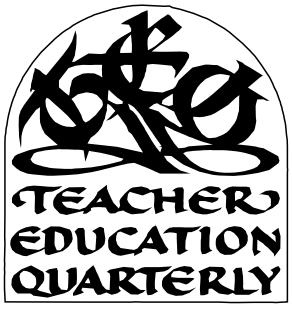


Assessment of Teacher Candidate Dispositions

Evidence of Reliability and Validity

By Hee-sook Choi, Nicholas F. Benson, & Nicholas J. Shudak

Since the National Council for Accreditation of Teacher Education (NCATE), now known as the Council for the Accreditation of Educator Preparation (CAEP), adopted its Standards 2000 that required "professional knowledge, skills, and dispositions necessary to help all students learn" for teacher candidates, literature has been replete with debates concerning the definition and assessment of professional dispositions (e.g., Burant, Chubbuck, & Whipp, 2007; Duplass & Cruz, 2010; Ruitenberg, 2011; Welch, Pitts, Tenini, Kuenlen, & Wood, 2010). A by-product of the ongoing and contentious debates over the definition of dispositions is reflected in the subsequent revisions of the NCATE/CAEP standards (NCATE 2008 and CAEP 2013 Standards); that is, although NCATE/CAEP continues to incorporate the assessment of teacher candidates' dispositions into its standards, it has gradually moved away from explicitly defining what dispositions are necessary for becoming an effective teacher. A case in point is that no glossary includes the term *professional dispositions* in its 2013 standards, a notable departure from the

Hee-sook Choi is professor emerita with the School of Education at the University of South Dakota, Vermillion, South Dakota. Nicholas F. Benson is an associate professor of educational psychology in the School of Education at Baylor University, Waco, Texas. Nicholas J. Shudak is chair of the Division of Curriculum and Instruction of the School of Education at the University of South Daokta, Vermillion, South Dakota. Email addresses: hchoi@usd.edu, Nicholas_Benson@baylor.edu, & Nicholas.Shudak@usd.edu

previous standards (see Appendix A for complete definitions included in the 2002 Edition Standards and 2008 Standards).

It is, then, no surprise that the definition of *dispositions* varies greatly among teacher education programs. It is up to individual teacher education programs to identify a set of dispositions they surmise to be critical for their candidates' success. The absence of a clear, universally accepted definition of what dispositions are, however, calls into question not only the assessment of dispositions but also the utility of assessing them. Critics contend that without definitional clarity, a reliable and valid method for measuring dispositions does not exist (Borko, Liston, & Whitcomb, 2007). So, then, is assessing dispositions nothing but a futile exercise for the sake of meeting the NCATE/CAEP requirements? Despite the "skepticism about the ability to assess dispositions as if they are comparable to skills and knowledge" (Duplass & Cruz, 2010, p. 144), and despite scant empirical data that dispositions can be assessed in a reliable and valid manner, all NCATE/CAEP accredited institutions continue to assess their teacher candidates' dispositions. Duplass and Cruz went on to say, "Whether dispositions can or cannot be defined or effectively assessed is apparently a moot point for NCATE, but it has left COEs [colleges of education] with the herculean task of implementing an ill-defined concept" (p. 144).

Dispositions Assessment in a Historical and Research Context

To understand what dispositions are or what a disposition might be, perhaps a first step is to juxtapose *dispositions* with two other words most commonly associated with dispositions: *knowledge* and *skills*. The field of teacher education has long held that there are ways to assess the identifiable, relevant, and important skills and knowledge necessary for teaching. Though content and pedagogical skills tests have experienced multiple iterations, overall they are considered reliable and valid, and some version of a content and skills test exists in every state for certification purposes. However, dispositions have been more difficult to define, let alone to assess validly and reliably. According to NCATE (2008), we can understand dispositions as those "professional attitudes, values, and beliefs demonstrated through both verbal and nonverbal behaviors as educators interact with students, families, colleagues, and communities" (p. 89). In other words, dispositions are those internal conditions (attitudes, values, beliefs, thoughts, etc.) that influence our external behaviors (actions and interactions with students and others). Historically, the struggle has been concerned with whether those internal conditions are naturally endowed, whether they can be implanted, and whether they can be shaped if endowed or implanted. After all, if dispositions cannot be shaped, then how teacher education programming promotes them, let alone assesses them, would be drastically impacted.

Arguably one of the first to struggle philosophically with something akin to dispositions was Plato. Perhaps most famously in *The Republic*, Plato (1991, pp. 124–125), through Socrates, argues for the necessity of a noble lie as a hypothetical

starting point to ordering the State. Through this lie, Socrates frames an argument for how best to order the soul of the State and the souls of its citizens. It is an argument for determining who gets to do what—rulers, guardians, artisans—within the State by having the founders of the State identify citizens who "God has framed"—or disposed—to serve in one way over others (i.e., identifying those who get to teach). More clearly, however, dispositions become evident in Socrates' allegory of the cave (Plato, 1991, pp. 253–261). Through the allegory, Socrates broaches the possibility that certain "natures" can be put into one's soul, that they can be circumcised from a soul, or that there is a process, or an art, by which to turn a soul from one direction and toward another—a process that might be painful yet necessary. But if we are to think of dispositions as a kind of virtue, it is in Plato's *Meno* (Plato, 1984) that Socrates struggles over whether virtue can be taught or if it is simply something with which a person is born.

In looking at dispositions as virtue, and in questioning the assessability of dispositions, educational philosopher Douglas McKnight would have us consider the work of Aristotle. According to McKnight (2004),

> Aristotle defined the term disposition as the nature of virtue or vice in relation to the agent and the possession of a particular frame of mind in any given ethical or moral situation. [Aristotle] explained [that] a disposition is one thread in a highly complex and pervasive ethical existence that begins with a child being inculcated into virtuous habits as defined by a community. (p. 214)

It is the lengthy process through which and the environment in which virtue is inculcated that gave McKnight pause as to the viability of assessing and even adjusting dispositions through college course work. Something developed over a lifetime cannot be changed over the course of a few semesters.

To give clearer insight into Aristotle's construction of dispositions and the concern about adjusting them, McKnight relied on the work of scholar Sarah Broadie. According to Broadie (as quoted in McKnight, 2004),

> a virtuous person is one who is such as to, who is disposed to, act well when occasion arises. And so far as 'acting well' implies not merely causing certain changes in the world, but doing so in the right frame of mind or with right motive, a disposition to act well is also a disposition to act in the right frame of mind. (p. 220)

As dispositions are possibly understood as "deeply rooted habits of thought and feelings" (p. 214) developed over a lifetime, they are not things easily dug up, displaced, or changed. Thus the project to assess, guided by the understanding that dispositions can be changed or adjusted, is very problematic, misguided, and even too late. For McKnight, the better proposition is to study dispositions rather than to assess them.

Bringing things into the 20th century, the work of Dewey and Combs informs. For Dewey (1922), the profit of education is the habits of mind that develop over time, in particular, the habit to engage in discriminating inquiry and action that thus

impel people to make right decisions within a given democratic context. For Dewey (1938/1997), educative experiences, as opposed to those miseducative experiences, align the interaction between what one thinks (internal) is right and what one does (external) that is right. Dewey's concern was with an entire pedagogical way of being in schools and society, which is analogous to our contemporary conversations about dispositions. For Dewey, these habits and alignments are transmissible through a carefully crafted environment.

Several years later, Arthur Combs and a team of psychologists engaged in a series of studies regarding those personal "perceptions" that might make someone a more effective teacher. The five categories Combs and his team identified are quite salient to our current concerns with dispositions. According to Combs et al. (1969), the five perceptions are of self, others, one's subject field, the purpose and processes of education, and a general category. Combs's study into perceptions was essentially a concern about how what one thinks as a teacher influences what one does as a teacher. Similarly to Dewey, it is a concern that links thinking with doing and a belief that thoughts and actions are adjustable, even if naturally endowed.

More contemporarily, and on a conceptual level, Jung and Rhoades (2008) challenged teacher educators to think differently about dispositions prior to any concerns with assessing them. According to Jung and Rhoades, "the task of evaluating and determining teacher dispositions in teacher education has not been easy due to conceptual and technical complications" (p. 649). To make dispositions assessment more valid and reliable, Jung and Rhoades challenged teacher educators to become clearer as to what they really want to assess. For example, in their view, teacher education has inordinately focused on dispositions as "character-related," even though many programs are becoming more and more "competency-related" (p. 647). The difference between the two orientations is the difference between finding how comfortable candidates feel about differentiation in classrooms and finding how competently candidates enact differentiation in classrooms. Clarifying what they refer to as character- versus competency-related dispositions will help determine the type of programming one needs and the types of assessments used throughout, giving teacher educators a more valid and holistic look at their candidates.

Brewer, Lindquist, and Altemueller (2011) studied whether candidates identified as having problematic dispositions can improve on them through an intervention process. Through qualitative methodology, and using an intervention they called the "Professional Dispositions Qualities: Preparing Reflective and Effective Practitioners" process, the authors found that candidates with problematic dispositions (e.g., resistance to feedback and correction and unwillingness to collaborate professionally with colleagues) can improve on their problematic dispositions after following the intervention. The authors did admit, however, that "not all individuals are as capable or motivated as our case examples" (p. 65), putting into question the validity of their intervention. If a candidate does not improve, is it because the intervention was not valid, because the candidate truly is unredeemable dispositionally, or because the

programming was not good and thus the intervention was necessary? Borrowing again from Jung and Rhoades (2008), it does appear that Brewer et al. (2011) were concerned with demonstrable competency-related dispositions evidenced in teaching contexts. The question the study does not address is whether those dispositions were intentional components of the teacher education programming.

According to Cummins and Asempapa (2013), choosing education as a major and choosing teaching as a career are seemingly indicators that a candidate is already disposed toward teaching. If that premise is true, as Cummins and Asempapa seemed to believe it is, then "dispositions can be viewed like other professional skills, mainly observable behaviors that are intentional and applied within an educational setting, [and] they can be taught and supported by educational experiences" (p. 105). To that end, Cummins and Asempapa conducted a qualitative study coupled with pre- and posttest analyses to examine whether teacher candidates' knowledge and understanding of the dispositions of professionalism, collaboration, and inclusion can be altered through teaching interventions in a course. Though some growth in all three dispositions was evidenced, results from the pre- to postassessment were statistically insignificant. Thus one of the questions the researchers left with the reader was whether they could have done a better job adjusting content after the pretests in light of the fact that some students came in with higher levels of knowledge and understanding of some dispositions (professionalism) over others (collaboration and inclusion). Though the results were not statistically significant, the authors were cautiously optimistic that growth through course work is possible.

Echoing Brewer et al. (2011), and though dispositions are only a recent accreditation requirement for teacher education programming, "dispositions are not a new research topic in the field of education. . . . [They have] been researched for many years with questionable outcomes" (p. 52). As the preceding review indicates, the discourse's concern with the internal quality of a person, with his or her affective and cognitive attributes that might influence or impact the external actions the person takes and the interactions he or she has in classrooms with students, is an ancient one. In other words, this is something that is seemingly eternal and with enduring importance; thus teacher education needs to continue taking it seriously, as our study does.

Statement of the Problem

Our teacher education program has used a dispositions measure since the inclusion of professional dispositions in the NCATE Standards 2000. It has gone through revisions, the latest of which took place shortly after the publication of the NCATE 2008 Standards. The 2008 Standards stipulate that other than "fairness and the belief that all students can learn," professional education units can "identify, define, and operationalize additional professional dispositions" in accordance with "their mission and conceptual framework" (p. 89). Similar to the procedures used

for assessing candidate dispositions by other professional education units (e.g., Almerico, Johnston, Henriott, & Shapiro, 2011), our teacher education program identified a set of six dispositions based on a review of related literature and a series of faculty input: responsibility, respect, integrity, caring/humanity, fairness, and the belief that all students can learn.

The current study is an attempt to inform professionals, such as university trainers and field-based supervisors, as to the utility of professional dispositions assessment as a means of gauging teacher candidate success in their professional roles. According to the *Standards for Educational and Psychological Testing* (*Standards*; American Educational Research Association, American Psychological Association, & National Council of Measurement in Education, 2014), test users must select measures that accurately reflect the measurement targets of interest and have validity for intended purposes. Are supervisors' ratings of professional dispositions supported by reliability and validity evidence? Do dispositions make a difference in practice? To address these questions, we examined evidence for reliability such as internal consistency, stability, and interrater agreement. To this end, we explored the structure of our dispositions rating form as to its consistency with its intended structure. We also examined whether these dispositions are measured consistently across raters and time. Finally, criterion-related evidence for validity was examined by correlating teacher candidate disposition ratings with teaching effectiveness, which was measured by direct observation of their levels of student engagement.

Method

Participants

Participants included all teacher education candidates enrolled during a fall semester ($N = 147$) at a rural, midwestern university. Of those, the students of 53 candidates were also observed to measure their levels of engagement while the candidate led the class. Demographic information for the participants is presented in Table 1.

Table 1

Descriptive Statistics for the Total Sample and the Student Engagement Subsample

			Race and ethnicity				
Sample	Age, M (*SD*)	Sex, %F	%W	%B	%H	%API	%O
Total[a]	24.6 (4.5)	70.7	95.2	0.7	1.4	1.4	1.4
Subsample[b]	23.85 (2.78)	83.0	90.6	0.0	3.8	1.9	3.8

Note. API = Asian or Pacific Islander; B = Black; F = female; H = Hispanic; O = other; W = White.
[a] $N = 147$.
[b] $n = 53$.

Instrumentation

The teacher education dispositions rating form consists of 19 items completed by university and field supervisors (see Appendix B). The items are rated on a 3-point scale ranging from 1 (*below expectations*) to 3 (*exceeds expectations*). This form was designed to measure the following six professional dispositions: responsibility, respect, integrity, caring/humanity, fairness, and the belief that all students can learn. The rating form is essentially a rubric consisting of four items each for responsibility, respect, and integrity. Caring/humanity comprises three items, and there are two items each for fairness and belief that all students can learn. For example, one of the responsibility items defines "does not identify or complete needed tasks without specific direction" as below expectations (1), "identifies and completes needed tasks with little or no direction" as meets expectations (2), and "demonstrates leadership and takes initiative in identifying and completing tasks" as exceeds expectations (3). Descriptive statistics, along with the percentage of missing data for each item, are presented in Table 2. These data represent ratings provided by university supervisors.

Student engagement was assessed using the Behavioral Observation of Students in Schools (BOSS; Shapiro, 2010). Descriptive statistics, along with the percentage of missing data for BOSS variables, are presented in Table 2. The BOSS measures two categories and five subcategories of behavior within the classroom environment. The first major category is academic engagement, which is split into two subcategories: (a) active engaged time (AET) and (b) passive engaged time (PET). AET is coded when a student is actively engaged in academic responding (e.g., reading aloud, making appropriate verbal comments, writing responses to classwork assignments), and PET is coded when a student appears to be passively engaged with instruction (e.g., listening to instruction, looking at relevant instructional stimuli). The second major category is nonengagement, which is split into three subcategories corresponding to the form of responses that appear to be incompatible with academic engagement: (a) off-task motor (OFT-M), (b) off-task verbal (OFT-V), and (c) off-task passive (OFT-P).

AET and PET are scored using momentary time sampling and are coded if responses corresponding to these categories are displayed at the beginning of a 15-second interval. OFT-M, OFT-V, and OFT-P are coded during the remainder of the 15-second intervals using the partial interval method. Evidence exists to support using the BOSS as a measure of academic engagement in classroom settings (Volpe, DiPerna, Hintze, & Shapiro, 2005). Notably, the BOSS was designed for observations that focus on a target student and a comparison peer. We modified administration procedures for our study so that a different student was observed at every interval. Observers were instructed to track students clockwise and code a different student every interval, coding the same student twice only after every student had been observed. As our goal was to assess teacher candidates' ability to

engage students, this modification was made to improve sampling and minimize the effects of internal student variables. We believe this modification helped ensure that observations reflected the extent to which the teacher candidates generally engaged students in academic learning.

Setting for BOSS Observations

BOSS observations were conducted within K–12 classrooms, each with eight or more K–12 students. Observations were arranged by a graduate assistant who was responsible for coordinating these observations, with the goal of selecting a random sample of teacher candidates.

Table 2
Descriptive Statistics for Dispositions Ratings and Behavioral Observation of Students in Schools Variables

Variable	*N*	*M*	*SD*	Sk (*SE*)	Ku (*SE*)	% Miss
Q1	147	2.61	.50	-0.60 (0.20)	-1.26 (0.40)	0
Q2	147	2.69	.46	-0.85 (0.20)	-1.30 (0.40)	0
Q3	147	2.61	.52	-0.75 (0.20)	-0.75 (0.40)	0
Q4	146	2.55	.55	-0.71 (0.20)	-0.56 (0.40)	1
Q5	146	2.56	.50	-0.25 (0.20)	-1.96 (0.40)	1
Q6	147	2.47	.51	-0.03 (0.20)	-1.64 (0.40)	0
Q7	147	2.40	.49	0.42 (0.20)	-1.84 (0.40)	0
Q8	145	2.62	.49	-0.50 (0.20)	-1.77 (0.40)	1
Q9	147	2.74	.47	-1.46 (0.20)	1.05 (0.40)	0
Q10	147	2.73	.45	-1.04 (0.20)	-0.94 (0.40)	0
Q11	147	2.45	.50	0.22 (0.20)	-1.97 (0.40)	0
Q12	145	2.62	.49	-0.50 (0.20)	-1.77 (0.40)	1
Q13	147	2.59	.49	-0.38 (0.20)	-1.88 (0.40)	0
Q14	147	2.54	.50	-0.15 (0.20)	-2.00 (0.40)	0
Q15	147	2.61	.50	-0.62 (0.20)	-1.22 (0.40)	0
Q16	147	2.57	.51	-0.44 (0.20)	-1.43 (0.40)	0
Q17	147	2.41	.53	-0.03 (0.20)	-1.12 (0.40)	0
Q18	147	2.49	.50	0.06 (0.20)	-2.02 (0.40)	0
Q19	147	2.51	.50	-0.06 (0.20)	-2.02 (0.40)	0
AET	53	41.45	23.10	0.36 (0.33)	-0.75 (0.64)	0
PET	53	39.99	20.06	-0.01 (0.33)	-0.37 (0.64)	0
TET	53	81.44	11.30	-0.50 (0.33)	-0.04 (0.64)	0
OFT-M	53	8.30	6.31	0.77 (0.33)	-0.05 (0.64)	0
OFT-P	53	7.52	7.48	0.67 (0.33)	-0.77 (0.64)	0
OFT-V	53	5.65	4.74	1.26 (0.33)	1.85 (0.64)	0
TOT	53	21.47	12.86	0.25 (0.33)	-0.77 (0.64)	0

Note. AET = active engaged time; OFT-M = off-task motor; OFT-P = off-task passive; OFT-V = off-task verbal; PET = passive engaged time; Q = question from the dispositions rating form; TET = total engaged time; TOT = total off-task.

Procedure

Using the dispositions rating form, both university and field supervisors rated their student teachers twice: at midterm and then at the end of the semester. The form was administered through the teacher education program's online student tracking system. Evaluations for various field experiences, including the dispositions rating form, are administered through this online database.

BOSS observations were 30 minutes in duration, consisting of 120 15-second intervals. Coding was completed by seven observers with classroom teaching experience who were hired by the teacher education program to serve as university supervisors for teacher candidates. These observers did not conduct observations on teacher candidates they were personally supervising. All observers attended a 1-hour training session covering the BOSS administration and scoring procedures, and all observers were asked to rate a video immediately following training. Interclass coefficients (ICC; Shrout & Fleiss, 1979) were calculated to examine interrater reliability. A two-way random effects model was used because all raters rated the same video and the behavior of students in the video was viewed as a random sample of behavior rather than as a fixed effect. ICC coefficients for BOSS scores ranged from .88 for ratings of academic engagement to .95 for OFT-V. These results suggest that raters understood the criteria for coding behaviors specified in the BOSS manual and produced reliable ratings across raters.

Data Analysis

Confirmatory factor analysis was conducted using Mplus Version 7.11 (Muthén & Muthén, 1998–2012). Because students start their programs of study at different times, some teacher candidates in our sample completed their student teaching experience in the fall, whereas others completed their student teaching experience in the spring. Thus not all teacher candidates received ratings for both the fall and spring semesters. A combined file ($N = 147$) consisting of ratings from the fall and spring was used for the CFA. For students who had ratings from both semesters, the spring rating was used. Raw scores for each of the 19 items were used as input. The weighted least squares with mean and variance adjustment (WLSMV) estimation method was used because dispositions are rated on a 3-point scale. However, estimates of skew and kurtosis are not reflective of severe issues with nonnormality. Missing data were addressed by using full-information maximum likelihood estimation, a preferred approach for addressing missing data (Enders, 2010; Schafer & Graham, 2002).

Model comparisons were used to determine the best fitting structural model for the professional dispositions rating form. As we used WLSMV estimation, a robust chi-square difference test ($\Delta\chi^2$) was used when comparing models. We also used the following measures of fit and criteria: (a) chi-square (χ^2; statistically nonsignificant values), (b) the comparative fit index (CFI; >.95), and (c) the root mean square error of approximation (RMSEA; <.05).

A bifactor modeling approach was used to allow for simultaneous examination of (a) the direct effects of a general dimension on each of the 19 items and (b) the direct effects of the six specific dispositions on the items intended to measure them. Bifactor models posit that there are two systematic and direct influences on test scores (Gignac, 2008; Reise, 2012). The use of a bifactor model allowed us to compare the relative importance of general versus specific effects on items. We conducted comparisons of nested models to test the structure of the dispositions rating scale. This involved sequentially removing latent disposition variables from the model to determine if their removal caused model fit to degrade. If the removal of a latent variable did not degrade model fit, as indicated by a statistically significant $\Delta\chi^2$, then the latent variable was considered to be superfluous.

The use of a bifactor model also allowed us to examine the model-based reliability of the general dimension as well as the six specific dispositions. Reliability was examined using coefficient omega total (ω_T; Lucke, 2005; Revelle & Zinbarg, 2009), which accounts for all common sources of variance; omega hierarchical (ω_H; Reise, 2012), which accounts for variance from the general dimension; and omega subscale (ω_S; Reise, 2012; Zinbarg, Yovel, Revelle, & McDonald, 2006), which accounts for variance attributable to specific dispositions. These omega indexes are calculated by placing construct-relevant variance in the numerator and construct-relevant variance plus error in the denominator. When more than one factor has direct effects on items, omega indexes provide more accurate estimates of reliability than do traditional methods of estimating reliability, such as coefficient alpha.

In addition to examining the structure of the dispositions rating form, reliability was examined across time and rater. A subsample of participants (n = 102) had disposition ratings completed at midterm as well as at the end of the semester. Items were summed to create disposition composites. Pearson correlation coefficients were obtained using SPSS Version 21 to estimate the stability of ratings for each disposition composite as well as a composite score derived by summing all 19 items. Teacher candidates who were on student teaching assignments (n = 82) had ratings from both a field-based supervisor and university supervisor that were completed at the end of the same semester. Pearson coefficients were obtained using SPSS Version 21 to estimate interrater reliability.

Finally, the percentage of time K–12 students were academically engaged (i.e., coded as AET or PET) or nonengaged (i.e., coded as OFT-M, OFT-V, or OFT-P) was calculated by dividing the number of intervals coded by 120. These percentages, which reflect BOSS ratings for teacher candidates obtained in the fall semester of their student teaching experience, were correlated with disposition variables. This allowed us to examine if dispositions are related to teaching performance.

Results

As shown in Table 3, a model that included a general dimension and six pro-

fessional dispositions (Model A) fit the data reasonably well. Removing the latent general dimension (Model B) caused the model to fit poorly. Model fit also degraded, as evidenced by a statistically significant $\Delta\chi^2$, with the latent caring/humanity (Model G) and responsibility (Model H) variables removed. Thus Model F, which is presented in Figure 1, provided the best fit to the data. This model includes a general dimension as well as the caring/humanity and responsibility dispositions. As shown in Table 4, the general dimension had strong effects on all items, whereas the effects of specific dispositions were modest.

Examination of omega estimates presented in Table 5 suggests that a total score derived by summing ratings across all 19 items has high internal consistency. Moreover, almost all of the reliable variance in the total score can be accounted for by a general dimension. The responsibility and caring/humanity dispositions were also found to have high internal consistency, as indicated by high ω_T values. However, as indicated by ω_H estimates that exceed .8, the bulk of variance in these specific dispositions is accounted for by the general dimension. Small ω_S values indicate that only small proportions of reliable, specific variance are uniquely attributable to the six dimensions. In other words, most of the covariance among rating scale items is accounted for by the common variance attributable to the general dimen-

Table 3

Fit of Alternative Structural Models

Model	χ^2 (*df*)	*p*	$\Delta\chi^2$ (*df*)	*p*	CFI	RMSEA (90% CI)
A. General dimension + 6 dispositions	178.15 (136)	<.01	–	–	.996	.046 (.024–.064)
B. Model A minus general dimension	9551.97 (155)	<.01	4107.556 (19)	<.01	.154	.642 (.631–.6530
C. Model A minus BACCL	178.78 (137)	<.01	0.69 (1)	.41	.996	.046 (.024–.063)
D. Model C minus fairness	180.32 (138)	<.01	2 (1)	.16	.996	.046 (.024–.063)
E. Model D minus respect	183.21 (142)	.01	3.17 (4)	.53	.996	.044 (.022–.062)
F. Model E minus integrity	184.97 (145)	.01	4 (3)	.26	.996	.043 (.021–.061)
G. Model F minus caring/humanity	192.99 (148)	<.01	11.57 (3)	<.01	.996	.045 (.025–.062)
H. Model F minus responsibility	219.91 (149)	<.01	32.42 (4)	<.01	.994	.057 (.04–.072)

Note. BACCL = belief that all children can learn; CFI = comparative fit index; RMSEA = root mean square error of approximation.

sion, and the results suggest that there is not sufficient reliable, specific variance to discriminate among six dimensions.

Test–retest and interrater reliability estimates are presented in Table 6. Results indicate that only responsibility and the composite score were rated somewhat consistently across time. Although some correlations were statistically significant, all interrater reliability estimates were small. Thus ratings obtained from field-based supervisors do not correspond well with ratings obtained from university supervisors. Finally, as shown in Table 7, none of the disposition variables, including the disposition total, correlated significantly with student engagement. Thus disposition ratings do not appear to predict the ability to engage students in the classroom.

Discussion

By and large, a one-factor model best explains the internal structure of our dispositions rating form, coupled with high internal consistency. Although responsibility and caring/humanity explain a nontrivial amount of variance, these factors do not have sufficient reliable, unique variance to support interpretation.

Table 4

Standardized Factor Loadings for Model C

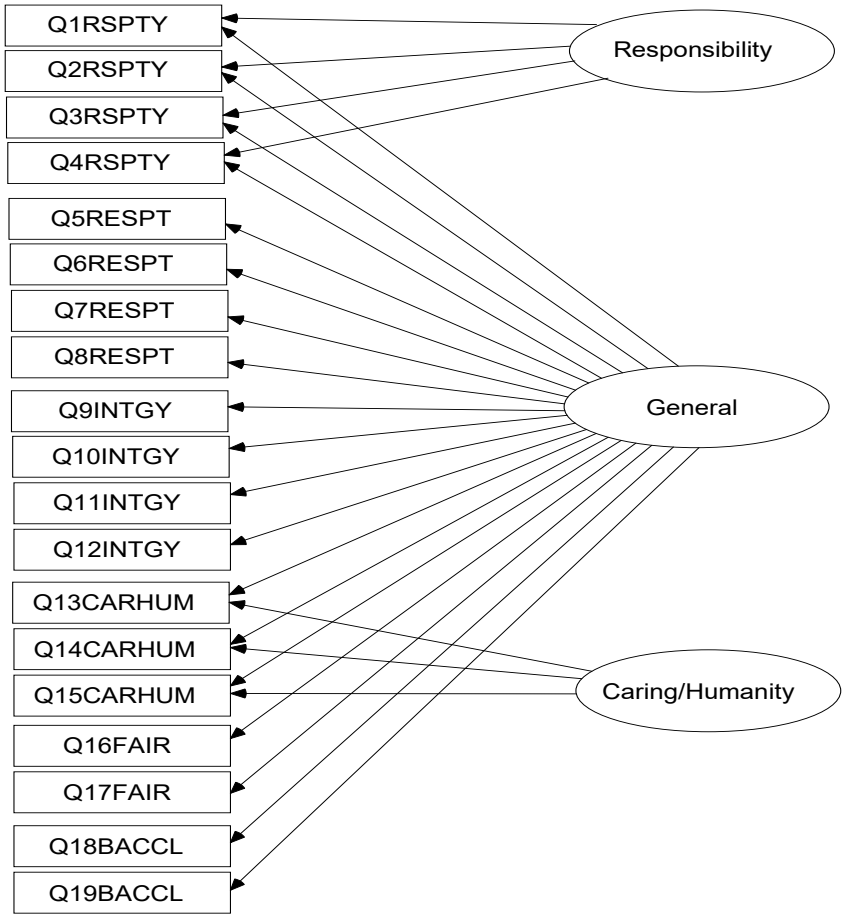
Item	General dimension	Responsibility	Caring/humanity
Q1_Responsibility	0.72	0.30	–
Q2_Responsibility	0.82	0.08	–
Q3_Responsibility	0.90	0.35	–
Q4_Responsibility	0.82	0.57	–
Q5_Respect	0.87	–	–
Q6_Respect	0.90	–	–
Q7_Respect	0.84	–	–
Q8_Respect	0.88	–	–
Q9_Integrity	0.94	–	–
Q10_Integrity	0.90	–	–
Q11_Integrity	0.93	–	–
Q12_Integrity	0.98	–	–
Q13_Caring/humanity	0.88	–	0.30
Q14_Caring/humanity	0.88	–	0.18
Q15_Caring/humanity	0.81	–	0.48
Q16_Fairness	0.98	–	–
Q17_Fairness	0.84	–	–
Q18_Belief that all children can learn	0.90	–	–
Q19_Belief that all children can learn	0.93	–	–

Note. Q = question from the disposition rating form.

Moreover, the results of this study do not support the two dispositions, "fairness" and "belief that all students can learn," required by the NCATE 2008 Standards to be separate and distinct from the others. Evidence for test–retest reliability was generally weak, with the exception of responsibility, whereas evidence for inter-rater reliability was poor. These findings suggest that conceptualizing dispositions to be a single, global dimension, rather than a set of distinguishable dimensions, may be a good starting point when developing a dispositions measure. The next step is to determine what dispositions signify within the context of teacher education programs and to define them in behavioral terms for assessment. To ensure that the specified dispositions are assessed reliably across time and raters, focus

Figure 1

Bifactor model with a general dimension and two specific dispositions (Model F).



should then be given to training university and field supervisors as to how to use the rating form with consistency.

More importantly, the results of this study indicate that dispositions ratings may not be good predictors of effective teaching. One important and observable

Table 5

Omega Estimates Based on Model C

Disposition	ω_H	ω_S	ω_T
General dimension	.98	.01	.99
Responsibility	.81	.13	.94
Caring/humanity	.83	.12	.94

Note. ω_H = omega hierarchical; ω_S = omega subscale; ω_T = omega total.

Table 6

Pearson Correlations Reflecting Test–Retest and Interrater Reliability for Professional Dispositions Ratings

Disposition	Test–retest	Interrater
Responsibility	.85**	.28*
Respect	.46**	.25*
Integrity	.44**	.19
Caring/humanity	.32**	.17
Fairness	.50**	.24*
BACCL	.43**	.17
Composite	.67**	.27*

Note. BACCL = belief that all children can learn.

*Correlation is significant at the .05 level. **Correlation is significant at the .01 level.

Table 7

Pearson Correlations Between Professional Dispositions and Behavioral Observation of Students in Schools Variables

	Responsibility	Respect	Integrity	Caring/ humanity	Fairness	BACCL	Disposition total
AET	.07	.16	.08	.05	.15	.20	.13
PET	-.06	-.09	-.05	-.05	-.13	-.15	-.09
TET	.04	.16	.08	.00	.08	.14	.09
OFT-M	.07	.04	.08	.15	-.07	-.10	.05
OFT-V	-.29*	-.25	-.26	-.13	-.23	-.24	-.28*
OFT-P	.09	.09	-.02	.23	.17	.01	.11
TOT	-.02	-.02	-.07	.16	-.02	-.13	-.02

Note. AET = active engaged time; BACCL = belief that all children can learn; OFT-M = off-task motor; OFT-P = off-task passive; OFT-V = off-task verbal; PET = passive engaged time; TET = total engaged time; TOT = total off-task.

component of effective teaching is the ability to engage students in learning tasks. Student engagement is frequently viewed as a means of increasing academic achievement and reducing problems like student boredom, alienation, and dropout (Fredricks, Blumenfeld, & Paris, 2004). Moreover, student engagement in learning is strongly correlated with academic achievement (Graden, Thurlow, & Ysseldyke, 1982). Our results indicate that there is no significant correlation between our teacher candidates' dispositions ratings and their ability to engage students in learning.

Given the continued debate over how to define dispositions, let alone how to assess them, our results beg the question, Is it necessary to assess something as elusive as dispositions? All NCATE/CAEP accredited institutions routinely assess their teacher candidates' dispositions multiple times during their field experiences. This has been an accepted practice even though there is a lack of a clear understanding of the necessary dispositions associated with teaching. What is truly disconcerting is that the dispositions we identified to be germane to teaching effectiveness (i.e., responsibility, respect, integrity, and caring/humanity) are fairly common and can be found in many institutions' dispositions instruments.

Overall, the findings of this study appear to support the skepticism regarding the assessment of dispositions in the teacher education field (e.g., Borko et al., 2007). Our findings challenge the utility of assessing dispositions; that is, unless the term *dispositions* is clearly understood and defined, it cannot be reliably and validly assessed. The ostensible lack of validity evidence to support the interpretation of scores derived from measures of professional dispositions is concerning, as evidence is needed to establish the objectivity and fairness of such measures given that they are commonly used to evaluate teacher candidates. As previously noted, the *Standards* (American Educational Research Association et al., 2014) clearly state that test users must select measures that accurately reflect the measurement targets of interest and have validity for intended purposes.

Our study does, of course, have limitations. Our sample comprised primarily White women in a rural setting in the northern plains region of the United States. Though the sample set is of a good size ($N = 147$), the teacher education candidates who were a part of this research are largely the same "type" of student. Perhaps a more demographically diverse set would provide different research results. Furthermore, and what might be a limitation of many teacher education programs in general, there was little consistency in terms of teaching about the particular professional dispositions on which the candidates were going to be assessed. In other words, the six dispositions being assessed were not programmatically embedded or developmentally taught throughout the candidates' programs. This is not to say that they were never a part of the candidates' course work and/or fieldwork; rather, it is to say that if teacher education candidates are going to be assessed on something in any kind of a high-stakes way, then that thing—dispositions—should be clearly and continuously found in course work and fieldwork. There is a lesson to be learned from this, too: Should dispositions be worth assessing, they must be

worth teaching about or developing in some programmatic way. And as intimated earlier, definitions in this regard will matter.

References

Almerico, G. M., Johnston, P., Henriott, D., & Shapiro, M. (2011). Dispositions assessment in teacher education: Developing an assessment instrument for the college classroom and the field. *Research in Higher Education Journal, 11*, 1–19.

American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

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

Brewer, R. D., Lindquist, C., & Altemueller, L. (2011). The dispositions improvement process. *International Journal of Instruction, 4*(2), 51–68.

Burant, T. J., Chubbuck, S. M., & Whipp, J. L. (2007). Reclaiming the moral in the dispositions debate. *Journal of Teacher Education, 58*, 397–411.

Combs, A. W., Soper, D. W., Goodling, C. T., Benton, J. A., Dickman, J. F., & Usher, R. H. (1969). *Florida studies in the helping profession* (Social Science Monograph No. 37). Gainesville, FL: University of Florida Press.

Cummins, L., & Asempapa, B. (2013). Fostering teacher candidate dispositions in teacher education programs. *Journal of the Scholarship of Teaching and Learning, 13*(3), 99–119.

Dewey, J. (1922, October 4). Education as politics. *The New Republic*, pp. 139–141.

Dewey, J. (1997). *Experience and education*. New York, NY: The Free Press. (Original work published 1938)

Duplass, J. A., & Cruz, B. C. (2010). Professional dispositions: What's a social studies education professor to do? *The Social Studies, 101*, 140–151.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Fredricks, J., Blumenfeld, P., & Paris, A. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*, 54–109.

Gignac, G. (2008). Higher-order models versus direct hierarchical models: g as superordinate or breadth factor? *Psychology Science Quarterly, 50*, 21–43.

Graden, J., Thurlow, M. L., & Ysseldyke, J. E. (1982). *Academic engaged time and its relationship to learning: A review of the literature* (University of Minnesota Monograph No. 17). Washington, DC: Institute for Research on Learning Disabilities, Office of Special Education.

Jung, E., & Rhoades, D. M. (2008). Revisiting disposition assessment in teacher education: Broadening the focus. *Assessment and Evaluation in Higher Education, 33*, 647–660.

Lucke, J. F. (2005). The á and the ù of congeneric test theory: An extension of reliability and internal consistency to heterogeneous tests. *Applied Psychological Measurement, 29*, 65–81. doi:10.1177/0146621604270882

McKnight, D. (2004). An inquiry of NCATE's move into virtue ethics by way of dispositions (Is this what Aristotle meant?). *Educational Studies, 35*, 212–230.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén and Muthén.

National Council for Accreditation of Teacher Education. (2008). *Professional standards*

for the accreditation of teacher preparation institutions. Washington, DC: Author.
Plato. (1984). *The dialogues of Plato* (Vol. 1). New Haven, CT: Yale University Press.
Plato. (1991). *The republic* (Complete and Unabridged Jowett Translation). New York, NY: Vintage Books.
Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*, 667–696. doi:10.1080/00273171.2012.715555
Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika, 74*, 145–154. doi:10.1007/s11336-008-9102-z
Ruitenberg, C. W. (2011). The trouble with dispositions: A critical examination of personal beliefs, professional commitments and actual conduct in teacher education. *Ethics and Education, 6*, 41–51.
Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177.
Shapiro, E. S. (2010). *Academic skills problems workbook*. New York, NY: Guilford Press.
Shrout, P. E., & Fleiss, J. L. (1979). Interclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428.
Volpe, R. J., DiPerna, J. C., Hintze, J. M., & Shapiro, E. S. (2005). Observing students in classroom settings: A review of seven coding schemes. *School Psychology Review, 34*, 454–474.
Welch, F. C., Pitts, R. E., Tenini, K. J., Kuenlen, M. G., & Wood, S. G. (2010). Significant issues in defining and assessing teacher dispositions. *The Teacher Educator, 45*, 179–201.
Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ùh. *Applied Psychological Measurement, 30*, 121–144. doi:10.1177/0146621605278814

Appendix A

NCATE Definitions of Dispositions

2002 Edition Standards

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. Dispositions are guided by beliefs and attitudes related to values such as caring, fairness, honesty, responsibility, and social justice. For example, they might include a belief that all students can learn, a vision of high and challenging standards, or a commitment to a safe and supportive learning environment. (p. 53)

2008 Standards

Professional attitudes, values, and beliefs demonstrated through both verbal and nonverbal behaviors as educators interact with students, families, colleagues, and communities. These positive behaviors support student learning and development. NCATE expects institutions to assess professional dispositions based on observable behaviors in educational settings. The two professional dispositions that NCATE expects institutions to assess are fairness and the belief that all students can learn. Based on their mission and conceptual framework, professional education units can identify, define, and operationalize additional professional dispositions. (p. 89)

Appendix B

Dispositions Rubric

	Below expectations (1)	Meets expectations (2)	Exceeds expectations (3)
RESPONSIBILITY	__1. Is not alert and/or present at designated teacher hours	__1. Is alert and present at designated teacher hours	__1. Spends additional time at school planning, preparing, and evaluating
	__2. Does not maintain professional dress and appearance according to district/school standards & or wears jeans	__2. Typically professional, appropriate, dress and appearance (according to district/school standards for spirit wear)	__ 2. Always maintains professional, appropriate, dress and appearance
	__3. Does not identify or complete needed tasks without specific direction	__3. Identifies and completes needed tasks with little or no direction	__3. Demonstrates leadership and takes initiative in identifying and completing tasks
	__4. Does not complete tasks in a timely manner	__4. Completes assigned tasks in a timely manner	__4. Always completes tasks in a timely manner
RESPECT	__5. Does not maintain positive communication	__5. Speaks positively of self and others	__ 5. Speaks positively resulting in motivation of others
	__6. Does not follow school and/or university policies and procedures	__6. Follows school/ university policies and procedures	__ 6. Follows school/university policies and procedures, providing leadership in those areas
	__7. Does not respond appropriately to diverse student populations	__7. Uses curriculum available and provides instruction in a manner that does not hinder learning for diverse populations, responds appropriately to cultural norms	__ 7. Honors diversity by purposively seeking out resources to enhance the learning of diverse populations
	__8. Does not accept constructive feedback	__8. Accepts/integrates constructive feedback	__ 8. Seeks constructive feedback
INTEGRITY	__9. Cannot be trusted, is not truthful	__9. Is truthful	__ 9. Can be trusted to be honest
	__10. Does not maintain confidentiality	__10. Maintains confidentiality with minimal prompting	__10. Maintains confidentiality in a manner that is consistent with professional, ethical, and legal standards
	__11. Frequently takes credit for others' ideas	__11. Acknowledges sources of ideas both written and spoken	__11. Provides formal reference list to acknowledge sources of ideas
	__12. Rude, abrasive, or dismissive in professional interactions	__12. Often implements plans as intended when collaborating with other professionals	__12. Consistently implements plans as intended when collaborating with other professionals

	Below expectations (1)	Meets expectations (2)	Exceeds expectations (3)
CARING/HUMANITY	__13. Does not foster positive interactions in the classroom	__13. Fosters positive interactions in the classroom	__13. Contributes toward building a community of learners
	__14. Is not approachable; P–12 students often avoid interaction with teacher candidate	__14. Is approachable; P–12 students seek teacher for academic help with teacher candidate	__14. Is approachable; P–12 students approach teacher candidate for academic help and social guidance
	__15. Uses students' names	__ 15. Converses with students daily on personally meaningful topics	__15. Acknowledges and supports students as individuals through conversations
FAIRNESS	__ 16. Follows school rules for managing behaviors inconsistently	__16. Applies school rules consistently	__16. Applies school rules; student displays trust that teacher applies rules consistently
	__17. Does not provide differentiated instruction	__17. Provides differentiated instruction for most learners most of the time	__17. Consistently plans for and provides effective and varied differentiated instruction equitably
BELIEF ALL STUDENTS CAN LEARN	__18. Does not encourage and motivate all students	__18. Encourages and motivates all students	__18. Helps students build intrinsic motivation for learning (evidenced by increased student self-monitoring or self-motivation to learn)
	__19. Fails to maintain high expectations for all student	__19. Maintains high expectations for all students	__19. Communicates consistent high expectations for all students