2	2	6	67%	3.00	1.83
11	4	6	4	4	4	5	100%	4.50	0.76
12	5	3	5	4	5	2	67%	4.00	1.15
13	2	1	5	2	4	2	67%	2.67	1.37
14	5	3	2		5	3	60%	3.60	1.20
15	1	1	5	4	5	6	67%	3.67	1.97
16	5	1	2	5	5	6	67%	4.00	1.83
17		1	5		2	2	75%	2.50	1.50
18	1	1	5	2	2	1	83%	2.00	1.41
19	1	1	2	2	2	1	100%	1.50	0.50
20	5	3	5	5	2	1	50%	3.50	1.61

The percent agreement shown in the table above reflects the percent of teachers who indicated that the item was in the disagree range (i.e., 1 to 3) or agree range (i.e., 4 to 6). As the data show, there was considerable agreement by the teachers on each item, with most of the teachers' ratings in either the disagree range or agree range. In addition, most of the standard deviations were below 1.50. When there was substantial divergence in the scores, it was typically due to one teacher's scores (i.e., Teacher #2, whose mean rating was 1.9, whereas the remainder of the teachers' ratings ranged from 3.1 to 3.55).

Furthermore, the means of the teacher ratings also provide evidence for which items should be reverse-scored. Specifically, items with a 3.0 or below (i.e., from slightly to strongly disagree on the 6-point scale), would be considered negative responses, and as such, one would expect that these items should be reverse-scored. According to the teachers' ratings, items 1, 2, 4, 7, 8, 10, 13, 17, 18, 19, and 20 should be reverse-scored. These correspond to a great extent with the expected reverse-scoring scheme detailed in the rubric presented in Section II. Items with means between 3.0 and 4.0 could be considered ambiguous in terms of their positivity/negativity, and include items 9, 14, 15, and 20. These items should be further examined for their content and wording, and perhaps revised or omitted prior to further testing. The teachers' written rationales for their responses may be used to clarify interpretations and to refine these items.

### *Reflective Practice*

Item Number	Teacher Number						% Agree	Mean	SD
	1	2	3	4	5	6			
1	1	1	3	3	2	1	100%	1.83	0.90
2	1	3	2	2	2	1	100%	1.83	0.69
3	6	6	6	6	5	6	100%	5.83	0.37
4	4	4	6	5	5	6	100%	5.00	0.82
5	2	1	2	2	2	1	100%	1.67	0.47
6	1	1	6	5	4	1	50%	3.00	2.08
7	6	1	6	6	5	6	83%	5.00	1.83
8	1	1	1	1	1	2	100%	1.17	0.37
9	6	6	6	5	5	6	100%	5.67	0.47
10		1	6	4	5	6	80%	4.40	1.85
11		4	6	6	3	6	80%	5.00	1.26
12		6	6	5	5	5	100%	5.40	0.49
13	3	5	2	3	4	4	83%	3.50	0.96
14	6	6	6	2	6	6	83%	5.33	1.49
15	4	1	1	1	1	1	83%	1.50	1.12
16	3	4	1	2	1	5	66%	2.67	1.49

Again, there was considerable agreement by the teachers on each item, and most of the standard deviations were below 1.50. Again, most of the teachers' ratings were in either the disagree range or the agree range, and often Teacher #2 was inconsistent in his or her ratings from the remainder of the teachers.

As was the case with the items for Professionalism, the means of the teacher ratings also provide evidence for which items should be reverse-scored. According to the teachers' ratings, items 1, 2, 5, 6, 8, 15, and 16 should be reverse-scored. These correspond to a great extent with the expected reverse-scoring scheme detailed in the rubric presented in Section II. Item 13 has a mean between 3.0 and 4.0 and could be considered ambiguous.

Thus, for both the Professionalism and Reflective Practice Dispositions, there appears to be substantial agreement by teachers in terms of the positivity/negativity of the items, and

this information is largely consistent with the a priori reverse-scoring detailed in the rubric in Section II.

The items which have means between 3.0 and 4.0, or items with standard deviations greater than 1.0 should be examined further because of their potential ambiguity, and may be further refined by examining their content and wording. As mentioned earlier, the teachers' written rationales for their responses may be helpful in clarifying problems with interpretation which may be resolved.

In addition to these ratings, the teachers were also asked to provide responses to two overall questions: (1) Do the vignettes and responses measure the dispositions of Professionalism and Reflective Practice, and (2) Do the responses target the dispositions we would like our candidates to exhibit? These two questions were intended to provide additional construct validity evidence as well as feedback for improving the survey.

One teacher indicated that the vignettes do measure these dispositions, although another teacher suggested that more questions about communication techniques and partnering should be included. Another teacher considered the vignettes to be very realistic, although another felt that the teachers in the vignettes seemed dated and isolated, rather than part of a team working on a shared curriculum.

One teacher expressed concern that some of the responses the teachers in the vignettes gave were inappropriate, and she would like to see more best-practices responses. However, given that some of the items were written to be reverse-scored, it is necessary to include some inappropriate responses. Another teacher commented that he felt that the vignettes seemed geared more toward upper level classes, such as middle or high school, rather than elementary school.

These comments will be reviewed and the surveys revised to the extent possible to address these comments.

### ***Planned Pilot Testing***

Due to time constraints at the end of the academic school year, it was not possible to conduct a full-scale pilot test involving undergraduate students enrolled in the Education Division. However, a pilot test of the survey is planned for the fall of 2003, using undergraduates who will be in their sophomore year and preparing to enter the professional teacher education program. The survey will be refined prior to that time, based on the feedback obtained from the teachers at the University of Hartford Magnet School in June of 2003.

Approximately 200 students will be asked to participate, providing approximately a 10:1 subject-to-item ratio for each vignette. This will allow the team to conduct factor analyses, in order to evaluate the extent to which the items do measure the dispositional aspects which have been hypothesized in the current scoring rubric (see Section II).

These analyses, when coupled with the hypothesized linkages between items and disposition aspects, will allow for the investigation of construct validity of the survey.

In addition, coefficient alpha will be calculated on the items on each vignette to examine the extent to which the items form a coherent scale, and to look for items which may need to be omitted or reverse-scored.

Once the survey has been piloted on a sample of students and the survey revised accordingly, the survey will be implemented on an online computer system and numeric responses will be automatically scored. Scores will be calculated by summing the responses to the items, subsequent to reverse-scoring certain items. Students will be given feedback about their relative status in one of three categories (as described in Section II). These categories will be developed based on normative data gathered during the planned pilot study.

The computerized scoring will enable fast, accurate feedback to candidates. The rationales will not be scored by computer, but instead will become part of the candidate's feedback report, and the candidate's advisor will use this to provide specific feedback about improving dispositions.