ADVANCES IN GLOBAL EDUCATION AND RESEARCH

GLOCER'21

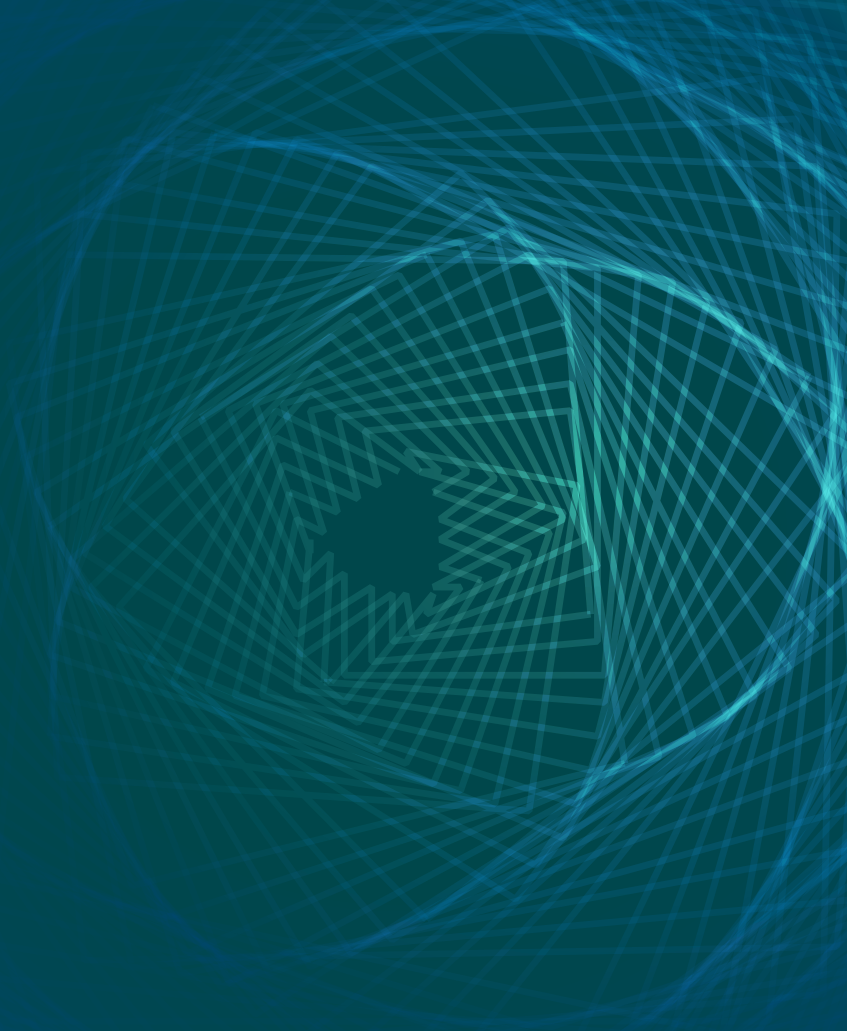
VOLUME 4

Editors:

Dr. Waynne B. James

Dr. Cihan Cobanoglu

Dr. Muhittin Cavusoglu



Co-Editors

***Dr. Waynne James*,** University of South Florida, USA
Dr. Cihan Cobanoglu, University of South Florida, USA
Dr. Muhittin Cavusoglu, Northern Arizona University, USA

ADVANCES IN GLOBAL EDUCATION AND RESEARCH: VOLUME 4

ISBN 978-1-955833-04-2

Authors are fully responsible for corrections of any typographical, copyrighted materials, technical and content errors.

https://digitalcommons.usf.edu/m3publishing/vol3/iss2021/52
DOI: https://www.doi.org/10.5038/9781955833042

Co-Editors

Dr. Waynne James, University of South Florida, USA
Dr. Cihan Cobanoglu, University of South Florida, USA
Dr. Muhittin Cavusoglu, Northern Arizona University, USA

ISBN 978-1-955833-04-2

This imprint is published by USF M3 Publishing, LLC
The registered company address is University of South Florida, 8350 N Tamiami Tr, Sarasota, FL 34243 USA.

Assistant Editor

Editor Assistants

****Authors are fully responsible for corrections of any typographical, copyrighted materials, technical and content errors.***

https://digitalcommons.usf.edu/m3publishing/vol3/iss2021/52
DOI: https://www.doi.org/10.5038/9781955833042

Beliefs About Teaching (BATS2) - Rasch Model Analysis of an Instrument Based on InTASC Critical Dispositions

W. Steve Lang[1], LaSonya Moore[1], and Judy R. Wilkerson[2]

[1]College of Education
University of South Florida, United States

[2]College of Education
Florida Gulf Coast University, United States

Abstract

This research is one of a chain of studies on the development and testing of a battery of teacher dispositions assessments designed to measure the InTASC critical dispositions (InTASC, 2013). Accredited programs in the US are required to use these standards and to assess teacher dispositions. Here, an alternate form of the Thurstone agreement scale (Beliefs About Teaching Survey or BATS, version 2, Form B) was validated with US undergraduate teacher education majors from two public universities (N=149) with this population. Results indicate that Form B of BATSv2 has comparable results to Form A and the first version of the instrument.

Keywords: teacher dispositions, Rasch model, teacher accreditation, InTASC

Recommended Citation: Lang, W. S., Moore, L., & Wilkerson J. R. (2021). Beliefs about teaching (BATS2) - Rasch model analysis of an instrument based on InTASC critical dispositions. In W. B. James, C. Cobanoglu, & M. Cavusoglu (Eds.), *Advances in global education and research* (Vol. 4, pp. 1–11). USF M3 Publishing.
https://www.doi.org/10.5038/9781955833042

Introduction

The affective components of teacher performance are described in the current version of the InTASC Standards (CCSSO, 2013) and CAEP Standards (2016a), which include the values and beliefs teachers called "dispositions." Less attention has been offered to the assessment of dispositions; and certainly to the level of instrument quality for the affective domain has less often been the target for professional assessment despite the recognition of its importance in overarching standards. For a variety of reasons, Educator Preparation Programs (EPPs) do not assess dispositions with the same effort; and consequentially they do not track improvement or report affective impact with the same frequency or consequences as cognitive evidence.

The knowledge and skills of teachers has been the subject of sophisticated research and application in recent years culminating in VAM scores, legislative mandates, and complex assessments. A common assumption is that all teachers should believe that all students can learn. Teachers should value the skills related to effective teaching that help children reach their potential. The extent to which these commitments are held is the overarching purpose of this research.

As in previous research, the focus here continues to be on standards-based assessment of teacher dispositions (Wilkerson & Lang, 2007), rather than morality-based assessment (Wilkerson, 2006).

Objective measurement of teacher dispositions, the affective component of teaching, provides a vehicle to determine the level of commitment and an opportunity to intervene when that commitment is questionable.

This study continues a long line of research in instrument development and validation that demonstrates the feasibility, utility, and practicality of a measurement process that yields valid and reliable scores of teacher candidates. Specifically, three research questions frame this study:

- Is there evidence that BATS2 (Form B) is valid and reliable in this setting?
- Can BATS2 be used diagnostically for each of the InTASC Standards?
- Is the development process robust across versions of the InTASC Standards?

Literature Review

The literature provides many examples of single assessments of teacher affect such as surveys, indices, observations, or interviews, (Richardson & Onwuegbuzie, 2003; Lund, Wayda, Woodward, & Buck, 2007; Schulte, Edick, Edwards, & Mackiel, 2004; Wasicsko, 2004; Jung & Vogt, 2006; Singh & Stoloff, 2008). None, however, use Rasch modeling, and none are systematically aligned with the national standards of teaching proposed by the Council of Chief State School Officers (CCSSO) through the InTASC Standards (2013).

Teacher education national accreditation requirements by the Council for Accreditation of Educator Preparation (CAEP, 2016a) include assessment of teacher dispositions using instruments yielding valid and reliable scores. 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).

There are five CAEP standards, each of which involves dispositions to some extent. The first CAEP Standard specifically requires use of the standards developed by the Council of Chief State School Officers (CCSSO, 2013). These are called the InTASC standards. Clinical settings are the topic of CAEP Standard 2; assessment at multiple points in time is the focus of CAEP Standard 3; the use of "valid and reliable data" is the focus for CAEP Standard 4; use of "empirical evidence that interpretations are valid and consistent" is the focus for Standard 5. Program assessments need to be evaluated as high quality and demonstrate usefulness to measure impact. CAEP states (Standard 3.3),

> Educator preparation providers establish and monitor attributes and dispositions beyond academic ability that candidates must demonstrate at admissions and during the program. The provider selects criteria, describes the measures used and evidence of the reliability and validity of those measures, and reports data that show how the academic and non-academic factors predict candidate performance in the program and effective teaching.

The 10 InTASC standards are grouped into four standards categories. Each category is comprised of a set of statements that incorporate knowledge, performances, and critical dispositions. All in all, there are a total of 43 critical dispositions statements, spread among the 10 InTASC Standards.

https://digitalcommons.usf.edu/m3publishing/vol3/iss2021/52
DOI: https://www.doi.org/10.5038/9781955833042

The authors of the InTASC Standards note the importance of differentiating knowledge, skills, and dispositions, explaining major changes in the 2013 Standards as follows:

> Another key point is that these standards maintain the delineation of knowledge, dispositions, and performances as a way to probe the complexity of the teacher's practice. The relationships among the three have been reframed, however, putting performance first—as the aspect that can be observed and assessed in teaching practice. The others were renamed. "Essential knowledge" signals the role of declarative and procedural knowledge as necessary for effective practice and "critical dispositions" indicates that habits of professional action and moral commitments that underlie the performances play a key role in how teachers do, in fact, act in practice.

Both CAEP and its predecessor, (the National Council for Accreditation of Teacher Education or NCATE) required the use of the InTASC Standards as the foundation for measuring dispositions, so the requirement has been in existence for some time. The instrument used int his research is part of a second generation of a battery of affective assessments called DAATS. The *Dispositions Assessments Aligned with Teacher Standards* (DAATS) Battery measures the commitment of teachers and teacher candidates to the standards-based skills of teaching, as they have evolved over time and across versions of the standards (Lang et.al, 2014; Wilkerson, 2012; Wilkerson & Lang, 2007; Wilkerson, 2006).

In this particular assessment development, BATS2 is designed from the start based on the InTASC Standards for the purpose of assessing *critical dispositions* to meet CAEP Standard 1.1 and 3.3 to monitor dispositions from selection through growth and into the graduate degrees. The psychometric model used is Rasch analysis (Linacre, 2017), which has been applied to disposition research in parallel applications.

This effort reports on one (BATS2) of a battery of instruments for the assessment of teacher dispositions. BATS2 has multiple forms (for assessment of dispositional gains), psychometric analysis (for validity, reliability, and scaling), and field-testing (for generalizability). Beyond the specific instrument, there is also a need to promote a model for affective assessment development in the context of professional standards that almost always include the disposition elements.

The DAATS Battery

The *Dispositions Assessment Aligned with Teacher Standards* (DAATS) model (Wilkerson & Lang, 2007) suggests various strategies for assessing affect. We have discussed both the process of designing assessments, as well as the specific battery extensively, and we refer readers to some of our earlier work for details on how the instruments have evolved and the results that they have achieved at various institutions (Wilkerson & Lang, 2004; Wilkerson & Lang, 2006; Lang & Wilkerson, 2008; and Englehart, Batchelder, Kelly, Wilkerson, Lang, & Quinn, 2011; Wilkerson & Lang, 2011). ASCD's support of multiple measures to make informed judgments about student learning in the cognitive domain and the success of education programs applies equally to the affective domain. The need for multiple measures was well supported by Herman, Baker, & Linn, (2004):

> No single test can tell all there is to know. As the directors of the National Center for Research on Evaluation, Standards, and Student Testing emphasize, "Multiple measures are needed to address the full depth and breadth of our expectations for student learning" (p. 2). Beyond the multiple-choice and short-answer items that are typical of current assessments, "other types of performance measures—essays,

> applied projects, portfolios, demonstrations, oral presentations, etc.—are needed to represent and guide students' progress" (p. 2).

There are a total of five affective assessment instruments which make up the DAATS Battery (Wilkerson & Lang, 2006). Each instrument is designed to measure all ten INTASC Standards. Using the instruments in the battery, students are placed on a version of the Krathwohl's affective taxonomy (Bloom & Krathwohl, 1956).

While field test results of only one instrument are presented herein; however, the first four are currently being field tested in their revised form for the updated InTASC Standards.

- Beliefs About Teaching Scale (BATS2): a pair of 50 item Thurstone agreement scales (Forms A and B), tested as a pre-admissions tool and monitoring/advising tool in both levels of internship.
- Experiences in Teaching Questionnaire (ETQ2): multiple forms that include constructed response items about prior experiences, to be field-tested as the reflective component of a Teacher Work Sample used in final internship.
- Situational Reflection Assessment (SRA2): 20 constructed response items, comprised of picture prompts (Slitkin, 2007), in a thematic apperception format. Four items being field tested in spring 2018 in first semester and Level 1 internship and a second form with four additional items planned.
- Classroom Behaviors Checklist (CBC2): 20-30 paired positive and negative behaviors evaluated as typically positive, typically negative or mixed. CBC is being field tested in three versions, with selected items repeated across instruments, for early field experiences and both levels of internship. This was formerly the Classroom Dispositions Checklist of CDC).
- K-12 Dispositions Impact (KIDS2): focus group with 10 clustered prompts measuring children's perceptions. This form has met with IRB resistance in the schools. A modified version using the Marzano framework is being considered.

Beliefs About Teaching Version 2 (BATS2)

All items in all versions of the Beliefs About Teaching Scale (BATS) were constructed in direct alignment with the InTASC Critical Dispositions statements. The item-writing team specifically targeted Krathwohl levels and critical dispositions statements for each item, attempting to balance the test forms by standard and difficulty (taxonomic level).

For example, Standard #2, Learning Differences, states: "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.
>
> I have a rule in my classroom: "We all speak proper English and ignore gestures, slang, or foreign languages."

https://digitalcommons.usf.edu/m3publishing/vol3/iss2021/52
DOI: https://www.doi.org/10.5038/9781955833042

BATS2 uses a Thurstone (1928) format of agree/disagree items. Thurstone scales lead to a dichotomous decision (either agree/disagree), while Likert scales typically use five-points ranging from strongly agree to strongly disagree and include a neutral midpoint. Roberts, Laughlin, and Wedel (1999) compared Likert and Thurstone scaling and concluded that the Thurstone scale was more appropriate when seeking positions that are extreme (e.g., high/low levels of commitment). In the case of teacher dispositions, users tend to expect higher levels of commitment.

Every item is coded by InTASC Standard, Core Area, and Krathwohl Taxonomy. A variation of the taxonomy (Wilkerson & Lang, 2011) classifies student affect into six levels. These include Unaware, Receiving, Responding, Valuing, Organizing, and Characterizing, with the "unaware" level added because the original taxonomy was designed for instruction and not assessment, naturally not accounting for the possibility that respondents might have no commitment.

These definitions of the Taxonomy were applied in the development of an interpretation guide to be provided to students with their BATS scores (measures) and originally published by Wilkerson and Lang (2011):

Unaware
- Has not considered the skill in any meaningful way.
- May be opposed to the skill.

Receiving
- Recognizes (is aware of) importance.
- Is beginning to think about it.
- May provide a promise to use it without evidence of having used it.

Responding
- Is emotionally ready to do something and makes an attempt.
- Gives a little extra effort, as time permits, to comply.
- Can easily be distracted from application.
- Has a beginning level of commitment or satisfaction.

Valuing
- Accepts worth and derives definite satisfaction from it.
- Feels a need and would commit continuing time and effort.
- Tolerates and may expect interferences.

Organization
- Plans, organizes, and schedules to ensure success with it.
- Determines inter-relationships among knowledge and skills.
- Adapts other aspects to fit it.
- Is uncomfortable with interferences or lack of time to finish.

Characterization
- Sees the skill as the center or driving force of all work.
- Helps others to see the skill's importance, lobbying for it.
- Integrates everything with it.

Previous versions of the DAATS Battery, including the BATS instrument, which were based on the 1992 InTASC Standards (CCSSO, 1992), demonstrated predictive validity and strong reliability (Wilkerson & Lang 2006, 2009; Lang, 2008), so the focus needed for BATS2 centered

on alignment with the revised standards. New items were added where new dispositions were introduced.

Based on lessons learned and new requirements for teacher accreditation, this research describes the creation and analysis of BATS2 (Form A and Form B) designed with validity and reliability in mind from the outset. Also, utility (a 10 minute administration time), and deliverability (on line) were considered important while producing useful data.

The intent was to provide scores that were normally distributed. These scores are scaled using the Rasch model for interval level data, standards-based, and diagnostic in terms of the InTASC Standards and Krathwohl Taxonomy. Teacher educators need to be able to use the scores with students to drive improvement at the student and program levels, while aggregated scores were appropriate for institutional reports. Respondents, especially those who are admitted to the teacher preparation program, are expected to be at the valuing level, with those above this level celebrated, and those below this level assisted. A main goal was to avoid an instrument that had limited construct validity, ceiling effects, or simple raw-score reporting so that the effort to administer a measure becomes a waste of time. A ceiling effect might be evidenced by a bunching effect at the top of the distribution of scores and a skewed distribution.

Problem Statement

The purpose of this research was to pilot a new version of the BATS2 and determine if the new version maintained previously obtained high levels of validity and reliability. Item analysis and instrument characteristics were the primary evidence under consideration.

Method

Subjects

In this study a total of 145 undergraduate teacher education students from a public university were given Form B of BATS2 via Survey Monkey or Canvas LMS in the spring, summer, and fall of 2019 and in spring of 2020. The average response time to complete Form B of BATS2 was nine minutes.

Item Creation and Analysis Method

Five faculty from two public institutions reviewed all of the items from the original BATS, modifying some and creating others. Each item was aligned or re-aligned with the InTASC Critical Dispositions for the 10 InTASC Standards organized into four groups. Following a pilot test, the items were reviewed again in light of an initial analysis and student comments. Winsteps software (Linacre, 2020) is a commonly used to calibrate item measures for the Rasch model of item response theory.

https://digitalcommons.usf.edu/m3publishing/vol3/iss2021/52
DOI: https://www.doi.org/10.5038/9781955833042

Results and Interpretations

Rasch Analysis and Interpretation of the Total Scale and Individual Items

When student responses were analyzed using the Rasch model of item response theory, item and person separation reliabilities were .93 and .62 respectively, with a Cronbach alpha (KR-20) of .67, indicating an excellent reliability for items and adequate reliability for persons, as expected for this homogeneous sample of respondents. A subsequent review indicates one item should be considered for revision, but all items were substantially difficult for this population so the misfit could be partially due to the relatively extreme measure (difficulty).

A variable (or Wright) map was produced and illustrated that the persons were normally distributed, as expected. Items, too, showed a normal distribution, indicating good coverage of the construct in both forms. The resulting distribution of items and persons was essentially as predicted and, therefore, supported confidence in construct validity.

Even though the overall performance of the BATS2 is excellent, there were a few misfitting items, as expected with this sample size. First, Linacre (2010) suggests that there are some situations in which misfitting items should remain in the scale, especially if removing them improves measurement of the persons. Hagell (2014) also 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). In the case of BATS, the items have been examined and tweaked multiple times, but the alignment with the InTASC critical dispositions and location on the Rasch ruler has remained a constant indicator of validity.

Given the Rasch analysis and distribution of measures, there is strong evidence that BATS2 is not only reliable internally but that it also shows evidence of construct validity for the population of students tested in this sample. The four misfitting items should be reviewed again; however, they do not have strong impact on the person measures. This is because the items are dropped for the majority of respondents – most respondents answered inconsistently with InTASC Standards and those who answered consistently appeared to do so almost randomly. Note that one of the 50 items classified as misfitting is possibly due to chance with a large sample size ($p < .05$), so again this is not alarming. The one item that may be problematic was "It is the teacher's job to teach the children, but the leadership's job to make decisions about the building and grounds." There is no obvious reason for this item to be misfitting. An examination of the source of the item from InTASC Critical Disposition 3(n): "The teacher is committed to working with learners, colleagues, families, and communities to establish positive and supportive learning environments." leads to the conclusion that the item has construct validity (InTASC, 2013, p.21).

There is an inherent difficulty when creating an affective scale for use in colleges of education, resulting from a cultural belief system that tends to be somewhat humanistic. In this believe system, success is expected because the population is generally wants to enter the teaching profession, making it 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). Stated in different words, if instruments that measure values are developed in situations where there is a narrow spread (a homogenous

population) the person reliability could be expected. This is different from skill-based assessment where the average item difficulty is .50 and skill level is more spread. Both may be normally distributed and appropriate to the construct, but if we were to develop designed to increase person separation reliability, that could result in an instrument that professional schools would resist using. Affective instruments of this type present challenges because an average measure that is too difficult will result in students and faculty not wanting to implement it. In the same vein, an average measure that is too lenient could result in a ceiling effect with a skewed distribution.

Analysis and Interpretation by InTASC Standard

While "level of commitment to the InTASC Standards" is treated and expected to be a unitary concept here from a measurement standpoint, it is, in fact, multi-dimensional. The ten standards comprising the full set of critical dispositions measured are different components of the values and beliefs of teachers – a unitary concept. However, although we measure teacher dispositions for all Standards, commitment can be different among standards. For example, it is possible for a teacher to be averse committed to assessment but deeply committed to planning.

The 10 InTASC Core Teaching Standards are organized into four general categories. The mean measures for each Standard are presented in parentheses:

Group One: The Learner and Learning contains
- #1 Learning Development (39.34)
- #2 Learning Differences (37.03)
- #3 Learning Environments (49.62)

Group Two: Content contains
- #4 Content Knowledge (56.01)
- #5 Application of Content (47.31)

Group Three: Instructional Practice
- #6 Assessment (56.35)
- #7 Planning for Instruction (47.31)
- #8 Instructional Strategies (47.15)

Group Four: Professional Responsibility
- #9 Professional Learning and Ethical Practice 56.80)
- #10 Leadership and Collaboration (64.03)

The mean measures by Standard range from 37.032 for Learning Differences to 64.02 for Standard 10, reflecting the curricular emphasis in both universities on diversity and the teacher candidate resistance at both universities to teamwork and collaboration. In the middle range, as expected, were Standards 4 (Content Knowledge), 6 (Assessment) and 7 (Planning), but also easy was learning development – another high area of focus. Experientially, then, this ordering of Standards is consistent with faculty's expectations, and this result is consistent with previous validity studies for the earlier version of BATS and for Form A of BATSv2.

https://digitalcommons.usf.edu/m3publishing/vol3/iss2021/52
DOI: https://www.doi.org/10.5038/9781955833042

Conclusions

In the beginning of this research, three research questions were posed, with conclusions drawn here.

- *Is there evidence that BATS2 (Form B) is valid and reliable in this setting?*

  The evidence continues to be very strong that BATS (both versions 1 and 2 and Forms A and B) demonstrate validity and reliability. The score distributions and statistics are as expected and are quite similar to previous studies of both the original version of BATS and BATSv2 Form A.

- *Can BATS2 be used diagnostically for each of the InTASC Standards?*

  Based on the Rasch analysis and a logical analysis of the placement of mean measures (high, medium, and low), Standards scores can be used to identify strengths and weaknesses in candidate dispositions for both student and program improvement purposes.

- *Is the development process robust across versions of the InTASC Standards?*

  The results of this current analysis are consistent with previous studies, as identified earlier in this article.

Limitations

The chief sampling limitation in this study is the lack of respondents at a beginning (low) level. Students entering a teacher education program should be assessed to increase the variability of person measures. Since respondents can provide the "expected" responses – those responses that they believe are expected by the test provider, validity on short, self-report instruments like BATS, that measure dispositions, may be less accurate than results that would be provided by multiple measures using a variety of item types, as in the full range of DAATS instruments or a sample of them. The samples are by nature convenience, because the instrument was used in courses volunteered by faculty and only tested in colleges of education. Data from other majors (e.g., accounting or other professions) and in service professionals would also be helpful for validation efforts, but it is difficult to obtain from a practical standpoint. All Rasch measures are dependent to a great extent on the quality and strength of the underlying latent construct. The InTASC Standards, including the Critical Dispositions, provide an appropriate source for instrument development, but they remain multi-dimensional, sometimes lacking concise language. As the dispositions themselves are read, it is easy to find overlap across the Standards, so this makes conclusions about individual standards somewhat suspect.

Significance, Implications, and Recommendations

Objective measurement of what pre-service teachers value and believe (their dispositions) regarding the nationally accepted teaching standards is just an initial step in the more important issue of trying to improve both their level of commitment and the programs in which they are enrolled. Because of the difficulty inherent in measuring with just one single instrument, the next

needed step is to expand on earlier research regarding the use of multiple measures of different item types.

What is more important, however, is that institutions face the bigger challenge that results from effective affective measurement – developing and implementing teaching strategies that have the capacity to improve teacher dispositions. This will require the development of advising/counseling and continuous monitoring procedures for those candidates who show measured deficits.

References

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.

Borko, H., Liston, D., & Whitcomb, J. (2007). Apples and fishes: The debate over dispositions in teacher education. *Journal of Teacher Education*, *58*(5), 359-364

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. (2013). InTASC Standards. Washington, DC: Author.

Drasgow, F., Chernyshenko, O. S., & Stark, S. (2010). 75 years after Likert: Thurstone was right! *Industrial and Organizational psychology: Perspectives on Science and Practice, 3,* 465-476.

Edwards, A. L. (1959). Social desirability and the description of others. *Journal of Abnormal and Social Psychology*, *59*, 434-436.

Engelhard, G., Jr. (1994). Examining rater errors in the assessment of written composition with a many-faceted Rasch model. *Journal of Educational Measurement*, *31*, 93–112.

Englehart, D. S., Batchelder, H. L., Jennings, K. L., Wilkerson, J. R., Lang, W. S., & Quinn, D. (2011). Teacher dispositions: Moving from assessment to improvement. *The International Journal of Educational and Psychological Assessment. 10*(1).

Fuller, D., Fitzgerald, K., & Lee, J.-S. (2008). The case for multiple measures. *INFO Brief, 52*. Washington, DC: Association for Supervision and Curriculum Development. Retrieved from http://www.ascd.org/publications/newsletters/ policy-priorities/winter08/num52/full/The-Case-for-Multiple-Measures.aspx

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

Jung, E., Rhodes, D., & Vogt, W. P. (2006). Disposition assessment in teacher education: A framework and instrument for assessing technology disposition. *The Teacher Educator*, *41*(4), 207-233.

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

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., Rea, D. C., Quinn, D., Batchelder, H. L., Englehart, D. S., & Jennings, K. J. (2014). Measuring teacher dispositions using the DAATS battery: A multifaceted Rasch analysis of rater effect. *Journal of Applied Measurement*, *15*(3), 240-215.

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

Linacre, J. M. (2017). Winsteps [Computer program Ver. 3.93.2]. Chicago, IL: winsteps.com.

Lund, J., Wayda, V., Woodard, R. & Buck, M. (2007, Winter). Professional dispositions: What are we teaching prospective physical education teachers? *The Physical Educator,* 38-47.

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

https://digitalcommons.usf.edu/m3publishing/vol3/iss2021/52
DOI: https://www.doi.org/10.5038/9781955833042

Richardson, D., & Onwuegbuzie, A. (2003). Attitudes toward dispositions related to the teaching of pre-service teachers, in-service teachers, administrators, and college/university professors. Retrieved from ERIC database (ED482689)

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.

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

Singh, D. K., & Stoloff, D. L. (2008). Assessment of teacher dispositions. *College Student Journal*, *42*(4), 1169-1180.

Taylor, R. L., & Wasicsko, M. M. (2000). *Dispositions to teach*. Retrieved from http://www.nku.edu/~education/educatordispositions/resources/The_Dispositions_to_Teach.pdf

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

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. Retrieved from http://tijepa.books.officelive.com/Documents/V9.2_TIJEPA.pdf