International Journal of Learning, Teaching and Educational Research
Vol. 17, No. 8, pp. 56-77, August 2018
https://doi.org/10.26803/ijlter.17.8.4

Beliefs about Teaching (BATS2) - Construction and Validation of an Instrument based on InTASC Critical Dispositions

W. Steve Lang and LaSonya Moore
University of South Florida St. Petersburg
St. Petersburg, Florida, USA

Judy R. Wilkerson, Christopher M. Parfitt, Jackie Greene, Diane Kratt, C. Dawn Martelli, Kyle LaPaglia, Vickie Johnston, Shelby Gilbert and Jason Zhang
Florida Gulf Coast University
Ft. Myers, Florida, USA

Lynette Fields
Pinellas County Schools
Largo, Florida, USA

Abstract. A team of researchers at two institutions revised and analyzed a battery of instruments to assess the Critical Dispositions (InTASC, 2013) required in the CAEP (2016a) accreditation standards for teacher education programs. This research presents initial findings for the revised version updating previous results from validity and reliability studies of the first version (Wilkerson & Lang, 2011). An in-depth study of one of the instruments, now in two forms, is presented. Version 2 was necessary because the standards providing an operational definition of the construct measured were updated. In this study, data were collected from teacher education students, in service teachers, and pre-school teachers (Form A = 1072; Form B = 372). Item analysis using Rasch modeling, results of a qualitative review of specific teacher candidates across multiple measures, and student/program improvement uses are discussed. The results indicated that evidence of validity and reliability is maintained in the new version, and student disposition measures were diagnostic and logical for students of different training and experience.

Keywords: teacher dispositions; InTASC standards; Rasch modeling.

1. Introduction

Competent teaching is believed to contribute to student learning (Boonen et al., 2014; Collinson, Killeavy, & Stephenson, 1999; Darling-Hammond, 2000; Heck, 2009; Ransdell, 2017), and dispositions have been connected positively to

improved student outcomes (Vaughn, 2012). School-based administrators not only want teachers to develop content and skills, but also appropriate professional dispositions, because teachers influence the development of their students, not only academically but also socially and emotionally (Osguthorpe, 2008). When administrators hire a teacher with effective dispositions, students learn and develop, parents are satisfied, and district administrators can focus on the business of education (Wasicsko, 2004).

Three approaches to assessing teacher dispositions are predominate in the literature: (a) research-based frameworks of Marzano and Danielson (Alexander, 2016; Donahue, 2016; Graziano, 2017; Marzano, 2012; Quinn, 2014; Sargent, 2014; Wilkins, 2017), (b) teacher evaluation domains (Alexander, 2016; Quinn, 2014; Sargent, 2014; Wilkins, 2017), and (c) the national standards developed by the Interstate Teacher Assessment and Support Consortium (InTASC) (CCSSO, 2013; Klute, Apthorp, Harlacher, & Reale, 2017; Lang et al., 2018a&b; Sargent, 2014). This study will focus on the third approach. Assessing teacher dispositions using the InTASC Standards is a requirement for the accreditation of educator preparation programs by the Council for Accreditation of Teacher Preparation (CAEP, 2016a), as it has been for decades, surviving the changes in name, standards, and processes as the National Council for Accreditation of Teacher Education (NCATE) evolved into CAEP. At present, a new group is attempting to take over the teacher preparation accreditation function, the Association for Advancing Quality in Educator Preparation (AAQEP) has put forth a new set of standards, the first of which is Completer Performance, requiring that "Candidates and completers exhibit the knowledge, skills, and professional dispositions of competent, caring, and effective professional educators" (AAQEP, 2018, p. 6). While AAQEP breaks the link with InTASC, it maintains the connection to dispositions, listing "dispositions and behaviors required for successful professional practice," as its sixth and final expectation. Despite the fact that the requirement to assess teacher dispositions has withstood the test of time, it remains today a source of confusion and difficulty.

2. Problem Statement

While teachers' knowledge and skills has been a compelling, but under-examined paradox (Moore, 2016), as it relates to the sophisticated inquiry and application has forced recent requirements. While there is substantial research on assessing teacher knowledge and skills, there is little on assessing teacher dispositions. The knowledge and skill component of the teacher education curriculum is typically well assessed using multiple measures, while the affective component receives much less attention and sophistication. The focus in this research is on measuring standards-based (InTASC) teacher dispositions, which is an accreditation requirement and also important to lead to comprehensive decisions about teacher effectiveness. This is an issue of importance to all educator preparation programs accredited both in the USA and internationally by CAEP.

In its *Accreditation Handbook,* CAEP (2016b) defines dispositions as "The habits of professional action and moral commitments that underlie an educator's

performance (InTASC Model Core Teaching Standards, p. 6.)" (p. 180). The focus here, as in previous research, continues to be standards-based "habits of professional action," rather than morality-based commitments (Wilkerson, 2006). Objective measurement of teacher dispositions provides a mechanism to determine the level of commitment, and that can lead to opportunities to celebrate or intervene, as appropriate. This research presents an application of the Rasch model of item response theory to meet that challenge rather than using the typical raw score approach.

3. A Summary of the Requirements from CAEP and InTASC

All five CAEP standards touch on assessing dispositions in credible ways. CAEP Standard 1 specifically requires a standards-driven approach, making use of the InTASC Standards. CAEP Standard 2 requires assessment in clinical settings; CAEP Standard 3 requires non-cognitive assessment specifically at multiple points in time from admission to graduation; CAEP Standard 4 requires "valid and reliable data;" CAEP Standard 5 requires "empirical evidence that interpretations are valid and consistent." All program assessments, both cognitive and affective, need to be evaluated as high quality and demonstrate usefulness to measure impact.

There are 10 InTASC standards, divided into four categories. Each InTASC Standard includes a set of statements delineating knowledge, performances, and critical dispositions. The categories and standards are:

1. "The Learner and Learning (Learner Development, Learning Differences, and Learning Environments)
2. Content Knowledge: Content Knowledge and Application of Content
3. Instructional Practice: Assessment, Planning for Instruction, and Instructional Strategies
4. Professional Responsibility (Professional Learning and Ethical Practice and Leadership/Collaboration"

There are a total of 43 critical dispositions statements, spread among the 10 InTASC Standards within the four categories.

4. The DAATS Model and Battery

The *Dispositions Assessments Aligned with Teacher Standards* (DAATS) is a theoretical model and design process described by Wilkerson and Lang (2007). The model suggests various strategies for assessing affect and presents the literature supporting each strategy along with design techniques useful for instrument development. It advocates for the use of multiple measures, similar to those needed in the cognitive domain (Fuller, Fitzgerald, & Lee, 2008; Herman, Baker, & Linn, 2004).

The creators of the DAATS model also advocate for the use of multiple standards sets and taxonomies in instrument design, including the InTASC Standards, Bloom and Krathwohl (1956) affective taxonomy, and the Standards of Educational and Psychological Testing (AERA, APA, NCME, 2013). The DAATS model presents strategies for quality control and interpretation with

various applications for validity and reliability testing and use of the Rasch model of item response theory (Rasch, 1960) where feasible. Extensive discussion of the literature, the model, the instruments, and the results has been presented previously (Wilkerson & Lang, 2004; Wilkerson & Lang, 2006; Lang & Wilkerson, 2008; and Englehart, Batchelder, Kelly, Wilkerson, Lang, & Quinn, 2011; Wilkerson & Lang, 2011; Wilkerson, 2012, Lang et al., 2018a; 2018b).

The DAATS battery (Wilkerson & Lang, 2006) consists of five instruments, all of which measure all 10 INTASC Standards. The battery includes:

- "Beliefs About Teaching Scale (BATS): a Thurstone agreement scale useful as a pre-admissions and progress monitoring tool.
- Experiences in Teaching Questionnaire (ETQ): a set of constructed response items about prior experiences, useful for progress monitoring and continuing development.
- Situational Reflection Assessment (SRA): constructed response items, using picture prompts (Slitkin, 2007), in a thematic apperception format.
- Classroom Behaviors Checklist (CBC): paired positive and negative behaviors useful for progress monitoring and formerly called the Classroom Dispositions Checklist of (CDC).
- K-12 Dispositions Impact (KIDS): focus group of clustered prompts measuring children's perceptions."

The first four instruments are in various stages of updating for the revised InTASC Standards and are being field tested. BATS2 now has two forms (with the aim of establishing gains), psychometric analysis (for validity, reliability, & scaling), and field-testing (for generalizability).

The battery attempts to place students on a modified version of the Krathwohl's affective taxonomy (Bloom & Krathwohl, 1956), with a scale point for "pre-receiving" or "unaware" added to the original five levels to pinpoint respondents who are clearly uncommitted to the measured dispositions (Wilkerson & Lang, 2011). The original taxonomy includes Receiving, Responding, Valuing, Organizing, and Characterizing, because the original taxonomy was designed for instruction and not assessment, omitting the possibility that respondents might have no discernable commitment. Behaviors typical of each level, including unaware, in this application to teaching are presented by Wilkerson (2012, Figure 3).

5. Beliefs About Teaching (BATS) -- Versions 1 and 2

The first version of the DAATS Battery, including BATS, was based on the 1992 InTASC Standards (CCSSO, 1992). The original principles, now called standards, are presented in the 2013 version (CCSSO, 2013). Version 1 instruments had already demonstrated predictive validity and exceptional reliability (Lang, 2008; Wilkerson & Lang 2006, 2009), so the work necessary for version 2 was centered on ensuring that existing items were appropriately aligned with the current standards, adding new items where new dispositions were introduced, and developing a second form of the test to provide opportunities for pre- and post-testing to check for gains.

BATS1 and BATS2 use a Thurstone (1928) format. Thurstone's technique requires a dichotomous decision (agree/disagree only), while Likert provides for a rating scale, typically five-points, from strongly agree to strongly disagree with a neutral midpoint. "Roberts, Laughlin, and Wedel (1999) examined the relationship between Likert and Thurstone agreement scaling, recommending the Thurstone scale when extreme positions (e.g., high/low levels of commitment) are of interest." (Wilkerson, 2012). In the case of teacher dispositions, high levels of commitment are the norm, but low levels are of particular interest for diagnostic purpose.

Feedback to students for BATS1 did not provide meaningful feedback to students and needed a better connection to the Krathwohl levels. Figure 1 illustrates the change implemented to meet that need. Scaled score ranges were created, using a stanine scale, with interpretations written based on the taxonomic level. For example, students above 79.17 are interpreted as being at the "characterizing" level, while students below 47.13 are interpreted as being "unaware." The 56.3-70.02 range represents the expected level – "valuing" in the taxonomy.

Scaled Score Range	**Interpretation**
Above 79.17	You are **deeply committed and passionate** about teaching and the "critical dispositions" of teaching, as defined in the InTASC Standards. You are likely to **characterize** much of your life around teaching and these beliefs, which may become the central and driving force in your life. You may want to be careful not to over-commit.
70.03 – 79.16	You are **strongly committed** to the "critical dispositions" of teaching, as defined in the InTASC Standards. Your passion is strong but balanced with other aspects of your life. You are likely to **organize** your life to ensure sufficient time to do all that you believe you need to do, setting aside time to plan conscientiously and systematically for your students.
56.3 - 70.02	You are **committed** to the "critical dispositions" of teaching, as defined in the InTASC Standards. You **value** these dispositions sufficiently to apply them consciously in your practice whenever you see an opportunity to do so. **This is your expected score range.**
51.72 - 56.29	You are **making progress** toward building your understanding of, and commitment to, the "critical dispositions" of teaching, as defined in the InTASC Standards. Your **response** to your beliefs is not yet systematic and may not be conscious. While it is clearly visible, it may occur haphazardly or by chance.
47.15 – 51.71	You **understand** the "critical dispositions" of teaching, as defined in the InTASC Standards. You act accordingly on occasion indicating that you have **received** them in your mind, but it is relatively rare.

Below 47.13	You do **not seem to be aware** of the "critical dispositions" of teaching, as defined in the InTASC Standards. They are not yet likely to influence you behavior in the classroom. Read them carefully and reflect on how they can be visible in your future practice.

Figure 1. Interpretation Guide Provided to Students

6. Method (Research Design)

The research design had three components. Consistent with Rasch item development techniques, items were developed based on a careful analysis of the language of the InTASC Standards and the Krathwohl Affective Taxonomy, with as many Critical Dispositions as possible and all six levels of the taxonomy targeted. Second a quantitative analysis was conducted using the Rasch model of item response theory, beginning with the evaluation of separation statistics and item and person fit measures, including the development of item maps. Third, qualitative examination of individual subjects was completed for low scoring students to determine if scores were as expected and what remediation strategies might be used. The Rasch approach established separation reliability and distributional validity. The qualitative work was used to support validity of the constructs measured by the instruments.

6.1 Item Creation

Five members of the faculty representing both institutions used items from the original version of BATS, authored new items, and aligned items with the Critical Dispositions of the 10 InTASC Standards organized into InTASC's four groups (InTASC, 2013). The item-writing team specifically targeted Krathwohl levels, as well as the InTASC statements for each item, and they also attempted to balance the tests by standard and taxonomic level for both forms.

For example, Standard #2, Learning Differences, includes: *"The teacher respects learners as individuals with differing personal and family backgrounds and various skills, abilities, perspectives, talents, and interests."* BATS2 items ask a student to Agree/Disagree with statements such as:

"I usually think about children's home life and environment so that I can tell if something is wrong." [Valuing]

"I have a rule in my classroom: 'We all speak proper English and ignore gestures, slang, or foreign languages.'" [Unaware]

The original test was 60 items in length and now consists of two sets of 50 items. The intent was to produce scores that were normally distributed, scaled using the Rasch model for interval-level data, standards-based, and diagnostic in terms of the InTASC Standards and Krathwohl Taxonomy. We wanted to avoid an instrument that suffered from little construct validity, ceiling effects, or simple raw-score reporting, but we also anticipated a ceiling effect with a bunching effect at the top of the distribution of scores and a skewed distribution because the sample population is typically homogeneous (teacher candidates

with sufficient GPA to be considered for, or already admitted to, a teacher preparation program). The social expectation is that most students will be correct on most items. The combination of sample characteristics and items of easy difficulty doom many instruments to poor validity or else resistance by faculty to use them. BATS2 was designed to include enough difficulty to form a normal distribution, but also enough easy items that students and faculty do not balk at its use.

After a pilot test, the items were reviewed again with adjustments made based on results. Though not reported here, five additional common items were added to both forms for future equating.

6.2 Participants

Students from the teacher education programs at two state universities were given Form A or Form B of BATS2 via Survey Monkey or Canvas LMS in the spring, summer, and fall of 2017, and in spring of 2018. Respondents represented lower division undergraduate (including certification & non-certification students), upper division undergraduate, and graduate students. Form A was taken by 1072 students and Form B was taken by 372 students in the analysis reported here. A subset of students took both forms for the purpose of future equating analysis. BATS2 was administered online, and the average time for completion was 9 minutes and 42 seconds.

6.3 Basic Rasch Analysis and Interpretation

Winsteps software (Linacre, 2017) was used to calibrate item measures according to the Rasch model of item response theory. To find model parameters, the Rasch model provides sophisticated and precise results based on mathematical models for the data. The basic model for dichotomous data is:

$$ø_{nil} = \frac{\exp(\beta_n - \delta_{i1})}{1 + \exp(\beta_n - \delta_{il})}$$

where $ø_{nil}$ is person n's probability of scoring 1 on item i, β_n is the ability of person n, and δ_{i1} is the difficulty level of item i. Therefore, the probability of success to answer a question correctly is governed by person ability and item difficulty. A summary of the calibration diagnostics are provided in Table 1:

Table 1. Summary Rasch Statistics, BATS2

	Form A, N=1072		Form B, N=372	
Persons	Separation Reliability	Outfit Z-Standard	Separation Reliability	Outfit Z-Standard
	.64+	.0	.66+	.1
	α=.69	1.1 (SD)	α=.69	.8 (SD)
Items	.99*	.0	.97*	.0
		3.9 (SD)		1.7 (SD)

+Indicates that the scale discriminates between persons adjusted for misfit in the data
*Similar to internal consistency indicating items create a well-defined variable

The internal consistency reliability (items) of both instruments is excellent and, the ability of the instrument to distinguish people on the scale is adequate. In Table 1, the expected Z-Standard is 0.0 and the expected SD is 1.0. The only value that is not on target in the two forms is the SD of 3.0 in the Outfit Z-standard. This indicates that some items in the set may be misfitting. A subsequent review indicated four items that may need revision, but all four were substantially difficult for this population so the misfit may be partially due to the relatively extreme measure. A review and rewrite of misfitting items did not significantly alter the instrument pattern for extreme items, so they appear to be a consequence of position and a sensitivity to sample size. Dropping the four misfitting items and rerunning the analysis (Form A) did not change the measures of the top and bottom students more than .2 logits, and the Person Separation improved to .64 (from .55). The person separation is likely affected by the circumstances of the sample, where the length of the test is moderate, the sample ability is moderate, and the categories per item are restricted. As the Winsteps (2018) program manual suggests:

> **"Person (sample, test) reliability** depends chiefly on
> 1) Sample ability variance. Wider ability range = higher person reliability.
> 2) Length of test (and rating scale length). Longer test = higher person reliability
> 3) Number of categories per item. More categories = higher person reliability
> 4) Sample-item targeting. Better targeting = higher person reliability"

The variable (or Wright) maps from Winsteps are provided in Figure 2 and Figure 3. The maps illustrate the distribution of person commitment (left) and item difficulty (right). At the bottom are the least committed persons and the easiest items. At the top are the most committed persons and most difficult items. The maps show normally distributed groups of persons, which is expected. Items show good coverage of the construct in both forms. The distribution of items and persons supports the confidence in validity.

```
TABLE 12.2 C:\Users\baywoof\Desktop\WinstepsCont ZOU959WS.TXT2 May 31 2018 15:37
INPUT: 1072 Person  50 Item  REPORTED: 1072 Person  50 Item  2 CATS WINSTEPS 4.0.0
-------------------------------------------------------------------------------
    MEASURE    Person - MAP - Item
    90                +
                      |
                    . |
                      |
                   .# |
                      T|  12AP9C4A
                   .# |T 15AP0C4A  16AP3C1D
    80                +
              .###### |
                      |  13AP8C3A
         .########## S|  49AP6C3A
       .#############  |  06AP0C4D
                      |
       .#############  |
    70 #############  +
         .########## M|  46AP5C2A
            .#######  |
            .#######  |S
             .######  |  01AP8C3D  32AP7C3D
               .####  |  18AP9C4A  31AP3C1D
                 .## S|  08AP6C3D  23AP5C2D
    60              .#  +  02AP7C3D  48AP5C2A
                    .#  |  11AP6C3A
                    .#  |  25AP8C3D
                    .#  |  40AP3C1D  44AP8C3D
                    .# T|  43AP9C4A
                    .#  |  47AP4C2D  50AP9C4A
                        |  27AP3C1D  42AP2C1A
    50               .  +M 28AP6C3A  35AP9C4D
                     .  |  10AP5C2D  19AP2C1D
                        |
                        |  20AP0C4A  24AP2C1D  33AP1C1A  38AP6C3D  41AP7C3D
                        |
                        |
    40                  +
                        |  04AP2C1A  34AP5C2D  39AP8C3A
                        |  17AP5C2A  36AP3C1A  37AP4C2A
                        |
                        |S 29AP6C3A
                        |
                        |  21AP3C1A  26AP1C1A  45AP0C4A
    30                  +  09AP5C2A  14AP9C4A  30AP8C3A
                    .   |  03AP1C1A  05AP5C2A
                        |  07AP1C1A  22AP1C1A
                        |
                        |
                    .   |
                        |
    20                  +
                        |T
                        |
                        |
                        |
                        |
                        |
    10                  +
                        |
                        |
                        |
                        |
                    .   |
                        |
     0                  +
  EACH "#" IS 10: EACH "." IS 1 TO 9
```

Figure 2. Variable Map of BATS2 Form A

```
TABLE 12.2 C:\Users\baywoof\Desktop\WinstepsCont ZOU066WS.TXT2 May 31 2018 10:55
INPUT: 372 Person  50 Item  REPORTED: 372 Person  50 Item  2 CATS WINSTEPS 4.0.0
-------------------------------------------------------------------------------
MEASURE    Person - MAP - Item
  100               +
                     |
                  .  |
                     |
                     |
                     |
                     |
   90             .#  +
                     |
                    T|
                     |
              .####  |
                     |
               .###  |
   80                +
            .##### S|
                     |T
       ###########   |  33BP9C4D
          .#######   |  24BP0C4A
         #########   |  22BP0C3D
                    M|  04BP6C3D  08BP4C2D  12BP1C1A  26BP7C3A
   70  .###########  +
        .#########   |
            .#####   |  41BP5C2A
           .######   |
            .##### S|S 44BP3C10  50BP2C1D
               ###   |
              .###   |  48BP1C1D
   60           .#   +  32BP8C3A
                 .   |  37BP8C3D  42BP3C1D
                 .   |
                .# T|
                 .   |  10BP4C2A  25BP0C4D  39BP1C1A  45BP6C3A
                 .   |  05BP6C3D  07BP5C2A  14BP8C3D  34BP6C3D  43BP1C1A
   50            .  +M 01BP9C4A  17BP8C3D  20BP6C3A  35BP2C1A
                     |
                 .   |  15BP8C3D  31BP3C1D
                 .   |  16BP3C1A  21BP4C2A  28BP5C2D  40BP1C1D
                     |  03BP7C3A
                     |
                     |  11BP3C1A  23BP2C1D  27BP6C3D  29BP9C4D  49BP2C1A
   40                +  09BP4C2A
                     |  02BP8C3A  36BP7C3A
                     |
                     |S 18BP5C2D  30BP7C3A
                     |  46BP1C1A
                     |
                     |
   30                +  06BP2C1D  47BP2C1A
                     |
                     |
                     |
                     |  13BP1C1A
                     |T
                     |
   20                +  19BP5C2A  38BP1C1A
   EACH "#" IS 4: EACH "." IS 1 TO 3
```

Figure 3. Variable Map of BATS2 Form B

Note that the scale score mean for item calibration was set to 50 on both forms and that the mean for persons is higher for both forms. Ranges are summarized in Table 2.

Table 2. Summary Statistics Scale Scores

BATS2	Mean Scale Score	High Scale Score	Low Scale Score	Raw Score Range (of 50)
Form A	68.69	87.40	46.66	23-47*
Form B	70.89	96.82	46.78	22-49

*One student dropped for an incomplete response

6.4 Discussion of Rasch Analysis

There are a few aspects to keep in mind about a few misfitting items in large samples. Linacre (2010) indicates that with some situations, misfitting items should not be removed if removing them does not improve measurement of the persons. Also, Hagell (2014) concludes in a discussion of Rasch model dimensionality and misfit that, "Statistical procedures and reliance on P-values and CIs cannot compensate for conceptual and theoretical considerations." (p. 463). The authors have continually examined and tweaked the language of a small number of items, but the alignment with InTASC's critical dispositions and locations on the Rasch ruler are primary indicators of validity.

Note that there is a practical dissonance when creating an affective scale for use in colleges of education where the culture is generally humanistic success and the student population is homogeneous. As Bond and Fox (2007) relate, "Person reliability requires not only ability estimates well targeted by the suitable pool of items, but also a large enough spread of ability across the sample so that the measures demonstrate a hierarchy of ability/development (person separation) on this construct." (p. 41). In other words, if value-measuring instruments are created where most of a homogenous population (a narrow spread) is usually consistent with most professional concepts, the person reliability might be expected to be relatively low compared to a skill-based assessment where the average item difficulty is .50 and skill level is wide. Both may be construct appropriate and score results might be normally distributed, but creating items that raised the person separation reliability would result in an instrument that professional schools would balk at using. Instruments of this type are challenging because an average measure that is too difficult will result in professional students/faculty lacking desire to use it. Likewise, an average measure that is too lenient results in a skewed distribution and ceiling effect.

Finally, a self-report instrument is susceptible to response sets where the respondent is aware of the expected answer even though they may not hold the assessed belief. Some pre-service teachers may also have difficulty taking an extreme value position. Again, these may be reasons that the more extreme items appear to misfit the most as students struggle internally with an honest answer vs. recognition of the "preferred" answer. As a screening device, BATS2 can still possess situational validity, but other instruments in the DAATS battery would be useful to confirm professional behavior and values.

Given the Rasch analysis and distribution of measures, we have evidence that BATS2 is reliable internally and possesses construct validity for population of students similar to those tested in this sample. The four (out of 100) misfitting

items will be reviewed again, but they do not have much effect on the person measures since the items are virtually dropped for majority; almost everyone answered inconsistently with InTASC and those who answered consistently did so almost randomly. In fact, four of 100 items total (Form A & B combined) showing as misfitting are possible due to chance with a large sample size ($p < .05$), so again this is not disconcerting.

6.5 Statistical Analysis of InTASC Standards and Groups

Next, we examined a few contrasts of the InTASC construct measured by the BATS2 instruments. The 10 InTASC Core Teaching Standards are organized into four general categories. In summary:

"Group One: The Learner and Learning contains
- #1 Learning Development
- #2 Learning Differences
- #3 Learning Environments

Group Two: Content contains
- #4 Content Knowledge
- #5 Application of Content

Group Three: Instructional Practice
- #6 Assessment
- #7 Planning for Instruction
- #8 Instructional Strategies

Group Four: Professional Responsibility
- #9 Professional Learning and Ethical Practice
- #10 Leadership and Collaboration"

While we treat "level of commitment to the InTASC Standards" as a unitary concept, it is, in fact, the summation of dispositions across the original principles. The 10 standards, which comprise the set of 43 critical dispositions measured, are different components of the values and beliefs of teachers. For example, it is possible for a teacher to be deeply committed to planning but less so to assessment or professional development.

The purposes of these analyses are not to exhaust possible conclusions for these specific programs, but to demonstrate by example how data might be used in accreditation or institutional improvement. Instruments should be sensitive enough to differentiate by common independent variables in order to possess practical utility. As such, we conducted illustrative parametric testing of several typical research questions that were of interest. Are there differences by InTASC standard? Are there differences by program level? Are their differences by clinical placement?

Table 3. Examples of BATS2 Analysis by InTASC Standards and Groups

TABLE 27.1 C:\Users\baywoof\Desktop\WinstepsCont ZOU809WS.TXT2 May 31 2018 15:10
INPUT: 1072 Person 50 Item REPORTED: 1072 Person 50 Item 2 CATS WINSTEPS 4.0.0

Subtotal specification is: ISUBTOTAL=$S5W1
ALL Item SCORES ARE NON-EXTREME

Item COUNT	MEAN MEASURE	S.E. MEAN	P.SD	S.SD	MEDIAN	MODEL SEPARATION	MODEL RELIABILITY	CODE
50	50.00	2.19	15.31	15.47	48.99	12.56	.99	*
5	32.01	3.23	6.46	7.23	28.82	3.23	.91	1
4	45.40	2.51	4.35	5.03	46.27	3.94	.94	2
6	52.90	7.19	16.07	17.61	52.56	14.28	1.00	3
2	45.11	7.56	7.56	10.70	45.11	6.50	.98	4
8	46.76	5.23	13.85	14.80	43.49	10.62	.99	5
6	53.69	5.67	12.67	13.88	53.57	12.80	.99	6
3	56.33	5.55	7.85	9.61	58.95	9.62	.99	7
6	53.68	6.82	15.24	16.70	55.59	13.71	.99	8
6	55.14	7.03	15.72	17.22	53.25	14.57	1.00	9
4	58.14	11.63	20.14	23.25	59.39	17.72	1.00	10

TABLE 27.1 C:\Users\baywoof\Desktop\WinstepsCont ZOU809WS.TXT2 May 31 2018 15:10
INPUT: 1072 Person 50 Item REPORTED: 1072 Person 50 Item 2 CATS WINSTEPS 4.0.0

Subtotal specification is: ISUBTOTAL=$S7W1
ALL Item SCORES ARE NON-EXTREME

Item COUNT	MEAN MEASURE	S.E. MEAN	P.SD	S.SD	MEDIAN	MODEL SEPARATION	MODEL RELIABILITY	CODE
50	50.00	2.19	15.31	15.47	48.99	12.56	.99	*
15	43.94	3.80	14.23	14.73	44.36	9.95	.99	1
10	46.43	4.28	12.85	13.55	43.49	10.07	.99	2
15	54.21	3.49	13.06	13.52	56.37	12.94	.99	3
10	56.34	5.89	17.68	18.64	53.25	16.04	1.00	4

* indicates all persons; 1-10 indicates the ten InTASC Standards; 1-4 indicates Critical Areas

Here, students scored more consistently with Standard 10 and Group 4, and they were more inconsistent with the dispositions in Standard 1 and Group 1. Without speculating why the results occurred, it is illustrative that for program improvement, BATS2 provides the opportunity to report and diagnose differential scores down to the individual student level. Rasch programming also allows ratio-level data to be analyzed parametrically. There were many possible contrasts, but a few that were of note are described here. In the results for one institution, there was a significant difference in scores from Preadmission to Senior to Masters students. This is summarized in Table 4, in which the scores were all in the expected direction from lowest to highest.

Table 4. Contrast Between Dispositions by Classes of Students

Person CODE	CODE	MEAN DIFFERENCE MEASURE	S.E.	t	Welch-2sided d.f.	Prob.
02	04	-4.87	.94	-5.18	125	.000
02	05	14.02	1.47	9.56	4	.001
04	05	18.89	1.46	12.97	4	.000

Person Codes: 02=Preadmission, 04=Senior, 05=Masters

In the other institution, preschool teachers were significantly different on dispositions for InTASC Groups 1 (Learning) at one particular school (labeled YA), compared to another local placement (AO), and to undergraduate students in general. This is illustrated in Table 5.

Table 5. Contrast Between Dispositions by Placement

Person CLASS	DGF SCORE	DGF SIZE	DGF S.E.	Person CLASS	DGF SCORE	DGF SIZE	DGF S.E.	DGF CONTRAST	JOINT S.E.	Rasch-Welch t	d.f.	Prob.	Item CLASS
| AO	-.03	3.03	2.76	YA	.04	-5.83	2.54	8.86	.37	2.36	318	.0187	1
| UG	-.02	1.90	1.36	YA	.04	-5.83	2.54	7.73	.29	2.69	557	.0074	1

AO=School 1, YA=School 2, FG=Institution 1, UG=Undergraduate

Why would one particular preschool setting differ from another school? These and other exploratory analyses raise questions and provide evidence for improvement.

6.6 A Holistic (Qualitative) View

The first small field test of using three instruments within the revised DAATS Battery was concluded in fall 2017 with groups of undergraduate and alternative certification students. The projective Situational Reflection Assessment (SRA) and observational Candidate Belief Checklist (CBC) instruments can be used for comparative validity. Findings indicate that for both high scoring and low scoring students on BATS2, there are revealing patterns that can point toward productive interventions. Two case studies are presented herein. The first student, Jim, scored well while the second student, Kathy, scored near the bottom of the pool.

Jim

Jim is a 43 year old career changer from real estate. He has been a musician at weddings and other venues, and now wants to teach secondary mathematics.

Beliefs About Teaching 2: Jim's BATS2 score was strong -- a Rasch measure (scale score) of 73 and 42 out of 50 items answered as expected. This score placed him in the "organizing" level of the Krathwohl Taxonomy, and is interpreted as a strong level of commitment to the InTASC critical dispositions. The interpretation for this score range is:

Your passion is strong but balanced with other aspects of your life. You are likely to **organize** your life to ensure sufficient time to do all that you believe you need to do, setting aside time to plan conscientiously and systematically for your students.

This score seemed higher than expected for him. The analysis revealed four items for which the responses were surprising, and three of the items were of mid-level difficulty (near the mean value of 50); only one was at a difficulty equivalent to Jim's level of commitment. Jim:

- Feels that creativity is best taught in art and music (item 23; measure of 60.27)
- Feels that a test at the end of each unit is the best strategy for assessment (item 8; measure of 58.57)
- It is okay for student to not understand why a lesson is meaningful as long as they learn content (item 13; measure of 77.20)
- Student motivation is best addressed through use of consequences (item 31; measure of 60.13)

Situational Reflection Assessment (SRA): For the prompt showing a child alone, curled up on the floor, Jim's reaction was that the child needs to learn how to express himself. For the picture of a professional meeting, Jim found himself most like the teacher he thought was reflective, thinking about a variety of perspectives and not focused on the workshop.

Candidate Behavior Checklist (CBC): Jim was rated as exhibiting mostly negative behaviors on the following items:

- Discourages or ignores students who are having trouble with material
- Is unimaginative or out-of-date, using only the textbook or basic adopted materials
- Controls the design of all formative and summative assessments tightly
- Assumes assessments are aligned with standards and district instructional materials
- Assesses without regard to individual student needs
- Rigidly adheres to a plan, missing important instructional opportunities
- Plans lessons for whole group strengths and needs

Interventions to help Jim:

- Watch the TEACH film: The film follows four teachers who illustrate how tenacity, innovation, and a passion drives these educators as they navigate the ups and downs in their classrooms. Reflection on each educator's strengths, how they improved their weaknesses, and how they modified curriculum to meet students' needs and interests.
- Sharing with other teachers: Teachers of Record shared a variety of formative and summative assessments used in their classrooms.
- Data Analysis key assessment: Reinforced for him how this assessment should help him to see students' strengths and weaknesses in order to modify curriculum.
- Observations: Encouraged him to observe teachers that use creativity in their classrooms Collaborated with the instructor of EDF 5443 Measurement & Evaluation for Teachers: Requested more emphasis on formative and summative assessments in the course. (others had this weakness)

Kathy

Kathy is a 35 year old, who has been teaching for two years. We do not know her previous career.

Beliefs About Teaching 2: Kathy's BATS2 score was weak—a Rasch measure (scale score) of 60 and 34 out of 50 items answered as expected. This score placed her near the bottom of the "valuing" level of the Krathwohl Taxonomy, and is interpreted as a commitment to the InTASC critical dispositions but not a strong one. The interpretation for this score range is:

You are committed to the "critical dispositions" of teaching, as defined in the InTASC Standards. You value these dispositions sufficiently to apply them consciously in your practice whenever you see an opportunity to do so. This is your expected score range.

Some of the responses that were surprising are listed below. All of the items with measures less than 50 (mean) were relatively easy for the pool of respondents; items in the 20-40 range were very easy. For example, for item 7, with a measure of 26, only 14 of 672 students answered inconsistently with the InTASC Standards.

- Does not enjoy adapting to different learning styles and watching children achieve (item 7: measure of 26.45)
- Need to be inflexible in the classroom in order to maintain structure and consistency (item 16: measure of 77.75)
- Believes that preventing children from expressing their opinions about school related topics is ok. (item 18: measure of 62.38)
- Believes test scores are a true test of a student's learning and not a realistic performance (item 28: measure of 46.49)
- Does not believe that student learning is connected to motivation or engagement (item 21: measure of 30.74)
- Does not set up learning centers so students can learn in different ways (item 38: measure of 43.43)
- Teachers do not need to share what works (item 45: measure of 31.23)
- Will not include current events in lessons (item 48: measure of 58.59)

Situational Reflection Assessment (SRA2): For the prompt showing a professional meeting, Kathy wrote that the teachers looked bored, like to whisper and make lit of situations at workshops, not seeing anything serious or dedicated in their behaviors.

Candidate Behavior Checklist (CBC): Kathy was rated as exhibiting mostly negative behaviors on the following items:

- Discourages or ignores students who are having trouble with material
- Discourages peer interaction and cooperative opportunities
- Misses opportunities to introduce current events into planned topics
- Assumes assessments are aligned with standards and district instructional materials
- Assesses without regard to individual student needs
- Plans lessons for whole group strengths and needs

Interventions to help Kathy:

- Watch the TEACH film: The film follows four teachers who illustrate how tenacity, innovation, and a passion drives these educators as they navigate the ups and downs in their classrooms. Reflection on each educator's strengths, how they improved their weaknesses, and how they modified curriculum to meet students' needs and interests.
- Sharing with other teachers: Teachers of Record shared a variety of formative and summative assessments used in their classrooms.
- Observations: Encouraged her to observe teachers using learning centers or grouping students to work collaboratively
- Assigned Cooperative Teacher: Kathy needs a CT who is strong in knowing students' individual needs and strengths

7. Results for Research Questions

When we started this project we had five research questions:

1.) Is there evidence that BATS2 (Forms A&B) are valid and reliable in this setting?
The evidence is very strong that BATS2 demonstrated validity and reliability with our students. The score distributions and statistics in our sample are within expected parameters of the Rasch model analysis.

2.) Can BATS2 be used diagnostically?
It is clear that BATS2 can be used diagnostically both by statistical analysis and confirmed by anecdotal (qualitative) examination of the scores by InTASC Standard, the program improvements identified, and the combined use of BATS2 with other DAATS instruments. The analyses offered by the Rasch model provide opportunities for exploration and diagnosis.

3.) Is the Krathwohl taxonomy evident in the BATS2 item analysis?
The items were created with the Krathwohl affective taxonomy in mind, concurrent validation with other instruments and qualitative analysis with other instruments in the DAATS battery have not yet been completed.

4.) Are person scores logical (graduate students more consistent with InTASC than lower division undergraduate, for example)?
Partial analysis of a subset of BATS2 scores supports logical person progression.

5.) Are Form A and Form B equivalent and useful for measuring growth?
The intention here was to use Form A and Form B of BATS2 in a pre-post test mode. The overall statistics from the Rasch model analysis would allow several methods for equating the two forms. One would be to have the same students take both forms, which we have already done with a small subset of students. Another would be to include a set of common items on both instruments to provide anchor items for equating. We have also completed this with a subset of administrations. Finally, since all items on both forms are classified by InTASC Critical Dispositions, it would be possible to align items without regard to difficulty measure. Regardless of methodology, the initial pilot of BATS2

indicated similarity across forms that appear to be no more than sampling variability, so we have great confidence that equivalence is supported.

7.1 Qualitative Results

BATS2 (and all DAATS2) items were constructed to provide operational definitions of InTASC Standards, with theory driving item writing. Judgmentally, items expected to be more difficult were more difficult; items expected to be easier were easier (construct validity). Students expected to be less committed were measured as less committed; those expected to be highest were measured as highest. The correspondence between faculty perceptions of students and DAATS results supported construct validity with version 1 (Englehart, et al., 2011; Lang & Wilkerson, 2008) and again here with version 2. Progressions from coursework to final internship and in alternative certification (Lang & Wilkerson, 2008) and progressions through degree level were evidence of predictive validity in version 1 and continue to be evidenced with version 2.

For us, another key question is: "Are the results useful in confirming quality and improving individual teachers and preparation programs?" The results described above on validity confirm that programs are "on track" – a relevant finding in terms of public perception and accountability. Finding students and program components to target for improvement is critical to utility. Internship coordinators for both Level I and Level II undergraduate students and the program coordinator for the alternative certification program all confirm that high scoring and low scoring students are consistent with their expectations. As of this writing, low scoring students who have transitioned from Level I to Level II are being counseled, and ten pre-admit students have been identified as needing to be monitored early in the program if admitted. In the alternative certification program, specific BATS item responses explained behaviors noted in class.

Continuing in-depth analysis of the results, both Rasch measures and unexpected responses to individual items are framing the remediation efforts. Faculty are meeting with low scoring students in the classic meaning of "assessment" sitting side-by-side to discuss. While it is necessary to include some very difficult items to which authors expect limited "correct" responses, there are still some items that were unexpectedly difficult and serve as a starting point for program improvement discussions. In these cases a large number of "incorrect" responses (a.k.a. inconsistent with the dispositions measured) were identified. These point to the need for discussions with students in the following areas:

1. The 3 R's (reading, 'riting, & 'rithmetic) are not the most important thing for children to learn.
2. Lesson plans need to be developed by teachers and not just the "experts."
3. It is always important for students to understand why a lesson is meaningful; learning the content is not the intended goal.
4. Rigidity is not ever best practice.
5. Students do need to learn to think, and that is the goal needed for all lessons.

6. Consequences are not the way to motivate.
7. The teachers' manual does not contain everything a teacher needs.
8. All lessons should be assessed with the data used to drive future instruction.

8. Limitations

Limitations of this research were inferred or mentioned in the text of the article, but should be reiterated here. Validity on short, self-report instruments that measure affect are always subject to subjects falsifying the expected response. By far and away the best use of BATS2 as initial, screening instrument is appropriate, but validation with multiple instruments from the DAATS battery and qualitative confirmation would be required for validation and best use.

The samples are by nature purposive, because they consist of the self-selected students taking courses in a college of education. Contrasting data from other majors and in service professionals would also be useful for validation but practically difficult to obtain.

Finally, all Rasch measures are dependent to a great extent on the strength and quality of the underlying latent construct. The InTASC Standards and the associated Critical Dispositions provide a reasonable start at instrument development, but like many similar sets of standards they are multi-dimensional, sometimes lack concise language, or contain overlapping concepts across Standards. As such, the item writers may still produce gaps, create misfitting items, and introduce ideas that were not in the original standards.

9. Conclusions

Some general conclusions are listed here:

1. The INTASC Standards provide a useful construct definition that can be measured holistically and by Standard.
2. The Thurstone agree/disagree scale contributes to the identification of strongly and weakly committed teachers.
3. The Bloom and Krathwohl affective taxonomy works in assessment, yielding proficiency levels with a credible category structure.
4. A well-designed measurement device leads to accurate and actionable decisions.
5. A qualitative analysis of individual items enhances Rasch score interpretations, making them more useful for evaluation at the individual and program levels.

10. Significance, Implications, and Recommendations

Objective measurement of what teachers believe in terms of the nationally accepted teaching standards is the first step in improving both their level of commitment and the programs that influence that commitment. Given the difficulty of measuring with a single instrument, the next step is to continue earlier research on the use of multiple measures from version 1 into version 2.

More important, though, is the work that lies ahead in determining teaching strategies that can influence, develop, or improve teacher dispositions and the development of advising, monitoring, and counseling out procedures for those who show measured deficits.

The use of multiple measures has promise of providing a strong measure of teacher commitment to the standards of the profession for those who are sufficiently committed to measuring dispositions to take the time to do it.

References

Alexander, S. D. (2016). *The relationship between teacher evaluation model, value-added model, and school grades* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global database. (UMI No. 10141740)

Association for Advancing Quality in Educator Preparation. (2018). *AAQEP Expectations Framework*. Retrieved from https://aaqep.org/wp-content/uploads/2018/07/AAQEP-Expectations-Framework-February-2018-2.pdf

Bloom, B. S., & Krathwohl, D. R. (1956). *Taxonomy of education objectives: the classification of educational goals, by a committee of college and university examiners*. New York, NY: Longman, Green.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. New York, NY: Routledge.

Boonen, T., Van Damme, J., & Onghena, P. (2013). Teacher effects on student achievement in first grade: which aspects matter most? *School Effectiveness and School Improvement*, 25(1), 126–152. doi:10.1080/09243453.2013.778297

Collinson, V., Killeavy, M., & Stephenson, H. J. (1999). Exemplary teachers: Practicing an ethic of care in England, Ireland, and the United States. *Journal for a Just and Caring Education, 5*(4), 349-66.

Council for the Accreditation of Educator Preparation. (2016a). *The CAEP Standards*. Washington, DC: Author. Retrieved from http://caepnet.org/standards/introduction

Council for the Accreditation of Educator Preparation. (2016b). *CAEP Accreditation Handbook.* Washington, DC: Author. Retrieved from http://caepnet.org/~/media/CAEP%20Accreditation%20Handbook_March%202016.pdf?la=en

Council of Chief State School Officers. (1992). InTASC Standards. Washington, DC: Author.

Council of Chief State School Officers. (2013). InTASC Standards. Washington, DC: Author.

Darling-Hammond, L. (2000). Teacher Quality and Student Achievement. Education Policy Analysis Archives, 8, 1. doi:10.14507/epaa.v8n1.2000

Donahue, B. P. (2016). *The implementation of a new teacher evaluation model: A qualitative case study of how teachers make sense of the Marzano teacher evaluation model* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global database. (UMI No. 10090271)

Englehart, D. S., Batchelder, H. L., Jennings, K. L., Wilkerson, J. R., Lang, W. S., & Quinn, D. (2012). Teacher dispositions: Moving from assessment to improvement. *The International Journal of Educational and Psychological Assessment. 9*(2), 26-44.

Graziano, S. K. (2017). *An exploration of teacher perception of the Marzano causal teacher evaluation model and its impact on professional practices* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global database. (UMI No. 10254663)

Hagell, P. (2014). Testing rating scale unidimensionality using the principal component analysis (PCA)/*t*-Test Protocol with the Rasch model: The primacy of theory over statistics. *Open Journal of Statistics*, *4*, 456-465. doi:10.4236/ojs.2014.46044

Heck, R. H. (2009). Teacher effectiveness and student achievement. *Journal of Educational Administration, 47*(2), 227-249. doi:10.1108/09578230910941066

Herman, J. L., Baker, E. L., & Linn, R. L. (2004). Accountability systems in support of students learning: Moving to the next generation. *The CRESST Line*, 1-7.

Klute, M., Apthorp, H., Harlacher, J., & Reale, M. (2017, February). *Formative assessment and elementary school student academic achievement: A review of the evidence*. Washington, DC: Institute of Education Sciences. Retrieved from https://ies.ed.gov/ncee/edlabs/regions /central/pdf/REL_2017259.pdf

Lang, W. S. (2008). The DAATS Model: Initial psychometric and statistical findings: A Top ten illustration. *Resources in Education*. Retrieved from ERIC database (ED502864).

Lang, W. S., Wilkerson, J. R., Moore, L. L., Fields, L. J., Parfitt, C. M., Greene, J. S., Kratt, D. M., Martelli, C. D., LaPaglia, K. E., Johnston, V. D., Gilbert, S. G., Zhang, J., & Wang, C. X. (2018a, February). Beliefs about teaching (BATS2): Construction and validation of an instrument based on InTASC critical dispositions. Paper presented at the meeting of the Eastern Educational Research Association, Clearwater, FL.

Lang, W. S., Wilkerson, J. R., Moore, L. L., Fields, L. J., Parfitt, C. M., Greene, J. S., Kratt, D. M., Martelli, C. D., LaPaglia, K. E., Johnston, V. D., Gilbert, S. G., Zhang, J., & Wang, C. X. (2018b, February). Measuring Teacher Dispositions Systematically Using Multiple Measures. Paper presented at the meeting of the Eastern Educational Research Association, Clearwater, FL.

Linacre, J. M. (2010). When to stop removing items and persons in Rasch misfit analysis? *Rasch Measurement Transactions, 23*(4), 1241.

Linacre, J. M. (2018). Winsteps [Computer program Ver. 3.93.2]. Chicago, IL: winsteps.com.

Marzano, R. J. (2012). The two purposes of teacher evaluation. *Educational Leadership*, *70*(3), 14-19.

Moore, L. (2016). Intrinsic and extrinsic motivators that impact teacher retention in challenging urban schools (Doctoral dissertation, University of Central Florida). Retrieved from http://stars.library.ucf.edu/etd/4944

Osguthorpe, R. D. (2008). On the Reasons We Want Teachers of Good Disposition and Moral Character. *Journal of Teacher Education*, 59(4), 288–299. doi:10.1177/0022487108321377

Quinn, A. E. (2014). Looking at the bigger picture with Dr. Robert Marzano: Teacher evaluation and development for improved student learning. *Delta Kappa Gamma Bulletin*, *81*(1), 12-18.

Ransdell, K. F. (2017). *Informing a school district's hiring procedure through examining the relationship between teacher dispositions and student voice* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global database. (UMI No. 10264781)

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) [foreword & afterword Wright, B. D.]. Chicago, IL: The University of Chicago Press.

Roberts, J. S., Laughlin, J. E., & Wedel, D. H. (1999). Validity issues in the Likert and Thurstone approaches to attitude measurement. *Educational and Psychological Measurement, 59*(2), 211-233. doi:10.1177/00131649921969811

Sargent, M. S. (2014). *An investigation of research-based teaching practices through the teacher evaluations in Indiana public schools* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global database. (UMI No. 3616908)

Thurstone, L. L. (1928). Attitudes Can Be Measured. *American Journal of Sociology*, 33(4), 529–554. doi:10.1086/214483

Vaughn, K. A. (2012). *Teacher dispositions and student achievement* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global database. (UMI No. 3505422)

Wasicsko, M. M. (2004). The 20-minute hiring assessment. *The School Administrator Web Edition.* Retrieved from http://aasa.org/publications/sa/2004_10/wasicsko.htm

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 Press.

Wilkerson, J. R., (2006, April). Measuring teacher dispositions: Standards-based or morality-based? *Teachers' College Record.* Retrieved from http://www.tcrecord.org/content.asp?contentid=12493

Wilkerson, J. R., & Lang, W. S. (2009). *Technical report: report to the Louisiana Board of Regents on measuring teacher dispositions with Wilkerson and Lang DAATS instruments*. Qualitative Research Team, Louisiana Board of Regents: Baton Rouge, LA.

Wilkerson, J. R., & Lang, W. S. (2011). Standards-based teacher dispositions as a necessary and measurable construct. *The International Journal of Educational and Psychological Assessment*, *7*, 34-54.

Wilkerson, J. R. (2012). Measurement and evaluation perspectives on scaling teacher affect with multiple measures. *The International Journal of Educational and Psychological Assessment 9*(2), 165-191.

Wilkins, B. L. (2017). *Teacher perspectives on the Marzano teacher evaluation model during year one of implementation* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global database. (UMI No. 10610499)