AACTE 2006
Measuring Teacher Dispositions Systematically Using INTASC Principles:
Building Progressive Measures of Dispositions

William Steve Lang
Judy R. Wilkerson
University of South Florida St. Petersburg

Statement of the Problem

Current Events

Scott W. McConnell believes in using corporal punishment; he wrote about it in a paper in his classroom management course; and there is nothing Le Moyne College could do to drop him from their program to protect his future students. The decision was made by the New York State Court of Appeals on Wednesday, January 18, 2006 and reported in the *Chronicle of Higher Education* two days later (Jacobsen, 2006). The college tried to remove him from their teacher education program because "his personal goals did not match the program's," but the Court ruled that his due process rights were violated. He is now back in college, having his court costs borne by a conservative sponsoring agency. The article goes on to say:

> Lawyers for Mr. McConnell hailed the ruling. "There is an attempt in teaching programs nationwide not only to indoctrinate the students but also to make sure only people with particular political views can graduate," said Christopher J. Hajec, a lawyer with the Center for Individual Rights, a Washington-based advocacy group, which represented Mr. McConnell. "Whether you agree with him or not, he definitely has the right to get his degree."

While the dollar amount expended by LeMoyne to defend itself has not been publicly released, it is not hard to imagine that hundreds of thousands of dollars have been expended – and lost. Perhaps if Le Moyne had a system in place that made use of national standards to measure students' dispositions in valid and reliable ways, Scott W. McConnell would be pursuing a different career. We have written on the subject of the convergence of psychometrics and legal decisions in the area of competency assessment (Wilkerson and Lang, 2003); this decision in New York State provides evidence that the same requirements for psychometrically sound assessment apply to dispositions and skills equally.

Avoiding lawsuits is a reasonable goal for disposition assessment. A better motivation would be for colleges to create effective disposition assessments in order that they diagnostically prepare individual students, improve programs, research new ideas in teacher education, and model positive attitudes about assessment to future educators. Typical observation of colleges that avoid sound assessment as a political is that they often create bad attitudes within the very students they are charges to teach!

NCATE Requirements

NCATE (2002) requires the measurement of dispositions as part of its accreditation requirements for teacher education programs. The first standard, entitled, "Candidate Knowledge, Skills, and Dispositions," requires that: "Candidates preparing to work in schools as teachers or other professional school personnel know and demonstrate the content, pedagogical, and professional knowledge, skills, and dispositions necessary to help all students learn. Assessments indicate that candidates meet professional, state, and institutional standards." (p. 14) At first, one might be tempted to blame NCATE for the conundrum faced by Le Moyne and other universities struggling

with how to measure dispositions. On the other hand, if one thinks about the need to help candidates become better teachers, accompanied by the need to have policies and procedures in place to say "no" to those who should not enter the profession, the NCATE standards give us the tool we need to do what we should do.

INTASC Principles

Fortunately, specific guidance is provided to the community by the common set of national standards developed by the Council of Chief State School Officers (CCSSO, 1992) and promulgated by the Interstate New Teacher Assessment and Support Consortium (INTASC) in the form of ten principles. Each of the principles includes indicators written at the knowledge, skill (performance), and dispositional levels, forming constructs that colleges are required to measure. In fact, the first NCATE standard (as cited above) requires the use of standards in the assessment process. For both the acceptable and target proficiency levels in the element related to dispositions, NCATE requires that the work of candidates reflect the dispositions delineated in professional, state, and institutional standards.

When we begin to conceptualize the INTASC Principles as symbiotic in nature, the need for measuring dispositions becomes clearer. If a teacher learns what elements comprise a good lesson plan and then demonstrates on multiple occasions that he/she has the appropriate level of skill to produce (and hopefully deliver) effective lesson plans, we are often lulled into believing that our job is done. They have the knowledge and can apply it, but what happens if they do not think it is important? No pre-graduation faculty evaluative judgment of "proficient in planning" will ever compensate for the damage that can be done by the teacher who thinks lesson planning is a boring waste of time. That teacher will just stand up and deliver. That is the fundamental reason why dispositions are, in the long run, more important than knowledge and skills. The assessment of dispositions helps us to answer the question, "Are they likely to do what we taught them to do when we are not watching them any longer?"

The INTASC Principles lay the foundation upon which we can build solid assessment devices for measuring teacher dispositions. Take for example the following sequence of elements of INTASC Principle #7 on planning:

- The teacher knows when and how to adjust plans based on student responses and other contingencies.(Knowledge)

- The teacher believes that plans must always be open to adjustment and revision based on student needs and changing circumstances.(Dispositions)

- The teacher responds to unanticipated sources of input, evaluates plans in relation to short- and long-range goals, and systematically adjusts plans to meet student needs and enhance learning.(Performances)

The teacher knows about it, believes in it, and does it. We are familiar with processes to assess knowledge. We can give a test. It is also not terribly difficult to observe a teacher's performance looking for his/her ability to adjust to unanticipated inputs. It is difficult to determine if the teacher believes in it enough to do it on his/her own and plan for it when no one is watching. But if we do not attempt to project whether the skills will continue to be applied in the "real" world, we have partially failed in our obligation to produce highly qualified teachers, leaving no child behind. Therein lies the challenge.

Literature Review

The NCATE Standards define dispositions (affect) as follows: "The ***values, commitments***, and professional ethics ***that influence behaviors*** toward students, families, colleagues, and communities and affect student learning, motivation, and development as well as the educator's own

professional growth." (p. 53, emphasis added), and even the Miriam-Webster On-line Dictionary helps us to see that dispositions are about tendencies to act rather than the skills themselves. Educators have often described constructs such as cognitive and affective objectives as different (Bloom, et al., 1956; Anderson and Krathwohl, 2001), thereby requiring different assessment techniques. They are different constructs for assessment purposes even though one can philosophically see a teacher as a composite of performance on several constructs simultaneously. The INTASC Principles (CCSSO, 1992) help articulate the differences between dispositions and skills by listing indicators for both constructs separately and at the same time align the knowledge, skills, and dispositions across principles.

Although there is literature on accreditation in general and measuring knowledge and skills, there is a paucity of literature on measuring teacher dispositions. Hopkins (1998) notes that the affective taxonomy has not had the impact on education that the cognitive taxonomy has had, partially because of the unique assessment problems associated with affective measurement.

Hopkins identifies the following affective measures: Scales, including ***Thurstone Attitude Scales****, Likert Scales, Rating Scales, and Semantic Differential Scales; ***Self-Report*** Inventories and ***Questionnaires*;** Interviews and ***Focus Groups****; ***Observations;*** and ***Projective techniques***. (Bold italics indicate the types of instruments currently under development by the session organizers, with asterisks indicating the instruments being field-tested and presented.)

A Battery of Assessments: Measurement Theory Applied

The Assessments described here represent initial data from a series of methods possible from less to more inference in the item types:

Less Inference			**More Inference**
Agree-Disagree Forced-Choice	***Questionnaire with Essay Answers***	***Focus Group with Kids***	Qualitative Text Analyses
Likert Response	Behavioral Checklist (Filled out by Peers)	Interview with Teacher Candidate	Abstract Projective (Ink Blot, TAT)
Historical Record (Fingerprint, etc.)	Scenario Analysis Essay Answers	Observation in Field	Trait Analysis of Handwriting, Verbals

The assessments highlighted above are the ones that are the subject of this report and analysis. Some of the others have been drafted or will be used in the future. All the instruments are measures of constructs that derive from the INTASC principles for Dispositions. All items are intended to measure the same constructs along a continuum of more to less of the dispositions defined in the principles. All item types are intended to calibrate on a ruler created using probabilistic conjoint scores estimated with the Rasch model.

Disposition Instruments Developed and Field-Tested at Increasing Levels of Inference (Progressive Measures)

Just like assessing knowledge and skills, we gain confidence that we have measured well when we progress through a series of well designed, progressive measures. To better understand the concept of progressive measures, we will look at a familiar example from knowledge and skills -- lesson planning and delivery as an example. Students can be tested to determine if they know the levels of Bloom's taxonomy, how to classify objectives and what the parts of a lesson plan are. Then they can be assessed on their ability to use the taxonomy to write a lesson plan (a product) which then they deliver (observation). The knowledge and skills applied by both the student and the professor become increasingly complex. They can guess the answers on the test, but we can score it easily by machine. They have lots of time to develop the lesson plan, which they can copy off the Internet or from a friend. We must apply a rubric to evaluate it, making judgments about quality. These judgments become more difficult in the observation than on the written work, but we may have more confidence that what we observe is real. These shifts in the assessment difficulty are what we refer to

as increasing levels of inference. In the case of dispositions, we would usually measure "consistency" with a stated attitude or belief associated with INTASC as our level of difficulty.

Scales that can be machine scored are at the lowest level of inference. As we start thinking about measuring dispositions, we know that when we ask students questions, it is easy for them to fake or guess the answers by anticipating what they think we want to hear. In a Thurstone agreement scale, the respondents have a 50% chance of guessing correctly or faking. This is not to say that scales do not work. There are still lots of respondents who cannot even guess or fake it -- they are oblivious to the dispositions they should have. So, scales help us to make a first cut at sorting those with dispositions appropriate to teaching and those who are clueless! Here is an example:

INTASC Principle	**Thurstone Statements**
3.4: The teacher is sensitive to community and cultural norms.	**Agree:** 3. I believe good teachers learn about the students' backgrounds and community so they can understand students' motivations. (98.2% Agreed)
N=1089	**Disagree:** 47**.** Many immigrants to the U.S. need school so they can learn the American way. (55.2% Disagreed)

Field test results of this instrument indicate that more teacher candidates respond correctly to the first question than the second one, which gives us some valuable information at the item level for what our students believe about community and cultural norms. Almost all know they needed to learn about students cultural backgrounds, but why was question 47 a toss-up. Here are the indicators described in INTASC which might apply to question 47 that make "Disagree" the consistent response:

3.2	The teacher appreciates and values human diversity, shows respect for students' varied talents and perspectives, and is committed to the pursuit of "individually configured excellence."
3.3	The teacher respects students as individuals with differing personal and family backgrounds and various skills, talents, and interests.
3.4	The teacher is sensitive to community and cultural norms.

In this case, we cannot be certain what caused the higher percentage of incorrect responses. Perhaps they did not really understand the question, so the results require some rational and empirical analysis, which will be the subject of the final paper in this series. Such methods help refine this basic analysis and make it more meaningful for interpreting our results for both candidates and programs. We are looking at the application of the Item Response Theory (Rasch model) of measurement to these data (Wright and Stone, 2004) and finding the diagnostic power of the Rasch Model to locate teachers who are on track and teachers who need some help with their values. For now, we know that not everyone believes what we want them to believe.

We also know that there are times when the results are clear, striking, and frightening without any further analysis. Such results lead to some unwelcome surprises – things we definitely need to investigate and work on. In our field test last year, we found that 23% of the respondents across three institutions did <u>not</u> believe that all children could learn. For us, that is a very alarming statistic.

Questionnaires and interviews are a little more difficult to score and a little more difficult to fake, so they provide the next level of useful assessment of dispositions. Unlike an agree/disagree scale, the respondents do not have a 50% chance of getting it right, but they can still give the socially acceptable answer. In each of these measures, the manner in which we pose the question is critically important to our ability to judge the amount of the construct they possess. With questionnaires and interviews, we can develop rubrics and anticipate likely responses. Here is an example:

INTASC Principle	**Questionnaire Item**
1.1: The teacher realizes that subject matter knowledge is not a fixed body of facts but is	How have you kept abreast of current developments in your field? For example, did you attend any workshops, subscribe to any journals, read or buy a new book? If so, describe in one to two

complex and ever-evolving. S/he seeks to keep abreast of new ideas and understandings in the field.	sentences something you learned and the source.

Sample responses here include the two following extremes, one candidate who articulates enthusiasm for learning more about the content area while another articulates satisfaction with the status quo:

- "I receive the *New York Times* online and have "education" as one of my highlighted topics. I … read what is going on in our country concerning education… I have bought three of E.D. Hirsch, Jr's books, *What your --- grader needs to know*… I have relearned things … forgotten that I will need … in the classroom." (Rated: Target)

- "I am only aware of developments in my field through school. What I have learned in school keeps me updated on what is going on in the school system." (Rated: Unacceptable)

At the next level of inference are focus groups and observations. These are more complex to analyse than simple questionnaires, often because there is interaction among the group members and conflicting evidence. Faking becomes extremely difficult at this point, because other people are involved. The complexity now, though, is that judgement has to be applied to sort good data from noise, and all this takes time. However, there is no substitute for first hand observation of a teacher's performance or for hearing what children have to say about their teacher. An example follows:

INTASC Principle	**Focus Group Questions**
5.2: The teacher understands how participation supports commitment, and is committed to the expression and use of democratic values in the classroom.	Group work: (1.2, 5.2) • Usually, when you work in groups, do group members tend to work alone and compile the work at the end, or do they tend to complete most/all components together? Does the teacher do anything to ensure that students work together? If so, what does he/she do? • When your groups do their work, do they attempt to reach consensus on group operations and products, or does one person tend to dominate? What does your teacher do if someone dominates the group?

Here is an example of a teacher with whom we would want to have a conversation, based on students' perceptions of her beliefs. Here is what five students of an intern said during a focus group:

- "Sometimes she tells us to work together. Sometimes she is loud about it."
- "If we sound like we are not working she yells at us to get to work. She does not yell loud, just sounds like it."
- "I think that the smart people get most of the attention. The dumber students don't get talked to as much as the smart ones."
- "Teacher talks a lot."
- "Sometimes she is not paying attention."

How would you rate these statements: Target, Acceptable, or Unacceptable? If Scott McConnell from Le Moyne College had demonstrated attitudes clearly inconsistent with INTASC, a good professor would challenge his expressed beliefs. If he had then continued to demonstrate those values and fail to improve with faculty intervention, we believe the evidence would be supported by courts and other faculty if a decision was made to deny him the program. Simply dismissing Scott since it "leaked out" that his values were unacceptable to the college was also unacceptable to the court. The question was not one of the court imposing values, but of fair process in assessment for Scott!

The DAATS Model:
Dispositions Assessments Aligned with Teacher Standards

The Dispositions Assessments Aligned with Teacher Standards (DAATS) model was designed to address the need for valid assessment systems comprised of standards-based instruments designed to determine candidates' consistency with the dispositions indicators of the INTASC Principles. It answers the basic question: Is the candidate committed to the values inherent in the skills that have defined as critical to effective teaching? The DAATS model consists of five steps. They are less linear than they appear, since designers need to revisit constantly the systems they are building. Ideas change; standards change; people change.

DAATS Step 1: Define purpose, use, propositions, content, and other contextual factors.

In this step, designers begin by determining why they need an assessment system (assessment purpose), the decisions they will need to make (use), what givens underlie their work (propositions), and what they want to know (assessment content). Each purpose and use are conceptualized and evaluated separately as a matter of validity. At the end of this step, designers analyse all the local factors that would affect the system, e.g., conceptual framework, resources, faculty resistance/cooperation, since these factors can impede or help them in their work.

DAATS Step 2: Develop a valid sampling plan.

A critical next step is the identification of all relevant standards and the alignment of standards with each other into assessment domains. In most cases, this will include the INTASC Principles and the institution's own conceptual framework. In some states, dispositions have been included in state standards, as well. When considered together, as a kind of content domain or a set of content domains, one can clearly see the similarities and differences between and among the perceptions of what is important from each group of professionals. Next, faculty members should visualize the competent teacher exhibiting the dispositions, brainstorm a series of items that elicit those attitudes, values, and beliefs, and then determine what methods would best be used to assess them. A blueprint linking items and methods can then be developed as a framework for instrument design. The costs and benefits for each method should be carefully considered.

DAATS Step 3: Create instruments aligned with standards and consistent with the sampling plan.

Using appropriate affective measurement techniques (e.g., writing statements that generate dissonance), the items should be written for each instrument. Assistance from measurement professionals may be helpful in this step. The instruments should be reviewed by a variety of stakeholders – teacher candidates, practicing teachers, school district personnel, etc., and then field tested. Clear directions need to be written.

DAATS Step 4: Design and implement the system and aggregate data for decision-making.

The data must be accumulated and managed for decision making, so decisions need to be made about how this will be done and what the procedures should be for counselling candidates. Rubrics need to be written for open-ended response items (e.g., questionnaires), and, anchor responses from the field test should be selected and used in developing these rubrics. A maintenance program is necessary and should be created to include training in the use of rubrics, collection of scored examples showing different levels on the rating scale, orientation of teachers being assessed, advising materials (including due process), and an appeals process. Formal review times to update and improve the tasks and the system should be established in advance. Identified people or committees responsible for data review is also important for the valid implementation of the system.

DAATS Step 5: Ensure credibility and utility of data.

There are increasing calls for ensuring the credibility of assessments, including validity, reliability, and fairness. Assessment designers should make use of the *Standards* (1999), including blueprints; a focus on job-relatedness; and evidence of validity (particularly content validity), reliability, and fairness. Logical as well as empirical data should be gathered. A plan to collect this evidence should be developed and implemented conscientiously.

Substeps of the model and a list of worksheets to be included in the book are included in Appendix A.

Conclusion

Measuring dispositions has become a very difficult issue for many institutions. Many institutions are also very focused on using disposition assessments to look at broad attributes such as lifelong learning or at professional behaviours such as punctuality and proper dress, but they are forgetting to look at some of the important dispositions in the INTASC Principles that can lead them to improving those fundamental attitudes that teachers need to have to ensure that they do what we have taught them to do because they want to do so. There is a solid rationale behind the INTASC Principles and the inclusion of dispositions, but they should not be viewed as the only dispositions that can be measured. Clearly, institutions should be comfortable adding dispositions that they value to the assessment system. Our primary concern here is that institutions use the established standards (INTASC Principles) as a minimum. Skipping them is unconscionable, since we need to ensure that teachers are ***likely*** to apply the skills they have learned in our colleges. Measuring dispositions may well prove to be one of the most important components of an assessment system.

The INTASC dispositions are a complex construct best measured using different types of instruments designed by measurement professionals in collaboration with teacher education faculty and other stakeholders. Such instruments need to take into consideration the importance of increasing levels of inference, so that a progressive set of measures helps us to build confidence in our decisions.

Dispositions are not only measurable, but they can be measured with results that are both valid and reliable. If one uses the INTASC Principles as the basis for designing instruments, evidence of construct validity should be present. Using a blueprint that ensures coverage of most or all of the dispositional statements in one form or another also adds to the evidence of content validity, so that we are more confidence that our decisions about candidate values are appropriate.

In this paper, we have presented a way to respond proactively to NCATE assessments without risking legal challenges or violating our sense of responsibility to our students. While we find it possible to be both liberal in our political views and committed to the use of standards-based assessment, we acknowledge that there are many who resist the use of standards and the accountability movement. Political views aside, the standards and modern measurement do provide a mechanism to ensure accountability and exclude those extreme groups such as the one that supported Mr. McConnell. If our decisions are clearly routed in accountability for standards, then the argument about political views driving our decisions about individual candidates becomes much weakened. Add to that the data to show that the instruments and processes we use are public, have due process considerations, and produce valid and reliable results, and we can prevail in court and produce better teachers at the same time.

Appendix A

Disposition Assessment Aligned with Teacher Standards --
The DAATS Model for Improved Teacher Assessment

DAATS Step 1: Define purpose, use, propositions, content, and other contextual factors.

DAATS Step 1A: Define the Purpose(s) and Use(s) of the System
DAATS Step 1B: Define the Propositions or Principles that Guide the System
DAATS Step 1C: Define the Content(s) of the System
DAATS Step 1D: Review Local Factors That Impact the System

Worksheets

Worksheet #1: Step 1: Purpose, Content, Use, Context

DAATS Step 2: Develop a valid sampling plan.

DAATS Step 2A: Analyse Standards and Indicators
DAATS Step 2B: Visualize the Teacher Demonstrating the Affective Targets
DAATS Step 2C: Select Assessment Methods at Different Levels of Inference
DAATS Step 2D: Build an Assessment Framework Correlating Standards and Methods

Worksheets

Worksheet #2.1: Organizing for Alignment (Version 1)
Worksheet #2.1: Organizing for Alignment (Version 2)
Worksheet #2.2: Visualizing the Dispositional Statements
Worksheet #2.3: Selecting Assessment Methods for INTASC Indicators
Worksheet #2.4: Assessment Methods for INTASC Indicators: Blueprint
Worksheet #2.5: Cost/Benefit and Coverage Analysis of Assessment Methods

DAATS Step 3: Create instruments aligned with standards and consistent with the sampling plan.

DAATS Step 3A: Draft items for each instrument
DAATS Step 3B: Review items

Worksheets

Worksheet #3.1: Creating Scales
Worksheet #3.2: Creating Questionnaires, Interviews, or K-12 Focus Group Protocols

Worksheet #3.3: Creating an Affective Behavior Checklist
Worksheet #3.4: Creating an Affective Behavior Rating Scale
Worksheet #3.5: Creating a Tally Sheet for Affective Observation:
Worksheet #3.6: Checklist for Reviewing Scale Drafts
Worksheet #3.7: Review Sheets for Questionnaires and Interviews
Worksheet #3.8: Review Sheets for K-12 Focus Group Protocols

Worksheet #3.9: Checklist for Reviewing Observations and Behavioral Checklists
Worksheet #3.10: Coverage Check
Worksheet #3.11: Rating Form for Stakeholder Review

DAATS Step 4: Design and implement the system and aggregate data for decision-making.

DAATS Step 4A: Develop Scoring Rubrics
DAATS Step 4B: Determine How Data Will Be Combined and Used
DAATS Step 4C: Develop Implementation Procedures and Materials

Worksheets

Worksheet 4.1: Explanation of Dichotomous Scoring Decisions
Worksheet 4.2: Rubric Design
Worksheet 4.3: Sample Format for Candidate/Teacher Tracking Form
Worksheet 4.4: Format for Data Aggregation
Worksheet 4.4: Decision Making Tool for Measurement Method
Worksheet 4.5: Sample Disposition Event Report
Worksheet 4.6: Management Plan

***DAATS* Step 5: Ensure credibility and utility of data.**

DAATS Step 5A: Create a Plan to Collect Evidence of Validity, Reliability, Fairness, & Utility
DAATS Step 5B: Implement the Plan Conscientiously

Worksheets and Samples

Worksheet #5.1: Assessment Specifications
Sample 1: Analysis of Appropriateness of Decisions for Teacher Failures
Sample 2: Expert Rescoring
Sample 3: Fairness Review
Sample 4: Analysis of Remediation Efforts and EO Impact

References

Anderson, L.W., & Krathwohl (Eds.). (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Bloom, B., Englehart, M. Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, Toronto: Longmans, Green.

Hopkins, K.D. (1998). *Educational and Psychological Measurement and Evaluation*. Allyn and Bacon: Boston.

Council of Chief State School Officers (1992). INTASC's *Model Standards for Beginning Teacher Licensing, Assessment and Development: A Resource for State Dialogue*. Washington, D.C.: Author. Retrieved Sept. 15, 2004 from http://www.ccso.org/projects/Interstate_New_Teacher_Assessment_and_Support_Consortium/

Miriam Webster On-Line Dictionary. Retrieved November 4, 2004 from http://www.m-w.com/

National Council for Accreditation of Teacher Education (2002). *Professional Standards for the Accreditation of Schools, Colleges, and Departments of Education.* Washington, D.C.: Author. Retrieved November 4, 2004 from http://www.ncate.org/2000/unit_stnds_2002.pdf

Wilkerson, J.R., & Lang, W.S. (December 3, 2003). Portfolios, the Pied Piper of teacher certification assessments: Legal and psychometric issues. *Education Policy Analysis Archives*, 11(45). www.epaa.asu.edu/epaa/v11n45/.

Wilkerson, J.R. & Lang, W.S. (2004). A standards-driven, task-based assessment approach for teacher credentialing with potential for college accreditation. *Practical Assessment, Research & Evaluation*, 9(12). Retrieved June 20, 2004 from http://PAREonline.net/getvn.asp?v=9&n=12

Wilkerson, J. and Lang, W.S. (July, 2004). Measuring Teacher Dispositions with Different Item Structure: An Application of the Rasch Model to a Complex Accreditation Requirement, International Objective Measurement Workshop (British for "Conference") in Cairns, Australia.

Wright, B. D. & Stone, M. H. (2004). *Making Measures*, Phaneron Press, Chicago.