Measuring Teacher Dispositions with Different Item Structures: An Application of the Rasch Model to a Complex Accreditation Requirement

W. Steve Lang

University of South Florida St. Petersburg

Judy R. Wilkerson

Florida Gulf Coast University

Paper presentation at the annual meeting of the

American Educational Research Association

New York, March 2008-03-11

Measuring Teacher Dispositions with Different Item Structures:

An Application of the Rasch Model to a Complex Accreditation Requirement

Abstract

The construct of dispositions is defined in national standards, and colleges of education are required to assess candidate dispositions to meet accreditation requirements. Similarly, there is a need to review teacher dispositions in making hiring decisions about teachers, although this need may not yet be realized. Measurement is virtually non-existent in the area of teacher dispositions. On-line reviews of college accreditation reports indicate that colleges are attempting to assess dispositions without the use of sound measurement techniques or adequate definitions of the construct. The end result, of course, is a reliance on face validity. Rasch measurement provides a much needed solution to scaling the dispositions needed for good teaching when a credible construct is operationally used and visibly defined. This paper presents early work in the development of five related instruments measuring ten standards-based principles related to dispositions. The instruments use different item structures and response formats, which are aggregated into a single disposition scale that includes sub-scale scores for each of the principles.

Objectives or Purposes

NCATE (2002) requires the measurement of dispositions as part of its accreditation requirements for teacher education programs. The first standard, entitled, "Candidate Knowledge, Skills, and Dispositions," requires that: "Candidates preparing to work in schools as teachers or other professional school personnel know and demonstrate the content, pedagogical, and professional knowledge, skills, and dispositions necessary to help all students learn. Assessments indicate that candidates meet professional, state, and institutional standards." Our first purpose is to help teacher preparation institutions to meet this need in a credible way.

In addition to the need experienced by teacher education institutions to measure dispositions, we suspect that over time, school districts will also realize that teachers who are committed to the profession, and all of its critical skills, are likely to have a greater impact on student learning. We aim to move districts in that direction with this work.

Beyond the bureaucratic needs to be accredited and to hire high quality teachers, the most fundamental purpose of this battery is "to find and fix" dispositional weaknesses in teacher candidates and teachers hired in the school districts. As in skills-based measures, we find that some teachers are strong in some skills and weak in others; they are also strong on some attitudes and weak on others. Professional development plans, at whatever level, need to focus on building on strengths and remediating weaknesses that are both skill and dispositional based.

Literature Review

The published literature on measuring teacher dispositions has been sparse for many years. This is probably at least in part due to the general assessment illiteracy that pervades our culture (Stiggins, 2000; Popham, 2004). A study by Schulte, Edick, Edwards, & Mackiel (2004) confirms that little guidance is provided about measuring teacher dispositions. There is instead much in the literature that is skill based rather than values-based (Darling-Hammond, 2000). There are some exceptions, Wasicsko's (2004) 20-Minute Hiring Assessment, Holt-Reynolds' (1991) biographical and metaphorical assessments, and several discussions of portfolios applied to disposition assessment; all are cited in (Schulte, et al., 2004). The Schulte team developed and validated a Teacher Disposition Index (TDI), which is similar in some ways to one of the four instruments modelled by Wilkerson & Lang (2007). There are many differences, though, and chief among them is the limitation to a single instrument type. Four recent books have been published on the subject:

1. *Teacher dispositions: Envisioning their role in education* (Koeppen & Davidson-Jenkins, 2007). This new book discusses the general context of teacher dispositions and includes a Personal Qualitative Inventory and several proposed rubrics for scoring dispositions.

2. *Teacher Dispositions: Building a teacher education framework of moral standards* (Sockett, 2006). This short book of philosophical essays advocates for the evaluation of morals and the development of a code of ethics for the profession. Limited advice for the development of assessment instruments is provided in one of the three essays by Mary Diez.

3. *Dispositions in Teacher Education* (Diez & Raths, 2007). This is an edited volume that describes a variety of perspectives on dispositions including several historical and theoretical descriptions of the concept in teacher education.

4. *Assessing Teacher Dispositions, Five Standards-Based Steps to Measuring Teacher Dispositions Using the DAATS* Model (Wilkerson & Lang, 2007b). A model for developing disposition assessments, based on the INTASC Principles, and using accepted affective measurement techniques is described in detail. In this paper presentation, we summarize the steps of the model and provide brief examples of the techniques recommended. Steps of the DAATS (Disposition Assessment Aligned with Teacher Standards) model are described below.

With the notable exception of Jim Raths and Wilkerson/Lang, the measurement community has been largely uninvolved in trying to solve the disposition dilemma. The second and third books cited above (Socket, 2006 and Diez and Raths, 2007) are the product of a movement within the American Association of Colleges of Teacher Education (AACTE), lead by those authors and several others, to promote the assessment of morality as the cornerstone construct defining teacher dispositions. This work is organized by a task force entitled TEAM-C, Teacher Education as A Moral Community (http://www.aacte.org/Programs/TEAMC/default.aspx). This

task force has "been charged to begin the engaging and challenging work of consensus around dispositions in teacher education." Given the name of the team, morality is clearly at the core of their work.

Some measurement professionals are disturbed at this direction. In this paper, we will present our attempts at providing an objective measurement solution to the problem. We approached this problem developing an evaluation model to help assessment designers use the standards of teaching as the operational definition of teacher performance and teacher values (Wilkerson and Lang, 2007a and 2007b). For both forms of assessment we envision the use of multiple measures. In the dispositions based process, the one we describe here, we envision multiple instruments, of different formats and at differing levels of inference, and this is substantially different from our vision of the skills-based area. We concluded that the *Rasch model* provided the most utility for scaling the construct (Linacre, 2003).

The common set of national standards developed by the Council of Chief State School Officers (1992) and promulgated by the Interstate New Teacher Assessment and Support Consortium (INTASC) in the form of ten principles serves as the basis for defining both constructs – competency and dispositions. Each of the principles includes indicators written at the knowledge, performance, and dispositional levels. Colleges are instructed to use these standards, but drift has occurred with the new focus on morality.

There are distinct advantages for institutions to consider the INTASC Principles when developing their processes for measuring teacher dispositions. It is also both a measurement and legal detriment to rely on generic traits, often linked to morality, ethics, caring, fairness, and/or social justice that leave units wide-open for attack in the measurement and legal worlds. This failure to attend to standards-based definitions and adequate due process procedures, when combined with a lack of solid data to diagnose and remediate, creates the potential for institutions to be attacked if they choose to deny graduation to a teacher candidate who exhibits dispositional deficits.

The INTASC Principles, when combined with appropriate measurement methods, provide an appropriate standards-based construct that allows college personnel to do what they need to do -- make decisions that are less likely to be successfully challenged while, at the same time, providing data that can be aggregated to improve the outcomes of their programs. Valid, reliable, and fair measurement can lead to information about what students are learning and what they are not learning. It can also serve as a predictor of future behavior and a tool for both individual and program improvement. Dispositions are different from knowledge and skills, and both require our serious attention.

Teacher educators commonly use tests, products and live observations to measure knowledge and skills. Appropriate measurement methods for affect (dispositions) include scales, questionnaires and interviews, focus groups, observations, and apperception tests, as well as documentation of inappropriate behavior when it occurs. In both cases, there is a need to focus the content of the measurement instruments on the standards that programs are seeking to demonstrate.

In general, levels of inference are dictated by how hard it is to score an instrument. If a machine or a relatively untrained rater can score with a high degree of accuracy, then the level of inference is low. As the level of inference increases, the difficulty of scoring and the need for rater training, rubrics, and examples increase. It becomes more difficult to make the judgment about the observed response. In the development of this scale, we are using five instruments of increasing levels of inference: a Thurstone scale, a teacher questionnaire, an apperception test, and an interview (focus group) of a sample of K-12 students, an adaptive behavior observation. As of this writing, the primary pilot includes the first three instruments in the battery. The methodology used for instrument design is described at length in Wilkerson and Lang (2007).

To facilitate reliable and meaningful scoring, we add a framework as part of the construct mapping: the Bloom and Krathwohl Affective Taxonomy (1956) provides a vehicle for making visible the different levels of attitude about the teacher standards. We add to the original Taxonomy a new level, which we call "unaware." In conclusion, the process we are developing uses the standards-based values of teaching (as articulated in the INTASC Principles), made visible through the Bloom and Krathwohl affective taxonomy, and measured using proper measurement methods. Finally the instruments rely on the *Standards of Educational and Psychological Testing* (AERA, APA, NCME, 1999) to ensure the psychometric credibility of the instruments starting first and foremost with the recommended underpinning of job-relatedness.

Even though comprehensive results are not presented here, the analysis and basis for psychometric development and statistical reporting were based on guidelines and recommendations from the following sources:

1. Bond, T. & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences (2nd).* Mahwah, NJ: LEA.

2. Linacre, J. M. (2003). *A user's guide to FACETS: Rasch-model computer programs.* MESA Press: Chicago.

3. Linacre, J. M. (2003). *A user's guide to WINSTEPS: Rasch-model computer programs.* MESA Press: Chicago.

4. Smith, E. & Smith, R., Eds. (2004). *Introduction to Rasch Measurement.* Maple Grove, MN: JAM.

5. Wilson, M. (2005). *Constructing measures: An item response modelling approach.* Mahwah, NJ: LEA.

Method

Instruments

The initial design is a battery of five instruments of increasing inference and field-tested the first three extensively. We include for information purposes a sixth form which is less an instrument and more a tracking process for problems. The instruments are:

Beliefs About Teaching Scale (BATS):

BATS is a 60-item Thurstone agreement scale. Items cover the ten Principles and range from easy to difficult, by design, with half the responses expected as "agree" and the other half "disagree" and six per Principle. Because of the guessing/faking factor and the ease of scoring, this instrument is low inference and provides for limited confidence. The assumption is that most candidates are "normal," and distributions in pilot test data are, in fact, normal. The candidates more than two standard deviations from the mean clearly exhibit extreme values when assessed judgmentally, running the risk of burn-out on the high end (retention risk) or harming children on the low end.

Experiences in Teaching Questionnaire (ETQ):

ETQ is a ten-item questionnaire that includes small sub-sets of questions targeting each of the INTASC Principles. It is scored manually and a little more difficult to fake, so it provides the next level of useful assessment of dispositions. Responses are rated using a five-item rating scale for each Principle, based on the Krathwohl Affective Taxonomy. Each set of items is hierarchically ordered from low on the Taxonomy (unaware or receiving) to high on the Taxonomy (characterizing), with the expectation that most teachers will reach the valuing level.

Situational Reflection Assessment (SRA):

The SRA is a modified thematic apperception test. In this instrument the goal is not to diagnose extreme behaviors but rather beliefs about specific skills of teaching. Hence, the prompts are more thematically specific than in a traditional TAT. We make use of emotional pictures, ambiguity of expression, story-telling, and projections into the future from graphic and verbal prompts. The graphic art has been designed by a professional artist (Slitkin, 2007), modelled after carefully selected images and verbal descriptions. In this instrument, we provide 20 prompts (pictures and questions), two per Principle. Picture development Gieser & Stein (1999) and scoring Teglasi (2001) were designed according to modern apperception design, but focused for the disposition construct. Responses are rated using a five-item rating scale for each Principle, based on the Krathwohl Affective Taxonomy.

K-12 Impact Disposition Scale (KIDS):

KIDS, a K-12 focus group, is at the next level of inference. Again there are ten sets of items, with flexibility allowed to the interviewer in selecting the subset. Required in each set, though, are one or two items that elicit specific responses about negative attitudes on the part of the teacher. When there are clear patterns of concern among children in a group and comments tend to reflect similar concerns, the interviewer can infer a problem may exist with the teacher's dispositions. Good data need to be sorted from noise; however, faking on the teacher's part is no longer an issue, making this instrument high inference and high confidence. The results are rated using the same basic scale.

Classroom Disposition Checklist (CDC):

The CDC is an adaptive behavior checklist that seeks to determine if the teacher exhibits affective-based behaviors appropriately in the classroom. It requires a high level of inference, since it requires judgment over time. It must be administered by someone who has observed the teacher for an extended period of time, thereby again reducing the opportunity for faking. The checklist includes a series of 50 paired behaviors (positive and negative) that are indicative of valuing a standards-based skill. There are an unequal number of behaviors by Principle, but all ten are covered. It is expected that not all of the 50 behaviors will have been observed with adequate frequency to render a judgment.

Disposition Event Report (DER):

The DER is a different application, not truly a part of the battery. It is more a form for recording a disturbing event, keeping track of remediation activities, and having a record of multiple incidents that could lead to a decision to deny a diploma or fail to rehire a teacher. There is no cut-score intended on these instruments, but this instrument helps to build a case for the preponderance of evidence of a teacher's commitment (or lack thereof) to teaching skills. The DER provides anecdotal evidence for the battery scores.

Attached in Appendix A is a sample for INTASC Principle 1 that demonstrates coverage across instruments of each Principle in the five instruments.

Item Formats

While the scoring for BATS is dichotomous, we use a rating scale for ETQ and SRA. The rating scale, based on the Bloom and Krathwohl Affective Taxonomy, is defined as:

0 = Unaware (considered to be dangerous to practice)
1 = Receiving (beginning level teacher candidates))
2 = Responding (acceptable for beginning teachers)
3 = Valuing (target for teachers at all levels)
4 = Organizing (target for teacher/leaders)
5 = Characterizing (highest level for leaders; not expected to be frequently used)

Rubrics

The rubrics being used to assess teacher responses to the SRA and ETQ have been piloted but have yet another round to go on their refinement. As we get closer to clearly defined standards for each rubric level, the data, even with minimal sample size, models better. Attached in Appendix B is the generic rubric for the levels of the Krathohl Taxonomy. That rubric defines not only the description of each rating scale point but also interpretation and intended use of the points. Appendix C is a sample of specific rubric for one INTASC Principle.

Sample and Data Sources

At present, we have analysed data on about 1800 respondents on various scales as item analysis and data collection is on-going. For this study, we have reanalysed the responses of 335 teacher candidates on the BATS, ETQ, and SRA with the latest refinement of the rubrics. The initial intent was to use the Krathwohl Taxonomy relatively loosely, but the data have indicated that the taxonomy provides a sound basis for making visible the levels of quality in the responses to both of these instruments.

Analysis

The instruments use two Rasch models for the scales: dichotomous (BATS) and rating scale (ETQ, SRA) as follows (Stone, 2003):

$$\ln\left(\frac{P_{ni}}{(1-P_{ni})}\right) = B_n - D_i$$

$$\ln\left(\frac{P_{nik}}{(1-P_{nik})}\right) = B_n - D_i - F_{ik}$$

All analyses were completed using three commonly available software programs:

1. *Winsteps:* Rasch measurement software for persons & items. Multiple choice questions, rating scales, and partial credit can be analysed with up to 255 categories per scale.

2. *FACETS:* Many-facet Rasch measurement for persons, items, judges, and tasks is software that assesses rater error and calibrates different variables on parallel scales.

3. *Systat 12:* SYSTAT is statistical analysis and graphical software.

The initial analysis with Winsteps groups (blocks) item types together in a single calibration. This is described by Linacre (2007) as:

$$\log\left(\frac{P_{nij}}{P_{ni(j-1)}}\right) = B_n - D_{gi} - F_{gj}$$

where P is a probability, and the Rasch parameters are Bn, the ability of person, Dgi, the difficulty of item i of grouping g, and Fgj, the Rasch-Andrich threshold between categories j-1 and j of grouping g.

Linacre (2007) also warns that calibration with different item types may result in one type "underfitting" and another type "overfitting". Because of this issue, individual item calibrations are reported below as separate analyses, but the overall model results are provided when appropriate.

Model Results

Here we present a highlight of the results of the pilot of most relevance and interest. That is followed by the technical presentation related to the measurement properties of the tests.

Table 1
Winsteps Separation Results of Pilot Results of BATS, ETQ, and SRA Calibration
BATS (60 items), ETQ (10 items), SRA (20 items)

```
 TABLE 3.1 Disposition Analysis  Spring 08 Pilot G ZOU796WS.TXT Mar  9 18:31 2008
INPUT: 335 persons  90 items  MEASURED: 335 persons  90 items  13 CATS      3.64.2
 -------------------------------------------------------------------------------

                 SUMMARY OF 261 MEASURED (NON-EXTREME) persons
 +-------------------------------------------------------------------------------+
 |           RAW                           MODEL          INFIT        OUTFIT    |
 |          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
 |-------------------------------------------------------------------------------|
 | MEAN      65.9      70.8       58.81    2.65      1.17     .5    .95    -.3 |
 | S.D.      17.7       5.4        9.64     .69       .71    2.3    .62    1.8 |
 | MAX.     110.0      81.0       90.31   10.33      6.36    9.9   4.71    9.9 |
 | MIN.      12.0      60.0       24.60    1.93       .39   -3.2    .27   -3.0 |
 |-------------------------------------------------------------------------------|
 | REAL RMSE   3.08  ADJ.SD    9.14  SEPARATION  2.97  person RELIABILITY  .90 |
 |MODEL RMSE   2.74  ADJ.SD    9.24  SEPARATION  3.37  person RELIABILITY  .92 |
 | S.E. OF person MEAN = .60                                                     |
 +-------------------------------------------------------------------------------+
                    MINIMUM EXTREME SCORE:     74 persons
                             VALID RESPONSES:  78.7%
           CRONBACH ALPHA (KR-20) person RAW SCORE RELIABILITY = .96

                   SUMMARY OF 90 MEASURED (NON-EXTREME) items
 +-------------------------------------------------------------------------------+
 |           RAW                           MODEL          INFIT        OUTFIT    |
 |          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
 |-------------------------------------------------------------------------------|
 | MEAN     191.2     205.3       50.00    1.70      1.07     .1   1.26     .8 |
 | S.D.      60.4      81.6       14.06     .48       .58    2.7    .99    3.6 |
 | MAX.     309.0     261.0       75.39    2.57      3.49    7.1   5.98    9.9 |
 | MIN.      50.0      44.0       29.36     .91       .48   -3.4    .18   -4.3 |
 |-------------------------------------------------------------------------------|
 | REAL RMSE   1.95  ADJ.SD   13.93  SEPARATION  7.14  item   RELIABILITY  .98 |
 |MODEL RMSE   1.77  ADJ.SD   13.95  SEPARATION  7.88  item   RELIABILITY  .98 |
 | S.E. OF item MEAN = 1.49                                                      |
 +-------------------------------------------------------------------------------+
                            UMEAN=50.000 USCALE=10.000
```

The Real Person Separation of .90 indicates that the scale discriminates between persons well. The Real Item Separation of .98 indicates that the items create a well defined variable. The mean person measure (58.81 with expected mean of 50) suggests that the items are moderately easy for the persons to assign high scores. The person outfit zstd of -.3 and SD of 1.8 indicate there is a little more variability in the fit of these persons than expected and the mean fit is slightly lower than expected. The item outfit zstd mean of .8 and SD of 3.6 would normally indicate some misfitting items, but that is confounded here with the grouped item types contributing to the expected misfit.

Table 2 Item Calibration Statistics of BATS, ETQ, and SRA (Simultaneous)

```
TABLE 14.1 Disposition Analysis  Spring 08 Pilot  ZOU796WS.TXT Mar  9 18:31 2008
INPUT: 335 persons  90 items  MEASURED: 335 persons  90 items  13 CATS      3.64.2
-------------------------------------------------------------------------------
person: REAL SEP.: 2.97  REL.: .90 ... item: REAL SEP.: 7.14  REL.: .98

item STATISTICS:  ENTRY ORDER
```

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEA CORR.	EXACT MATCH OBS%	EXACT MATCH EXP%	DISPLACE	item	G
1	222	261	38.69	1.94	.78	-1.7	.79	-1.1	.82	91.6	87.7	-.02	BAT01RC03	1
2	197	261	46.27	1.59	.97	-.3	1.01	.1	.70	81.6	80.0	-.02	BAT02RC07	1
3	228	261	36.27	2.08	.73	-1.9	.76	-1.2	.85	93.5	89.4	-.03	BAT03VA10	1
4	241	261	29.36	2.57	.50	-3.1	.25	-3.5	.93	94.6	92.6	-.03	BAT04RC03	1
5	214	261	41.47	1.79	.82	-1.6	.82	-1.2	.79	88.9	85.4	-.02	BAT05VA02	1
6	241	261	29.36	2.57	.48	-3.3	.18	-4.2	.93	94.6	92.6	-.03	BAT06RP04	1
7	132	261	59.22	1.32	1.01	.2	.95	-.6	.50	60.5	63.3	-.01	BAT07RC10	1
8	236	261	32.37	2.34	.58	-2.8	.38	-3.1	.90	96.2	91.3	-.03	BAT08OR03	1
9	209	261	43.01	1.72	.86	-1.3	.89	-.8	.76	87.0	83.8	-.02	BAT09RC08	1
10	223	261	38.31	1.96	.80	-1.5	.84	-.8	.82	92.0	88.0	-.02	BAT10OR02	1
11	241	261	29.36	2.57	.49	-3.2	.21	-3.9	.93	94.6	92.6	-.03	BAT11VA01	1
12	235	261	32.91	2.30	.62	-2.6	.55	-2.0	.89	95.8	91.1	-.03	BAT12RC07	1
13	121	261	61.14	1.32	1.12	3.2	1.09	1.1	.44	55.9	63.0	-.01	BAT13VA08	1
14	215	261	41.14	1.81	.85	-1.3	.94	-.3	.79	89.3	85.7	-.02	BAT14RC08	1
15	238	261	31.24	2.43	.53	-3.1	.32	-3.3	.92	96.2	91.8	-.03	BAT15RC08	1
16	143	261	57.29	1.33	1.08	2.0	1.05	.7	.51	57.9	64.7	-.01	BAT16RC05	1
17	185	261	49.10	1.49	.98	-.2	1.04	.5	.66	77.8	76.1	-.02	BAT17RC01	1
18	240	261	30.01	2.52	.48	-3.4	.18	-4.3	.93	95.4	92.3	-.03	BAT18RC09	1
19	186	261	48.88	1.50	.92	-1.0	.90	-1.0	.68	78.5	76.5	-.02	BAT19RP10	1
20	123	261	60.80	1.32	1.04	1.1	.99	-.1	.47	59.0	63.0	-.01	BAT20VA09	1
21	116	261	62.02	1.32	1.06	1.8	1.01	.1	.44	54.8	63.2	-.01	BAT21VA03	1
22	233	261	33.94	2.23	.64	-2.4	.90	-.3	.88	95.4	90.6	-.03	BAT22VA04	1
23	50	261	75.39	1.63	1.07	.8	1.12	.6	.25	80.8	81.1	-.01	BAT23RP09	1
24	225	261	37.52	2.01	.73	-2.0	.74	-1.4	.84	92.3	88.5	-.02	BAT24RP06	1
25	235	261	32.91	2.30	.57	-2.9	.34	-3.5	.90	95.8	91.1	-.03	BAT25VA06	1
26	226	261	37.11	2.03	.70	-2.3	.60	-2.2	.85	92.7	88.8	-.03	BAT26VA10	1
27	179	261	50.40	1.45	.99	-.1	.97	-.3	.64	75.5	74.3	-.02	BAT27RP03	1
28	184	261	49.32	1.48	.94	-.8	.92	-.8	.67	77.4	75.8	-.02	BAT28RP10	1
29	116	261	62.02	1.32	1.18	4.8	1.16	1.7	.41	51.0	63.2	-.01	BAT29RC04	1
30	233	261	33.94	2.23	.67	-2.2	.69	-1.4	.88	95.4	90.6	-.03	BAT30RC05	1
31	236	261	32.37	2.34	.60	-2.6	.50	-2.2	.90	96.2	91.3	-.03	BAT31VA05	1
32	238	261	31.24	2.43	.52	-3.2	.25	-3.9	.92	96.2	91.8	-.03	BAT32RP02	1
33	211	261	42.41	1.75	.82	-1.6	.78	-1.5	.78	87.7	84.5	-.02	BAT33RC04	1
34	217	261	40.48	1.84	.83	-1.3	.89	-.6	.80	90.0	86.3	-.02	BAT34RP05	1
35	241	261	29.36	2.57	.49	-3.2	.22	-3.8	.93	94.6	92.6	-.03	BAT35RC03	1
36	197	261	46.27	1.59	.95	-.6	.92	-.7	.71	82.4	80.0	-.02	BAT36RP06	1
37	234	261	33.44	2.27	.62	-2.5	.60	-1.8	.89	95.4	90.9	-.03	BAT37RC10	1
38	194	261	47.01	1.56	.97	-.3	1.01	.2	.69	80.5	79.0	-.02	BAT38RP04	1
39	231	261	34.91	2.17	.71	-2.0	.78	-.9	.86	94.6	90.1	-.03	BAT39RP06	1
40	238	261	31.24	2.43	.53	-3.1	.30	-3.5	.92	96.2	91.8	-.03	BAT40RC02	1
41	197	261	46.27	1.59	.98	-.2	1.06	.6	.70	83.1	80.0	-.02	BAT41RC02	1
42	225	261	37.52	2.01	.78	-1.6	.81	-1.0	.83	92.3	88.5	-.02	BAT42RC02	1
43	222	261	38.69	1.94	.76	-1.9	.76	-1.4	.83	91.6	87.7	-.02	BAT43RC09	1
44	150	261	56.03	1.35	.97	-.7	.92	-1.0	.56	69.0	66.1	-.02	BAT44RC01	1
45	196	261	46.52	1.58	.96	-.5	1.00	.1	.70	82.0	79.7	-.02	BAT45RC08	1
46	225	261	37.52	2.01	.73	-2.0	.66	-1.9	.84	92.3	88.5	-.02	BAT46VA01	1
47	178	261	50.61	1.45	.99	-.1	.99	.0	.63	75.1	73.9	-.02	BAT47RC05	1
48	189	261	48.20	1.52	.94	-.8	.99	-.1	.68	79.7	77.4	-.02	BAT48RC06	1
49	187	261	48.65	1.50	.91	-1.1	.89	-1.1	.68	78.9	76.8	-.02	BAT49RC01	1
50	224	261	37.92	1.98	.72	-2.2	.67	-1.9	.84	92.0	88.3	-.02	BAT50RC05	1
51	232	261	34.44	2.20	.64	-2.5	.53	-2.3	.88	95.0	90.4	-.03	BAT51RC06	1
52	230	261	35.38	2.14	.67	-2.4	.58	-2.2	.87	94.3	89.9	-.03	BAT52RC07	1
53	109	261	63.25	1.33	1.14	3.7	1.15	1.5	.40	56.7	63.8	-.01	BAT53RC04	1
54	229	261	35.83	2.11	.63	-2.7	.52	-2.7	.87	93.9	89.6	-.03	BAT54RC09	1
55	232	261	34.44	2.20	.61	-2.7	.47	-2.8	.89	95.0	90.4	-.03	BAT55RC01	1
56	228	261	36.27	2.08	.70	-2.2	.69	-1.6	.86	93.5	89.4	-.03	BAT56RC08	1
57	212	261	42.10	1.76	.84	-1.4	.82	-1.2	.78	88.1	84.8	-.02	BAT57RC07	1

58	203	261	44.71	1.65	.88	-1.3	.87	-1.0	.74	85.1	81.9	-.02	BAT58RP07 1
59	231	261	34.91	2.17	.63	-2.6	.49	-2.7	.88	94.6	90.1	-.03	BAT59RC09 1
60	93	261	66.14	1.36	1.08	1.7	1.06	.6	.37	63.2	66.7	-.01	BAT60RC07 1
61	238	136	69.88	.92	1.34	2.8	2.17	7.2	.36	33.8	41.3	.01	ETQ01**01 2
62	309	137	64.20	.91	1.07	.7	1.46	3.4	.46	42.3	41.4	.00	ETQ02**02 2
63	289	134	65.41	.92	1.28	2.3	1.98	6.3	.32	42.5	41.6	.00	ETQ03**03 2
64	251	135	68.77	.92	1.17	1.5	2.01	6.4	.38	38.5	41.3	.01	ETQ04**04 2
65	301	136	64.78	.92	1.23	1.9	1.58	4.1	.44	43.4	41.5	.00	ETQ05**05 2
66	258	133	67.83	.92	1.05	.5	1.71	4.8	.31	48.9	41.5	.01	ETQ06**06 2
67	273	135	66.90	.92	1.76	5.4	3.17	9.9	.35	32.6	41.6	.00	ETQ07**07 2
68	228	134	70.52	.93	1.09	.8	1.95	6.0	.35	36.6	41.3	.01	ETQ08**08 2
69	231	133	70.11	.93	1.02	.2	1.84	5.4	.43	46.6	41.3	.01	ETQ09**09 2
70	266	133	67.25	.92	1.35	2.7	2.72	9.6	.31	39.8	41.6	.01	ETQ10**10 2
71	76	60	66.30	1.59	2.43	6.3	5.98	9.9	-.34	20.0	48.1	-.04	SRA01**01 3
72	167	105	69.44	1.13	1.04	.3	1.09	.7	.39	37.1	41.0	.00	SRA02**01 3
73	75	44	57.78	1.95	1.90	3.4	2.45	4.7	.11	36.4	48.8	-.05	SRA03**02 3
74	191	106	66.57	1.14	1.16	1.3	1.30	2.2	.40	39.6	41.9	.00	SRA04**02 3
75	70	46	61.18	1.86	2.02	4.0	3.18	6.4	.03	34.8	49.8	-.05	SRA05**03 3
76	194	106	66.19	1.14	1.53	3.8	1.70	4.7	.28	34.9	42.0	.00	SRA06**03 3
77	78	45	57.55	1.93	2.20	4.3	2.62	5.2	.23	26.7	48.4	-.06	SRA07**04 3
78	171	106	69.12	1.12	1.05	.5	1.29	2.2	.35	44.3	41.0	.00	SRA08**04 3
79	75	46	59.42	1.88	1.81	3.3	2.31	4.4	.22	21.7	49.7	-.05	SRA09**05 3
80	191	105	66.29	1.14	1.01	.1	1.02	.2	.44	44.8	41.9	.00	SRA10**05 3
81	88	45	53.68	2.00	3.01	6.1	3.22	6.5	-.03	24.4	45.5	-.06	SRA11**06 3
82	168	106	69.50	1.12	1.20	1.6	1.43	3.2	.48	31.1	40.9	.00	SRA12**06 3
83	93	46	52.67	1.99	2.86	5.8	2.99	6.0	-.01	37.0	47.1	-.06	SRA13**07 3
84	205	105	64.43	1.16	1.35	2.6	1.60	4.0	.42	33.3	42.7	.00	SRA14**07 3
85	86	45	54.48	1.98	3.49	7.1	3.82	7.6	-.12	24.4	46.1	-.06	SRA15**08 3
86	167	106	69.62	1.12	1.06	.5	1.14	1.2	.42	36.8	40.8	.00	SRA16**08 3
87	76	46	59.06	1.89	2.68	5.7	3.96	8.0	-.01	39.1	49.5	-.05	SRA17**09 3
88	71	53	70.81	1.61	1.50	2.6	2.10	4.8	.20	32.1	40.8	.00	SRA18**09 3
89	63	48	65.00	1.81	2.29	5.1	3.89	7.5	.09	35.4	50.0	-.05	SRA19**10 3
90	161	106	70.36	1.12	.99	.0	1.09	.8	.41	43.4	40.7	.00	SRA20**10 3
MEAN	191.2	205.3	50.00	1.70	1.07	.1	1.26	.8		68.7	69.9		
S.D.	60.4	81.6	14.06	.48	.58	2.7	.99	3.6		25.6	20.2		

Examination of the pilot sample predicts that the dichotomous items will underfit and the rating scale items will overfit. That appears to be the case. The poorest fit values are located in the apperception test (SRA). These items had the highest inference rating scale and were the smallest sample size in the pilot. SRA items are also the most extreme (difficult) on the battery to date, so their ptbis estimates with this sample of persons are questionable. As the pilot sample size increases and level of consistency with INTASC in the sample grows, the stability of the item fit will allow empirical analysis. For the purpose of item revision, this sample size and range was sufficient (Linacre, 1994).

Table 3 Map of Persons and Items

```
TABLE 1.2 Disposition Analysis  Spring 08 Pilot G ZOU796WS.TXT Mar  9 18:31 2008
INPUT: 335 persons  90 items  MEASURED: 335 persons  90 items  13 CATS      3.64.2
--------------------------------------------------------------------------------

      persons - MAP - items
              <more>|<rare>
  90             .  +
                    |
                    |
                    |
                    |
                    |
                    |
  80                +
                 . T|T
                    |
                 .  |  BAT23R
                    |
                 .  |
                 .  |  SRA18*
  70             .#  +  ETQ01*  ETQ08*  ETQ09*  SRA02*  SRA12*  SRA16*  SRA20*
                  . S|  ETQ04*  SRA08*
                .##  |  ETQ06*  ETQ07*  ETQ10*  SRA04*
               .###  |  BAT60R  ETQ03*  SRA01*  SRA06*  SRA10*
               .###  |S ETQ02*  ETQ05*  SRA14*  SRA19*
              .####  |  BAT53R
               .###  |  BAT13V  BAT20V  BAT21V  BAT29R  SRA05*
  60          .####  +  SRA09*
                .## M|  BAT07R  SRA17*
              .####  |  BAT16R  SRA03*  SRA07*
                .##  |  BAT44R
                 .#  |  SRA11*  SRA15*
                 .#  |  SRA13*
                  .  |
  50              .  +M BAT27R  BAT28R  BAT47R
                  . S|  BAT17R  BAT19R  BAT48R  BAT49R
                  .  |  BAT38R  BAT45R
                  .  |  BAT02R  BAT36R  BAT41R
                     |  BAT58R
                  .  |  BAT09R  BAT33R
                     |  BAT05V  BAT14R  BAT57R
  40                T+  BAT34R
                  .  |  BAT01R  BAT10O  BAT43R  BAT50R
                     |  BAT24R  BAT26V  BAT42R  BAT46V
                     |S BAT03V  BAT52R  BAT54R  BAT56R
                     |  BAT22V  BAT30R  BAT39R  BAT51R  BAT55R  BAT59R
                  .  |  BAT08O  BAT12R  BAT25V  BAT31V  BAT37R
                  .  |  BAT15R  BAT32R  BAT40R
  30              .  +  BAT04R  BAT06R  BAT11V  BAT18R  BAT35R
                  .  |
                  .  |
                  .  |
                  .  |
                     |
                     |T
  20                 +
              <less>|<frequ>
 EACH '#' IS 6.
```

Table 4
Calibration Statistics of BATS, ETQ and SRA (Separate Analyses)

Beliefs About Teaching Scale

TABLE 14.1 Disposition Analysis Spring 08 Pilot ZOU790WS.TXT Mar 9 19:49 2008
INPUT: 335 persons 60 items MEASURED: 335 persons 60 items 2 CATS 3.64.2

person: REAL SEP.: 1.35 REL.: .65 ... item: REAL SEP.: 4.72 REL.: .96

item STATISTICS: ENTRY ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEA CORR.	EXACT MATCH OBS%	EXACT MATCH EXP%	item
1	222	243	47.46	2.35	1.02	.2	.84	-.6	.87	90.9	91.4	BAT01RC03
2	197	243	57.14	1.71	1.08	.8	1.00	.1	.75	80.2	81.6	BAT02RC07
3	228	243	43.65	2.73	1.09	.5	1.45	1.4	.89	93.8	93.8	BAT03VA10
4	241	243	22.46	7.13	1.01	.3	.80	.0	.97	99.2	99.2	BAT04RC03
5	214	243	51.28	2.05	.96	-.2	.85	-.7	.84	88.5	88.2	BAT05VA02
6	241	243	22.46	7.13	.90	.1	.14	-1.5	.98	99.2	99.2	BAT06RP04
7	132	243	71.29	1.36	1.02	.6	1.04	.7	.53	62.1	63.3	BAT07RC10
8	236	243	35.40	3.89	.93	-.1	.57	-.9	.95	97.1	97.1	BAT08OR03
9	209	243	53.24	1.92	.98	-.1	.84	-.9	.81	86.0	86.2	BAT09RC08
10	223	243	46.90	2.40	1.04	.3	.99	.1	.87	91.4	91.8	BAT10OR02
11	241	243	22.46	7.13	.93	.1	.39	-.7	.98	99.2	99.2	BAT11VA01
12	235	243	36.82	3.65	.94	-.1	.61	-.9	.94	96.7	96.7	BAT12RC07
13	121	243	73.31	1.36	1.23	5.5	1.31	5.2	.46	51.0	63.0	BAT13VA08
14	215	243	50.86	2.08	1.03	.3	.97	.0	.83	88.1	88.6	BAT14RC08
15	238	243	31.88	4.56	.87	-.2	.30	-1.6	.97	97.9	97.9	BAT15RC08
16	143	243	69.23	1.38	.97	-.8	.93	-1.2	.58	66.3	64.8	BAT16RC05
17	185	243	60.36	1.58	1.07	.8	1.07	.6	.70	74.5	77.0	BAT17RC01
18	240	243	26.60	5.85	.87	-.1	.15	-1.8	.98	98.8	98.8	BAT18RC09
19	186	243	60.11	1.58	.98	-.2	.96	-.3	.72	76.5	77.4	BAT19RP10
20	123	243	72.95	1.36	1.04	1.1	1.12	2.2	.50	63.4	63.0	BAT20VA09
21	116	243	74.24	1.36	1.00	.0	1.04	.8	.49	63.4	63.2	BAT21VA03
22	233	243	39.21	3.28	.95	-.1	1.00	.1	.93	95.9	95.9	BAT22VA04
23	50	243	88.21	1.66	1.05	.5	1.22	1.6	.28	80.7	80.1	BAT23RP09
24	225	243	45.70	2.51	.98	.0	.77	-.8	.89	92.6	92.6	BAT24RP06
25	235	243	36.82	3.65	.92	-.1	.50	-1.3	.95	96.7	96.7	BAT25VA06
26	226	243	45.05	2.58	.96	-.1	.75	-.8	.90	93.0	93.0	BAT26VA10
27	179	243	61.80	1.53	1.08	1.1	1.15	1.4	.67	74.5	74.9	BAT27RP03
28	184	243	60.60	1.57	1.04	.5	1.09	.8	.70	75.7	76.6	BAT28RP10
29	116	243	74.24	1.36	1.10	2.4	1.13	2.4	.47	57.6	63.2	BAT29RC04
30	233	243	39.21	3.28	1.04	.2	.90	-.1	.93	95.9	95.9	BAT30RC05
31	236	243	35.40	3.89	1.01	.2	.80	-.3	.94	97.1	97.1	BAT31VA05
32	238	243	31.88	4.56	.90	-.1	.45	-1.1	.96	97.9	97.9	BAT32RP02
33	211	243	52.49	1.96	.98	-.1	1.09	.5	.82	86.8	87.0	BAT33RC04
34	217	243	49.97	2.14	1.03	.2	.95	-.2	.84	88.9	89.4	BAT34RP05
35	241	243	22.46	7.13	.92	.1	.20	-1.3	.98	99.2	99.2	BAT35RC03
36	197	243	57.14	1.71	1.02	.2	1.04	.3	.75	81.9	81.6	BAT36RP06
37	234	243	38.08	3.45	1.00	.1	1.13	.4	.93	96.3	96.3	BAT37RC10
38	194	243	58.00	1.67	1.11	1.1	1.19	1.4	.73	79.4	80.4	BAT38RP04
39	231	243	41.19	3.02	1.05	.3	1.03	.2	.91	95.1	95.1	BAT39RP06
40	238	243	31.88	4.56	.90	-.1	.42	-1.2	.96	97.9	97.9	BAT40RC02
41	197	243	57.14	1.71	1.00	.1	.94	-.4	.76	81.9	81.6	BAT41RC02
42	225	243	45.70	2.51	1.06	.4	1.14	.6	.88	92.6	92.6	BAT42RC02
43	222	243	47.46	2.35	.99	.0	.90	-.3	.87	90.9	91.4	BAT43RC09
44	150	243	67.89	1.39	.94	-1.2	.96	-.6	.60	68.7	66.1	BAT44RC01
45	196	243	57.43	1.69	.98	-.1	.92	-.5	.75	80.7	81.2	BAT45RC08
46	225	243	45.70	2.51	.92	-.3	.68	-1.2	.89	92.6	92.6	BAT46VA01
47	178	243	62.03	1.52	.99	-.2	.95	-.4	.69	74.1	74.5	BAT47RC05
48	189	243	59.34	1.61	.99	-.1	.95	-.4	.73	79.0	78.5	BAT48RC06
49	187	243	59.85	1.59	.95	-.5	.97	-.2	.72	79.4	77.7	BAT49RC01
50	224	243	46.31	2.45	.91	-.4	.78	-.8	.89	92.6	92.2	BAT50RC05
51	232	243	40.24	3.14	.92	-.2	.71	-.7	.93	95.5	95.5	BAT51RC06
52	230	243	42.06	2.91	.97	.0	.77	-.6	.91	94.7	94.7	BAT52RC07
53	109	243	75.53	1.36	1.07	1.7	1.13	2.2	.45	63.8	63.9	BAT53RC04

54	229	243	42.88	2.81	.89	-.4	.65	-1.1	.92	94.2	94.2	BAT54RC09
55	232	243	40.24	3.14	.87	-.4	.49	-1.6	.93	95.5	95.5	BAT55RC01
56	228	243	43.65	2.73	.96	-.1	.92	-.1	.90	93.8	93.8	BAT56RC08
57	212	243	52.09	1.99	1.08	.6	1.42	1.9	.80	86.4	87.4	BAT57RC07
58	203	243	55.30	1.80	.94	-.5	.84	-1.0	.79	84.4	83.9	BAT58RP07
59	231	243	41.19	3.02	.88	-.4	.56	-1.4	.93	95.1	95.1	BAT59RC09
60	93	243	78.56	1.39	1.01	.2	1.12	1.8	.42	69.1	66.5	BAT60RC07
MEAN	201.6	243.0	50.00	2.73	.99	.2	.87	.0		85.9	86.2	
S.D.	43.7	.0	15.17	1.53	.07	.9	.29	1.2		12.2	11.6	

UMEAN=50.000 USCALE=10.000

Experiential Teaching Questionnaire

TABLE 14.1 Disposition Analysis Spring 08 Pilot ZOU540WS.TXT Mar 9 19:48 2008
INPUT: 335 persons 10 items MEASURED: 137 persons 10 items 6 CATS 3.64.2

person: REAL SEP.: 2.13 REL.: .82 ... item: REAL SEP.: 2.22 REL.: .83

item STATISTICS: ENTRY ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEA CORR.	EXACT MATCH OBS%	EXACT MATCH EXP%	item
1	238	136	52.63	.98	1.01	.1	.99	.0	.69	44.1	42.6	ETQ01**01
2	309	137	46.08	.98	.85	-1.3	.85	-1.3	.69	51.8	43.2	ETQ02**02
3	289	134	47.56	.99	.86	-1.2	.86	-1.2	.68	48.5	43.2	ETQ03**03
4	251	135	51.39	.98	.92	-.7	.91	-.8	.66	48.1	42.5	ETQ04**04
5	301	136	46.80	.98	1.06	.5	1.07	.6	.64	45.6	43.1	ETQ05**05
6	258	133	50.20	.99	1.07	.7	1.11	1.0	.45	42.9	42.9	ETQ06**06
7	273	135	49.27	.98	1.40	3.1	1.38	2.9	.67	37.0	42.9	ETQ07**07
8	228	134	53.35	.99	.91	-.7	.90	-.9	.62	36.6	42.4	ETQ08**08
9	231	133	52.95	1.00	.86	-1.2	.87	-1.1	.68	48.9	42.4	ETQ09**09
10	266	133	49.79	.99	1.02	.2	1.03	.3	.65	47.4	42.7	ETQ10**10
MEAN	264.4	134.6	50.00	.99	1.00	-.1	.99	-.1		45.1	42.8	
S.D.	27.1	1.4	2.47	.00	.16	1.3	.16	1.3		4.8	.3	

Situational Reflection Assessment

TABLE 14.1 Disposition Analysis Spring 08 Pilot ZOU460WS.TXT Mar 9 19:47 2008
INPUT: 335 persons 20 items MEASURED: 153 persons 20 items 5 CATS 3.64.2

person: REAL SEP.: 1.48 REL.: .69 ... item: REAL SEP.: 1.40 REL.: .66

item STATISTICS: ENTRY ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEA CORR.	EXACT MATCH OBS%	EXACT MATCH EXP%	item
1	76	60	54.28	1.54	1.17	1.0	1.12	.7	.38	36.7	41.3	SRA01**01
2	167	105	50.95	1.15	.84	-1.4	.89	-.9	.58	45.7	41.9	SRA02**01
3	75	44	49.73	1.75	.66	-2.0	.65	-2.1	.49	50.0	41.9	SRA03**02
4	191	106	48.05	1.16	.91	-.7	.89	-.8	.61	46.2	43.0	SRA04**02
5	70	46	52.47	1.70	.72	-1.7	.73	-1.6	.55	39.1	40.2	SRA05**03
6	194	106	47.64	1.16	1.31	2.3	1.26	1.9	.50	42.5	43.0	SRA06**03
7	78	45	49.48	1.73	.93	-.3	.92	-.3	.59	40.0	42.0	SRA07**04
8	171	106	50.69	1.15	1.00	.0	1.02	.2	.48	54.7	41.9	SRA08**04
9	75	46	51.02	1.70	.74	-1.5	.74	-1.5	.59	43.5	40.8	SRA09**05
10	191	105	47.76	1.17	.93	-.5	.96	-.2	.53	39.0	43.0	SRA10**05
11	88	45	46.42	1.77	.75	-1.4	.79	-1.1	.68	48.9	43.4	SRA11**06
12	168	106	51.08	1.14	1.09	.8	1.09	.7	.58	39.6	41.9	SRA12**06

```
|    13     93     46   45.64    1.77| .86   -.7| .84   -.8|  .54| 47.8  43.6| SRA13**07|
|    14    205    105   45.82    1.19|1.03    .3|1.00    .1|  .64| 40.0  44.2| SRA14**07|
|    15     86     45   47.05    1.76|1.02    .2|1.01    .1|  .55| 37.8  42.0| SRA15**08|
|    16    167    106   51.22    1.14| .92   -.6| .97   -.2|  .57| 42.5  41.9| SRA16**08|
|    17     76     46   50.73    1.71|1.04    .3|1.03    .2|  .47| 37.0  40.9| SRA17**09|
|    18     71     53   52.73    1.63|1.33   1.8|1.28   1.4|  .42| 39.6  41.1| SRA18**09|
|    19     63     48   55.22    1.68|1.57   2.8|1.49   2.5|  .24| 43.8  39.0| SRA19**10|
|    20    161    106   52.02    1.14|1.02    .2|1.02    .2|  .48| 43.4  41.6| SRA20**10|
|-----------------------------------+---------+---------+-----+-----------+----------|
| MEAN   123.3   73.8   50.00    1.46| .99   -.1| .99   -.1|     | 42.9  41.9|          |
| S.D.    52.1   29.1    2.66     .28| .22   1.3| .20   1.1|     |  4.7   1.2|          |
+-----------------------------------------------------------------------------------------+
```

Figure 1 Correlation and Scatterplot of Measures of BATS, ETQ, SRA

Pearson Correlation Matrix

	BATS	SRA	ETQ
BATS	1.000		
SRA	-0.014	1.000	
ETQ	-0.054	0.566	1.000

Pairwise Frequency Table

	BATS	SRA	ETQ
BATS	243		
SRA	131	142	
ETQ	100	39	107

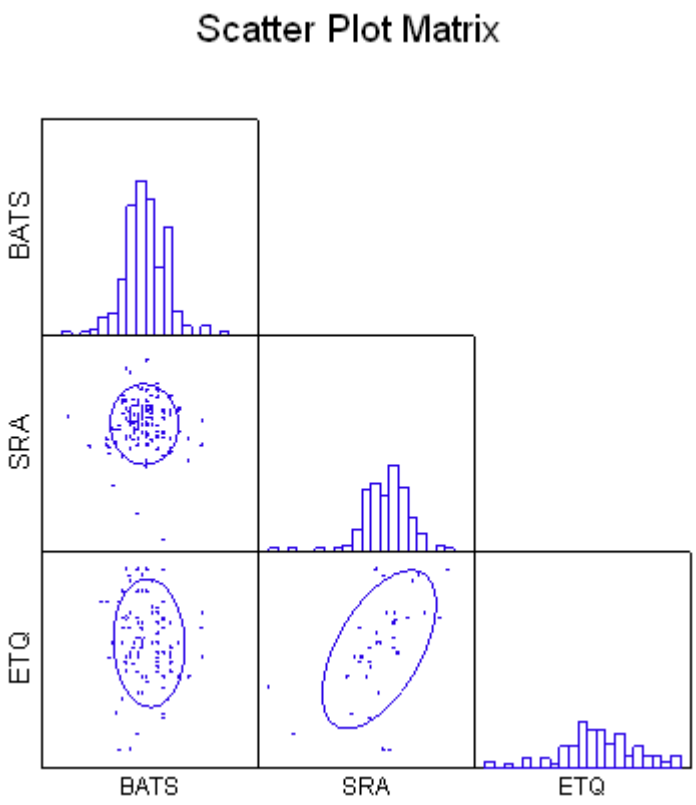


Figure 2 Threshold Analysis for Rating Scales for ETQ and SRA

ETQ

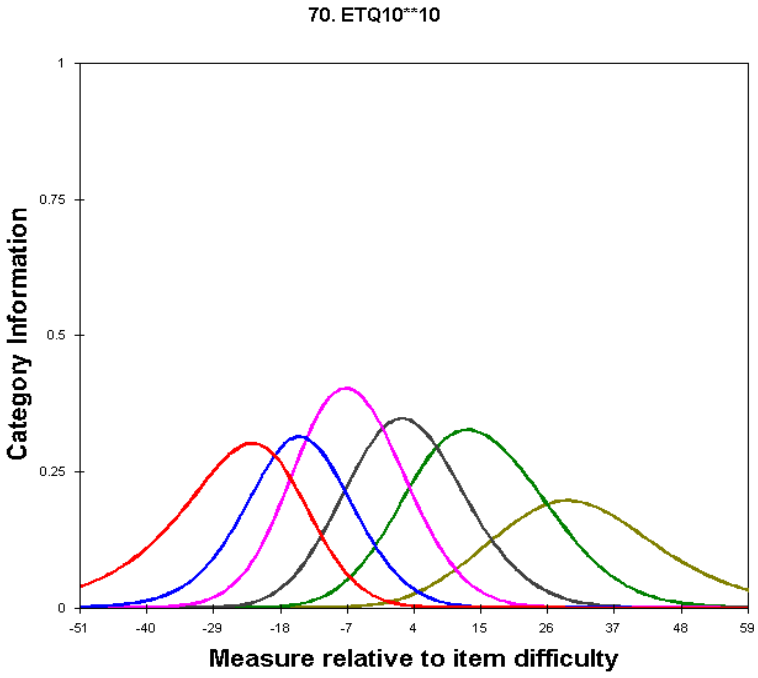
SUMMARY OF CATEGORY STRUCTURE. Model="R"
FOR GROUPING "2" item NUMBERS: 61-70

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	STRUCTURE CALIBRATN	CATEGORY MEASURE	
0	0	157	6	-14.17	-21.2	1.49	1.45	NONE	(-32.96)	0
1	1	223	9	-12.29	-10.9	.89	1.18	-18.70	-18.11	1
2	2	390	15	-7.96	-7.17	1.39	3.45	-14.46	-6.75	2
3	3	215	8	-5.88	-4.55	1.44	3.94	.12	4.51	3
4	4	67	3	-5.42	-2.23	1.47	3.12	8.28	18.03	4
5	5	5	0	-1.83	-.13	1.13	1.12	24.77	(37.03)	5
MISSING		1543	59	-8.53						

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

CATEGORY LABEL	STRUCTURE MEASURE	S.E.	SCORE-TO-MEASURE AT CAT.	----ZONE----		50% CUM. PROBABLTY	COHERENCE M->C	C->M	ESTIM DISCR	
0	NONE		(-32.96)	-INF	-25.97		17%	7%		0
1	-18.70	1.06	-18.11	-25.97	-12.25	-22.66	37%	32%	-.30	1
2	-14.46	.74	-6.75	-12.25	-1.19	-12.76	44%	74%	.23	2
3	.12	.76	4.51	-1.19	10.66	-.85	43%	30%	.45	3
4	8.28	1.27	18.03	10.66	28.42	9.75	0%	0%	.74	4
5	24.77	4.50	(37.03)	28.42	+INF	26.38	0%	0%	.98	5

M->C = Does Measure imply Category?
C->M = Does Category imply Measure?



SRA

TABLE 3.3 Disposition Analysis Spring 08 Pilot G ZOU158WS.TXT Mar 9 3:25 2008
INPUT: 335 persons 90 items MEASURED: 335 persons 90 items 13 CATS 3.64.2

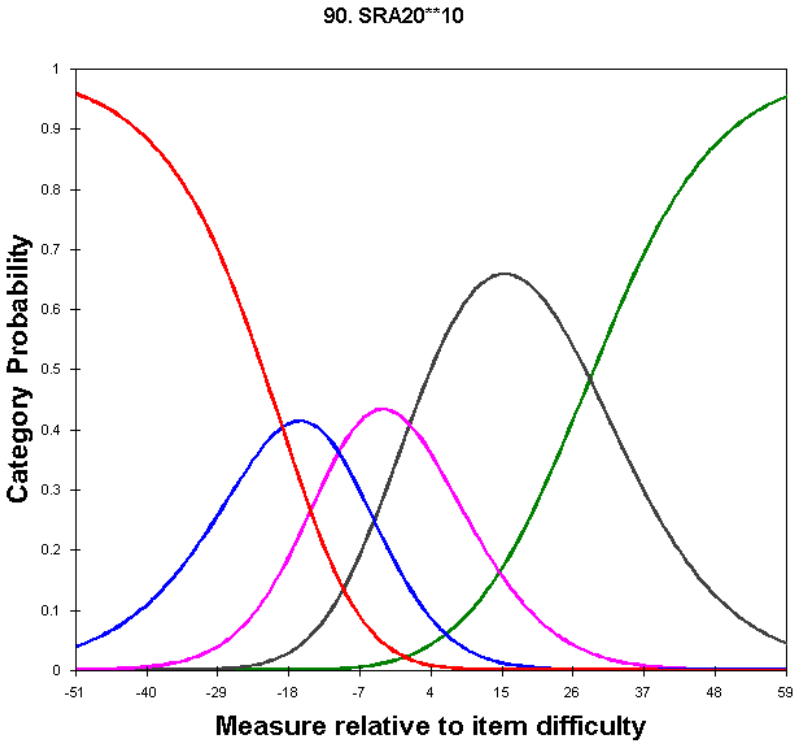
SUMMARY OF CATEGORY STRUCTURE. Model="R"
FOR GROUPING "3" item NUMBERS: 71-90

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	STRUCTURE CALIBRATN	CATEGORY MEASURE	
0	0	213	4	-10.65	-18.1	1.70	1.67	NONE	(-31.63)	0
1	1	345	7	-9.12	-9.50	1.12	1.17	-18.14	-15.81	1
2	2	455	9	-7.48	-5.28	2.21	4.31	-9.94	-3.54	2
3	3	317	6	-4.63	-2.36	1.52	2.40	-.16	14.85	3
4	4	17	0	-.23	.29	1.04	1.02	28.24	(39.65)	4
MISSING		3853	74	-3.97						

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

CATEGORY LABEL	STRUCTURE MEASURE	S.E.	SCORE-TO-MEASURE AT CAT.	----ZONE----		50% CUM. PROBABLTY	COHERENCE M->C	C->M	ESTIM DISCR	
0	NONE		(-31.63)	-INF	-24.19		9%	4%		0
1	-18.14	.89	-15.81	-24.19	-9.62	-21.23	33%	31%	-.26	1
2	-9.94	.64	-3.54	-9.62	3.88	-9.69	35%	67%	-.72	2
3	-.16	.69	14.85	3.88	29.69	2.07	56%	11%	.07	3
4	28.24	2.47	(39.65)	29.69	+INF	28.78	0%	0%	.99	4

M->C = Does Measure imply Category?
C->M = Does Category imply Measure?

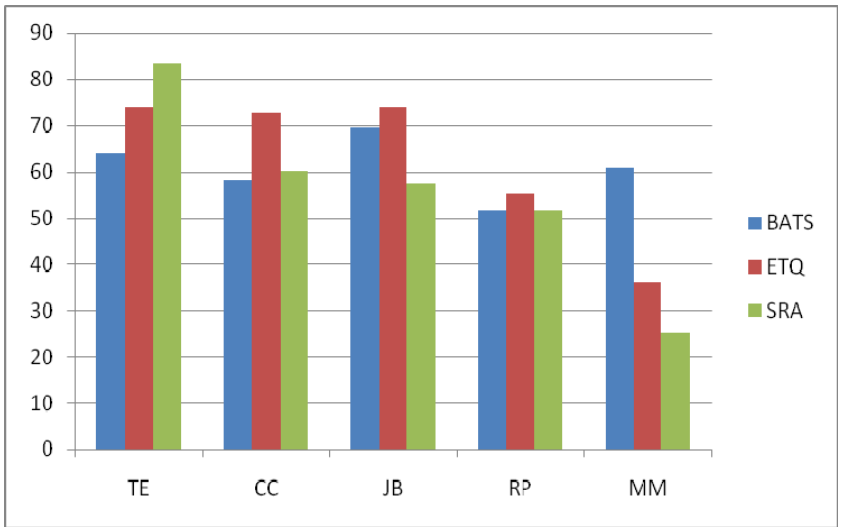


Applied Results and Utility

Individual Case Study Results Demonstrating Construct "Validity"

Here we present the results obtained on four practicing teachers and one teacher candidate. We selected these students because they show distinct differences when measured on each of the three instruments. We start with charted data on the five teachers to be described, and we will reference these charts in the discussion that follows. We reference our teachers by initials and code numbers and report their instrument scores and overall measure. Each teacher will be discussed separately. We also provide a bar chart to highlight the relationship between the scores.

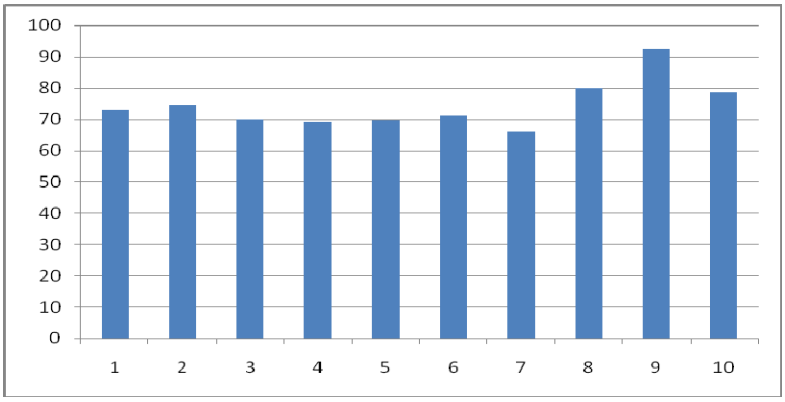
Teacher	BATS	ETQ	SRA	Overall Measure
TE (255)	63.82	74.02	83.38	73.29
MM (196)	60.82	36.18	25.02	47.44
RP (260)	51.61	55.25	51.65	52.89
CC (252)	58.19	72.59	60.25	64.02
JB (250)	69.43	74.02	57.47	67.29



Case Study #1: TE

TE is a master's degree student who is highly articulate and enthusiastic about the teaching profession. Her BATS score is among the highest in the sample, exceeded by only ___ candidates in the sample, one of whom was JB who scored correctly on three additional items. However, it somewhat underestimates her level of commitment, as is evidenced by her even

higher SRA score, which drives her overall battery score to the top of the distribution, about two logits (10 = 1 logit) higher than everyone else in the group discussed herein. TE shows a clear pattern of growth across the three instruments, achieving the highest? score in the sample on SRA, the test requiring the highest level of inference. Her writing is loaded with enthusiasm and emotion, with expressions such as "relish the opportunity," "shrinking in the chair," "wishing profusely," "peer taunting," "bear-hugs," "a child with the weight of the world on his shoulders," and so forth. At one point she writes: "This child would break my heart. This child has broken my heart." There is no doubt that, as a teacher, she sees the world through the eyes of the children she teaches, connecting deeply with them in meaningful ways. She makes connections across the standards, and provides detailed descriptions of strategies she uses. Her scores for each Principle are all consistently high with a notably higher score on the standard related to reflection and continuous improvement – as would be expected from the depth of her writing and thought on ETQ and SRA.



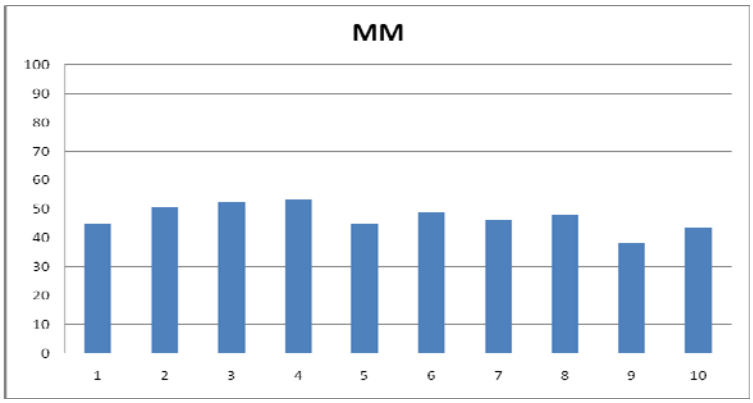
TE's measure is 73, well above the mean measure of about 59. Her fit statistic (not shown) of 1.06 is expected. We interpret these results, both judgmentally and empirically, as indicative of a model teacher.

Case Study #2: MM

MM is an undergraduate student, the only one whose initials are not real but bear the acronym of the nickname assigned by his professor – "Motorcycle Mama." He wants to teach elementary children but arrives late in class, helmet under his arm, with tattoos from wrist to shoulder, ankle to knee. When asked how he would handle a child having a tantrum (making an obscene gesture with his middle finger), MM wrote the following:

> Well, I'd force myself to transform into a fierce orange dog with large cold eyes and long white curved fangs. I'd leap onto his desk, raise the hair on my back, show my teeth, and would growl, 'Would you like to repeat that?' Would I want to be his teacher? Why not? It does not bother my self-worth any. That child can't grow up doing that, so a swift immediate response would end that behavior permanently in my eyes. (Student also drew a picture of a child sitting at a desk going "eeek" and a dog standing on the desk with a "grrrr!")

There were a number of unusual responses for MM, but, of particular note here is that his BATS score is high and his unusual views only show in the higher inference tests where he plummets to a 25 on the SRA, 2.5 logits below the mean. The ETQ score, too, is very low, and his overall battery score is among the lowest in the sample with a 27 % ile rank.

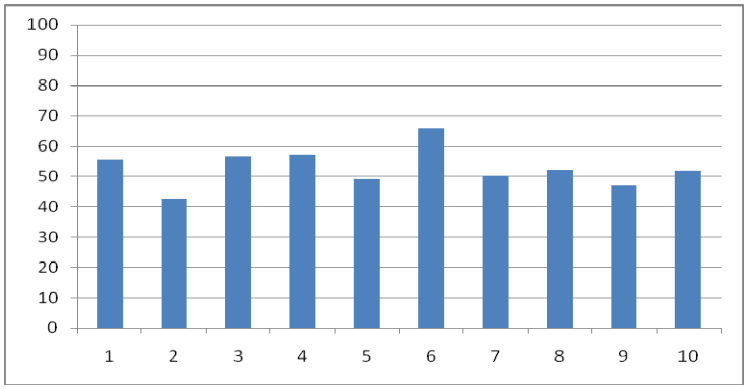


Not surprisingly, these low scores have brought all sub-scale scores on the INTASC Principles to nearly equivalent scores below the mean score of 59. BATS did not seem to impact these scores. Of particular note here is that many institutions are asking us to allow them to use BATS by itself. We have resisted because we suspected that it had the potential to underestimate the highest affect and overestimate the lowest affect students. These two case studies provide evidence to support our conclusion. It is only with the three item formats that we obtain a more stable picture of the teacher. In MM's case, we would recommend that the college consider suggesting that he leave the program. This may not be necessary, since he has failed his courses this semester, having had a fist fight in the local bar and ended up in the hospital during exam week. The scores were obtained before MM's behaviors surfaced and until recently, he had a A-GPA.

Case Study SP: RP

RP also is a master's degree student who teaches calculus and is a low scoring teacher. Unlike MM, he is certified. His responses are terse and without emotion. There is little variability in his sub-scores on the Principles, so he is generally low on the affective side. This may be acceptable in the calculus classroom. It may be useful to know him, though, as one who is not likely to go the extra mile for his students or his fellow teachers. For example, rather than the typical nurturing reSPonse of seeking social services, shoes, and foods for the impoverished Asian child, he wants to befriend and chat with him; he fails to detect the lack of collaboration among children preparing for a debate. The frantic teacher who can't grade her work is just "a bit frustrated. A little more disturbing, though, is his statement that he does not need to plan his lessons because he has taught for three years, so he plan in the shower before school. When children do not do well on tests, it is because the county requires too much work of them; he bears no responsibility for the results. His BATS responses indicate that he has no interest in continuing to learn about his field, that children need to stay on task and behave, and that there is no need to reflect about instructional successes or failures. He would not join in on a curriculum planning committee and prefers to plan alone because he is the only calculus teacher in the school. Given some of these statements and some of his BATS responses, it is not surprising

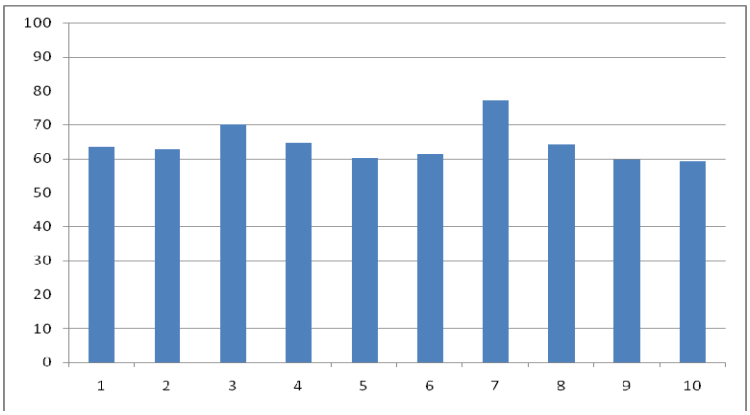
that his scores are evenly matched across the three instruments as well as the sub-scale scores on the Principles, as reported below.



His scores all tend to be lower than the mean of 59, sometimes as much as one logit. He has a good fit statistic of .99. In terms of the future, it would be useful to keep an eye on him in the classroom and make sure that his students' social and emotional needs are met and that his colleagues are not alienated.

Case Study #4: CC

CC is also a master's student and very bright. She is the type of student who does not stop thinking and contributing. She continually returns to discussion boards to post more, and she reads incessantly. She scored one of the only 5s on the ETQ because of an extreme commitment to reading. (She would not be happy if there was even one day that she did not learn something new.) The reading, though, was not related to teaching or her content area. She SPecifically expressed a desire to lobby with other teachers to read SPecific materials, learning about Native Americans and sustainability, for example. In the picture asking about a teacher's impact on math with a pointed finger, she ignored the component of the question addressing content. Her BATS scores indicate some issues with creativity and a desire to plan alone, although she will be a good planner (bar #7). In the SRA she was not able to put herself into the place of the teacher or the student on several occasions, remaining aloof from the pictures, as she remained aloof from the content. She wanted to mediate a planning session rather than participate in it; yet, she showed much empathy for children in other questions, sensitive to diverse learning needs and to developmental aspects. She had one of the most unusual responses about the inappropriately dressed teacher, not commenting on it in terms of children. These anomalies in her responses indicate that she is responding honestly but needs some counselling or mentoring in the school to check on some of the beliefs that are inconsistent with effective schooling while supporting and reinforcing those that are. Her sub-scale scores on the Principles do not show much variability, except for the peak in Planning (#7).



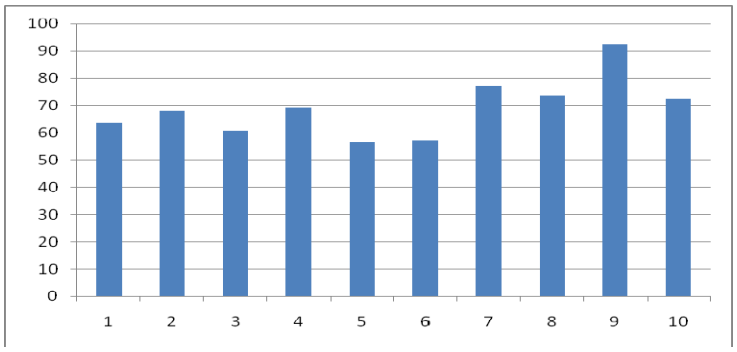
CC's battery score is well above the mean at 64, and her scores are also somewhat misfitting. She should have some assistance from a peer teacher or mentor to make sure she stays on track, focusing on the content she is supposed to teach and engaging more with her colleagues and students.

Case Study #5: JB

JB is an anathema. She is a master's student who has written some of the most bizarre responses. In response to the picture of the impoverished Asian child, she wrote:

> Any teacher that has had the proper education to be qualified to be a teacher should be able to teach this child. If this child were in my class I would inquire about his home conditions because he does not seem to have a proper home life based on the way that he has come dressed to school. Again, he would not even make it onto the bus or through the front door for that matter of our school because he is against code. (He has no shoes on and his shirt is un-buttoned.) While I do not believe that he does not deserve the same education. I do believe that the law is law and he would be against dress code.

Regarding the scantily clad teacher, JB wanted to know where her nametag was, since she would need that for security purposes in the school. She did eventually note the inappropriate attire. There are repeated references to following instinct over theory and a generally attitude toward the children. In the picture where children are chatting, she sees them as taunting each other and the teacher (who is smiling) having to discipline them. Yet, at other times, she provides sensitive comments about the "distinguished profession of teacher" and that "Everyone smiles in the same language." So, while she would bar a poor child from the school, she sees all children as smiling in the same language. Because of issues such as these, her scores fluctuate widely, and she has one of the most misfitting scores in the sample. The deep thinking patterns evident in many of the responses confirm the high point on her sub-scale scores – her high point is continuous improvement and reflection (bar #9), but her lack of respect for children shows as a low point on #5 (learning environment and motivation) and #6 (communication).



JB scores very well on BATS, the highest in this set of case studies. Her intelligence helps her guess her way through many of the items and write effective examples in the ETQ, but the more questionable attitudes are caught in the more sensitive instrument (SRA). Her battery score remains high at 67, but it is the fit statistic that shows the issues. She has much potential as a teacher, but she would be well-served by a friend in the school to soften her and a few days visiting the homes of minority and low economic scale children.

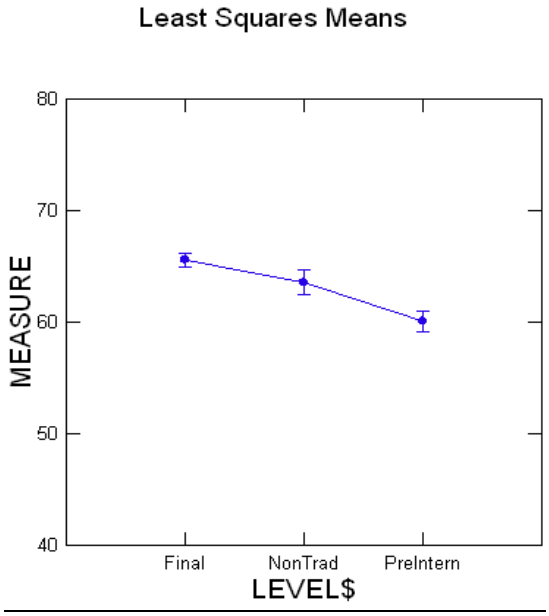
Measuring dispositions has become a very difficult issue for many institutions. Many institutions are also very focused on using disposition assessments to look at broad attributes such as lifelong learning or at professional behaviours such as punctuality and proper dress but forgetting to look at some of the important dispositions in the INTASC Principles that can lead them to improving those fundamental attitudes that teachers need to have to ensure that they do what we have taught them to do because they want to do so. More frightening is the current movement away from standards-based assessment in favour of assessing morality. There is a solid rationale behind the INTASC Principles and the inclusion of dispositions in assessment systems.

The INTASC dispositions are a complex construct best measured using different types of instruments designed by measurement professionals in collaboration with teacher education faculty and other stakeholders. Such instruments need to take into consideration the importance of increasing levels of inference, so that a progressive set of measures helps us to build confidence in our decisions. Dispositions are not only measurable, but they can be measured with results that are both valid and reliable. Rasch measurement provides a vehicle to accomplish this.

Aggregated Results Demonstrating Construct Validity

It is expected that, over time, and with experience and exposure to teaching, teacher candidates and teachers would construct meaning for their values and show growth in their commitment to the Principles of teaching. Stated differently, during their teacher preparation programs, students appear to be acquiring increasing commitment to the skills of teaching, as operatinally defined in the INTASC Principles. This graph is also confirmed by an ANOVA and the points plotted are the least-squared means from that analysis. Also, the variability of teacher candidates as final interns is the smallest. The variablility of the nontraditional (alternative certification) candidates is the greatest. As students progress to final internship, they become more consistent and more homogeneous in their consistency with INTASC principles. This is evident in the standard

errors indicated in the points plotted above by the brackets above and below each plotted point. Figure 3 shows least squares means for three populations with increasing levels of commitment.



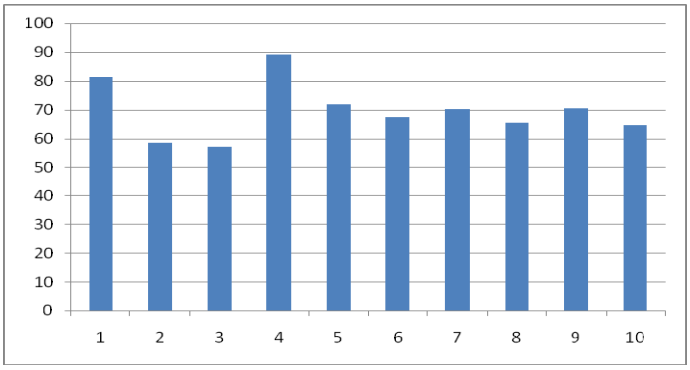
Even with small samples, over the Junior and Senior years of teacher education, these results indicate a change in consistency with INTASC principles that is meaningful.

Sub-Scores Useful for Targeted Diagnostics

It is possible to create sub-scores for the sub-constructs – each of the INTASC Principles. This is both important and useful, since many teachers are likely to be committed to some Principles and not others. They might like to plan but not to assess, for example. Figure 4 shows the DPF measure for each Principle for an individual student. This sample indicates that the student (GC001) values learning more about the discipline taught (DPF of 81.23) than working with children of different stages of learning and development (DPF of 58.47). Faculty would need to monitor his/her adjustments for these children.

item	OBSERVATIONS		BASELINE		DPF	DPF	DPF	person	
CLASS	COUNT	AVERAGE	EXPECT	MEASURE	SCORE	MEASURE	S.E.	Number	Name
01	8	1.50	1.25	68.30	.25	81.23	9.78	1	xx001*****
02	8	1.13	1.37	68.30	-.25	58.47	6.42	1	xx001*****
03	7	.86	1.06	68.30	-.20	57.21	8.17	1	xx001*****
04	6	1.33	1.07	68.30	.26	89.10	19.11	1	xx001*****
05	7	1.43	1.34	68.30	.09	71.80	7.95	1	xx001*****
06	7	1.29	1.31	68.30	-.02	67.51	6.97	1	xx001*****
07	7	1.14	1.09	68.30	.05	70.19	7.40	1	xx001*****
08	7	1.00	1.07	68.30	-.07	65.56	7.39	1	xx001*****
09	8	1.13	1.06	68.30	.07	70.54	6.49	1	xx001*****
10	8	1.00	1.12	68.30	-.12	64.67	6.19	1	xx001*****

Figure 5 shows the manner in which we report these results to students in a bar chart, so they quickly grasp their weaker areas.



1. Content Knowledge
2. Development and Learning
3. Diversity
4. Critical Thinking
5. Learning Environment and Motivation
6. Communication
7. Planning
8. Assessment
9. Reflection and Continuous Improvement
10. Collegiality and Professionalism

Rater Reliability

Both classical and Rasch analyses of rater effect have been computer. Both confirmed acceptable levels. Figure 6 provides the FACETS output for this analysis, conducted PRIOR to redesign and correction of the rubrics. Subsequent analyses should indicate even less difference among raters. This FACETS analysis show of the rater effects for the ETQ ratings from one institution. The rulers can be reversed for interpretation. In this case, Students 17 and 6 are the most consistent with INTASC. Meanwhile, Learning Environment & Motivation were the most difficult for students to agree with INTASC on the ETQ questionnaire. Judge number 1 was the "harshest" and had the highest standard for consistency with INTASC as a rater (even though the judges were relatively close). Krathwohl stages range from a rating of 1 (Receiving) to 5 (Characterization). The Rasch model produces a linear transformation of Cohen's Kappa except the expected result is 0 instead of 1. In this case the Rasch Kappa values are:
Judge 1's inter-rater Kappa = .26, Judge 2's inter-rater Kappa = .20, Judge 3's inter-rater Kappa = .28. Since 0 is the expected value, this is a moderate, but not excellent set of values. It is likely that rater training or rubric improvement would result in better consistency.

```
Experiential Teacher Questionnaire  02-01-2008 17:55:27
Table 6.0  All Facet Vertical "Rulers".

Vertical = (1A,2A,3A) Yardstick (columns,lines,low,high)= 60,15,-1,1
-------------------------------------------------------------------------------------------
|Measr|+candidate      + ETQ Question INTASC                                  |+Rater|KRATH|
-------------------------------------------------------------------------------------------
+   1 +                +                                                     +      + (5) +
|     | 17  6          |                                                     |      | ---  |
|     | 18             |                                                     |      |      |
|     |                |                                                     |      |      |
|     |                |                                                     |      |      |
|     | 4              |                                                     |      |      |
|     |                |                                                     |      |      |
|     |                |                                                     |      |      |
|     |                | Learning Environment & Motivation                   |      |      |
|     | 22             |                                                     |      |      |
|     |                |                                                     |      |  3   |
|     | 9              | Development & Learning                              |      |      |
|     |                | Collegiality & Professionalism, Planning            |      |      |
|     | 21  23  7      |                                                     |      |      |
|     | 12  3          |                                                     | 1    |      |
*   0 * 11             * Diversity                                           * 2    *      *
|     |                |                                                     | 3    | ---  |
|      | 19  20  5   8   | Reflection & Continuous Improvement                    |      |     |
|     | 14             | Assessment, Critical Thinking                       |      |      |
|     |                | Communication, Content Knowledge                    |      |      |
|     | 10  15  16     |                                                     |      |      |
|     |                |                                                     |      |      |
|     |                |                                                     |      |      |
|     |                |                                                     |      |  2   |
|     |                |                                                     |      |      |
|     | 13             |                                                     |      |      |
|     |                |                                                     |      |      |
|     |                |                                                     |      |      |
|     |                |                                                     |      |      |
|     |                |                                                     |      |      |
+  -1 +                +                                                     +      + (0) +
-------------------------------------------------------------------------------------------
|Measr|+candidate      |+Question                                              |+Rater|KRATH|
-------------------------------------------------------------------------------------------
```

FACETS analysis of the Rasch model allows instant and easy understanding of rater effects. It is possible to have an excellent inter-rater reliability and fail to show that judges are too lenient or too harsh.

<u>DIF Analyses to Compare Institutional Consistency with INTASC</u>

We compared three institutions that are very similar in consistency with INTASC principles with a few (possibly random-Type I) “significant” differences. On the other hand, one institutions differs from the other two with regard to Krathwolh development at the Receiving, Valuing, and Organizing level. *Whether this is “good” or “bad” is unknown*, but it seems that the type of student, the development of the student, and the differences in groups of student could lead to improvement if faculty are aware of the student dispositions. Without measures aggregated results, important information will never be known nor addressed

TABLE 33.1 Disposition Rating Analysis 08 ZOU143ws.txt Feb 1 10:21 2008
INPUT: 228 persons, 90 items MEASURED: 154 persons, 88 items, 12 CATS 3.57.2

Institution differences by Krathwolh

person CLASS	DIF MEASURE	DIF S.E.	person CLASS	DIF MEASURE	DIF S.E.	DIF CONTRAST	JOINT S.E.	t	d.f.	Prob.	item CLASS
#1	.93	.54	#2	2.95	1.15	-2.02	1.27	1.60	I#2	.1098	RC
#1	.93	.54	#3	-6.81	1.20	7.73	1.31	-5.89	I#2	.0000	RC
#2	2.95	1.15	#3	-6.81	1.20	9.75	1.66	-5.88	I#2	.0000	RC
#1	-.49	.90	#2	2.88	1.90	-3.36	2.10	1.60	I#2	.1096	RP
#1	-.49	.90	#3	-.55	1.73	.07	1.95	-.03	I#2	.9730	RP
#2	2.88	1.90	#3	-.55	1.73	3.43	2.56	-1.34	538	.1816	RP
#1	.31	1.02	#2	3.95	2.08	-3.64	2.32	1.57	I#2	.1163	VA
#1	.31	1.02	#3	-5.25	2.14	5.56	2.37	-2.35	I#2	.0190	VA
#2	3.95	2.08	#3	-5.25	2.14	9.20	2.98	-3.08	493	.0022	VA
#1	-4.21	3.67	#2	1.93	7.28	-6.14	8.15	.75	232	.4519	OR
#1	-4.21	3.67	#3	7.78	4.21	-11.99	5.59	2.15	230	.0329	OR
#2	1.93	7.28	#3	7.78	4.21	-5.85	8.41	.70	88	.4886	OR

Institution differences by INTASC Principle

person CLASS	DIF MEASURE	DIF S.E.	person CLASS	DIF MEASURE	DIF S.E.	DIF CONTRAST	JOINT S.E.	t	d.f.	Prob.	item CLASS
#1	.71	.80	#2	-1.89	1.73	2.59	1.90	-1.36	838	.1734	01
#1	.71	.80	#3	-1.01	1.45	1.72	1.65	-1.04	846	.2991	01
#2	-1.89	1.73	#3	-1.01	1.45	-.87	2.25	.39	326	.6987	01
#1	.06	.86	#2	1.51	1.83	-1.45	2.02	.71	843	.4749	02
#1	.06	.86	#3	-1.26	1.54	1.32	1.76	-.75	851	.4557	02
#2	1.51	1.83	#3	-1.26	1.54	2.76	2.39	-1.16	328	.2487	02
#1	-.53	.81	#2	-.39	1.80	-.14	1.97	.07	842	.9453	03
#1	-.53	.81	#3	1.88	1.45	-2.40	1.66	1.45	851	.1481	03
#2	-.39	1.80	#3	1.88	1.45	-2.27	2.31	.98	327	.3260	03
#1	-.02	.79	#2	1.18	1.68	-1.20	1.86	.64	725	.5192	04
#1	-.02	.79	#3	-1.06	1.61	1.04	1.79	-.58	724	.5627	04
#2	1.18	1.68	#3	-1.06	1.61	2.23	2.32	-.96	271	.3370	04
#1	.13	.78	#2	.25	1.70	-.12	1.87	.06	834	.9505	05
#1	.13	.78	#3	-.73	1.48	.86	1.68	-.51	841	.6077	05
#2	.25	1.70	#3	-.73	1.48	.98	2.26	-.43	327	.6655	05
#1	.04	.81	#2	1.13	1.84	-1.09	2.01	.54	833	.5889	06
#1	.04	.81	#3	-.98	1.54	1.02	1.74	-.59	844	.5564	06
#2	1.13	1.84	#3	-.98	1.54	2.11	2.40	-.88	325	.3802	06
#1	.76	.78	#2	-2.84	1.54	3.60	1.72	-2.09	823	.0370	07
#1	.76	.78	#3	-.18	1.55	.94	1.73	-.54	819	.5884	07
#2	-2.84	1.54	#3	-.18	1.55	-2.66	2.18	1.22	316	.2236	07
#1	-1.07	.77	#2	2.78	1.57	-3.85	1.75	2.20	833	.0281	08
#1	-1.07	.77	#3	1.62	1.52	-2.69	1.70	1.58	841	.1139	08
#2	2.78	1.57	#3	1.62	1.52	1.16	2.19	-.53	328	.5961	08
#1	-.16	.78	#2	.31	1.66	-.47	1.83	.26	835	.7961	09
#1	-.16	.78	#3	.32	1.54	-.48	1.72	.28	844	.7805	09
#2	.31	1.66	#3	.32	1.54	-.01	2.26	.00	327	.9975	09
#1	.02	.76	#2	-1.52	1.70	1.54	1.87	-.83	842	.4080	10
#1	.02	.76	#3	1.03	1.48	-1.01	1.66	.61	849	.5425	10
#2	-1.52	1.70	#3	1.03	1.48	-2.56	2.25	1.13	327	.2574	10

.

Data for Individual Diagnosis

The Rasch results are useful for showing why students obtain a “misfitting” score. In this example, we see a student who had a number of unexpected inconsistent answers according to the Rasch model’s estimate of the student overall pattern. Here are a subset of the unexpectedly inconsistent responses. Is it possible that this student has a sensitivity to backgrounds different

from their own? **This pattern would likely never be observed without a statistical analysis of calibrated assessment items!**

- I really enjoy meeting the parents and family of my students and seeing them inadvertently outside of the school setting.
- Teacher's who express their personal values should know that it makes some students with different backgrounds uncomfortable.
- I constantly ask students to describe their point of view so that I can understand their different perspectives.

The student's results are presented in Figure 6.

```
NUMBER - NAME ----------------- MEASURE - INFIT (MNSQ) OUTFIT - S.E.
                                  60.60     1.8   B      2.8     2.42
10     30      50      70      90     110     130     150
|------+-------+-------+-------+-------+-------+-------|  NUM   item
                    1 .2.                                  61  ETQ01  01
                    1.2.                                   68  ETQ08  08
                    1.2.                                   69  ETQ09  09
                    1    .3.                               64  ETQ04  04
                    2                  (5)                 67  ETQ07  07
                    2  .3.                                 63  ETQ03  03
                    0      .1.                             20  BAT20VA09
                    2  .3.                                 62  ETQ02  02
                    0     .1.                              21  BAT21VA03
                    0   .1.                                 7  BAT07RC10
               .0.  1                                      16  BAT16RC05
               .0.  1                                      44  BAT44RC01
             .0.    1                                      27  BAT27RP03
            .0.     1                                      38  BAT38RP04
            .0.     1                                      19  BAT19RP10
            .0.     1                                      28  BAT28RP10
           .0.      1                                      45  BAT45RC08
           .0.      1                                      36  BAT36RP06
           .0.      1                                      49  BAT49RC01
10     30      50      70      90     110     130     150
```

Conclusions and Importance

The INTASC Principles provide a basis for construct and content validity. If we apply appropriate measurement techniques and a useful evaluation design model, dispositions can be measured with results that are both valid and reliable. The INTASC dispositions are a complex construct best measured using different item structures. Dispositions present, by their very nature, some special assessment challenges. They are difficult to operationally define and difficult to assess without respondents faking their responses. The importance of this work is in its potential to provide a new application for Rasch measurement and an opportunity to measure a construct that can have a major impact on children's learning. Clearly, the teacher who hates to teach is likely to do a terrible job.

The challenge of this battery of assessments concerns the scaling of different item types on the same construct. Because the range and characteristics of dichotomous items, rated

writing, apperception, focus groups, and observations are so different, care is taken to both justify the construct map and scoring rubrics while overlooking predictable misleading trends in fit statistics. The use of this variety of assessment instruments is important, since each instrument type provides a unique perspective on the teacher's commitment to the standards of teaching.

A scaled set of measures is desirable since diagnosis of individual teacher commitments is the primary target. But also important are data that can be aggregated at the program or unit levels to refine curricula, allowing for growth across candidates over time. In this paper, we have looked not only at the DAATS battery from the perspective of Rasch modeling but also from its other psychometric properties and its practical utility. We concluded the following:

1. Evidence of construct validity when the three instruments are combined is present based on a judgmental and empirical analysis of individual cases.

2. The fit statistic is sensitive to anomalies in student responses, allowing users to review teacher responses that highlight areas needing improvement.

3. BATS is subject to faking, as would be anticipated in a survey instrument. It can overestimate and underestimate the construct when looked at in isolation. Use of BATS as a single measure should be avoided, despite many institutional requests for a "quick fix" to the accreditation requirement.

4. Teachers with more experience improve their commitments to teaching skills.

5. Sub-scores for each Principle are useful for targeted diagnostics.

6. Rubrics are sufficiently well constructed as to yield reliable results.

7. It is possible to determine areas of consistency and inconsistency across programs using DIF analysis, and these results can be used for program improvement.

8. Keyforms analyses are useful for identifying patterns of misfitting data for individual teachers, showing results that might otherwise remain undetected.

This paper represents the proverbial tip of the iceberg. With carefully constructed instruments and data analysis, disposition assessment has the possibility of informing teacher education in ways that are self-correcting for students, program revealing for faculty, institutionally meaningful for accreditation, and job-related predictions for teacher certification and future performance on the job.

Selected References

American Educational Research Association, American Psychological Association, and National Council of Measurement in Education (1999). *Standards for educational and psychological testing.*

Anderson, L.W. (1988a). Attitudes and their measurement. In J.P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook. (*pp. 421-426). Oxford, England: Pergamon.

Anderson, L.W. (1988b). Likert Scales. In J.P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook. (*pp. 427-428). Oxford, England: Pergamon.

Anderson, L.W. (1988c). Guttman Scales. In J.P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook. (*pp. 428-430). Oxford, England: Pergamon.

Darling-Hammonds, L. (2000). Teacher quality and student achievement: a review of state policy evidence. *Education Policy Analysis Archives, (8)1,* Retrieved April 21, 2005 from http://epaa.asu.edu/epaa/v8n1/

Elder, C., McNamara, T., & Congdon, P. (2004). Rasch techniques for detecting bias in performance assessments: an example comparing the performance of native and non-native speakers on a test of academic English. In *Introduction to Rasch Measurment,* Smith, E. & Smith, R. (eds). Maple Grove, MN, JAM.

Flores, J.G. & Alonso, C.G. (1995). Using focus groups in educational research. *Evaluation Review,* 19 *(1),* 84-101.

Council of Chief State School Officers (1992). *Model Standards for Beginning Teacher Licensing, Assessment, and Development: a resource for state dialogue.* Washington, D.C.: Author. Retrieved April 24, 2005, from http://www.ccsso.org/projects/Interstate_New_Teacher_Assessment_and_Support_Consortium/Publications

Daniel, L. and O'Farrell, C. (2005, February). Decisions about dispositions using progressive measures at a large state urban institution. In Wilkerson, J. (Chair), *Measuring teacher dispositions with credibility: a multi-institutional perspective,* Symposia presented at the Annual Meeting of the American Association of Colleges of Teacher Education, Washington D.C.

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions,* 7:4, 328.

National Council for Accreditation of Teacher Education (2002). *Unit standards.* Washington, D.C.: Author. Retrieved April 24, 2005 from http://www.ncate.org/institutions/standards.asp?ch=8

Schulte, L., Edick, N., Edwards, S., & Mackiel, D. (2004). The development and validation of the Teacher Dispositions Index. *Essays in Education, 12, Winter 2004.* Retrieved April 15, 2005 from http://www.usca.edu/essays/vol12winter2004.html

Smith, E.V. & Smith, R.M. (2004). *Introduction to the Rasch model: theory, models and applications.* Maple Grove, MN, JAM.

Stalling, J.A. & Mohlman, G.G. (1988). Classroom Observation Techniques. In J.P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook. (*pp. 469-173). Oxford, England: Pergamon.

Phillips, J.L. (1988). Semantic Differential. In J.P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook. (*pp. 430-432). Oxford, England: Pergamon.

Thurstone, L.L. (1928). Attitudes can be measured. *American Journal of Sociology.* 33, 529-554.

Toldson, I. and Jacobs, R. (2005, February). Analyses and interventions with candidates identified as having dispositional deficits at an HBCU. In Wilkerson, J. (Chair), *Measuring teacher dispositions with credibility: a multi-institutional perspective,* Symposia presented at the Annual Meeting of the American Association of Colleges of Teacher Education, Washington D.C.

Walsh, J.J. (1988). Projective Testing Techniques. In J.P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook. (*pp. 432-436). Oxford, England: Pergamon.

Wolf, R.M. (1988a). Rating scales. In J.P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook. (*pp. 478-481). Oxford, England: Pergamon.

Wolf, R.M. (1988b). Questionnaires. In J.P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook. (*pp. 496-500). Oxford, England: Pergamon.

Wilkerson, J.R. & Lang, W. S. (2007a). *Assessing teacher competency: Five standards-based steps to valid measurement using the CAATS model.* Thousand Oaks, CA: Corwin.

Wilkerson, J.R. & Lang, W. S. (2007b). *Assessing teacher dispositions: Five standards-based steps to valid measurement using the DAATS model.* Thousand Oaks, CA: Corwin.

Appendix A

Scoring Guide for INTASC Principle 1 – ETQ and SRA

NOTICE

NO PARTS OF THIS DOCUMENT MAY BE REPRODUCED, DISTRIBUTED, OR ADAPTED WITHOUT THE EXPRESS WRITTEN PERMSSION OF THE AUTHORS.

ETQ 1: How have you kept abreast of current developments in your field? For example, did you attend any workshops, subscribe to any journals, read or buy a new book? If so, describe in one to two sentences something you learned and the source.

- **Unaware**: This teacher might say that s/he has done nothing to learn more about the content and is satisfied with what is being learned in school. S/he might discuss learning that is off-topic (e.g., safety, testing, salaries), completion of degree requirements, storage of materials and files, attendance at staff meetings that are not described or related to improving content knowledge, or discussions with colleagues about unrelated things (e.g., social topics).

  Example #1: I have a subscription of NEA National Educator Association journal. I learned how Florida teacher salaries compare with that of other states that definitely we are not in this profession for the money and how many teachers in our country are struggling with payment of stifling student loans compared to what they earn as new teachers. (245)

  Example #2: I have kept abreast of current developments in the reading filed by completing all of the competencies for the reading endorsement. I am now waiting on the DOE to add the endorsement to my certificate! (228)

- **Receiving:** This teacher is willing to learn more. S/he might read a few books, purchase books that are not yet read, read a book from a course and describe its contents in a matter-of-fact way, or attend meetings discussing what was learned but without enthusiasm. Typically, s/he cites a limited set – one book, one book and one workshop, a couple of one or the other. There is often a promise that s/he will do something in the future.

  Example #1: Yes I have been reading books that a colleague gave to me about children with special needs and how to identify and help them. (230)

  Example #2: I am not yet a teacher but when I become one I will certainly keep abreast of current developments. I will subscribe to teaching journals attend

workshops and read new books in order to stay on top of the latest developments. (155)

- **Responding:** This teacher has expressed a clear desire to learn and focuses on-topic, as evidenced by use of something learned or multiple opportunities to learn taken. S/he might comment on a strategy learned in a book or a workshop and how it was (or would be) used or provide a list of books read, workshops attended, etc. S/he may even go out of the way to find sources. One strategy might be tried with success noted.

  Example #1: I usually search the internet and read different articles about the subject. I like to know how other teachers teach the subject. I like to learn from great teacher with experience. I go to different sites on my subject which is math. I haven't subscribed to any journals yet. (146)

  Example #2: I discuss lesson with teacher that have currently attended workshops in reading and math. I have attended a workshop on science and using small and whole group activities that allow student to move about the classroom. I don't subscribe to any journals but I do read some journals that the school library receives. I buy books often in addition to games and manipulatives to use for student interactions. (78)

- **Valuing:** This teacher engages actively in reading, attending workshops, seeking current information, etc. There is clear enthusiasm about learning expressed. The teacher sees the utility, gravitates toward it, and uses strong words such as "fascinating," "memorable," or "enjoyable." Strategies discovered and used are more specific, with results cited. There must be either enthusiasm or direct ties to students and learning to achieve this level – preferably both.

  Example #1: When I become a teacher I do plan to attend workshops, subscribe to a journal, join an organization and read/buy books. Today I looked at several online resources pertaining to concept mapping. This was very fascinating because it was entirely new to me! I can see where concept maps would enable a visual learner to grasp a lot of technical information in an organized yet enjoyable and memorable way! (161)

  Example #2: In order to keep updated on new developments in my current field I have read a variety of books. Because I am working in a first grade classroom the reading process is a vital component of student's curriculum. spalding phonics is used in my school. Modeling is an effective strategy that helps students develop and memorize new material. The Writing Road to Reading by Romalda Bishop spalding has helped me gain knowledge of spalding phonics. Interacting with your students helps them to understand and develop new techniques faster and effectively. (112)

- **Organizing:** This teacher actively engages in professional development extensively and has integrated learning into teaching that cuts across the INTASC Principles. S/he might

make connections to Critical Thinking (#4), Human Development (sp), Diversity (#2), Motivation (#5), Assessment (#8), or others.

> Example #1: I attended a workshop on expository and narrative writing. They taught us how to teach students the correct way to write these two types of writing. One way was giving the students an example of each writing. We could read them a story as an example or as a class make us our own. Then the students would write one of their own. Another method of keeping up with current developments in my field is reading books. A teacher let me borrow a book called Classroom Instruction that Works. In this book it explained tips of classroom instruction that works. One thing that I learned was how to provide evidence that students have learned certain knowledge; I can do this by providing feedback. After teaching a lesson it is important to give students information about how well they are doing and what they can do to improve their performance. This book also taught me to provide appropriate recognition. It is crucial to recognize each student's efforts, especially when those efforts lead to success. specific praise is vital in teaching. (103)
>
> Note integration of INTASC 1, 3, 4, 6, and 8
>
> Example #2: I currently read articles, pursue my education in obtaining my master's degree, and also attend conferences. This week I am attending the Florida Education technology Conference held in Orlando. I will use this to give feedback to my administrators about what new technology we could introduce into our classrooms. Additionally, I do peruse articles and occasionally use them in my classroom. I have integrated a nutrition and fitness theme into my classroom after reading some articles and information about healthy eating habits and a teacher's influence on them. I addition I am reading textbooks and articles quite often through the courses I take. These have influenced anything form parent communications to inquiry learning in my classroom. (254)
>
> Note integration of INTASC 1, 7, 10

- **Characterizing:** This teacher is so enthusiastic about learning from reading, workshops, or other activities that s/he promotes it to others. It is evident that this is a driving force in his/her life, and s/he would feel deprived without the time to develop extensively.

> Example #1: I would not be happy if I could not learn something new each day. I also believe teachers must be of the same mind set if they are serving their students. I do attend workshops, read journals and books , watch "Book TV" C-span, and attend seminars on a regular basis. I can recommend many books and journals but believe everyone should read: "Radical Evolution " by Joel Garreau: For a good understanding of what is happening globally and domestically. Subscribe to Orion magazine. Attend a LEEF conference (one coming this March) to balance what you may have read in Radical Evolution" and because: "In the end we will conserve only what we love we will love only what we

understand we will understand only what we are taught." Baba Dioum "The Courage to Teach " by Parker J. Palmer (His sections on Practicing Paradox in the Classroom and "Holding The Tension of Opposites") And finally learn about American Indian (AKA Native American) ways and teachings because there is hope when we can understand the concepts of "Sustainability" and Preservation practices as those of indigenous peoples. (252)

SRA Prompt 1*: *Professional Meeting

These teachers are at a workshop about the content they teach. Why did they go there? How do they feel about it? Gender and ethnicity aside, which teacher do you think is most like you? Why?

Note: Consistent with TAT theory, this picture is designed to draw out inner emotions that are positive or negative. The teacher who is committed to learning more about the discipline will be intrigued by the thoughtful expressions on the faces of these teachers and want to be like them or will be dismayed by their lack of enthusiasm and resent their unwillingness to learn. The teacher who is not committed to additional development will commiserate with these unhappy teachers.

Unaware: This teacher sees the meeting as the presentation of bad news. The teachers are being punished, distraught, surprised, worried, and forced to attend. They are disinterested and distracted by personal matters. S/he wants no part of it.

Example #1: The teachers went there to find out some bad news about what they teach. Their faces look somewhat distraught and they look worried about something. The teacher that represents me the most is probably the one in the back because he is the most calm out of all of them and I can stay pretty calm for the most part. (148)

Example #2: They look like they are telling each other secrets. If I had to pick who I was like it would be the person in the back listening because it does not look like they want to share their ideas and I would not impose on their private conversation. (123)

Receiving: The teacher will see these teachers as being forced to attend. S/he may be worried about having to go to a meeting like this, but s/he will be resigned to it. S/he acknowledges that it is something necessary.

Example #1: I think they went there because they had to. I do not think they feel really positive about the workshop. None of those teachers are like me. I would not act as

unprofessional as they are and if I had any issues I would write them down, and I would wait for the workshop to be over. (77)

Example #2: I see these teachers being at a math workshop. I would be the woman on the left side of the picture because I have major math anxiety. This teacher is worried about the new materials that are being presented and how she will present them to her class in a similar manner. (99)

Responding: The teacher is willing to go to the meeting and will consider trying what is being learned. S/he has an open mind and will listen politely, stay calm and relaxed, and look for something positive that can be applied.

Example #1: They went because they had to. They are not happy about going. I am most like the woman. I just sit and listen while others around me talk. I like to get the information that is being presented. (156)

Example #2: They probably went there to learn different strategies about the subject they teach. It looks like one is bored and the other one is listening carefully. The teacher that I think is most like me is the one that is listening carefully. (153)

Valuing: The teacher at this level wants to go to the meeting and wants to learn. S/he will take advantage of any opportunities provided and is enthusiastic about them. There is often a direct connection to how this learning will help students. An enthusiastic term is typically visible.

Example #1: These teachers went to a workshop on the content they teach because they care about what it is they are teaching their students. The teachers look like they are contemplating a problem or issue. They look like they might be in deep discussion or trying to come up with solutions. The teacher that is most like me would probably be the male in the middle. He looks the least frustrated and he looks like the only teacher that is happy to be there. In order for anything to be accomplished the teachers have to want to work together and be happy that they have an opportunity to find a solution to an issue. (155)

Example #2: The man and woman with their hand on their face seem like they don't think they need to be there. They look like they are thinking "I will never use this. This is too complicated for my students. I don't need this info." They probably went to the workshop because it was mandated my their school district. The one most like me is the teacher in the back. When I hear colleagues/fellow students talk about what they think is a waste of time, I wonder why they became teachers. Teachers are lifelong learners and the best way to reach your students is to always learn about HOW to reach them. The guy in back is probably wondering how they became so negative and is glad he never developed that attitude. (69)

Organizing: This teacher sees growth as a habit of mind. There is a deeper level of commitment that shows a goal for continuous learning and application. The teacher might

search for opportunities and greet them as something that cannot be missed because they have such impact potential.

> Example #1: The teachers went there in hopes to learn a bit more about the content they teach and some insights on what they are doing right and wrong while teaching the content. The looks on their faces looks like they are very interested and also shocked with what is being said to them. I think that the teacher with his hand over his mouth is most like me. The reasoning for this is because he looks very interested in what is being said, and so he is thinking already how he can apply these things to his ways of teaching. Once I hear something new about how I am conducting something or even communicating with someone, I always want to know how I can improve the situation to make it better. (163)

Characterizing: This teacher would organize other teachers to go to meetings, lobbying for them in the school.

> Example #1: These are teachers who are attending a workshop that I am facilitating. I wrote a district grant on new ways to introduce emerging literacy through role-playing, and after a trial period I'm going to share with the rest of the staff. I have always wanted to learn more myself, but I mostly enjoy the challenge of engaging others and sharing my successes. That's what makes us professionals instead of simply clerks, right? I can see that I've got some interested and some worried, but just like the kids that I teach, I'll find a way to reach everyone if I can. I don't consider myself an administrator, but I can't help being a leader or mentor to others because I simply enjoy learning new ways to do my job of teaching better and better so much that it comes naturally. Sometimes I think the administration doesn't like if when I'm pushy, but if it's part of being the best that I can be, then I'll push a few buttons!

SRA Prompt 2: Teaching Math

What is this teacher saying and feeling about math? Is she like you? How will the students feel about math because of this teacher?

Note: Consistent with TAT theory, this picture is designed to draw out inner emotions that are positive or negative. The teacher who has a positive outlook will see this teacher in a positive light, encouraging and inspiring students and having an impact on their learning, seeing past the pointed finger. Other teachers will see her/him as a threat and disengage, making little or no connections to teaching math or inspiring learners.

Unaware: The teacher has a negative view of teachers teaching content. S/he may directly express a fear of math, connecting with a student being pointed at." The teacher makes no

association between this picture and learning or math. S/he can't tell anything or may simply see the pointed finger as gun.

> Example #1: A teacher should never show a weapon at school or show they condone such behavior at school. Zero tolerance for this behavior should be demonstrated by teachers and students. (193)
>
> Example #2: This picture addresses direction. It may not be negative though the teacher is pointing at someone? Perhaps the teacher is pointing at something? I find it impossible to make a judgment in this case even if "it is not polite to point." (252)

Receiving: This teacher recognizes that the pictured teacher's manner impacts the ability/desire of students to learn math, but there is no emotion or reaction to this and no acknowledgement of whether the impact is good or bad.

> Example #1: I would guess that the teacher is explaining a problem and calling on a student to answer. She appears anxious. She is also pointing to a student in an awkward manner. She is not like me in that I am more friendly and relaxed. I think students would be uncomfortable in her class.(258)
>
> Example #2: The teacher appears to be pointing an accusatory finger at the students and whatever point she is trying to make may appear to be at gun point. Most students will be resistant to whatever point she is trying to get across and this resistance may transfer into a dislike for math. I try not to be accusatory in my teaching methods and am more inclined to be inclusive in whatever I have to teach. (221)

Responding: This teacher recognizes the impact of demeanor on learning content and cites one or more results that occur as a result of the interpretation. S/he suggests a strategy that makes content delivery and demeanor more meaningful.

> Example #1: He is pointing his finger in an accusatory manner at a student and asking why the student cannot rattle off the correct answer. I don't think he is like me. My students frequently need questions/ directions repeated and you have to give them wait time to answer. He might make his students less likely to want to participate in class. (211)
>
> Example #2: This teacher looks like she is yelling at one of her students who may not have answered correctly. She is pointing and singling out that student, possibly embarrassing them for having the incorrect response. Students may fear math because of this teacher. I do not point at my students, and, if they are struggling with an answer, I ask them if they'd like to have a friend or neighboring student help them. (186)

Valuing: This teacher typically sees the teacher in the picture as inspiring students, telling them that they are providing correct answers and moving them forward at a rapid pace. They might see the finger as a non-verbal, Bingo!" Praise is offered. Occasionally, the respondent becomes angry at what they perceive to be a nasty teacher who discourages students. The impact of

demeanor is discussed vividly, with the responding teacher connecting directly and strongly to students and learning.

> Example #1: This teacher is excited about math! She is going whoo-hoo at a student who just a question right and seems to be proud of her students. The students will respond positively to her because she is outgoing and theatrical. Hopefully, this is something I will be able to do to encourage student responses. (168)

> Example #2: This teacher seems to like teaching math because she seems very excited to know that a student got the question correct. The teacher is pointing to the students and saying "Yes, you are coooorreect!" Yes I would say this teacher is like me because I get excited when students get questions that are hard. I am proud of them. The students will enjoy math because the teacher is excited. (217)

Organizing: This teacher generalizes to other contexts and emotions, seeing both good and bad in the picture, pointing how different students might react to this teacher, and making suggestions about how to improve the learning environment and motivation (INTASC #5) or other INTASC standards.

> Example #1: This teacher is lecturing from the front of the classroom. The teacher has written some concepts on the board in order to provide a short instructional period and is now pointing to a student to ask that student to perform the problem on the board. While some students relish the opportunity to get a math problem right up front, many students react by avoiding eye contact, shrinking in the chair, and wishing profusely that he/she is not the one called on to possibly embarrass himself in front of the class. While I think it is important to have public opportunities to present and speak in front of the class I try to stay away from the situations when students are legitimately afraid of being called upon.

Characterizing: This teacher is driven by the need to instil enthusiasm about content in students and would lobby for professional development on strategies to do so.

Appendix B

Taxonomic Levels Adapted from Bloom, Krathwohl, and Masia (1964) for DAATS Scale Development

Rubric for Dispositions Assessments

Taxonomic Levels	Typical Teaching Behaviors at Each Taxonomic Level
Receiving	• Recognizes (is aware of) importance. • Is beginning to think about it.
Responding	• Is emotionally ready to do something and makes an attempt. • Gives a little extra effort as time permits to comply. • Can easily be distracted from applying the value. • Has beginning level of commitment or satisfaction.
Valuing	• Accepts worth and derives definite satisfaction from it. • Feels a need and would commit continuing time and effort. • Tolerates and may expect interferences.
Organization	• Plans, organizes, and schedules to ensure success. • Determines inter-relationships. • Adapts other aspects to fit. • Is uncomfortable with interferences or lack of time to finish.
Characterization	• Sees the value as the center or driving force of all work. • Helps others to see the world in this light, lobbying for it. • Integrates everything with it.

Rationale

Items in the DAATS instruments have been written with Bloom and Krathwohl's Affective Taxonomy in mind, although the focus is clearly different – assessment rather than instruction. The intent is to make visible the affective domain of the INTASC standards in an ordered and credible way.

Here we provide an operational definition for each of the five Bloom, Krathwohl, and Masia (1964) levels in the affective domain taxonomy – receiving, responding, valuing, organizing, characterizing. Note that we have added a sixth initial level called, "unaware." It describes negative characteristics and behaviors that could occur prior to entering the receiving stage – if "receiving" is ever attainable for the teacher. It serves as a warning sign in this measurement process but it is certainly not appropriate for an instructional objective – the intent

of the taxonomy. Teachers measured at this level have the potential to impede student competence and self-confidence.

As this operational definition of the Taxonomy is applied to the INTASC Principles, it is important to note that most teachers will be higher on some Principles and lower on others. For those who have "unaware" ratings on any item, further investigation is recommended, typically followed by counselling. For those who have multiple ratings of "unaware," resulting in low scale scores, another profession might be a better choice. These are difficult decisions that should be driven by a combination of judgmental and empirical evidence.

In the following operational definition of the taxonomy to measuring teacher dispositions, we discuss for each of the six levels, the following:

- Typical characteristics and behaviours
- Typical response patterns
- Interpretation of the rating
- Use of the rating

Taxonomic Levels

Unaware

Typical Characteristics and Behaviors:

This teacher does not understand or is not committed to the Principle. S/he may exhibit a behaviour that is the opposite of the skill. S/he may be lazy, attend to important details with a "lick and a promise," or use a "one-size-fits-all" approach, waiting for the bell ring at the end of the day. S/he may be isolated from colleagues and parents. S/he rarely reflects adequately on successes and failures and does not look back. S/he may feel limited responsibility for other people's problems (including both children and colleagues). Typically, there are incidents that bring this teacher to the attention of faculty or administrators as exhibiting some questionable behaviours. S/he may impart a negativism to students.

Typical Response Patterns:

For objective self-report measures, the easiest items are incorrectly answered (BATS). Subjective self-report responses (ETQ and SRA) or off-topic and unrelated to the question posed or the skill being assessed or may include a negative statement, such as "I wouldn't…" or "I don't want to…" or "It's not my job…" or "I don't like to…" Observations (CDC) typically result in the identification of negative behaviours. Children (KIDS) may see this teacher as disrespectful, unfair, boring, controlling, unimaginative, didactic, mean, oblivious, rude, negative, uninviting, insensitive, rigid, inflexible, or lazy.

Interpretation of Ratings:

Some "unaware" ratings will be easy to resolve with further investigation and possibly counselling and remediation. If the scoring remains at the "unaware" level, however, at the point of certification or re-employment, this teacher should be viewed as one who might impede student competence and self-confidence.

Use of Ratings:

A teacher with multiple indications of lack of awareness could be denied certification or employment, especially if the Rasch scale score is very low. In a teacher training context, counselling and opportunities for remediation should be provided. If the teacher is hired, a monitoring and professional development plan should be in place.

Receiving

Typical Characteristics and Behaviors:

This teacher recognizes (is aware of) the importance of, or need for, the skill, but the evidence is non-committal and non-emotional – just matter-of-fact reporting. It is only lip-service at this point – a promise of future commitment. There may be an air of resignation. This could be a "socially acceptable" response pattern, but the benefit of the doubt takes the decision to "receiving" and not "unaware" because this teacher tolerates and pays attention to the requirements of the skill.

Typical Response Patterns:

The easy objective self-report responses typically are correct (BATS). Subjective self-report responses (ETQ and SRA) are typically on-topic, simplistic, and brief and might include statements such as: "I wouldn't mind…" or "I think this is important because…" or "I thought about this but I haven't tried it yet." Or "I'm so busy, and I wish I could …." Or "this is something we have to do." Sometimes they are lengthy but fall short of stating a strategy or action, containing only an acknowledgment of a condition. Observations (CDC) typically show a pattern of mixed results. Children (KIDS) may see still see many of the negative attributes that are overcome with practice (e.g., nervous, uncomfortable, rigid, inflexible, didactic), but they will also see the teacher as respectful and nice.

Interpretation of Ratings:

"Receiving" ratings are expected for teacher candidates in the early stages of their training. They are not expected at the point of graduation or hiring. By the end of the first year or two of teaching, teachers should be beyond this level on all Principles.

Use of Ratings:

A teacher with multiple indications of "receiving" early in the program should be acknowledged as "in progress." A teacher with multiple indications of "receiving late in a preparation program, at the point of hire, or during orientation should have a plan for professional development. Mentoring from a more advanced teacher may be necessary.

Responding

Typical Characteristics and Behaviors:

This teacher is trying and is committed in some small way. There is some level of positive recognition of the meaning of the standard, perhaps through early stages of experimentation and engagement. This teacher might be easily distracted for something more intriguing. S/he is testing the waters but not yet sure it is worth it. Some low level activity is typically attempted and reported. It may be just an early attempt but there is a promise, expressed or inferred, of continued growth. The teacher appears emotionally ready or willing; s/he doesn't mind doing it. The effort may be misdirected in terms of skill but not affect – a wrong solution or a naïve conclusion.

Typical Response Patterns:

The easy and some moderately difficult objective self-report responses are correct (BATS). Subjective self-report responses (ETQ and SRA) are typically on-topic and may still be simplistic and brief, but an action is cited. The teacher will be neutral and not show enthusiasm yet and might say, "want to…" or "like to." Observations (CDC) may still show a pattern of mixed results. Children (KIDS) may see less negative attributes, beginning to describe the teacher as less didactic and more interesting and aware of them as individuals, but they are not likely to describe high energy levels, originality, or enthusiasm yet.

Interpretation of Ratings:

"Responding" is an indication of potential but not a promise of it. "Valuing" is the target level, with an expectation that "responding" will move up to "valuing" with some additional experience. A mixture of "Responding" and "Valuing" ratings are expected for teacher candidates in the later stages of their training and for teachers seeking initial employment. A preponderance of "responding" ratings could be a predictor that progression may be slow.

Use of Ratings:

A certified teacher with multiple indications of "responding" should be encouraged to continue to experiment and grow. Support systems should be provided, as seen necessary and available, possibly from peers. Teachers with "responding" values are ready to become self-sufficient, and would usually benefit from mentors, immersion into "unknown" experiences, and positive support from administrators. Supervising teachers need to recognize "responding" and encourage it when it occurs.

Valuing

Typical Characteristics and Behaviors:

This teacher has accepted the worth or value of the skill, wants to use it, spends time and energy on it, and derives satisfaction from so doing. S/he may have routines already built or use multiple strategies. S/he will set aside time to prepare for, implement, practice, or reflect on the skill. S/he gets to the root of the problem, offering long term support and solutions and working with others. This teacher may "go the extra mile" or "light a fire" in students. This teacher may still tolerate or expect interferences, but they have become annoying. S/he celebrates and respects children, greeting difficult situations with enthusiasm. Discussions are empathetic, enthusiastic and child-centered, with connections to results and learning. The teacher is emotionally involved with children and issues, encourages and supports them, and feels their pain and joy. This teacher may resent those who are not committed.

Typical Response Patterns:

All or most of the objective self-report responses are correct (BATS), since it is difficult to tap higher levels in an objective format. Subjective self-report responses (ETQ and SRA) are on-topic and specific, sometimes quite detailed. specific strategies (and usually more than one) are noted with implications for children made and some form of enthusiasm expressed with strong words or expressions, such as "great," "wonderful," wow," and "I want to…" "Observations (CDC) are mostly positive with negative ratings rare. Children (KIDS) are enthusiastic about the teacher, listing attributes and signing praises. They are enthusiastic about what they are learning and feel well supported and cared for.

Interpretation of Ratings:

"Valuing" is an indication of strong affect -- the target level for practicing teachers.

Use of Ratings:

The teacher should be encouraged to grow at his/her own pace through staff development and other outlets. This teacher could mentor other teachers or teacher candidates.

Organizing

Typical Characteristics and Behaviors:

This teacher is a role model for this value for children and teachers. The skill is a habit of mind. S/he integrates and synthesizes this skill other skills in ways that are systematic, frequent, studied, and creative. Combinations might be (a) content and critical thinking; (b) diversity, learning environment, and communication; or the (c) collegial aspects of planning, professionalism, and assessment. This teacher places a high priority on the skill, dedicating extra time and effort to it, searching for opportunities to apply it in day-to-day events, deriving personal meaning and self-worth from accomplishments. S/he is uncomfortable if deprived of those opportunities. This teacher is late to bed and early to rise.

Typical Response Patterns:

Subjective self-report responses (ETQ and SRA) are creative and detailed, showing integration across disciplines and/or extensive effort/thought or continuous involvement. Negative ratings on observations (CDC) are rare. Children (KIDS) are enthusiastic and may show evidence of sharing the value with the teacher, emulating his/her communication techniques, advocacy positions, self-assessment, or content curiosity.

Interpretation of Ratings:

"Organizing" is the level of the teacher/leader. For excellent teachers, one would expect a combination of "valuing" and "organizing," since one could not reasonably reach this level for all skills. Skills assessed at this level are a driving force for the teacher.

Use of Ratings:

The teacher may help at the school improvement level as a team leader, mentor other teachers or teacher candidates, and provide staff development within the school or the district.

Characterizing

Typical Characteristics and Behaviors:

This teacher views every aspect of the profession through the lens of this skill. It is the driving force, life style, or mission of his/her work. S/he is tireless in working toward this value. It would be rare to see a teacher reach this level on more than one skill. S/he has vision and instils it in others, making personnel sacrifices and expecting others to do the same. S/he would lobby for and promote school-based decisions and activities based on this value.

Typical Response Patterns:

Most subjective self-report responses (ETQ and SRA) reference this skill in some way. If it were diversity, for example, every response would talk about children from different backgrounds and how their needs could be, or were being, met. If it were planning, all discussions would indicate near compulsive planning, and this would be observable in the classroom as well. Negative ratings on observations (CDC) are rare. Children (KIDS) may say, "She is always talking about (or doing)…"

Interpretation of Ratings:

"Characterising" is the highest level of the teacher/leader.

Use of Ratings:

The teacher could mentor other teachers or teacher candidates and provide staff development within the school or the district, but care needs to be taken that this teacher's internal drive does not lead to burn out.

Appendix C

Measuring INTASC Principle 1 -- Five Instruments

<u>Principle #1:</u> The teacher understands the central concepts, tools of inquiry, and structures of the discipline(s) he or she teaches and can create learning experiences that make these aspects of subject matter meaningful for students.

1. The teacher realizes that subject matter knowledge is not a fixed body of facts but is complex and ever-evolving. S/he seeks to keep abreast of new ideas and understandings in the field.
2. The teacher appreciates multiple perspectives and conveys to learners how knowledge is developed from the vantage point of the knower.
3. The teacher has enthusiasm for the discipline(s) s/he teaches and sees connections to everyday life.
4. The teacher is committed to continuous learning and engages in professional discourse about subject matter knowledge and children's learning of the discipline.

<u>Beliefs About Teaching Scale (BATS</u>**)**

#	Ans.	Prin.	Tax	Item
11	A	1	V	One of the most important tasks in my class is to make the subject meaningful to students.
17	D	1	RC	Sometimes students may not understand why a lesson is meaningful, but that is okay as long as they learn what they have to learn.
44	D	1	RC	I expect my school district to provide all the training I need to keep up-to-date.
46	A	1	V	I constantly ask students to describe their point of view so that I can understand their different perspectives
49	D	1	RC	A well-written teacher's manual has all the details you need to give a great lesson.
55	D	1	RC	Making content practical and connected to everyday life is a great goal, but it is too time-consuming to be realistic.

<u>Experiences in Teaching Questionnaire (ETQ)</u>

ETQ Question 1: How have you kept abreast of current developments in your field? For example, did you attend any workshops, subscribe to any journals, read or buy a new book? If so, describe in one to two sentences something you learned and the source.

Situational Reflection Assessment (SRA)

Test Form 1 (odds) -- SRA Prompt 1: *Professional Meeting*

These teachers are at a workshop about the content they teach. Why did they go there? How do they feel about it? Gender and ethnicity aside, which teacher do you think is most like you? Why?

Test Form 2 (evens) -- SRA Prompt 2: *Teaching Math*

What is this teacher saying and feeling about math? Is she like you? How will the students feel about math because of this teacher?

Candidate Dispositions Checklist (CDC

KEY:

P = Positive (Typically positive – rarely negative)
M = Mixed (Multiple observations of both positive and negative)
N = Negative (Typically negative – rarely positive)
ND = No Decision or No Data

INTASC #1: Content

	Positive Dispositions	Rating				Negative Dispositions
1	Makes knowledge accessible to all students.	P	M	N	ND	Presents only to the most competent students.
2	Actively encourages diverse viewpoints about the content.	P	M	N	ND	Makes discouraging statements or facial expressions about diverse views expressed by children about content.
3	Voluntarily seeks out recent developments or materials in the content area(s) to share with students.	P	M	N	ND	Is unimaginative or out-of-date, using only the textbook or basic adopted materials.
4	Actively seeks out connections to everyday life.	P	M	N	ND	Misses opportunities to connect content with real life and/or current events.
5	Possesses vibrant facial expressions during presentation and discussion of content area(s).	P	M	N	ND	Looks bored or unhappy during discussion of <u>any</u> content area.

K-12 Impact Dispositions Scale (KIDS)

1a: Do you think your teacher likes to teach? How do you know?
1b: Does your teacher encourage (want) you to explore what you are learning on your own in centers, on the computer, or in books and magazines?
1c: Does your teacher ever share with you something s/he has recently read or learned (6-12 only)?
1d: Does your teacher ask you what your favorite subject is? What did (would) you say?
1e: Does your teacher show you interesting web sites and books about new developments in the things you are studying? (6-12 only)
1f: Does your teacher tell or show you why the work you do is important for you to learn?
1g: Does your teacher ever say that s/he doesn't like something she is teaching?