2008

Exploration of a Taxonomic Framework To New Instrument Development and Items Types:

Dimensional Disaster or Informed Instruments?

Judy R. Wilkerson, Florida Gulf Coast University
jwilkers@jgcu.edu
W. Steve Lang, University of South Florida St. Petersburg
ws.lang@knology.net

IOMW
New York City

Introduction

U.S. teacher educators are faced with serious challenges to demonstrate the quality of the graduates they prepare. These challenges are expressed in the country's fixation on accountability and testing. We are witnessing the growth of standardized tests and alternative routes to certification as panaceas intended to solve the teacher shortage crisis in most states. National accreditation and state program approval agencies are attempting to stem the tide through the ever-increasing demand for standards-based assessment data documenting teacher quality. The common set of standards used nationally for accreditation was developed by the Council of Chief State School Officers and promulgated by the Interstate New Teacher Assessment and Support Consortium (INTASC) in the form of ten principles. Each of the principles includes indicators written at the knowledge, skill, and dispositional levels, forming constructs that colleges are required to measure.

Sadly, neither the profession nor the accreditors have realized the need for objective measurement. This is largely a function of what Stiggins bemoans as assessment illiteracy (2000). They are satisfied, at best, with ordinal scales for poorly constructed criteria on ill-defined tasks, or, at worst, with counting papers in portfolios constructed without regard to any form of psychometric consideration (Wilkerson and Lang, 2003).

In this paper, however, the focus was on assessment of dispositions. Searches for measurement of this elusive construct have been fruitless, and teacher educators are clamouring for answers to this requirement since it is being imposed by the National Council for the Accreditation of Teacher Educators (NCATE). At a workshop these authors conducted last year for the American Association of Colleges of Teacher Education (Wilkerson, et al., 2003), virtually all participants saw dispositions as their burning issue.

The measurement community has been largely uninvolved in trying to solve the disposition dilemma. Teacher educators who have attempted disposition assessment generally resort to one or more disconnected measures with single item structures. These are often poorly conceptualized, failing to adequately sample from the construct to be measured. This paper presents an attempt at providing an objective measurement solution to the problem. The approach to this problem started by envisioning multiple instruments, of different formats and at differing levels of inference. The construct map included the content from the INTASC principles and was framed by the stages of affective response described by Krathohl's Taxonomy. As exploration of the results by sub-grouping items as confounding or informing the results was the primary question of this paper. The original plan includes six instrument types, but three are considered here. Comments on planned Rasch models and dimensionality:

As described by Wright and Stone (1999, “The way to proceed in understanding multidimensionality is first to construct unidimensional variables upon which to make measures and then evaluate the relationship among the measures generated from these singular variables.” (p. 181).

“Confusing a number of attributes into a single generic score makes confident predictions from that score more hazardous and the score a less useful summary of ability or achievement. But carefully constructed tests that make good measurement estimates of single attributes might be sufficient for a number of thoughtfully decided purposes. For special or difficulty situations, collecting additional estimates of other appropriate attributes is essential.” (Bond & Fox, 2007, p. 34).

“The art of assessing dimensionality is to find the smallest number of ability domains such that the are both statistically well-defined and substantively meaningful.” (Briggs & Wilson, 2004, p. 323).

Standards based assessment of the INTSAC principles must consider the ten principles as subscores at the least. Dispositions are affective constructs, so high inference items types require extensive scoring, while self-report is subject to faking. For our purposes, the way to achieve confidence in the assessment was to use multiple item types that confirm the score validity. This mandated that the INTASC principle construct needed a framework so that scoring for both dichotomous and various rating conditions was meaningful as an attribute and statistically reasonable. In this pilot, we explored the characteristics of the initial calibration of the new instruments to assess teacher dispositions.

Method

Instruments

This was a field test of three instruments that make use of three different item structures (agree/disagree statements, self-report questions, and apperception). The total scale to be created combines all instruments. The belief scale is composed of 60 items, with sub-scales for each of the ten INTASC Principles. The self report questionnaire contains ten questions, also aligned to INTASC Principles. The apperception instrument presents 20 stimulus cards, two each for the 10 INTASC principles. The instruments range from low to high inference. The item design and scoring guides were developed as framed by the Krathwohl Affective Taxonomy. The three instruments considered here are the Belief Scale Aligned with Teacher Standards (BATS), Experiential Teacher Questionnaire (ETQ), and the Situational Reflection Assessment (SRA).

Sample

At present, we have analysed data on 335 respondents from the teacher education programs at three colleges. With a few missing cases, teacher candidates ranging from pre-internship to final internship to inservice graduate students were assessed. The sample came from teacher education programs at three universities in Florida.

Analysis

We are using three Rasch models for the scales: dichotomous, rating scale, and FACETS as follows (Stone, 2003; Linacre, 2007):

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

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

$$\log\left(\frac{P_{nmijk}}{(P_{nmijk(k-1)})}\right) = B_n - A_m - D_i - C_j - F_{ik}$$

Results

The following (Figure 1) is the logistic ruler produced from the initial analysis. The 60 Thurstone (agree-disagree) statements were scored dichotomously and are indicated as a D. The two rated instrument items are indicated by an X. In this calibration, all items were pooled as one scale and coded as item groups in Winsteps.

That is followed by the Separation Table from Winsteps (Table 2). The OUTFIT ZSTD of -0.3 and standard deviation of 1.9 imply more variability than expected, but Linacre (2007) warns that “the fit statistics stratify by item type; so that, say, all the Yes-No items overfit, and all the Frequency items underfit” (p. 89). Figure 1 is a from a joint calibration of the data.

Table 3 contains the INTASC Principle statistics across all three instruments in the battery. In this case, each principle is represented by 6 Thurstone items, 1 questionnaire item, and 2 apperception items. The questionnaire and apperception items are scored on a rating of 0 to 6 based on modification of the Krathwohl affective taxonomy. The ratings are:

0=Unaware
1=Receiving
2=Responding
3=Valuing
4=Organizing
5=Characterizing

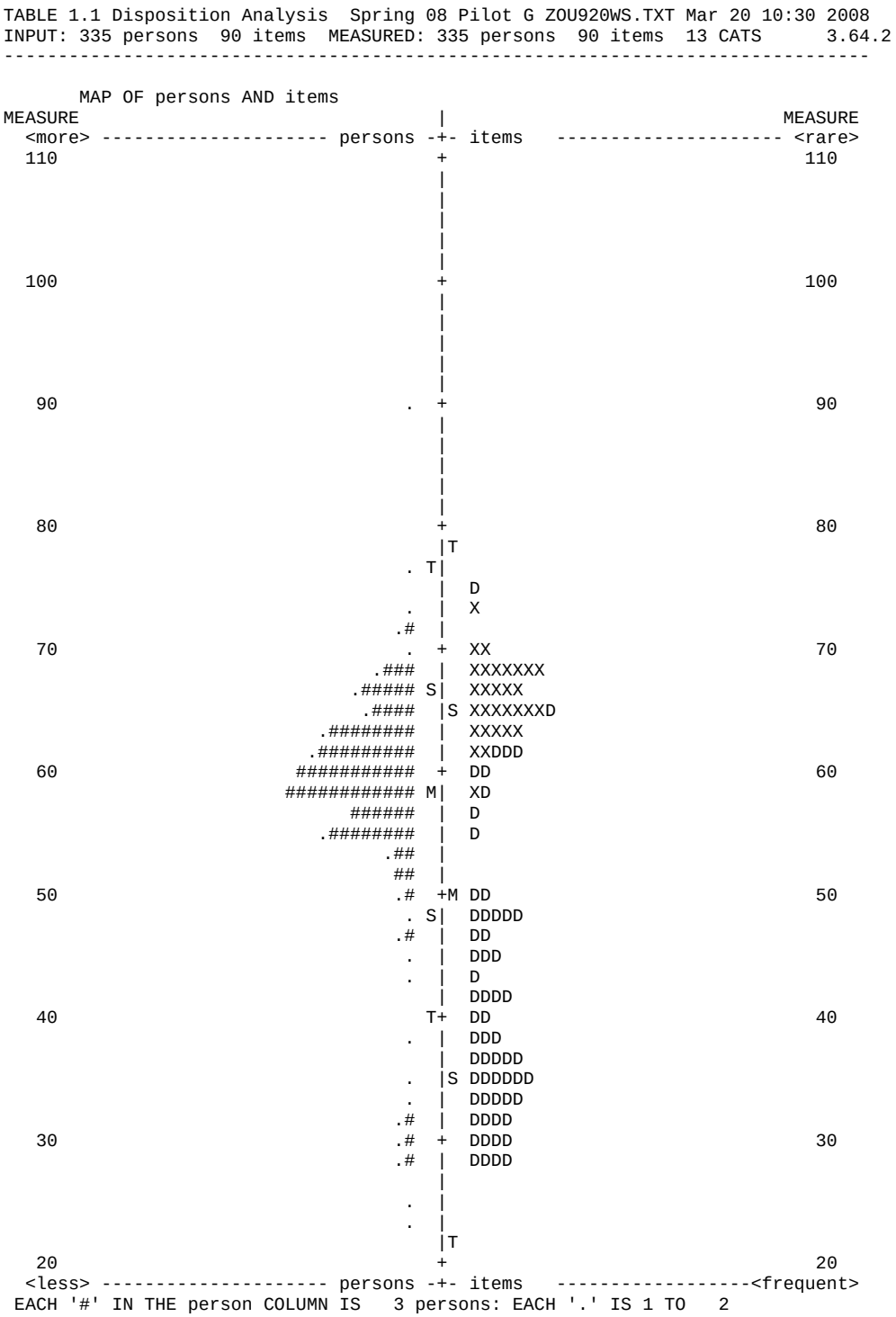

Figure 1. Map of Items and Persons for BATS, ETQ, and SRA

Table 1. Summary Statistics for BATS, ETQ, and SRA as Separate Calibrations

```
TABLE 3.1 Disposition Analysis  Spring 08 Pilot G ZOU920WS.TXT Mar 20 10:30 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      68.0      72.3       58.08    2.57      1.16     .5    .96    -.3 |
| S.D.      18.6       5.8        9.28     .69       .70    2.5    .64    1.9 |
| MAX.     110.0      81.0       89.58   10.33      5.65    9.9   4.27    9.6 |
| MIN.      12.0      60.0       24.07    1.94       .39   -3.4    .27   -3.0 |
|------------------------------------------------------------------------------|
| REAL RMSE   3.00  ADJ.SD    8.78  SEPARATION  2.93  person RELIABILITY  .90 |
|MODEL RMSE   2.66  ADJ.SD    8.89  SEPARATION  3.34  person RELIABILITY  .92 |
| S.E. OF person MEAN = .58                                                    |
+------------------------------------------------------------------------------+
  MINIMUM EXTREME SCORE:     74 persons
        VALID RESPONSES:  80.3%

     SUMMARY OF 335 MEASURED (EXTREME AND NON-EXTREME) persons
+------------------------------------------------------------------------------+
|           RAW                          MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|------------------------------------------------------------------------------|
| MEAN      52.9      69.6       41.94    6.05                                 |
| S.D.      32.6       7.2       31.40    6.56                                 |
| MAX.     110.0      81.0       89.58   18.31                                 |
| MIN.        .0      60.0      -14.98    1.94                                 |
|------------------------------------------------------------------------------|
| REAL RMSE   9.00  ADJ.SD   30.08  SEPARATION  3.34  person RELIABILITY  .92 |
|MODEL RMSE   8.92  ADJ.SD   30.10  SEPARATION  3.37  person RELIABILITY  .92 |
| S.E. OF person MEAN = 1.72                                                   |
+------------------------------------------------------------------------------+
person RAW SCORE-TO-MEASURE CORRELATION = .94 (approximate due to missing data)
CRONBACH ALPHA (KR-20) person RAW SCORE RELIABILITY = .97 (approximate due to missing data)

     SUMMARY OF 90 MEASURED (NON-EXTREME) items
+------------------------------------------------------------------------------+
|           RAW                          MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|------------------------------------------------------------------------------|
| MEAN     197.1     209.6       50.00    1.64       .99    -.1   1.18     .8 |
| S.D.      51.1      73.9       14.40     .48       .36    2.5    .83    3.8 |
| MAX.     309.0     261.0       74.52    2.55      2.12    5.9   4.00    9.9 |
| MIN.      50.0      52.0       28.97     .90       .50   -3.2    .19   -4.3 |
|------------------------------------------------------------------------------|
| REAL RMSE   1.77  ADJ.SD   14.30  SEPARATION  8.09  item   RELIABILITY  .98 |
|MODEL RMSE   1.71  ADJ.SD   14.30  SEPARATION  8.39  item   RELIABILITY  .99 |
| S.E. OF item MEAN = 1.53                                                     |
+------------------------------------------------------------------------------+
UMEAN=50.000 USCALE=10.000
item RAW SCORE-TO-MEASURE CORRELATION = -.53 (approximate due to missing data)
18862 DATA POINTS. APPROXIMATE LOG-LIKELIHOOD CHI-SQUARE: 20432.72
```

Table 2. Results of the Calibration by Principle and Instrument Type

TABLE 27.1 Disposition Analysis Spring 08 Pilot ZOU034WS.TXT Mar 20 10:38 2008
INPUT: 335 persons 90 items MEASURED: 335 persons 90 items 13 CATS 3.64.2

Subtotal specification is: ISUBTOTAL=$S8W2

ALL SCORES ARE NON-EXTREME

item COUNT	MEAN MEASURE	S.E. MEAN	OBSERVED S.D.	MEDIAN	REAL SEPARATION	CODE
90	50.00	1.53	14.40	48.07	8.09	**
9	50.55	5.07	14.35	48.41	8.27	01
9	46.05	4.71	13.32	40.87	7.29	02
9	48.15	5.41	15.29	49.69	8.11	03
9	52.52	5.02	14.21	61.21	8.28	04
9	49.24	4.64	13.11	49.90	7.60	05
9	47.57	4.88	13.79	45.61	7.51	06
9	50.09	4.38	12.40	45.61	7.10	07
9	50.66	4.87	13.77	45.86	7.90	08
9	53.09	6.09	17.23	60.00	9.07	09
9	52.10	5.09	14.40	48.63	8.53	10

UMEAN=50 USCALE=10

The INTASC Principles are coded as:

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

TABLE 27.14 Disposition Analysis Spring 08 Mar 20 10:38 2008
INPUT: 335 persons 90 items MEASURED: 335 persons 90 items 13 CATS

Subtotal specification is: ISUBTOTAL=$S1W3

ALL SCORES ARE NON-EXTREME

item COUNT	MEAN MEASURE	S.E. MEAN	OBSERVED S.D.	MEDIAN	REAL SEPARATION	CODE
90	50.00	1.53	14.40	48.07	8.09	***
60	42.00	1.39	10.71	38.14	5.48	BAT
10	66.27	.70	2.10	66.27	1.83	ETQ
20	65.87	.76	3.31	65.48	1.86	SRA

UMEAN=50 USCALE=10

Table 3. Summary of the Calibration for Each Instrument Type

TABLE 27.17 Disposition Analysis Spring 08 Pilot ZOU034WS.TXT Mar 20 10:38 2008
INPUT: 335 persons 90 items MEASURED: 335 persons 90 items 13 CATS 3.64.2

"BAT" SUBTOTAL FOR 60 NON-EXTREME items

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	201.6	261.0	42.00	1.88	.80	-1.3	.74	-1.4
S.D.	43.7	.0	10.71	.39	.19	1.7	.27	1.5
MAX.	241.0	261.0	74.52	2.55	1.19	5.0	1.19	2.1
MIN.	50.0	261.0	28.97	1.32	.50	-3.2	.19	-4.3

REAL RMSE	1.92	ADJ.SD	10.53	SEPARATION	5.48	item	RELIABILITY .97
MODEL RMSE	1.92	ADJ.SD	10.54	SEPARATION	5.50	item	RELIABILITY .97

S.E. OF item MEAN = 1.39
MEDIAN = 38.14

TABLE 27.18 Disposition Analysis Spring 08 Pilot ZOU034WS.TXT Mar 20 10:38 2008
INPUT: 335 persons 90 items MEASURED: 335 persons 90 items 13 CATS 3.64.2

"ETQ" SUBTOTAL FOR 10 NON-EXTREME items

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	264.3	134.6	66.27	.91	1.22	1.8	1.91	5.6
S.D.	27.1	1.4	2.10	.00	.20	1.4	.47	2.1
MAX.	309.0	137.0	69.15	.92	1.72	5.2	3.00	9.9
MIN.	228.0	133.0	62.98	.90	1.02	.2	1.36	2.8

REAL RMSE	1.01	ADJ.SD	1.84	SEPARATION	1.83	item	RELIABILITY .77
MODEL RMSE	.91	ADJ.SD	1.89	SEPARATION	2.08	item	RELIABILITY .81

S.E. OF item MEAN = .70
MEDIAN = 66.27

TABLE 27.19 Disposition Analysis Spring 08 Pilot ZOU034WS.TXT Mar 20 10:38 2008
INPUT: 335 persons 90 items MEASURED: 335 persons 90 items 13 CATS 3.64.2

"SRA" SUBTOTAL FOR 20 NON-EXTREME items

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	149.9	92.8	65.87	1.28	1.45	2.8	2.16	5.2
S.D.	32.9	13.3	3.31	.13	.34	1.8	.96	3.2
MAX.	202.0	105.0	72.83	1.67	2.12	5.9	4.00	9.9
MIN.	69.0	52.0	57.97	1.15	1.04	.4	1.04	.4

REAL RMSE	1.57	ADJ.SD	2.92	SEPARATION	1.86	item	RELIABILITY .77
MODEL RMSE	1.29	ADJ.SD	3.05	SEPARATION	2.37	item	RELIABILITY .85

S.E. OF item MEAN = .76
MEDIAN = 65.48

Table 4. Item Statistics for BATS, ETQ, and SRA

```
TABLE 10.1 Disposition Analysis  Spring 08 Pilot  ZOU604WS.TXT Mar 20 10:58 2008
INPUT: 335 persons  90 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

         "BATS" item STATISTICS:  MISFIT 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	G
3	228	243	43.65	2.73	1.09	.5	1.45	1.4	A .89	93.8	93.8	BAT03VA10	1
57	212	243	52.09	1.99	1.08	.6	1.42	1.9	B .80	86.4	87.4	BAT57RC07	1
13	121	243	73.31	1.36	1.23	5.5	1.31	5.2	C .46	51.0	63.0	BAT13VA08	1
23	50	243	88.21	1.66	1.05	.5	1.22	1.6	D .28	80.7	80.1	BAT23RP09	1
38	194	243	58.00	1.67	1.11	1.1	1.19	1.4	E .73	79.4	80.4	BAT38RP04	1
27	179	243	61.80	1.53	1.08	1.1	1.15	1.4	F .67	74.5	74.9	BAT27RP03	1
42	225	243	45.70	2.51	1.06	.4	1.14	.6	G .88	92.6	92.6	BAT42RC02	1
29	116	243	74.24	1.36	1.10	2.4	1.13	2.4	H .47	57.6	63.2	BAT29RC04	1
37	234	243	38.08	3.45	1.00	.1	1.13	.4	I .93	96.3	96.3	BAT37RC10	1
53	109	243	75.53	1.36	1.07	1.7	1.13	2.2	J .45	63.8	63.9	BAT53RC04	1
20	123	243	72.95	1.36	1.04	1.1	1.12	2.2	K .50	63.4	63.0	BAT20VA09	1
60	93	243	78.56	1.39	1.01	.2	1.12	1.8	L .42	69.1	66.5	BAT60RC07	1
33	211	243	52.49	1.96	.98	-.1	1.09	.5	M .82	86.8	87.0	BAT33RC04	1
28	184	243	60.60	1.57	1.04	.5	1.09	.8	N .70	75.7	76.6	BAT28RP10	1
2	197	243	57.14	1.71	1.08	.8	1.00	.1	O .75	80.2	81.6	BAT02RC07	1
17	185	243	60.36	1.58	1.07	.8	1.07	.6	P .70	74.5	77.0	BAT17RC01	1
39	231	243	41.19	3.02	1.05	.3	1.03	.2	Q .91	95.1	95.1	BAT39RP06	1
7	132	243	71.29	1.36	1.02	.6	1.04	.7	R .53	62.1	63.3	BAT07RC10	1
21	116	243	74.24	1.36	1.00	.0	1.04	.8	S .49	63.4	63.2	BAT21VA03	1
30	233	243	39.21	3.28	1.04	.2	.90	-.1	T .93	95.9	95.9	BAT30RC05	1
10	223	243	46.90	2.40	1.04	.3	.99	.1	U .87	91.4	91.8	BAT10OR02	1
36	197	243	57.14	1.71	1.02	.2	1.04	.3	V .75	81.9	81.6	BAT36RP06	1
14	215	243	50.86	2.08	1.03	.3	.97	.0	W .83	88.1	88.6	BAT14RC08	1
34	217	243	49.97	2.14	1.03	.2	.95	-.2	X .84	88.9	89.4	BAT34RP05	1
1	222	243	47.46	2.35	1.02	.2	.84	-.6	Y .87	90.9	91.4	BAT01RC03	1
4	241	243	22.46	7.13	1.01	.3	.80	.0	Z .97	99.2	99.2	BAT04RC03	1
BETTER FITTING OMITTED													
9	209	243	53.24	1.92	.98	-.1	.84	-.9	z .81	86.0	86.2	BAT09RC08	1
24	225	243	45.70	2.51	.98	.0	.77	-.8	y .89	92.6	92.6	BAT24RP06	1
49	187	243	59.85	1.59	.95	-.5	.97	-.2	x .72	79.4	77.7	BAT49RC01	1
52	230	243	42.06	2.91	.97	.0	.77	-.6	w .91	94.7	94.7	BAT52RC07	1
16	143	243	69.23	1.38	.97	-.8	.93	-1.2	v .58	66.3	64.8	BAT16RC05	1
56	228	243	43.65	2.73	.96	-.1	.92	-.1	u .90	93.8	93.8	BAT56RC08	1
5	214	243	51.28	2.05	.96	-.2	.85	-.7	t .84	88.5	88.2	BAT05VA02	1
26	226	243	45.05	2.58	.96	-.1	.75	-.8	s .90	93.0	93.0	BAT26VA10	1
44	150	243	67.89	1.39	.94	-1.2	.96	-.6	r .60	68.7	66.1	BAT44RC01	1
58	203	243	55.30	1.80	.94	-.5	.84	-1.0	q .79	84.4	83.9	BAT58RP07	1
12	235	243	36.82	3.65	.94	-.1	.61	-.9	p .94	96.7	96.7	BAT12RC07	1
11	241	243	22.46	7.13	.93	.1	.39	-.7	o .98	99.2	99.2	BAT11VA01	1
8	236	243	35.40	3.89	.93	-.1	.57	-.9	n .95	97.1	97.1	BAT08OR03	1
46	225	243	45.70	2.51	.92	-.3	.68	-1.2	m .89	92.6	92.6	BAT46VA01	1
51	232	243	40.24	3.14	.92	-.2	.71	-.7	l .93	95.5	95.5	BAT51RC06	1
35	241	243	22.46	7.13	.92	.1	.20	-1.3	k .98	99.2	99.2	BAT35RC03	1
25	235	243	36.82	3.65	.92	-.1	.50	-1.3	j .95	96.7	96.7	BAT25VA06	1
50	224	243	46.31	2.45	.91	-.4	.78	-.8	i .89	92.6	92.2	BAT50RC05	1
6	241	243	22.46	7.13	.90	.1	.14	-1.5	h .98	99.2	99.2	BAT06RP04	1
40	238	243	31.88	4.56	.90	-.1	.42	-1.2	g .96	97.9	97.9	BAT40RC02	1
32	238	243	31.88	4.56	.90	-.1	.45	-1.1	f .96	97.9	97.9	BAT32RP02	1
54	229	243	42.88	2.81	.89	-.4	.65	-1.1	e .92	94.2	94.2	BAT54RC09	1
59	231	243	41.19	3.02	.88	-.4	.56	-1.4	d .93	95.1	95.1	BAT59RC09	1
15	238	243	31.88	4.56	.87	-.2	.30	-1.6	c .97	97.9	97.9	BAT15RC08	1
55	232	243	40.24	3.14	.87	-.4	.49	-1.6	b .93	95.5	95.5	BAT55RC01	1
18	240	243	26.60	5.85	.87	-.1	.15	-1.8	a .98	98.8	98.8	BAT18RC09	1
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		

Table 4. (Continued)

```
TABLE 10.1 Disposition Analysis  Spring 08 Pilot  ZOU146WS.TXT Mar 20 11:00 2008
INPUT: 335 persons  90 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

         "ETQ" item STATISTICS:  MISFIT ORDER

+-----------------------------------------------------------------------------------------+
|ENTRY    RAW                  MODEL|   INFIT  |  OUTFIT  |PTMEA|EXACT MATCH|            |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.| OBS%  EXP%| item      G |
|------------------------------------+----------+----------+-----+-----------+------------|
|    67    273    135   49.26     .99|1.40   3.1|1.37   2.9|A .67| 37.0  42.9| ETQ07**07 2 |
|    66    258    133   50.19     .99|1.07    .7|1.11   1.0|B .45| 43.6  42.9| ETQ06**06 2 |
|    65    301    136   46.78     .99|1.06    .5|1.06    .6|C .64| 45.6  43.1| ETQ05**05 2 |
|    70    265    133   49.88     .99|1.01    .1|1.01    .2|D .65| 49.6  43.2| ETQ10**10 2 |
|    61    238    136   52.62     .99|1.01    .1| .99    .0|E .69| 44.1  42.6| ETQ01**01 2 |
|    64    251    135   51.38     .99| .92   -.7| .91   -.7|e .66| 48.1  42.5| ETQ04**04 2 |
|    68    228    134   53.35     .99| .92   -.7| .90   -.8|d .62| 36.6  42.5| ETQ08**08 2 |
|    69    231    133   52.94    1.00| .87  -1.2| .87  -1.1|c .68| 48.9  42.4| ETQ09**09 2 |
|    63    289    134   47.54     .99| .86  -1.2| .86  -1.2|b .68| 48.5  43.2| ETQ03**03 2 |
|    62    309    137   46.07     .98| .85  -1.3| .85  -1.3|a .69| 51.8  43.3| ETQ02**02 2 |
|------------------------------------+----------+----------+-----+-----------+------------|
| MEAN   264.3  134.6   50.00     .99|1.00   -.1| .99   -.1|     | 45.4  42.9|            |
| S.D.    27.1    1.4    2.47     .00| .15   1.2| .15   1.2|     |  4.9    .3|            |
+-----------------------------------------------------------------------------------------+

TABLE 10.1 Disposition Analysis  Spring 08 Pilot  ZOU854WS.TXT Mar 20 10:56 2008
INPUT: 335 persons  90 items  MEASURED: 193 persons  20 items  5 CATS       3.64.2
-------------------------------------------------------------------------------
person: REAL SEP.: 1.55  REL.: .71 ... item: REAL SEP.: 2.04  REL.: .81

         "SRA" item STATISTICS:  MISFIT ORDER

+-----------------------------------------------------------------------------------------+
|ENTRY    RAW                  MODEL|   INFIT  |  OUTFIT  |PTMEA|EXACT MATCH|            |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.| OBS%  EXP%| item      G |
|------------------------------------+----------+----------+-----+-----------+------------|
|    88     69     52   52.60    1.73|1.51   2.6|1.45   2.1|A .41| 34.6  44.9| SRA18**09 3 |
|    76    192    105   46.89    1.23|1.47   3.2|1.40   2.8|B .49| 41.9  45.3| SRA06**03 3 |
|    89     89     87   59.00    1.38|1.35   2.3|1.25   1.7|C .45| 43.7  44.1| SRA19**10 3 |
|    82    166    105   50.75    1.21|1.23   1.8|1.22   1.6|D .58| 39.0  44.6| SRA12**06 3 |
|    90    159    105   51.79    1.21|1.14   1.1|1.14   1.1|E .47| 42.9  44.3| SRA20**10 3 |
|    78    169    105   50.31    1.21|1.11    .9|1.13   1.0|F .48| 54.3  44.6| SRA08**04 3 |
|    84    202    104   45.02    1.25|1.13   1.0|1.10    .8|G .64| 40.4  46.4| SRA14**07 3 |
|    86    165    105   50.89    1.21|1.03    .3|1.07    .6|H .57| 41.9  44.6| SRA16**08 3 |
|    71    141    100   51.84    1.24|1.06    .5|1.05    .4|I .43| 50.0  44.3| SRA01**01 3 |
|    80    190    104   46.87    1.23|1.02    .2|1.05    .4|J .54| 39.4  45.3| SRA10**05 3 |
|    74    189    105   47.34    1.22|1.02    .2| .99    .0|j .62| 45.7  45.4| SRA04**02 3 |
|    72    166    103   50.36    1.22| .93   -.5| .98   -.1|i .57| 46.6  44.7| SRA02**01 3 |
|    87    116     84   53.40    1.34| .89   -.8| .87   -.9|h .58| 50.0  43.5| SRA17**09 3 |
|    77    141     85   48.97    1.33| .82  -1.3| .81  -1.4|g .65| 48.2  44.4| SRA07**04 3 |
|    83    163     83   44.70    1.38| .80  -1.4| .82  -1.3|f .53| 55.4  45.0| SRA13**07 3 |
|    85    140     83   48.86    1.35| .81  -1.4| .81  -1.4|e .64| 47.0  44.5| SRA15**08 3 |
|    81    143     84   48.41    1.34| .78  -1.7| .79  -1.5|d .73| 52.4  44.4| SRA11**06 3 |
|    79    128     86   51.65    1.32| .70  -2.5| .69  -2.5|c .63| 53.5  43.7| SRA09**05 3 |
|    75    130     86   51.30    1.32| .61  -3.3| .62  -3.1|b .60| 55.8  43.7| SRA05**03 3 |
|    73    139     85   49.04    1.33| .58  -3.5| .59  -3.4|a .50| 57.6  44.2| SRA03**02 3 |
|------------------------------------+----------+----------+-----+-----------+------------|
| MEAN   149.9   92.8   50.00    1.30|1.00   -.1| .99   -.2|     | 47.0  44.6|            |
| S.D.    32.9   13.3    3.14     .12| .25   1.8| .23   1.7|     |  6.3    .6|            |
+-----------------------------------------------------------------------------------------+
```

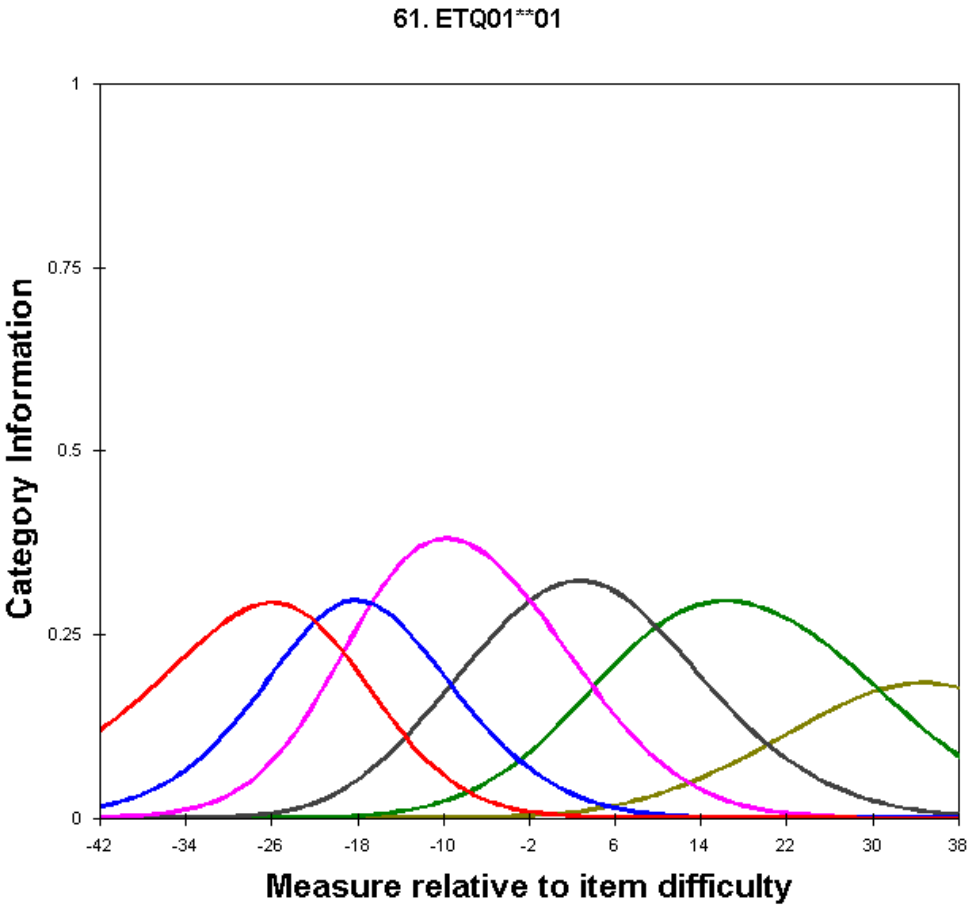

Figure 2. Threshold for the ETQ Rating Scale based on Krathwohl

0=Unaware
1=Receiving
2=Responding
3=Valuing
4=Organizing
5=Characterizing

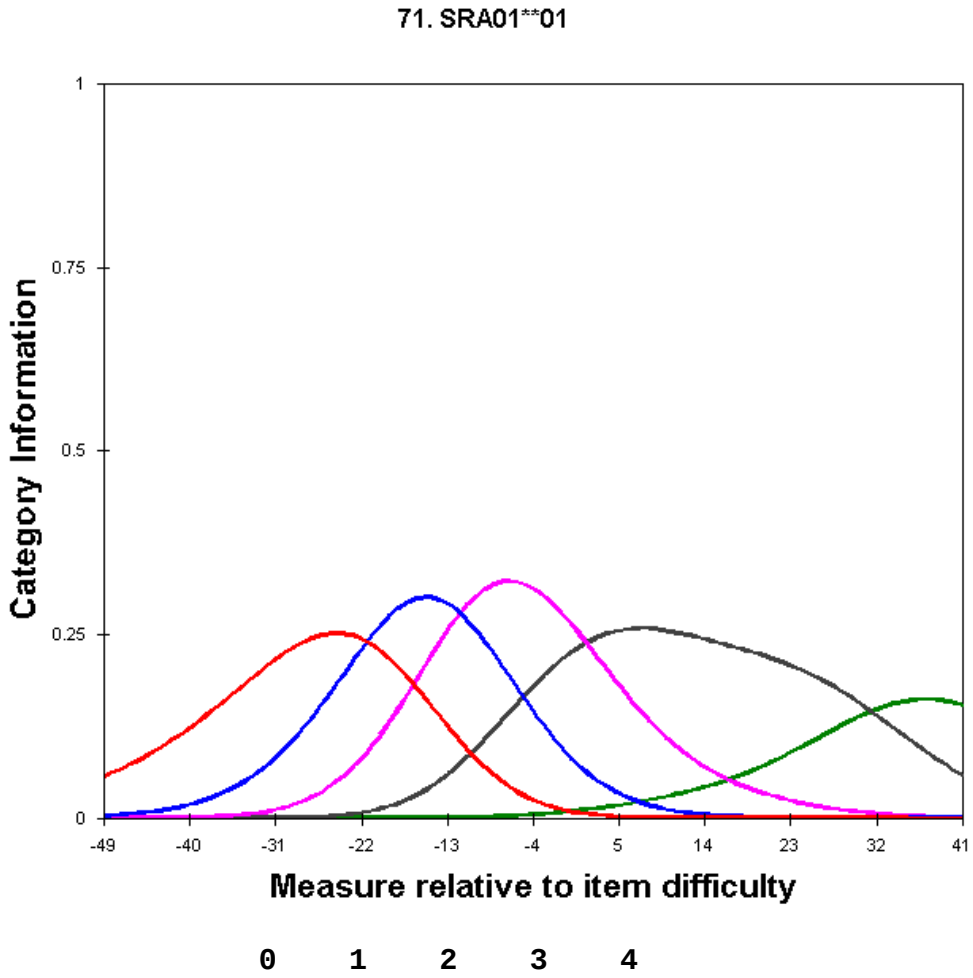

Figure 3. Threshold for the SRA Rating Scale based on Krathwohl

0=Unaware
1=Receiving
2=Responding
3=Valuing
4=Organizing
5=Characterizing
(No 5'S observed)

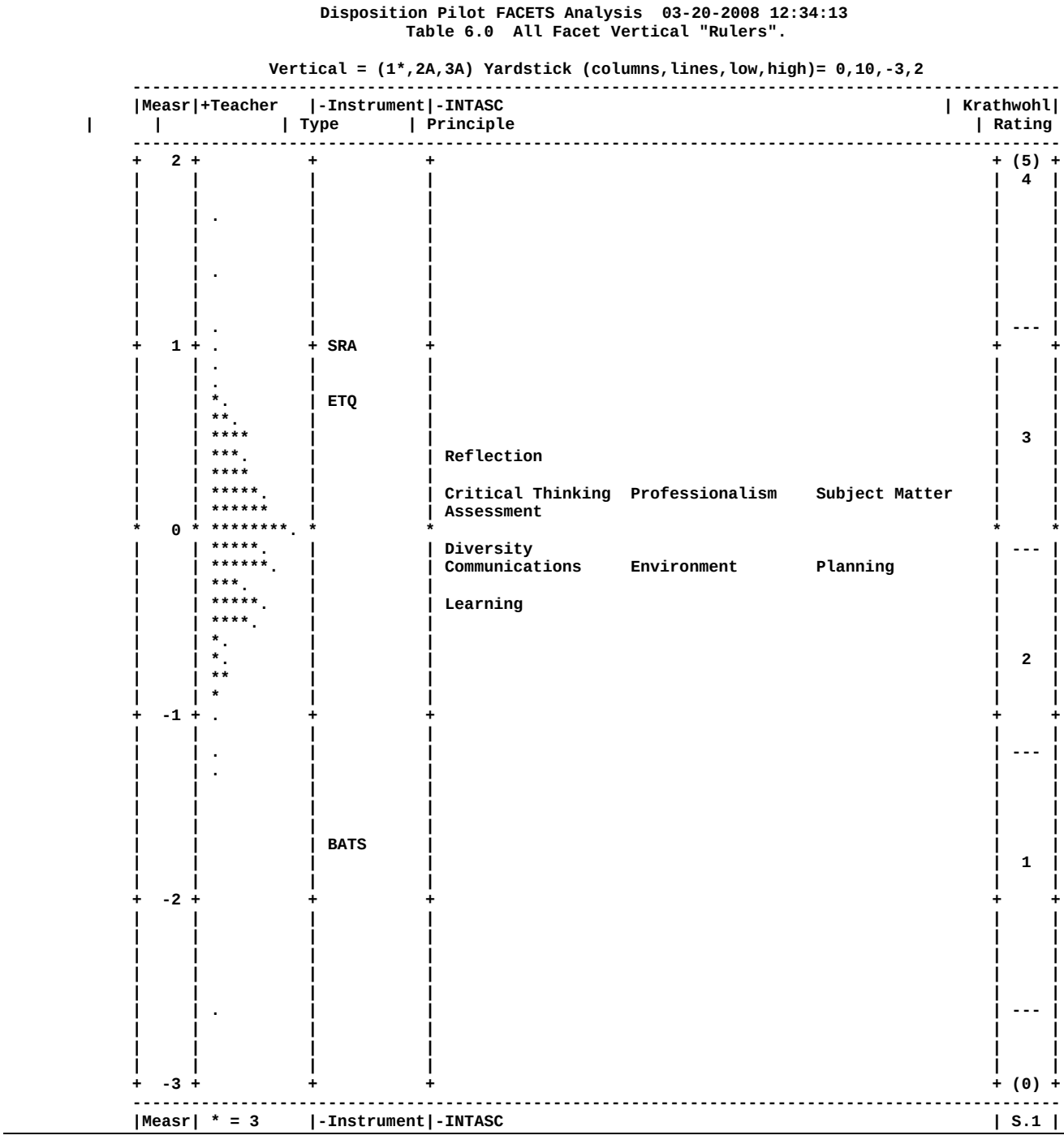

Figure 4. FACETS Analysis of Ten INTASC Principles and Three Instrument Types

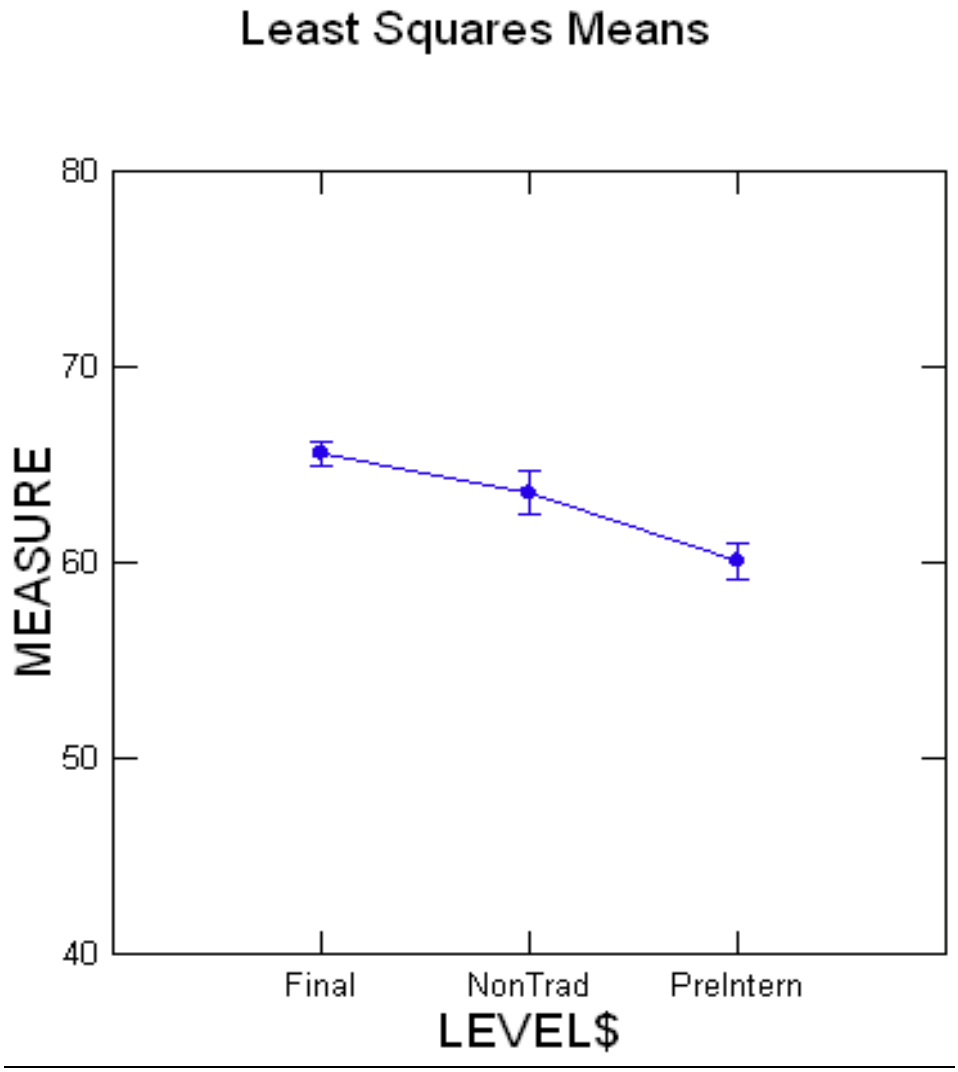

Figure 5. Analysis of Disposition Scores Across Placement

This figure is a subsample of the pilot where class standing was available. 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.

Table 5. Results of the FACETS Analysis of BATS, ETQ, and SRA

Disposition Pilot FACETS Analysis 03-20-2008 12:34:13
Table 7.2.2 Instrument Measurement Report (arranged by fN).

Obsvd Score	Obsvd Count	Obsvd Average	Fair-M Avrage	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	N	Instrument
2492	1276	2.0	1.93	.71	.03	1.02	.4	1.05	1.3	.98	2	ETQ
10749	12916	.8	.85	-1.74	.02	1.01	.3	1.04	2.1	.99	1	BATS
2711	1749	1.6	1.62	1.04	.03	.90	-3.5	.93	-2.3	1.12	3	SRA
5317.3	5313.7	1.4	1.47	.00	.03	.97	-.9	1.01	.4			Mean (Count: 3)
3841.8	5379.1	.5	.46	1.24	.00	.05	1.9	.06	2.0			S.D.

RMSE (Model) .03 Adj S.D. 1.24 Separation 46.71 Separation Reliability 1.00
Fixed (all same) chi-square: 7359.3 d.f.: 2 significance (probability): .00
Random (normal) chi-square: 2.0 d.f.: 1 significance (probability): .16

Disposition Pilot FACETS Analysis 03-20-2008 12:34:13
Table 7.3.1 INTASC Measurement Report (arranged by mN).

Obsvd Score	Obsvd Count	Obsvd Average	Fair-M Avrage	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Nu	INTASC
1352	1552	.9	.69	.40	.05	1.07	1.4	1.17	4.1	.85	9	Reflection
1504	1605	.9	.80	.23	.05	1.06	1.3	1.26	5.7	.78	4	Critical Thinking
1525	1606	.9	.82	.19	.05	.83	-3.9	.87	-3.0	1.14	1	Subject Matter
1527	1601	1.0	.82	.17	.05	1.02	.3	1.01	.2	1.00	10	Professionalism
1572	1603	1.0	.86	.08	.05	.92	-1.6	1.03	.5	1.03	8	Assessment
1669	1604	1.0	.94	-.14	.05	1.00	.0	1.08	1.5	.97	3	Diversity
1680	1605	1.0	.95	-.16	.05	.97	-.5	1.12	2.1	.95	5	Environment
1641	1556	1.1	.96	-.17	.05	1.17	3.1	1.21	3.4	.86	7	Planning
1697	1602	1.1	.96	-.21	.05	.86	-2.8	.79	-4.1	1.18	6	Communications
1785	1607	1.1	1.04	-.40	.05	.81	-3.7	.77	-3.9	1.18	2	Learning
1595.2	1594.1	1.0	.88	.00	.05	.97	-.6	1.03	.7			Mean (Count: 10)
117.8	20.1	.1	.10	.24	.00	.11	2.3	.16	3.3			S.D.

RMSE (Model) .05 Adj S.D. .23 Separation 4.88 Separation Reliability .96
Fixed (all same) chi-square: 243.4 d.f.: 9 significance (probability): .00
Random (normal) chi-square: 9.0 d.f.: 8 significance (probability): .34

Disposition Pilot FACETS Analysis 03-20-2008 12:34:13
Table 8.1 Category Statistics.

Model = ?,?,?,R

DATA Score	DATA Category Used	DATA Counts %	DATA Cum. %	QUALITY CONTROL Avge Meas	QUALITY CONTROL Exp. Meas	QUALITY CONTROL OUTFIT MnSq	STEP CALIBRATIONS Measure	STEP CALIBRATIONS S.E.	EXPECTATION Measure at Category -0.5	MOST PROBABLE from	.5 Cumul. Probabil. at	Cat PEAK Prob
0	426	14%	14%	-1.40	-1.48	1.1			(-3.34)	low	low	100%
1	859	28%	42%	-1.24	-1.17	.9	-2.02	.06	-1.80	-2.62	-2.02	-2.33	40%
2	1011	33%	76%	-.93	-.92	1.2	-1.21	.04	-.68	-1.22	-1.21	-1.20	40%
3	603	20%	96%	-.64	-.66	1.0	-.28	.05	.47	-.13	-.28	-.18	46%
4	117	4%	100%	-.13	-.36	.8	1.12	.09	1.87	1.13	1.12	1.11	47%
5	9	0%	100%	-.05	.00	1.0	2.38	.34	(3.63)	2.84	2.38	2.59	100%
									(Mean)		(Modal)	(Median)	

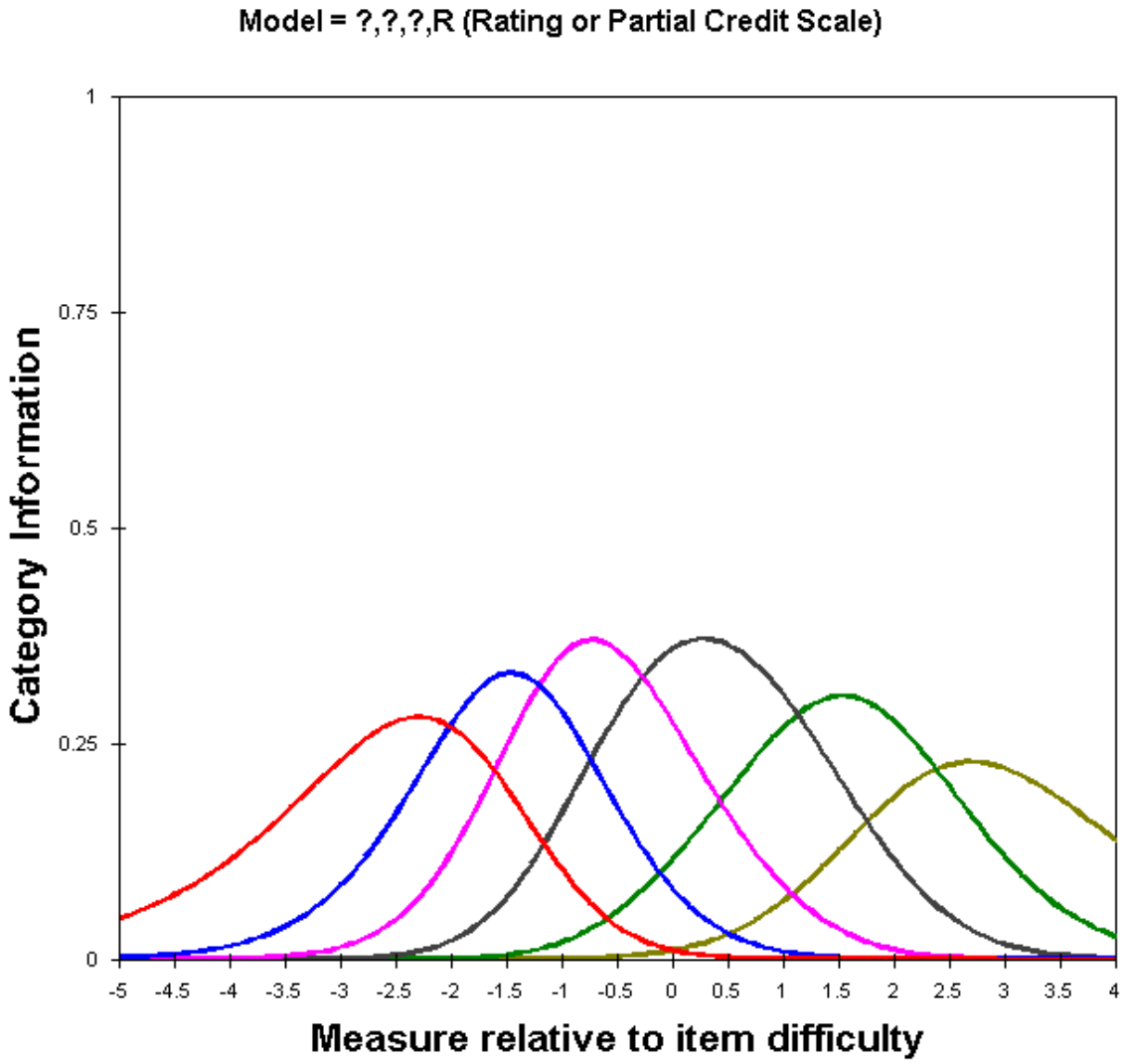

Figure 6. Threshold Analysis of Disposition Scores in FACETS

Conclusions and Discussion

The INTASC dispositions are a complex construct best measured using different item structures that presuppose a dimensional difficulty for Rasch models. Accreditation assessment often presupposes the aggregation of results by multiple overlapping frameworks. The importance of this work is in its potential to provide a new application for Rasch measurement. In this case, the initial study seeks to examine the results as one scale that is also informative as mapped originally on multidimensional framework and reportable along multiple informative scales.

As suggested by Wight & Stone (1999), the focused scale can still be deconstructed and provide useful information if the unifying elements were originally applied as a single unifying dimension. Even though there appears to be an interest in the mathematics of complex Rasch models as expressed in Davier & Carstensen's (2007) *Multivariate and Mixture Distribution Rasch Models: Extensions and Applications*, the original mapping of the construct with forethought along multiple dimensions appears to allow an acceptable Rasch model that also provides reasonable results while using traditional calibration methods.

In this case, the most useful representation comes from Figure 4. The BATS Thurstone scale is a screening instrument that is subject to faking and generally utilized with beginning education majors who are pre-internship. It appears to be the easiest on the FACETS analysis. ETQ is a reflective questionnaire that is aimed at teacher candidates near their final internship as they have their first experiences as teachers. It appears to be the next most difficult. Finally, the SRA is a projective instrument (apperception test) that would range in scores into the responses expected by experienced inservice teachers. The placement of the SRA meets that expectation.

The sample are mostly students at the end of their preservice experience. The authors had suggested that a typical student would be at the "valuing" level about the time of graduation and certification. The mid-range of our sample is in the valuing interval, which is expected. The sample contained a few beginning graduate students who scored into the "organizing" interval, and some lower division students were in the "receiving" and "responding" intervals.

In conclusion, there appears to be ample evidence that the original mapped construct is useful for instrument development using the Rasch model, even though the overfit and underfit with different item types is confounding. The category order is consistent empirically with this sample size and expected ability. The ten INTASC Principles also calibrate in an expected range and order.

The primary limitation here was the range restriction due to a lack of inservice and graduate students, rubric revision that was necessary during the pilot due to rater error, and a minimal but adequate sample size for pilot work.

Selected Bibliography

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

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

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

Davier, M. & Carstensen, C. (2007). *Multivariate and Mixture Distribution Rasch Models: Extensions and Applications.* NY: Springer

Krathwohl, D., Bloom, B. & Masia, B. (1956). *Taxonomy of educational objectives. Handbook II: affective domain.* New York, McKay.

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

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

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

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

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

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

Wright, B. & Stone, M. (1999). *Measurement Essentials (2nd).* Wilmington, DE: Wide Range.

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