Validity Issues in Assessing Dispositions: The Confirmatory Factor Analysis of A Teacher Dispositions Form

Chunling Niu   University of Kentucky
Kimberlee Everson Western Kentucky University
Sylvia Dietrich   Western Kentucky University
Cassie Zippay   Western Kentucky University

Critics against the inclusion of dispositions as part of the teacher education accreditation focus on the dearth of empirical literature on reliably and validly accessing dispositions (Borko, Liston, & Whitcomb, 2007). In this study, a confirmatory factor analysis (CFA) was performed to test the factorial validity of a teacher dispositions form (3 factors and 12 indicators) employed by a southern American university to assess teacher candidates from 2006 to 2016. The initial CFA results revealed highly unsatisfactory model fit statistics. Further model modifications were then implemented to remove two less relevant indicators (Diversity and Collaboration) which significantly improved the model fit.

Since its first introduction as one of the new standards by the National Council for Accreditation of Teacher Education (NCATE) in 2000, the role of dispositions in teacher education has been a topic of heated debates (NCATE, 2002). Controversies mainly center on two questions in the existing related literature in teacher education. First, exactly what are dispositions? Teacher educators and researchers have used similar but confusing terms regarding teachers' dispositions in the past decade; it has been difficult to generate a consensus within the teaching profession on an operational definition of dispositions as a construct that can be taught, assessed, and improved.

Second, how are dispositions measured? Partly due to the lack of clarity in defining dispositions, many critics complain that it is nearly impossible to create a reliable scale to conduct any meaningful empirical research related to dispositions, such as investigating the relationships between teachers' dispositions ratings and their teaching effectiveness (Lay, 2016; O'Neill, Hansen, & Lewis, 2014). Dispositions ratings attempt to identify the abilities of teacher candidates to interact confidently and respectfully with others who are either similar or dissimilar to their belief and value systems. However, the lack of psychometric validation of dispositions measures has made it challenging to integrate dispositions into teacher preparation constructively and convincingly.

Theoretical Framework

Historically, a qualified teacher must possess the "right kind" of "character, values, and beliefs," relevant knowledge, and pedagogical skills satisfactorily assessed by various "experts" such as members of the clergy and faculty in particular subjects (Murray, 2007, p. 381). The equivalent of teachers' "character, values, and beliefs" has been referred to as "attitudes" in the published

NCATE *Standards* until 2000 when the term "dispositions" was first used and broadly defined as follows (as cited in Borko, Liston, & Whitcomb, 2007, p. 360):

> 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. (NCATE, 2002, p. 53; 2006, p. 53)

The vagueness of such a definition and the lack of operational guidelines in assessing dispositions led to a variety of issues in the attempts of many teacher preparation programs to adopt dispositions as an effective program admission and exit standard along with teaching knowledge and skills. For instance, Murray (2007) argued that dispositions cited in the teacher education literature should be considered as no more than a "superfluous construct" (2007, p. 381). Likewise, Jung and Rhodes (2008, p. 647) also noted that "the current approach to dispositions assessment in the United States focuses on personal characteristics and character-related dispositions and is frequently used as a sorting device to identify those who appear to be inadequately disposed to teaching." In 2007, to address such methodological concerns about dispositions assessment, NCATE made several significant revisions to its operational definition as found in the Glossary of NCATE Terms (NCATE, 2016). The clarity of the definition of "professional dispositions" has been significantly improved in the following two ways: (1) "observable behaviors in educational settings" were clearly specified as the indicators to measure relevant professional dispositions; and (2) "*fairness* and the belief that all students can learn" were identified as the two main categories of the teacher dispositions to be assessed.

While the operational definition of dispositions has become less ambiguous, a number of "methodological obstacles" still remain in dispositions assessment, such as low inter-rater agreement, the tendency of the raters to rate either too high or too low, and psychometric properties of the dispositions scales falling below the acceptable level (Choi, Benson, & Shudak, 2016; Welch et al., 2010). Despite the lack of solid empirical evidence to support the validity and reliability of various self-developed dispositions scales, teacher preparation programs (e.g.,The Renaissance Group) are still using these instruments without rigorous validation processes. Many of the self-developed or adapted dispositions scales merely went through faculty approval and basic calculation of the internal consistency of the scales using Cronbach's Alpha without careful examination of validity (Almerico, Johnson, Henriott, & Shapiro, 2011). In this sense, the adaptation and utilization of the teacher dispositions scale in this study was no exception. Thus this study focused on conducting a confirmatory factor analysis (CFA) to consider one form of construct validity, factorial validity, for the scale adapted and used by a southern American university since 2006 to assess teacher candidates' professional dispositionss (See Appendix A).

The purposes of this study were (1) to add validity evidence for the use of an original 3-factor teacher candidate dispositions form using the CFA method, (2) to understand the extent to which the model fit with the longitudinal dispositions data, and (3) to improve the model fit of the current dispositions form by modifying the CFA model, as needed.

Methods

Sampling and Protocol. To examine the factorial validity of the 3-factor structure of the Teacher Candidate Dispositions Form, a confirmatory factor analysis (CFA) was conducted to examine whether the scale structure fit well with the 10-year longitudinal data; if the results yielded unsatisfactory model fit statistics, further investigation would be carried out using modification ot the model based on theoretical considerations and model modification indices to find out what might have caused the poor model fit.

According to Hoyle, confirmatory factor analysis (CFA), also known as restricted factor analysis or the measurement model, is often used "in a deductive mode to test hypotheses regarding unmeasured sources of variability responsible for the commonality among a set of scores" (2000, p. 465). To put it simply, CFA is a suitable statistical method to test if (or how well) a scale developed from an existing theory accurately reflects the intended factor structure.

As shown in Table 1, the general procedures in performing CFA to validate the Dispositions Form included the following five steps:

Table 1. *The General CFA Procedures to Validate the Dispositions Form*

Steps	Procedures	Purpose
Step 1	Data clean-up & random selection of 3 different samples (containing 152, 196, and 349 observations)	Use different samples to test modified models if needed
Step 2	An initial CFA was performed on Sample 1 ($N = 152$)	To produce model fit statistics for the original 3-factor instrument (Model 1)
Step 3	Model modifications made to the original Factor structure of the scale	Model modifications (Model 2)
Step 4	A CFA of the modified model (Model 2) was performed on Sample 2 ($N = 196$)	To discern any improvement occurred in the model fit (Model 2)
Step 5	Repeat Step 3 & 4 on Sample 3 ($N = 349$) if Model 2 model fit statistics still proved unsatisfactory	Further model modification and test of model fit if needed

Step 1. Data preparation and clean up of three samples containing 152, 196, and 349 observations that were randomly selected from the original dataset with 18,769 observed cases (since clean-up of the full data set was both time-consuming and unnecessary for the purpose of this study). The purpose of this procedure was to use different samples to test modified models if needed.

Step 2. An initial CFA was performed on Sample 1 to produce model fit statistics for the 3-factor instrument as it is.

Step 3. If the initial model fit statistics proved unacceptable, modifications would have been made to the original structure of the scale based on theoretical considerations and modification indices.

Step 4. A CFA of the modified model was performed on Sample 2 to see whether any improvement occurred in the model fit.

Step 5. If the model fit statistics from Step 4 still proved unsatisfactory, Step 3 and 4 would be repeated based on Sample 3.
All CFA analyses were performed using the Stata software (Version 14).

Measure. The Teacher Candidate Dispositions Form (hereafter referred to as the Form) was adapted from the dispositions rubric (including 5 values: Learning & Knowledge, Diversity, Collaboration, Professionalism, and Personal Integrity) used in Wayda & Lund (2005, p. 36) and has been used by the School of Teacher Education to assess teacher candidates' professional dispositions since 2006. The Form was initially approved by all program faculty and then adopted through the Professional Education Council (PEC) process; however, no formal validation research had been conducted to examine the validity and reliability of the scale until the current study.

The 5-point-Likert-Scale Form has 3 factors (i.e., Learning Attitudes, Professionalism, Personal Integrity, in addition to Diversity and Collaboration each measured by a single indicator) measured by twelve indicators (See Figure 1). The 10-year longitudinal dataset based on the Form yielded 18,769 observed cases for the teacher candidates (TCs). Each TC was rated multiple times by multiple raters (e.g., facilitators, faculty, P-12 practitioners, self, and peers) throughout their teacher preparation program. For convenience, this study only focused on faculty ratings since they accounted for the majority of the observations in the dataset.

Figure 1. The Original Factor Structure of the Dispositions Form.
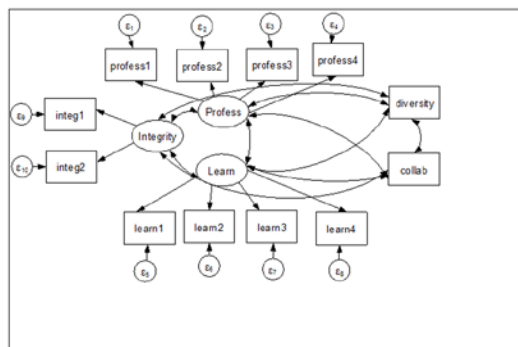
Results

The descriptive statistics for each of the three selected samples showed that the means of all three samples were around 4, suggesting an unusually large number of higher ratings on the Likert scale from 1 to 5.

The skewness and kurtosis of the data were also calculated, suggesting all three samples were notably skewed to the high end of the scale, and were not normally distributed.

An initial CFA was performed on Sample 1 using the robust maximum likelihood estimation method with Setorra-Bentler correction for data non-normality (See Figure 2), and yielded unsatisfactory model fit statistics of RMSEA=0.192, CFI=0.815, TLI=0.749, considering the conventional cutoff values are .06, .95, and .95 for RMSEA, CFI, and TLI respectively (Hu & Bentler, 1999). The overall model chi square value $X^2(48) = 312.10$, $p < 0.001$.

Figure 2. The Initial CFA on the Dispositions Form (Model 1).
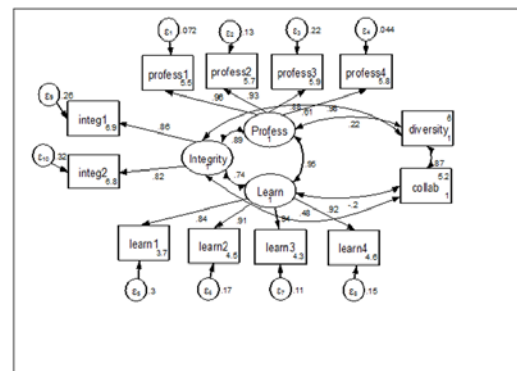
Based on the model modification indices combined with theoretical considerations, covariance between Collaboration and Professionalism and covariance between Learning Attitudes and Diversity, as well as correlated error terms of two indicators of Personal Integrity (Profess2 and Integ2) were added to CFA Model 2 (See Figure 3). Another CFA analysis of Model 2 was

performed on Sample 2. This modification resulted in substantially improved model fit statistics (RMSEA=0.065, CFI=0.966, TLI=0.951), although the RMSEA (0.065) value was still slightly above the cutoff value .06. The overall model chi square value for Model 2 was $X^2(45) = 80.57$, $p < 0.001$.

Figure 3. The Modified Model of the Dispositions Form (Model 2).
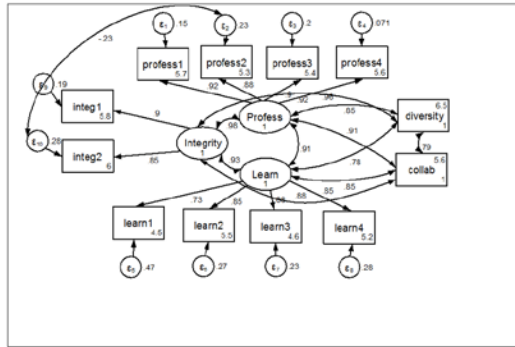
Comparison of factor loadings and covariances/correlations between Model 1 and 2 revealed that Diversity and Collaboration, each measured by a single indicator, had the weakest covariances/correlations with the 3 exisitng factors (Professionalism, Learning, and Personal Integrity) compared to the other indicators (See Table 2).

Table 2. The Confirmatory Factor Analysis Results for the Disposition Form Model 1 (Satorra-Bentler Estimation) (N = 152)

	Satorra-Bentler Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Measurement						
profess1 <-						
Profess	1	(constrained)				
_cons	4.456376	.0709368	62.82	0.000	4.317342	4.595409
profess2 <-						
Profess	.9241113	.0096243	96.02	0.000	.9052481	.9429746
_cons	4.442953	.0675976	65.73	0.000	4.310464	4.575442
profess3 <-						
Profess	.8506109	.018125	46.93	0.000	.8150867	.8861352
_cons	4.409396	.0652983	67.53	0.000	4.281414	4.537378
profess4 <-						
Profess	.9677601	.0062834	154.02	0.000	.9554448	.9800763
_cons	4.483221	.067746	66.18	0.000	4.350442	4.616001
learn1 <-						
Learn	1	(constrained)				
_cons	4.302013	.0860824	49.98	0.000	4.133295	4.470732
learn2 <-						
Learn	.9187252	.0146570	62.68	0.000	.8899965	.9474539
_cons	4.402685	.0712152	61.82	0.000	4.263105	4.542264
learn3 <-						
Learn	.9972424	.0139088	71.70	0.000	.9699817	1.024603
_cons	4.395973	.0742694	59.19	0.000	4.250408	4.541538
learn4 <-						
Learn	.908219	.0186247	48.76	0.000	.8717154	.9447227
_cons	4.362416	.0694716	62.79	0.000	4.226254	4.498578
integ1 <-						
Integrity	1	(constrained)				
_cons	4.469799	.0690409	64.74	0.000	4.334481	4.605116
integ2 <-						
Integrity	.9856912	.0704239	14.00	0.000	.8476629	1.12372
_cons	4.877181	.0700996	68.30	0.000	4.439789	4.714674
mean(divers-y)	4.456376	.068332	65.22	0.000	4.322448	4.590304
mean(collab)	4.463097	.0663307	67.28	0.000	4.333066	4.593109
var(e.profe-1)	.0474631	.0149041			.02664	.0978236
var(e.profe-2)	.0808045	.0157002			.0552139	.118266
var(e.profe-3)	.1265376	.0214245			.0908029	.1763354
var(e.profe-4)	.0261969	.0082868			.0140926	.0486976
var(e.learn1)	.3951932	.076548			.2703545	.5776772
var(e.learn2)	.1884796	.0266771			.1140844	.2201609
var(e.learn3)	.118709	.0245513			.0791479	.1780444
var(e.learn4)	.1366399	.0240203			.0988626	.1919223
var(e.integ1)	.109028	.0323671			.0609315	.1950896
var(e.integ2)	.1477762	.0192817			.1144292	.1908387
var(diversity)	.8470574	.0277203			.4953374	.6041776
var(collab)	.7386194	.0180378			.7040987	.774832
var(Profess)	.6092605	.0181766			.5746466	.6459382
var(Learn)	.9411904	.0239442			.8954115	.9893098
var(Integrity)	.3160425	.0287093			.2644979	.3776319
cov(divers-y, collab)	.5546222	.0101032	54.90	0.000	.5348203	.5744242
cov(divers-y, Profess)	.1285286	.0123489	10.41	0.000	.1043253	.152732
cov(divers-y, Integrity)	.2532182	.0176841	14.32	0.000	.218558	.2878764
cov(collab, Learn)	-.1651235	.0132989	-12.42	0.000	-.191189	-.1390581
cov(collab, Integrity)	.2320304	.0164024	18.06	0.000	.2018423	.2622186
cov(Profess, Learn)	.7185765	.0143415	50.10	0.000	.6904678	.7466853
cov(Profess, Integrity)	.389259	.0165696	23.49	0.000	.3567832	.4217348
cov(Learn, Integrity)	.4058042	.0202281	20.06	0.000	.3661578	.4454507

Thus we decided to conduct a CFA model that removed these two indicators from the original scale of the Form (See Figure 4). This final modification produced significantly improved model fit statistics (RMSEA=0.066, CFI=0.976, TLI=0.965, $X^2(31) = 75.81$, $p < 0.001$).

Figure 4. The Final Modified Model of the Dispositions Form (Model 3).
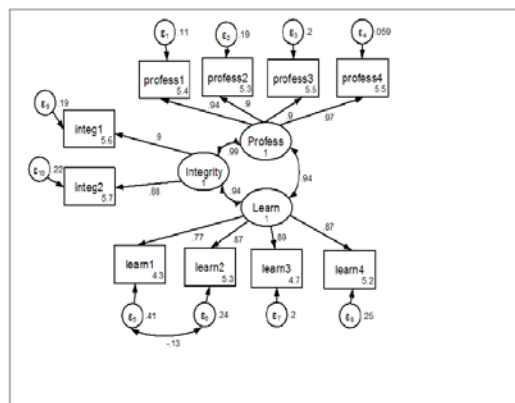
In conclusion, the results of this CFA validation study of the Form showed that it was very hard to measure

certain constructs such as Diversity and Collaboration accurately by a single survey item/indicator; and by so doing, the overall psychometric properties, especially the content and construct validity, of the dispositions scale are significantly weakened.

Discussion

Supported by previous dispositions studies (Brewer, Lindquist, & Altemueller, 2011; Ruitenberg, 2010), the current study contributes to the existing knowledge base by adding important empirical evidence from a local case study which found that dispositions instruments in teacher education need rigorous validation for effective dispositions assessment. Furthermore, this study reveals an existing tendency in the current approach adopted by many American teacher education institutions to oversimplify complex dispositions constructs in the development and use of their own dispositions scales. The results of this study point out the potential risks that threaten the validity of the overall dispositions scale. Theoretically, it seemed highly questionable that the Diversity and Collaboration dispositions could be validly observed and measured by a single indicator. Just like Professionalism, Learning, and Personal Integrity, both Diversity and Collaboration are complex concepts and embody a rich array of values, beliefs, and observable behaviors,which could hardly warrant acceptable face and content validity of such measurements (Castro, 2010; Reiter & Davis, 2011). Empirically, a single indicator could not even be considered a factor/latent variable in factor analysis, thus it was impossible to measure the statistical reliability of the data collected for the two dispositions (Diversity and Collaboration).

Several meaningful empirical attempts have recently tackled this issue. For instance, Kapner (2013) developed a new Valuing Diversity rubric for assessors to validly identify and give numeral scores on teacher candidates' performance related to valuing diversity. Street (2014) adopted the Diversity Disposition Index (DDI) survey (developed and validated by Schulte, Edwards, & Edick, 2008) to measure the construct of self-perceived teachers' dispositions towards diversity based on a 3-factor structure: *Educators' Skills in Helping Students Gain Knowledge* (18 items), *Educators' Beliefs and Attitudes about Students and Teaching/Learning* (16 items), and *Educators' Connections with the Community* (9 items). Future research directions include (1) modifying the Dispositions Form to include an adequate number of indicators based on theoretical considerations for the two dispositions, Diversity and Collaboration, to improve the content and construct validity of the measurement; (2) running the dyadic model analysis to test if there is a significant difference between the faculty raters and P-12 practitioner raters for the same teacher candidates; and (3) conducting a growth model analysis based on the longitudinal data.

References

Almerico, G., 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.

Borko, H., Liston, D., & Whitcomb, J. A. (2007). Apples and fishes. *Journal of Teacher Education*, *58*(5), 359-64.

Brewer, R. D., Lindquist, C., & Altemueller, L. (2011). The dispositions improvement process. *Online Submission*, *4*(2), 51-68.

Castro, A. J. (2010). Themes in the research on preservice teachers' views of cultural diversity implications for researching millennial preservice teachers. *Educational Researcher*, *39*(3), 198-210.

Choi, H. S., Benson, N. F., & Shudak, N. J. (2016). Assessment of teacher candidate dispositions: evidence of reliability and validity. *Teacher Education Quarterly*, *43*(3), 71.

Hoyle, R. H. (2000). Confirmatory factor analysis. *Handbook of applied multivariate statistics and mathematical modeling*, 465-497.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55.

Jung, E., & Rhodes, D. M. (2008). Revisiting dispositions assessment in teacher education: Broadening the focus. *Assessment & Evaluation in Higher Education*, *33*(6), 647-660.

Kapner, L. S. (2013). *Assessing Dispositions Towards Diversity in Math and Science Submissions of the Performance Assessment for California Teachers*(Doctoral dissertation, University of Southern California).

Lay, C. S. (2016). *The effectiveness of assessing teacher dispositions in the teacher selection processes of Taiwan's junior high schools* (Doctoral dissertation, Morehead State University).

Murray, F. B. (2007). Dispositions: A superfluous construct in teacher education. *Journal of teacher education*, *58*(5), 381-387.

National Council for Accreditation of Teacher Education (NCATE). (2002). *Professional standards for the accreditation of schools, colleges, and departments of education*. Washington, DC: NCATE.

National Council for Accreditation of Teacher Education (NCATE). (2006). *Professional standards for the accreditation of schools, colleges, and departments of education*. Washington, DC: NCATE.

National Council for Accreditation of Teacher Education (NCATE). (2016). *Professional standards for the accreditation of schools, colleges, and departments of education*. Washington, DC: NCATE.

O'Neill, J., Hansen, S., & Lewis, E. (2014). *Dispositions to teach: Review and synthesis of current components and applications, and evidence of impact*. Retrieved from https://education.govt.nz/assets/Uploads/DispositionsReportFinal100914.pdf

Reiter, A. B., & Davis, S. N. (2011). Factors influencing pre-service teachers' beliefs about student achievement: evaluation of a pre-

service teacher diversity awareness program. *Multicultural Education*, *18*(3), 41.

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*(1), 41-52.

Street, A. (2014). *A comparison of dispositions for educating diverse students*(Doctoral dissertation, Dallas Baptist University).

Wayda, V., & Lund, J. (2005). Assessing dispositionss: An unresolved challenge in teacher education. *Journal of Physical Education, Recreation & Dance*, *76*(1), 34-41.

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*(3), 179-201.

Dr. Chunling Niu is Research Administrative Principal Coordinator at Training Resource Center, College of Social Work, University of Kentucky. Her research interests include school performance and improvement, teacher preparation, interdisciplinary studies involving school intervention and child welfare.

Dr. Kimberlee Everson is an Assistant Professor of the Educational Administration, Leadership, and Research Department at Western Kentucky University. Dr. Everson is a quantitative methodologist whose research focuses on teacher evaluation.

Dr. Sylvia Dietrich serves as the Director of the School of Teacher Education at Western Kentucky University. Early childhood and personnel preparation, including clinical models comprise her research interests.

Dr. Cassie Zippay is the clinical model project director for Western Kentucky University's School of Teacher Education. Her research interests include culturally responsive instruction and clinical teacher preparation.

Appendix A. The Teacher Candidate Dispositions Form

Scoring Scale: (X=Not Observed; 1=Little/No Value for Dispositions; 2=Improvement Needed in Value for Dispositions; 3=Moderate Value for Dispositions; 4=Above Average Understanding Value for Dispositions; 5=Well-established and Consistent Behavior which Demonstrates Value for Dispositions)

Dispositionss	Descriptors	Holistic Scoring Scale
Values Learning: Attendance	♦ Contacts Instructor ♦ Attends Regularly ♦ On Time ♦ Stays Full Time	X 1 2 3 4 5 **Comments**
Values Learning: Class Participation	♦ Attentive in Class ♦ Engaged/Interested in Activities ♦ Responds Appropriately to Questions ♦ Participates in Discussions	X 1 2 3 4 5 **Comments**
Values Learning: Class Preparation	♦ Reliable ♦ Flexible ♦ Work Completed on Time ♦ Shows Diversity of Curriculum Design	X 1 2 3 4 5 **Comments**
Values Learning: Communication	♦ Work Shows Effort ♦ Uses Correct Grammar ♦ Expresses Ideas/Self Well ♦ Listens Thoughtfully and Responsibly	X 1 2 3 4 5 **Comments**
Values Personal Integrity: Emotional Control	♦ Controls Temper ♦ Shows Enthusiasm ♦ Takes Personal Responsibility for Own Behaviors ♦ Respects Others ♦ Takes Criticism Openly	X 1 2 3 4 5 **Comments**
Values Personal Integrity: Ethical Behavior	♦ Keeps Confidentiality ♦ Uses Truth and Honesty ♦ Trustworthy/Keeps Promises ♦ Shows Dignity and Integrity	X 1 2 3 4 5 **Comments**
Values Diversity	♦ Provides Equal Learning for All ♦ Treats Students in a Non-discriminatory Manner ♦ Endeavors to Understand Community and Home Cultures	X 1 2 3 4 5 **Comments**
Values Collaboration	♦ Is Collegial, Cooperates with Faculty/Peers ♦ Demonstrates Flexibility ♦ Exhibits Openness to Change ♦ Displays Willingness to Revise	X 1 2 3 4 5 **Comments**
Values Professionalism: Respects School Rules and Policies	♦ Dresses Appropriately ♦ Protects the Health, Safety, and Emotional Well-Being of Students	X 1 2 3 4 5 **Comments**
Values Professionalism: Commitment to Self-Reflection and Growth	♦ Shows Commitment to Reflection and On-Going Learning ♦ Exhibits Initiative, Self-direction	X 1 2 3 4 5 **Comments**
Values Professionalism: Shows Involvement and Professional Development	♦ Seeks to Improve Teaching Skills ♦ Assumes Responsibility for Classroom Climate ♦ Assumes Responsibility for Quality of Instruction	X 1 2 3 4 5 **Comments**
Values Professionalism: Professional Responsibility	♦ Demonstrates Belief that All Children Can Learn at High Levels ♦ Assumes Responsibility for Personal Actions ♦ Uses Educational Technology Appropriately ♦ Practices Ongoing Assessment	X 1 2 3 4 5 **Comments**