

A Technical Summary of Responses to the
GSTEP Self-Assessment for Accomplished Teaching Instrument

Stephen E. Cramer

Hye Jeong Choi

Vickie B. Bishop

Georgia Center for Assessment

College of Education

University of Georgia

With

Linda S. Gilbert

Georgia Systemic Teacher Education Program (GSTEP)

(Executive Summary and Interpretation Sections)

July 7, 2006

Executive Summary

The purpose of this paper is to provide a technical summary of the data collected in the online trial of the GSTEP Self-Assessment for Accomplished Teaching Instrument. This instrument is based on the GSTEP Framework for Accomplished Teaching (now the Georgia Framework for Accomplished Teaching). It is a self-assessment for teachers that uses a rubric-based scale.

The primary goal of the study was to examine the validity of the instrument, particularly as it relates to showing differences in teacher development. The study was conducted by Test Scoring and Reporting Services (TSARS). An online version of the instrument was created, including demographic information, and hosted on the GSTEP server.

Of the 353 completed self-assessments collected, the sample was split between practicing and pre-service teachers. There was variation across participants: pre-service teachers do rate themselves lower than experienced teachers. However, all rate themselves very highly using this scale. These findings suggest that the instrument does have content validity, but suggests additional attention to the scale/rubric and the anchors provided to various ratings.

In terms of the instrument content, the items composing each of the six elements had good correlations (Cronbach's alpha coefficients ranged from 0.80 to 0.89). However, the three factor analyses of items (one for all teachers, one for pre-service, one for experienced) showed varying patterns. The factor analyses suggest that experienced teachers have a more complex view of

teaching than beginners. A logical next step would be to revisit the organization of the constructs (elements) using the factor analysis to inform the organization further.

Additional participant comments indicate that participants generally found the instrument easy to use, though there were some concerns about wording and length, and that participants were concerned about potential use beyond self-assessment.

In terms of additional study, collecting data to look at criterion-related validity is one option. Another particularly interesting avenue that could be explored further involves examining the actual use of the instrument as a self-assessment and the consequences of use, possibly using Messick's Unified Validity Framework (see Appendix).

A Technical Summary of Responses to the
GSTEP Self-Assessment for Accomplished Teaching Instrument

The purpose of this paper is to provide a technical summary of the data collected in the online trial of the GSTEP Self-Assessment for Accomplished Teaching Instrument. This instrument was initially developed by members of the GSTEP consortium at Valdosta State University and further refined through extensive collaboration with researchers at other GSTEP institutions, most notably the University of Georgia (UGA).

In order to obtain a sample of respondents, the instrument was developed into an online survey hosted by the Georgia Center for Assessment at UGA. Teachers and teacher educators were invited to participate in the survey, and to recommend or require that their students also complete the instrument, either for their own professional development or as a requirement for class or practice teaching.

In order to comply with the UGA Institutional Review Board's very stringent privacy guidelines, prospective respondents were offered multiple opportunities to opt out of the process. This analysis represents results from only those who persevered through the entire survey and ranked themselves on every indicator.

The following analyses are based on 353 completed self-assessments collected between March 31, 2005 and December 1, 2005. However, not all of the demographic information was

completed by all respondents, so frequencies noted below do not add to 353.

The characteristics of the sample provided a reasonably good fit with the general in-service and pre-service teacher population. The sample was overwhelmingly female, with large numbers coming from Early Childhood, Middle School, and English Education programs. More than half of the respondents listed a UGA affiliation. The sample was evenly split between practicing and pre-service teachers, with a small (N=12) group of National Board Certified teachers. Overall years of experience in teaching was 6.5, with National Board teachers reporting 17.7 years and practicing teachers reporting 11.2. Since 39% of the sample reported less than one year of experience, the distribution of experience is severely skewed. Table 1 provides additional details.

The goal of the analysis was to answer several technical questions about the data produced by the instrument. As proposed earlier, these questions include:

1. Does the instrument generate variance across candidates (i.e., do different people look different)?
2. Do individual Indicators have appropriate item characteristics, such as frequency distribution of responses and Indicator-Category correlation?
3. Do the categories represent empirical as well as conceptual dimensions, when explored through a factor analytic approach?

The balance of the paper explores answers to these questions.

Question 1: Does the instrument generate variance across candidates (i.e., do different people

look different)?

To answer this question, we calculated the mean, standard deviation, and frequency distribution of each of the 41 indicators. A summary can be seen in Tables 2 and 3.

Not all Indicators received 353 ratings, since one of the response options was “Not Able to Rate.” Mean ratings ranged from a high of 4.37 on Indicator VI-C, Codes of professional conduct, to a low of 3.42 on Indicator IV-B, Use of pre-assessment data. Standard deviations ranged from .72 to 1.08. All Indicators except for 2, 6, 7, 25, 32, 37, 38, and 39 had values from 1 to 5; the smallest value for these items was 2. The distribution of responses was uniform across Indicators, being generally negatively skewed, i.e., very small numbers of ratings of 1 and 2.

The modal rating was 4 for all items, and the median rating was 4 or 5, indicating a high level of accomplishment. Considering that about half the sample was pre-service, however, this value seems inflated. To check on this, we ran a set of frequency distributions of Indicator score by Respondent Type, including National Board teachers with all practicing teachers, since there were so few of the former. The results showed clear response differences between pre-service and practicing teachers that were significant at an alpha of .05 for all Indicators except “III-A. Learning Community.” Pre-service teachers were more likely to rate themselves lower than practicing teachers, but their median response was still a 4 on the 1 to 5 scale (See Table 4).

Seven Indicators--18, 21, 25, 26, 34, 40, and 41--showed NATR rates of greater than 10%. In all

cases, the great majority of NATR ratings came from pre-service teachers, indicating that these items assess an aspect of teaching that they have not yet experienced, and not a flaw in the items themselves.

We also looked for instances of pattern responding, defined as all Indicators being rated the same. This occurred 15 times: 1 all 3's, 5 all 4's and 9 all 5's. This does not seem an excessive number of incidents, although it is hard to credit anyone being accomplished at level 5 on all of the Indicators.

In general, then the instrument appears to generate acceptable levels of variance on all items, although the distribution of responses is quite skewed, and the self-assessments seem quite high for both practicing and pre-service teachers.

Question 2: Do individual Indicators have appropriate item characteristics, such as Indicator-Category correlation?

Each of the Indicators is nested within a category. Table 5 lists the Indicators by category, with brief descriptors. A testable hypothesis is that the items within a category should correlate with a subscale constructed as the sum of those responses. Accordingly, we calculated a score for Content and Curriculum, Knowledge of Students and their Learning, Learning Environments, Assessment, Planning and Instruction, and Professionalism and correlated this score with each indicator. In all cases, the item correlated significantly with its category. We also correlated the

categories with each other, and found correlations ranging from .44 to .69, indicating that while the categories are related, as one would expect, they appear to be assessing somewhat different aspects of accomplished teaching.

The initial correlations of the categories with Years of Experience were low, due to the inclusion of the pre-service teachers, which skewed the distribution severely. We removed all pre-service teachers and all respondents with fewer than one year's experience. The correlations improved, although, while significant, are still not strong. Years of Experience correlates best with the Content and Curriculum category. See Table 6.

We further calculated Cronbach's alpha statistic for each category (See Table 7). Alpha coefficients ranged from 0.80 to 0.89, indicating that the category scores have good reliability in the context of self-assessment feedback.

The category scores are further validated by comparing the responses of pre-service and practicing teachers. A test of the hypothesis that practicing teachers are, as a group, more accomplished than pre-service teachers shows that in all cases, practicing teachers have significantly ($p < .05$) higher category scores than pre-service teachers. See Table 8.

Question 3: Do the categories represent empirical as well as conceptual dimensions, when explored through a factor analytic approach?

To explore this question, responses from all respondents to the 41 Indicators were factor analyzed using principal axis factoring. The solution yielded six factors, based on the eigenvalue-greater-than-one principle, and was rotated using the Oblimin method with Kaiser Normalization to maximize the differences among factors. These six factors account for 60.1% of the total variance.

The factor solution yields 5 very clear factors, which might be labeled Planning & assessment, Knowledge and learning, Professionalism, Concerns beyond the classroom, and Knowledge and skills. One additional factor does not seem to lend itself to easy description. See Table 9.

The fact that the empirical factor structure and the *a priori* categories do not match completely does not necessarily indicate a need to reorganize the assessment. However, this is an indication that developers should look carefully at the organization of the instrument to ensure that the categories possess a high level of conceptual integration.

In an effort to explore the underlying structure of the assessment more fully, we performed separate factor analyses for more experienced and less experienced (i.e., pre-service) teachers (See Tables 10 and 11). The identifiable factors derived from this analysis showed some significant differences. Less experienced teachers' responses grouped into only 3 identifiable areas: Understandings, Professionalism, and Classroom activities. More experienced teachers' responses grouped into 5 identifiable areas: Assessment and learning environment, Professionalism, Planning, Content, Students and learning, Managing and motivating. This

suggests that the instrument is capturing an expected outcome: that experienced teachers have more complex and better-defined conception of teaching than those who are just beginning in the profession.

Comments

In addition to the actual self-assessment responses, we also examined comments made by respondents after completing the instrument. Of the 353 total respondents, 212 (60%) respondents made comments about the survey.

Overall, the respondents were positive about the instrument and its usefulness. The majority found the instrument easy to complete (n= 151, 71%). Five respondents commented that it was user friendly, and nine respondents used it for self reflection and evaluation. Nineteen respondents remarked that the instrument was too long, thirteen found it difficult to complete, and eight commented that the wording was confusing. Twelve (all pre-service) commented that the items were not applicable to them.

Of the 107 practicing teacher respondents, 90 commented that the instrument was easy to use; five commented that it was user friendly; and one commented that the tool was a quality instrument. Five respondents stated they used the instrument for self reflection and evaluation. Two respondents commented that the wording was confusing; six commented that the instrument was too long; two found it difficult; one stated that he did not like it; and one found it tedious.

The majority of the practicing teachers found the instrument easy to complete and user friendly. However, a few practitioners voiced concern about the scale, the length, and the wording. Several practicing respondents commented that they were uncomfortable with the survey until they realized that it was purely self evaluation. One respondent also wanted to change her response and voiced her frustration at not being able to do so. The majority of the national board candidates found the survey easy to complete and user friendly.

Several respondents used the survey to verbalize frustration. One respondent used the survey to voice his frustration with education administrators – “Teaching is a second career since retiring from 35 years in the private sector. Most alarming in education are administrators who cannot or will not think outside the box.” Another respondent also used the survey to voice frustration “very simple; as with most of the garbage that currently passes for professional education material, the survey was long on “education crap” and short on actual substance.” And one person took the opportunity to comment that he “would like the golf prize.”

On the other hand, of the 95 pre-service respondents, only 61 commented that the instrument was easy to use; two commented that it was straightforward, and one commented that it was thorough. Four used the instrument to assist in self reflection and evaluation. Thirteen respondents commented that the instrument was too long; eleven found it difficult to use; and six respondents commented that the wording was confusing. Four commented that the wording was redundant and two found it overwhelming. Twelve pre-service respondents commented that the

items were not applicable to them as pre-service practitioners.

The majority of the pre-service respondents found the instrument easy to complete; however, this group of respondents verbalized more issues with the survey. Several respondents expressed difficulty in getting started and in obtaining a password. Seven respondents commented that they had difficulty applying the standards to their practice as student teacher. Several respondents also complained that the survey was too long, too wordy, repetitive, and confusing.

Overall, respondents found the instrument easy to use—almost half the comments used the words “easy,” “simple,” or “user-friendly.” There was a clear difference by respondent type however, with the some pre-service teachers finding the instrument hard to use, excessively wordy, and inapplicable to themselves, comments that occurred scarcely at all within the practicing teacher group.

Based on these responses, we would make a few recommendations:

1. Examine the scale to determine if the parameters are accurate and if the wording can be decreased and the repetitiveness reduced. Also revisit the survey and determine if the spacing is confusing.
2. Ask for specific feedback about the survey “Describe how you can apply these standards to your practice.”
3. Pre-service teachers need clarification and direction on the survey and its applicability to them.
4. Offer additional instruction for the pre-service teachers on the logistics – how to obtain a password, how to change answers, etc.

Interpretation

Based on this report, TSARS and GSTEP offer the following interpretations concerning this study's results and potential next steps.

In terms of the validity of the instrument, the four traditional kinds of validity (construct, content, face, and criterion-related) need to be discussed separately. Construct validity, which is hypothesis-driven, was the primary focus of this study. The study did confirm construct validity as described by Question 1: Does the instrument generate variance across candidates? Practicing and pre-service teachers *do* rate themselves differently at a statistically significant level. Content validity addresses the accuracy of the content, which was previously established by the expertise brought to the instrument construction. Face validity involves the reactions of the participants, which was also incorporated into the development process and confirmed by participant comments from this study. Criterion-related validity addresses the comparative or predictive power of the instrument (that is, does it correlate to other measures OR does it predict later performance). There is currently no comparative data with which to assess criterion-related validity. Criterion-related validity is one possible avenue for future study.

In terms of the instrument content, the items composing each of the six elements had good correlations (Cronbach's alpha coefficients ranged from 0.80 to 0.89). However, the three factor analyses of items (one for all teachers, one for pre-service, one for experienced) showed varying patterns. The factor analyses suggest that experienced teachers have a more complex view of teaching than beginners. A logical next step would be to revisit the organization of the constructs

(elements) using the factor analysis to inform the organization further.

In terms of the instrument scale, though the responses were significantly different across participant types, participants rated themselves very highly overall. This suggests attention to the scale/rubric and the anchors provided to various ratings. (We understand this work is already in progress with the “Continuum.”).

In terms of additional questions, a particularly interesting avenue that could be explored further is the actual use of the instrument as a self-assessment and the consequences of use. Asking teachers using the survey how they have applied or could apply these standards to their practice would be one way to address this issue. Examining the actual use of the instrument through Messick’s Unified Validity Framework (see Appendix) is also an option.

Finally, we noted issues specific to online survey administration (such as logistical difficulties obtaining a password, etc.) that would be pertinent only if online administration is chosen for future iterations of the instrument.

Table 1: Description of the Sample

Gender	N	Percent
female	248	80.78
male	59	19.22

Institution	N	Percent
Albany State	2	0.66
UGA	176	58.09
Valdosta State	42	13.86
Other GA institutions	48	15.84
Non-Georgia institutions	35	11.55

Program	N	Percent
Art Ed	7	2.30
Business Ed	18	5.90
Dance Ed	1	0.33
Early Childhood Ed	100	32.79
English Ed	43	14.10
Family and Consumer Sci	8	2.62
Foreign Language Ed	2	0.66
Health/Physical Ed	12	3.93
Marketing Ed	1	0.33
Mathematics Ed	4	1.31
Middle School Ed	57	18.69
Music Ed	5	1.64
Science Ed	11	3.61
Social Science Ed	10	3.28
Special Ed	20	6.56
Technology Ed	4	1.31
Trade & Industrial Ed	2	0.66

Respondent Type	N	Percent
National board	12	3.95
Practicing (excl. Nat Brd)	143	47.04
Pre-service	149	49.01

Years of Experience	Mean	Std Dev
Overall	6.53	9.09
National Board	17.66	13.41
Practicing	11.15	9.23
Pre-service	0.57	1.73

Table 2: Summary Statistics by Indicator

	Indicator	N	Minimum	Maximum	Mean	Std Dev
I1	I-A. Knowledge of major concepts	352	1	5	4.04	0.73
I2	I-B. Subject-specific knowledge	351	2	5	3.93	0.78
I3	I-C. Current subject area	347	1	5	3.82	0.96
I4	I-D. Relate content to other subject areas	352	1	5	4.16	0.81
I5	I-E. Use a wide variety of resources	351	1	5	3.82	0.89
I6	I-F. Reflect state, national standards	345	2	5	4.08	0.84
I7	II-A. Hold high expectations	350	2	5	3.95	0.73
I8	II-B. Understand how learning occurs	352	1	5	3.72	0.84
I9	II-C. Sensitive, alert, and responsive	352	1	5	4.16	0.83
I10	II-D. Environmental influences	352	1	5	3.99	0.87
I11	II-E. Stage of development, etc	349	1	5	3.78	0.94
I12	II-F. Relationships with families	331	1	5	3.66	1.08
I13	III-A. Learning community	350	1	5	3.94	0.87
I14	III-B. Manage time, space, activities	351	1	5	3.93	0.82
I15	III-C. Classroom management	350	1	5	3.82	0.91
I16	III-D. Human motivation and behavior	352	1	5	3.89	0.79
I17	III-E. Students' uniqueness	345	1	5	3.65	1.01
I18	III-F. Access resources	338	1	5	3.61	0.98
I19	III-G. Effective communication	351	1	5	4.02	0.82
I20	IV-A. Understand measurement	349	1	5	3.65	0.91
I21	IV-B. Use pre-assessment data	342	1	5	3.42	1.00
I22	IV-C. Appropriate classroom assessment	351	1	5	3.75	0.82
I23	IV-D. Learners in self-assessment	343	1	5	3.59	0.99
I24	IV-E. Valid, equitable grading	345	1	5	4.19	0.75
I25	IV-F. Data to communicate student progress	331	2	5	3.85	0.85
I26	IV-G. Accurate and up-to-date records	333	1	5	3.93	0.88
I27	IV-H. Committed to using assessment	346	1	5	3.82	0.92
I28	V-A. Clear, defensible rationales	351	1	5	3.92	0.85
I29	V-B. Knlg of content, curriculum, students, etc	352	1	5	3.97	0.76
I30	V-C. Variety of instructional strategies	352	1	5	4.03	0.73
I31	V-D. Monitor and adjust strategies	351	1	5	4.03	0.78
I32	V-E. Vary roles in the instructional process	347	2	5	3.92	0.85
I33	V-F. Resources, materials and technology	349	1	5	3.85	0.86
I34	V-G. Planning as a collegial activity	337	1	5	3.79	1.01
I35	VI-A. Examine and extend knowledge	347	1	5	3.68	0.96
I36	VI-B. Laws on rights and responsibilities	349	1	5	4.15	0.84
I37	VI-C. Codes of professional conduct	347	2	5	4.37	0.75
I38	VI-D. Reflect on teaching and learning	349	2	5	4.21	0.78
I39	VI-E. Opportunities to learn	341	2	5	4.08	0.72
I40	VI-F. Advocate for learning	328	1	5	3.90	0.91
I41	VI-G. Leadership and support roles	327	1	5	3.93	0.89

Table 3: Frequency distributions for all Indicators

I1	Percent	Cumulative
Percent		
1	0.3	0.3
2	1.1	1.4
3	19.3	20.7
4	53.1	73.9
5	26.1	100.0

I2	Percent	Cumulative
Percent		
2	3.1	3.1
3	24.7	27.9
4	48.0	76.1
5	23.9	100.0
Missing (NATR)	0.3	

I3	Percent	Cumulative
Percent		
1	1.4	1.4
2	7.7	9.2
3	23.9	33.4
4	39.8	73.8
5	25.9	100.0
Missing (NATR)	1.4	

I4	Percent	Cumulative
Percent		
1	0.3	0.3
2	2.3	2.6
3	17.9	20.5
4	40.3	60.8
5	39.2	100.0

I5	Percent	Cumulative
Percent		
1	1.1	1.1
2	5.1	6.3
3	27.3	33.6
4	42.9	76.6
5	23.3	100.0
Missing (NATR)	0.3	

I6	Percent	Cumulative
Percent		
2	4.8	4.9
3	16.5	21.7
4	42.6	65.2
5	34.1	100.0
Missing (NATR)	2.0	

I7	Percent	Cumulative
Percent		
2	2.0	2.0
3	23.0	25.1

4	52.0	77.4
5	22.4	100.0
Missing (NATR)	0.6	

I8	Percent	Cumulative
Percent		
1	0.9	0.9
2	6.0	6.8
3	29.8	36.6
4	47.2	83.8
5	16.2	100.0

I9	Percent	Cumulative
Percent		
1	0.3	0.3
2	2.3	2.6
3	19.3	21.9
4	37.5	59.4
5	40.6	100.0

I10	Percent	Cumulative
Percent		
1	0.9	0.9
2	4.0	4.8
3	21.0	25.9
4	43.2	69.0
5	31.0	100.0

I11	Percent	Cumulative
Percent		
1	1.4	1.4
2	8.2	9.7
3	24.1	34.1
4	42.3	76.8
5	23.0	100.0
Missing (NATR)	0.9	

I12	Percent	Cumulative
Percent		
1	3.7	3.9
2	10.8	15.4
3	22.2	39.0
4	34.9	76.1
5	22.4	100.0
Missing (NATR)	6.0	

I13	Percent	Cumulative
Percent		
1	0.9	0.9
2	4.5	5.4
3	21.9	27.4
4	44.9	72.6
5	27.3	100.0
Missing (NATR)	0.6	

Missing (NATR) 0.3

I14	Percent	Cumulative
Percent		
1	0.3	0.3
2	3.4	3.7
3	24.7	28.5
4	45.5	74.1
5	25.9	100.0
Missing (NATR)	0.3	

I15	Percent	Cumulative
Percent		
1	1.1	1.1
2	5.4	6.6
3	29.0	35.7
4	38.6	74.6
5	25.3	100.0
Missing (NATR)	0.6	

I16	Percent	Cumulative
Percent		
1	0.3	0.3
2	3.4	3.7
3	25.3	29.0
4	48.6	77.6
5	22.4	100.0

I17	Percent	Cumulative
Percent		
1	2.8	2.9
2	7.7	10.7
3	33.0	44.3
4	32.4	77.4
5	22.2	100.0
Missing (NATR)	2.0	

I18	Percent	Cumulative
Percent		
1	2.0	2.1
2	9.4	11.8
3	32.4	45.6
4	32.7	79.6
5	19.6	100.0
Missing (NATR)	4.0	

I19	Percent	Cumulative
Percent		
1	0.6	0.6
2	3.1	3.7
3	19.9	23.6
4	46.3	70.1
5	29.8	100.0

I20	Percent	Cumulative
Percent		
1	1.4	1.4
2	7.1	8.6
3	34.4	43.3
4	38.4	81.9
5	17.9	100.0

Missing (NATR) 0.9

I21	Percent	Cumulative
Percent		
1	3.4	3.5
2	13.1	17.0
3	33.0	50.9
4	34.7	86.5
5	13.1	100.0

Missing (NATR) 2.8

I22	Percent	Cumulative
Percent		
1	0.6	0.6
2	6.0	6.6
3	28.1	34.8
4	48.3	83.2
5	16.8	100.0

Missing (NATR) 0.3

I23	Percent	Cumulative
Percent		
1	1.1	1.2
2	12.5	14.0
3	31.3	46.1
4	33.0	79.9
5	19.6	100.0

Missing (NATR) 2.6

I24	Percent	Cumulative
Percent		
1	0.3	0.3
2	0.6	0.9
3	16.8	18.0
4	43.2	62.0
5	37.2	100.0

Missing (NATR) 2.0

I25	Percent	Cumulative
Percent		
2	6.5	6.9
3	22.4	30.8
4	44.0	77.6

5	21.0	100.0
Missing (NATR)	6.0	

I26	Percent	Cumulative
1	0.3	0.3
2	5.1	5.7
3	22.4	29.4
4	39.5	71.2
5	27.3	100.0
Missing (NATR)	5.4	

I27	Percent	Cumulative
1	0.9	0.9
2	7.4	8.4
3	25.3	34.1
4	40.3	75.1
5	24.4	100.0
Missing (NATR)	1.7	

I28	Percent	Cumulative
1	0.6	0.6
2	3.4	4.0
3	26.4	30.5
4	42.3	72.9
5	27.0	100.0
Missing (NATR)	0.3	

I29	Percent	Cumulative
1	0.3	0.3
2	2.3	2.6
3	21.9	24.4
4	50.9	75.3
5	24.7	100.0

I30	Percent	Cumulative
1	0.3	0.3
2	2.0	2.3
3	17.6	19.9
4	54.8	74.7
5	25.3	100.0

I31	Percent	Cumulative
1	0.3	0.3
2	2.8	3.1
3	18.8	21.9
4	50.0	72.1
5	27.8	100.0
Missing (NATR)	0.3	

I32	Percent	Cumulative
1	0.3	0.3
2	2.8	3.1
3	18.8	21.9
4	50.0	72.1
5	27.8	100.0
Missing (NATR)	0.3	

2	4.5	4.6
3	25.9	30.8
4	40.9	72.3
5	27.3	100.0
Missing (NATR)	1.4	

I33	Percent	Cumulative
1	0.9	0.9
2	4.3	5.2
3	27.8	33.2
4	42.6	76.2
5	23.6	100.0
Missing (NATR)	0.9	

I34	Percent	Cumulative
1	2.0	2.1
2	9.9	12.5
3	19.0	32.3
4	40.1	74.2
5	24.7	100.0
Missing (NATR)	4.3	

I35	Percent	Cumulative
1	2.3	2.3
2	7.4	9.8
3	30.1	40.3
4	38.6	79.5
5	20.2	100.0
Missing (NATR)	1.4	

I36	Percent	Cumulative
1	0.6	0.6
2	2.8	3.4
3	16.5	20.1
4	40.3	60.7
5	38.9	100.0
Missing (NATR)	0.9	

I37	Percent	Cumulative
2	1.1	1.2
3	12.5	13.8
4	33.5	47.8
5	51.4	100.0
Missing (NATR)	1.4	

I38	Percent	Cumulative
2	2.3	2.3
3	15.1	17.5

4	41.8	59.6
5	40.1	100.0
Missing (NATR)	0.9	

2	5.7	7.6
3	17.9	26.8
4	43.8	73.8
5	24.4	100.0
Missing (NATR)	6.8	

I39	Percent	Cumulative
Percent		
2	2.3	2.3
3	15.1	17.9
4	52.6	72.1
5	27.0	100.0
Missing (NATR)	3.1	

I41	Percent	Cumulative
Percent		
1	0.6	0.6
2	5.1	6.1
3	21.9	29.7
4	38.4	70.9
5	27.0	100.0
Missing (NATR)	7.1	

I40	Percent	Cumulative
Percent		
1	1.4	1.5

Table 4: Indicator responses by respondent type. Missing (NATR) not included.

I1		Respondent Type		
Rating	practice	pre-serv		Total
1	0.00%	0.00%		
2	1.30%	1.34%		
3	12.34%	25.50%		
4	47.40%	59.06%		
5	38.96%	14.09%		
Total	154	149		303

I2		Respondent Type		
Rating	practice	pre-serv		Total
2	3.25%	4.05%		
3	16.23%	31.08%		
4	50.00%	47.97%		
5	30.52%	16.89%		
Total	154	148		302

I3		Respondent Type		
Rating	practice	pre-serv		Total
1	1.30%	0.69%		
2	1.30%	15.28%		
3	13.64%	34.72%		
4	48.05%	32.64%		
5	35.71%	16.67%		
Total	154	144		298

I4 Respondent Type			
Rating	practice	pre-serv	Total
1	0.00%	0.67%	
2	0.65%	3.36%	
3	8.44%	27.52%	
4	37.66%	42.95%	
5	53.25%	25.50%	
Total	154	149	303

I5 Respondent Type			
Rating	practice	pre-serv	Total
1	1.30%	0.68%	
2	1.30%	8.78%	

I7 Respondent Type			
Rating	practice	pre-serv	Total
2	1.30%	0.68%	
3	16.88%	30.61%	
4	53.90%	50.34%	
5	27.92%	18.37%	
Total	154	147	301

I8 Respondent Type			
Rating	practice	pre-serv	Total
1	0.00%	1.34%	
2	2.60%	8.72%	
3	21.43%	36.91%	
4	54.55%	44.97%	
5	21.43%	8.05%	
Total	154	149	303

3	24.03%	29.73%	
4	40.91%	43.92%	
5	32.47%	16.89%	
Total	154	148	302

I6 Respondent Type			
Rating	practice	pre-serv	Total
2	4.61%	5.48%	
3	12.50%	19.18%	
4	36.84%	46.58%	
5	46.05%	28.77%	
Total	152	146	298

I9 Respondent Type			
Rating	practice	pre-serv	Total
1	0.00%	0.67%	
2	1.30%	2.68%	
3	17.53%	20.13%	
4	31.17%	46.98%	
5	50.00%	29.53%	
Total	154	149	303

I10		Respondent Type		
Rating	practice	pre-serv	Total	
1	0.65%	1.34%		
2	2.60%	4.70%		
3	10.39%	29.53%		
4	46.75%	43.62%		
5	39.61%	20.81%		
Total	154	149	303	

I11		Respondent Type		
Rating	practice	pre-serv	Total	
1	1.30%	2.04%		
2	5.19%	9.52%		
3	18.83%	29.93%		

I13		Respondent Type		
Rating	practice	pre-serv	Total	
1	1.29%	1.36%		
2	3.87%	4.76%		
3	19.35%	23.13%		
4	40.65%	50.34%		
5	34.84%	20.41%		
Total	155	147	302	

	4	42.86%	43.54%	
	5	31.82%	14.97%	
Total	154	147	301	

I12		Respondent Type		
Rating	practice	pre-serv	Total	
1	3.95%	5.26%		
2	6.58%	13.53%		
3	17.76%	29.32%		
4	40.79%	35.34%		
5	30.92%	16.54%		
Total	152	133	285	

I14		Respondent Type		
Rating	practice	pre-serv	Total	
1	0.65%	0.67%		
2	3.23%	0.67%		
3	15.48%	32.21%		
4	46.45%	46.98%		
5	34.19%	19.46%		
Total	155	149	304	

I15		Respondent Type		
Rating	practice	pre-serv	Total	
1	1.29%	0.68%		
2	0.65%	8.84%		
3	18.71%	40.14%		
4	38.71%	40.14%		

	5	40.65%	10.20%	
Total	155		147	302

	4	32.47%	35.77%	
	5	28.57%	13.14%	
Total	154		137	291

I16 Respondent Type

Rating	practice	pre-serv	Total
1	0.65%	0.67%	
2	1.94%	4.03%	
3	18.06%	30.87%	
4	47.10%	51.01%	
5	32.26%	13.42%	
Total	155	149	304

I19 Respondent Type

Rating	practice	pre-serv	Total
1	0.65%	0.67%	
2	2.58%	3.36%	
3	12.26%	30.20%	
4	48.39%	44.30%	
5	36.13%	21.48%	
Total	155	149	304

I17 Respondent Type

Rating	practice	pre-serv	Total
1	5.16%	2.11%	
2	2.58%	13.38%	
3	30.32%	36.62%	
4	32.26%	32.39%	
5	29.68%	15.49%	
Total	155	142	297

I20 Respondent Type

Rating	practice	pre-serv	Total
1	0.65%	2.03%	
2	8.39%	6.76%	
3	25.16%	42.57%	
4	41.29%	35.14%	
5	24.52%	13.51%	
Total	155	148	303

I18 Respondent Type

Rating	practice	pre-serv	Total
1	2.60%	2.19%	
2	5.84%	11.68%	
3	30.52%	37.23%	

I21 Respondent Type

Rating	practice	pre-serv	Total
1	2.61%	2.82%	

2	8.50%	18.31%	
3	31.37%	38.03%	
4	38.56%	30.99%	
5	18.95%	9.86%	
Total	153	142	295

1	0.66%	0.00%	
2	0.00%	0.68%	
3	11.84%	19.18%	
4	35.53%	54.79%	
5	51.97%	25.34%	
Total	152	146	298

I22 Respondent Type

Rating	practice	pre-serv	Total
1	0.65%	0.68%	
2	3.23%	8.11%	
3	19.35%	37.16%	
4	53.55%	43.24%	
5	23.23%	10.81%	
Total	155	148	303

I25 Respondent Type

Rating	practice	pre-serv	Total
1	0.66%	0.00%	
2	5.96%	6.77%	
3	14.57%	30.83%	
4	44.37%	48.87%	
5	34.44%	13.53%	
Total	151	133	284

I23 Respondent Type

Rating	practice	pre-serv	Total
1	1.30%	1.40%	
2	12.99%	11.19%	
3	27.92%	34.97%	
4	31.17%	39.16%	
5	26.62%	13.29%	
Total	154	143	297

I26 Respondent Type

Rating	practice	pre-serv	Total
1	0.66%	0.00%	
2	4.61%	5.11%	
3	16.45%	29.20%	
4	40.13%	44.53%	
5	38.16%	21.17%	
Total	152	137	289

I24 Respondent Type

Rating	practice	pre-serv	Total
--------	----------	----------	-------

I27 Respondent Type			
Rating	practice	pre-serv	Total
1	0.00%	1.39%	
2	7.10%	6.94%	
3	18.71%	30.56%	
4	39.35%	45.14%	
5	34.84%	15.97%	
Total	155	144	299

I30 Respondent Type			
Rating	practice	pre-serv	Total
1	0.65%	0.00%	
2	1.94%	1.34%	
3	11.61%	22.82%	
4	52.26%	59.73%	
5	33.55%	16.11%	
Total	155	149	304

I28 Respondent Type			
Rating	practice	pre-serv	Total
1	0.00%	1.35%	
2	1.29%	4.05%	
3	20.65%	31.08%	
4	39.35%	48.65%	
5	38.71%	14.86%	
Total	155	148	303

I31 Respondent Type			
Rating	practice	pre-serv	Total
1	0.00%	0.00%	
2	2.58%	2.01%	
3	11.61%	24.16%	
4	50.97%	53.69%	
5	34.84%	20.13%	
Total	155	149	304

I29 Respondent Type			
Rating	practice	pre-serv	Total
1	0.00%	0.00%	
2	1.94%	2.68%	
3	16.13%	26.85%	
4	47.10%	53.02%	
5	34.84%	17.45%	

I32 Respondent Type			
Rating	practice	pre-serv	Total
2	1.97%	6.80%	
3	20.39%	29.93%	
4	36.84%	48.30%	
5	40.79%	14.97%	

Total	152	147	299
-------	-----	-----	-----

I33 Respondent Type			
Rating	practice	pre-serv	Total
1	0.65%	1.35%	
2	1.30%	7.43%	
3	22.08%	33.11%	
4	42.86%	43.92%	
5	33.12%	14.19%	
Total	154	148	302

I34 Respondent Type			
Rating	practice	pre-serv	Total
1	3.25%	1.46%	
2	10.39%	9.49%	
3	16.23%	21.90%	
4	36.36%	51.09%	
5	33.77%	16.06%	

I37 Respondent Type			
Rating	practice	pre-serv	Total
2	0.00%	2.05%	
3	3.92%	19.18%	
4	28.76%	43.15%	
5	67.32%	35.62%	

Total	154	137	291
-------	-----	-----	-----

I35 Respondent Type			
Rating	practice	pre-serv	Total
1	1.29%	2.05%	
2	3.23%	11.64%	
3	24.52%	39.73%	
4	42.58%	33.56%	
5	28.39%	13.01%	
Total	155	146	301

I36 Respondent Type			
Rating	practice	pre-serv	Total
1	0.00%	0.00%	
2	1.29%	4.79%	
3	7.10%	23.97%	
4	36.77%	47.95%	
5	54.84%	23.29%	
Total	155	146	301

Total	153	146	299
-------	-----	-----	-----

I38 Respondent Type			
Rating	practice	pre-serv	Total
2	1.30%	2.03%	
3	11.69%	18.24%	

	4	34.42%	52.03%
	5	52.60%	27.70%
Total	154	148	302

I39 Respondent Type

Rating	practice	pre-serv	Total
2	0.65%	5.04%	
3	5.81%	22.30%	
4	54.84%	55.40%	
5	38.71%	17.27%	
Total	155	139	294

I40 Respondent Type

Rating	practice	pre-serv	Total
1	1.95%	1.54%	
2	4.55%	8.46%	
3	11.04%	26.15%	
4	43.51%	50.77%	
5	38.96%	13.08%	
Total	154	130	284

I41 Respondent Type

Rating	practice	pre-serv	Total
1	1.30%	0.00%	
2	5.84%	5.43%	
3	15.58%	30.23%	
4	34.42%	48.84%	
5	42.86%	15.50%	
Total	154	129	283

Table 5: Indicators by category

- I. CONTENT AND CURRICULUM:
 - I-A. knowledge of major concepts
 - I-B. subject-specific knowledge
 - I-C. current in subject area
 - I-D. relate content to other subject areas
 - I-E. use a wide variety of resources
 - I-F. reflect state, national standards
- II. KNOWLEDGE OF STUDENTS AND THEIR LEARNING
 - II-A. hold high expectations
 - II-B. understand how learning occurs
 - II-C. sensitive, alert, and responsive
 - II-D. environmental influences
 - II-E. stage of development, etc
 - II-F. relationships with families
- III. LEARNING ENVIRONMENTS
 - III-A. learning community
 - III-B. manage time, space, activities
 - III-C. classroom management.
 - III-D. human motivation and behavior
 - III-E. students' uniqueness
 - III-F. access resources
 - III-G. effective communication
- IV ASSESSMENT
 - IV-A. understand measurement
 - IV-B. use pre-assessment data
 - IV-C. appropriate classroom assessment
 - IV-D. learners in self-assessment
 - IV-E. valid, equitable grading
 - IV-F. data to communicate student progress
 - IV-G. accurate and up-to-date records
 - IV-H. committed to using assessment
- V. PLANNING AND INSTRUCTION
 - V-A. clear, defensible rationales
 - V-B. knowledge of content, curriculum, students, etc
 - V-C. variety of instructional strategies
 - V-D. monitor and adjust strategies
 - V-E. vary roles in the instructional process
 - V-F. resources, materials and technology
 - V-G. planning as a collegial activity.
- VI. PROFESSIONALISM
 - VI-A. examine and extend knowledge
 - VI-B. laws on rights and responsibilities
 - VI-C. codes of professional conduct
 - VI-D. reflect on teaching and learning
 - VI-E. opportunities to learn
 - VI-F. advocate for learning
 - VI-G. leadership and support roles

Table 6: Correlations of years of experience and category scores
All correlations significant at alpha = .05.

Category	Years of Experience
Content and Curriculum	0.45
Knowledge of Students and their Learning	0.25
Learning Environments	0.30
Assessment	0.29
Planning and Instruction	0.39
Professionalism	0.31

Table 7: Cronbach alpha reliability by category

Category	Alpha
Content and curriculum	.80
Knowledge of students and their learning	.85
Learning environments	.85
Assessment	.89
Planning and Instruction	.87
Professionalism	.87

Table 8: Comparison of practicing and pre-service teachers by category, all comparisons significant at alpha=.05

Category	Practicing	Pre-service
Content and curriculum	4.19	3.79
Knowledge of students and their learning	4.06	3.69
Learning environments	4.02	3.66
Assessment	3.95	3.61
Planning and Instruction	4.10	3.77
Professionalism	4.28	3.81

Table 9: Factor structure of indicators within factors, all respondents

Factor	Indicator		Indicator description
			Planning & assessment
1	i13	III-A.	learning community
1	i14	III-B.	manage time, space, activities
1	i19	III-G.	effective communication
1	i22	IV-C.	appropriate classroom assessment
1	i24	IV-E.	valid, equitable grading
1	i27	IV-H.	committed to using assessment
1	i28	V-A.	clear, defensible rationales
1	i30	V-C.	variety of instructional strategies
1	i31	V-D.	monitor and adjust strategies
1	i32	V-E.	vary roles in the instructional process
			(To be defined)
2	i3	I-C.	current in subject area
2	i5	I-E.	use a wide variety of resources
2	i17	III-E.	students' uniqueness
2	i18	III-F.	access resources
2	i26	IV-G.	accurate and up-to-date records
2	i33	V-F.	resources, materials and technology
2	i35	VI-A.	examine and extend knowledge
			Knowledge and learning
3	i2	I-B.	subject-specific knowledge
3	i4	I-D.	relate content to other subject areas
3	i7	II-A.	hold high expectations
3	i8	II-B.	understand how learning occurs
3	i9	II-C.	sensitive, alert, and responsive
3	i10	II-D.	environmental influences
3	i11	II-E.	stage of development, etc
3	i15	III-C.	classroom management.
			Professionalism
4	i36	VI-B.	laws on rights and responsibilities
4	i37	VI-C.	codes of professional conduct
4	i38	VI-D.	reflect on teaching and learning
4	i39	VI-E.	opportunities to learn
4	i40	VI-F.	advocate for learning
4	i41	VI-G.	leadership and support roles
			Outside the classroom
5	i6	I-F.	reflect state, national standards
5	i12	II-F.	relationships with families
5	i20	IV-A.	understand measurement
5	i21	IV-B.	use pre-assessment data
5	i23	IV-D.	learners in self-assessment
5	i25	IV-F.	data to communicate student progress
5	i34	V-G.	planning as a collegial activity.

			Knowledge and skills
6	i1	I-A.	knowledge of major concepts
6	i16	III-D.	human motivation and behavior
6	i29	V-B.	knowledge of content, curriculum, students, etc

Table 10: Factor structure of indicators within factors, less experienced teachers

Factor	Indicator		Indicator description
			Understandings
1	i16	III-D.	human motivation and behavior
1	i19	III-G.	effective communication
1	i20	IV-A.	understand measurement
1	i23	IV-D.	learners in self-assessment
split 1 6	i08	II-B.	understand how learning occurs
			(To be defined)
2	i03	I-C.	current subject area
2	i07	II-A.	hold high expectations
2	i10	II-D.	environmental influences
2	i11	II-E.	stage of development, etc
2	i12	II-F.	relationships with families
2	i17	III-E.	students' uniqueness
2	i18	III-F.	access resources
2	i34	V-G.	planning as a collegial activity.
2	i35	VI-A.	examine and extend knowledge
			Professionalism
3	i24	IV-E.	valid, equitable grading
3	i36	VI-B.	laws on rights and responsibilities
3	i37	VI-C.	codes of professional conduct
3	i38	VI-D.	reflect on teaching and learning
3	i39	VI-E.	opportunities to learn
3	i41	VI-G.	leadership and support roles
			(To be defined)
4	i04	I-D.	relate content to other subject areas
4	i09	II-C.	sensitive, alert, and responsive
split 4 5	i26	IV-G.	accurate and up-to-date records
			Classroom activities
5	i01	I-A.	knowledge of major concepts
5	i02	I-B.	subject-specific knowledge
5	i05	I-E.	use a wide variety of resources
5	i13	III-A.	learning community
5	i14	III-B.	manage time, space, activities
5	i22	IV-C.	appropriate classroom assessment
5	i30	V-C.	variety of instructional strategies
5	i32	V-E.	vary roles in the instructional process
split 4 5	i26	IV-G.	accurate and up-to-date records
split 5 6	i27	IV-H.	committed to using assessment
			(To be defined)
6	i06	I-F.	reflect state, national standards
6	i21	IV-B.	use pre-assessment data
6	i28	V-A.	clear, defensible rationales
6	i29	V-B.	knowledge of content, curriculum, students, etc

6	i31	V-D.	monitor and adjust strategies
6	i33	V-F.	resources, materials and technology
6	i40	VI-F.	advocate for learning
split 1 6	i08	II-B.	understand how learning occurs
split 5 6	i27	IV-H.	committed to using assessment

Table 11: Factor structure of indicators within factors, more experienced teachers

Factor	Indicator	Indicator description
		Assessment, learning environment
1	i06	I-F. reflect state, national standards
1	i13	III-A. learning community
1	i14	III-B. manage time, space, activities
1	i20	IV-A. understand measurement
1	i22	IV-C. appropriate classroom assessment
1	i23	IV-D. learners in self-assessment
1	i27	IV-H. committed to using assessment
split 1 3	i32	V-E. vary roles in the instructional process
		Professionalism
2	i34	V-G. planning as a collegial activity.
2	i38	VI-D. reflect on teaching and learning
2	i39	VI-E. opportunities to learn
2	i40	VI-F. advocate for learning
2	i41	VI-G. leadership and support roles
		Planning
3	i19	III-G. effective communication
3	i24	IV-E. valid, equitable grading
3	i28	V-A. clear, defensible rationales
3	i31	V-D. monitor and adjust strategies
3	i36	VI-B. laws on rights and responsibilities
split 1 3	i32	V-E. vary roles in the instructional process
split 3 8	i37	VI-C. codes of professional conduct
		(To be defined)
4	i05	I-E. use a wide variety of resources
4	i21	IV-B. use pre-assessment data
		(To be defined)
5	i12	II-F. relationships with families
5	i17	III-E. students' uniqueness
5	i18	III-F. access resources
5	i25	IV-F. data to communicate student progress
5	i26	IV-G. accurate and up-to-date records
5	i35	VI-A. examine and extend knowledge
split 5 6	i33	V-F. resources, materials and technology
split 5 8	i16	III-D. human motivation and behavior
		Content
6	i01	I-A. knowledge of major concepts
6	i02	I-B. subject-specific knowledge
6	i03	I-C. current subject area
6	i04	I-D. relate content to other subject areas
6	i29	V-B. knowledge of content, curriculum, students, etc
split 5 6	i33	V-F. resources, materials and technology

			Students and learning
7	i07	II-A.	hold high expectations
7	i08	II-B.	understand how learning occurs
7	i09	II-C.	sensitive, alert, and responsive
7	i10	II-D.	environmental influences
7	i11	II-E.	stage of development, etc
7	i30	V-C.	variety of instructional strategies
			Managing and motivating
8	i15	III-C.	classroom management.
split 5 8	i16	III-D.	human motivation and behavior
split 3 8	i37	VI-C.	codes of professional conduct

Appendix:

Beyond Classical Validity: Messick's Unified Validity Framework

A major step forward from the traditional approaches to examining validity came with the work of Samuel Messick (Messick, 1989). Messick proposed that construct validity (taken broadly to encompass content- and criterion-based evidence) was an essential but only an initial facet in the complete examination of validity. Below, we present Messick's view of what he refers to as a Unified Validity Framework, and explore how the GSTEP Instrument may be discussed in these terms.

Messick's presentation is based on his "progressive matrix," a two-by-two table showing the various intersections of evidence, consequences, interpretation, and utility. In the table below, each cell has a unique contribution (in italic), but the notion of the progressive matrix is that each issue must also be interpreted in light of the preceding considerations.

Table X: Messick's Unified Validity Framework

	Test Interpretation	Test Use
Evidential Basis	<i>Construct Validity</i>	Construct Validity + <i>Relevance/Utility</i>

Consequential Basis	Construct Validity + <i>Value Implications</i>	Construct Validity + Relevance/Utility + Value Implications + <i>Social Consequences</i>
---------------------	---	---

In terms of the UVF, the concept of construct validity is at the intersection of the evidential basis for the test, generated through data collection and analysis to test hypotheses derived from the theory of the construct, and the interpretation of the test scores to make inferences and/or predictions about the examinee. Thus, inferences based on the scores from a test cannot be accurate if the test lacks construct validity. The discussion above provides some support for the construct validity of the instrument, both in terms of content (based on widely held conceptions of the task of teaching) and hypothesis testing.

This is, of course, merely the beginning. Moving to the next cell (upper right), we need to consider the relevance and utility of the instrument. Although it may be capturing valid insights about a teacher’s craft, of what utility is it? We see at least two answers to this question. First, the Instrument serves as a means for reflection for the teacher: What do I know? What can I do? What do I still have to learn?

Second, when used by a supervising teacher or mentor, the Instrument could serve as structure for feedback and discussion. In our initial planning, we had hoped to recruit a sizeable number of mentor/mentee pairs, in order to be able to examine how these diads used the Instrument together. Unfortunately, we had insufficient data to be able to make

any conclusions about this use. We believe that it is not necessary for the mentor and mentee to agree on the Assessment. Indeed, it might be of greater utility in engendering discussion if they disagreed. We hypothesized that those items with the greatest disagreement might be the most useful. This hypothesis was not able to be tested.

It is conceivable that the instrument might be used for screening purposes, although we would not advocate this. This is common practice in business settings. Since the instrument is a self-assessment, there would be a higher probability for false positives (inflated ratings) than false negatives (over-stringent self-assessments).

The third aspect of the UVF relates to the values dimension. The authors of the instrument have done a good job in wording the items in a non-judgmental way, stating the skill or practice and distinguishing between people who seldom, usually, or consistently perform to standard. None of the responses to the open-ended items suggested that any respondent found any of the descriptors derogatory in any way. Most assessments are couched in valuing terms, of course. Higher scores are seen as preferable. But, paradoxically, the items showing the greatest development needs would be the most valuable in the teacher's professional improvement, as they indicate directions for study and reflection.

It is essential for the success of the Framework Assessment that it be viewed as a diagnostic, rather than as a judgmental, instrument. A person undergoing an MRI may be anxious about the findings of the test, but they do not view it as making a judgment on

their value. Similarly, the Framework Assessment needs to be put forward as a process to assist the teacher, not to judge them as potentially wanting. In a very real sense, every teacher, even those who have received National Board recognition, is in some location on a developmental path, never at perfection.

Finally, the social consequences of the use of the Instrument should be considered before it is adopted widely and formally. Could it have the potential effect of causing pre-service teachers to de-select from the profession? Could this de-selection be focused in some identifiable group?

Perhaps the answer to both of these lies in how the Assessment is presented. To the extent that it is personal and private, teachers are most likely to use it as intended. If it becomes something that is imposed by authority, it may be still useful as a judgmental screening instrument, but much of its power to point toward needed changes may be sacrificed. At its best, the instruments should foster a greater reflectivity towards practice on the part of pre-service and in-service teachers alike.