Abstract Title Page

Title: A study examining the dimensionality of Core Competencies measure in teacher preparation programs: Challenges and lessons

Authors and Affiliations:

Ruhan Circi Kizil, University of Colorado Boulder
Derek Briggs, University of Colorado Boulder
Kent Seidel, University of Denver
Kathy Green, University of Denver

Abstract

Background / Context

There is a natural expectation that teachers have an effect on the knowledge, skills, and behaviors of their students. Similarly, preparatory programs are expected to have an effect on the knowledge, skills, and behaviors of prospective teachers. The increasing attention on the quality of professional development is a consequence of the increasing emphasis on teacher effectiveness in systems of educational accountability. Unfortunately, the evidence that teacher preparation programs have an impact on teacher quality is often limited. Estimates of teacher effectiveness at increasing student achievement appear to differ very little between teachers coming from different preparatory programs (Koedel et al, 2012).

Progress in research on this topic will remain rather limited in its influence on practice until more proximal measures of teacher education outcomes can be established. The dearth of variables to measure the impact of teacher preparation program on teacher skills constitutes a measurement problem.

Purpose / Objective / Research Question / Focus of Study

We developed an instrument that attempts to measure the specific knowledge, skills, and behaviors that teachers need to help students learn. We refer to these knowledge, skills, and behaviors as "core competencies" (CCs). Our hypothesis is that in order for teacher candidates to achieve at least some minimal level of proficiency with the CCs, it should be the case that they have been taught explicitly and practiced as part of a program of systematic professional development.

As a part of a three year IES-funded project, the big picture motivating questions for the present study were as follows:

1. What is the best characterization of the dimensional structure of the CC survey?
2. How does the choice of dimensional structure change inferences about differences in quality among teacher preparation programs in Colorado?

In keeping the theme of the Spring SREE conference, a focus of our presentation will be examining whether a dimensional structure discovered with one sample of teachers can be replicated with a new sample of teachers.

Setting

This study utilized data collected from all teacher preparation programs in Colorado.

Population / Participants / Subjects

There were two groups of participants in this study. One group represents novice teachers who are in their first three years of teaching (graduates) and the other group represents respondents who were just completing their preparation programs at the time of survey administration (candidates). The graduate survey was administered to Colorado teachers who completed one of 21 teacher preparation programs. Both traditional and alternative programs are represented in the study; 17 can be classified as "traditional" routes to certification, and

remaining 4 as the alternative routes. A total of 648 graduates from 18 programs responded to at least some portion of the graduate survey. A total of 355 candidates from 13 programs responded to at least some portion of the candidate survey.

Intervention / Program / Practice

The Survey of Enacted Curriculum developed by researchers at the University of Wisconsin-Madison (Blank et al., 2000; Porter, 2002; WCER, 2003) and existing teacher observation protocols such as the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2007; Pianta et al., 2007) constituted an initial basis for the development of the CC constructs. After survey design meetings and 10 cognitive interviews during the pilot study of initial survey, a set of items were associated with the following collection of 8 CCs

1. Demonstrating mastery of and pedagogical expertise in content taught (CC1).
2. Managing the classroom environment to facilitate learning for students(CC2).
3. Developing a safe, respectful environment for a diverse population of students. (CC3)
4. Planning and providing effective instruction (CC4).
5. Designing and adapting assessments, curriculum & instruction (CC5).
6. Engaging students in higher order thinking and expectations (CC6).
7. Supporting academic language development and English Language Acquisition(CC7).
8. Reflection and professional growth(CC8)

Each CC had anywhere from 4-8 statements (items) associated with it. Different questions were posed to respondents for each statement: "How important do you find this to be in your current teaching?" (response scale 0-4) and "OVERALL, how well did your program prepare you to do this in your teaching?" (response scale 1-4). Scales based on the latter item responses for each CC were of principal interest in the analyses described below.

Research Design

Although the treatment or intervention of interest can be defined in terms of the teacher preparation program that are at the heart of this study, our focus is on the instrumentation being used to measures outcomes. The respondents to our survey are self-selected, so all comparisons are based on a convenience sample and observational data.

Data Collection and Analysis

Both surveys were administered using the survey software Qualtrics via the internet. Those respondents with more than 80% of item responses missing were eliminated from the analysis. This reduced the sample size from 648 to 479 cases from the graduate survey, and from 355 to 227 cases for the candidate survey.

Three approaches are used for exploring dimensional structure. Exploratory factor analysis (EFA) served as a starting point for examining the factor structure of the instrument, and then confirmatory factor analysis (CFA) and bi-factor analysis was used to test the hypothesized factor structure and to explore alternatives (Bollen, 1989; Reise et al, 2007).

First, a series of exploratory factor analyses were conducted for establishing a coherent subset of latent variables underlying the survey responses. In successive EFAs, as we increased the number of factors, we checked the individual item loadings to look for items that seemed to load similarly even as new factors were added.

A confirmatory analysis is conducted next. Four models were compared based on considering a variety of fit measures, and model comparisons are based on incremental differences in fit. Lastly, we specified a bi-factor model showing some appeal because it may serve to remove the influence of a general attitude that candidates and graduates have toward the programs where they received their preparation.

The comparisons between programs are made by using ANOVA and pairwise analysis by using overall composite scores, factor scores and CC-specific factor scores as the outcome variables of ANOVA.

Findings / Results

In EFA analysis, the chi-square test of model fit (H_0: the model fit the data) was consistently rejected ($p< 0.001$) for factor structures changing from 1 to 8. In other words, none of these factor structures fit the data well in a statistical sense. The successive examination of the factor loadings helped us to flag items with potential problems. This led us to revisit the wording of the items and the rationale for each item's inclusion within a hypothesized CC. This resulted in the decision to exclude 14 items from the graduate survey and 11 items from the candidate survey.

For CFA, the examination of four models

- Model 1= 8 hypothesized factors based on the 8 CCs (45 items for the graduate survey and 37 for the candidate survey).
- Model 2= 1 hypothesized factor, which probably represents some overall perception the respondents have toward their preparatory programs.
- Model 3 = 8 hypothesized CC factors, but items flagged as problematic after our EFA analyses were removed (31 items for the graduate survey and 26 for the candidate survey).
- Model 4 = 1 hypothesized factor (31 items for the graduate and 26 for the candidate survey)

was lead to Model 3 to be favored in both surveys.

--Insert Table 1--

In graduate survey Model 3, the covariance matrix predicted by the model explained about 88.4% of the total variability (GFI=0.884). For the candidate survey, although the criterion for the exact-fit hypotheses was not satisfied, again Model 3 (χ^2_{CM3} (271)=410.6, p=0.001) showed an improvement relative to other models. In a relative sense, the CFA analyses suggest that an 8 factor solution is preferable to a 1 factor solution.

Lastly, we experimented with a bi-factor analysis with just the restricted 31 items from the graduate survey responses. For the bi-factor analysis, not surprisingly, all survey items have higher values for the general higher order factor than on the CC-specific factors. Of greater interest are the CC-specific factor loadings after the influence of the general factor has been removed. In particular, the items associated with CC7 (supporting academic language development and English language acquisition) have had the highest partial factor loadings.

To further explore the robustness of an 8 factor solution with a restricted subset of items, we examine the solution shows population invariance. It is important to appreciate that both candidates and graduates could be conceptualized as coming from the same larger population of teachers with different levels of experience. As such one might expect to see the item to factor loadings for each survey sample to be strongly associated. Establishing factorial invariance is a necessary condition in order to accurately investigate group differences in mean scores and patterns of association with other variables. If the scales are not equivalent, findings about group differences or correlations from one survey to the next become difficult to interpret, because items from one sample to the next do not have the same relationships to the hypothesized CCs. Because our examination of population invariance resulted in moderate correlation ($r = 0.43$), questions about the invariance of the factor structure by teacher sample are appeared and we used the graduate survey results for comparison of teacher programs.

We compared the programs in three cases with ANOVA after Bonferroni and Benjamini and Hochberg corrections. As expected the latter lead to more significant difference among programs because of its less conservative nature We begin by focusing on the use of an overall CC composite (computed by taking the average across 31 items) as the outcome measure of interest. ANOVA result indicated that somewhere among the entire set of means for 18 programs there is at least one difference that is unlikely to be explained by chance ($p = .005$). There is found significant differences on CC1 (Demonstrating mastery of and pedagogical expertise in content taught), on CC5 (Designing and adapting assessments, curriculum & instruction), and on CC2 (Managing the classroom environment).

In connection to CFA, factor scores are computed and used in subsequent analyses for program comparisons. All CCs showed significant differences for at least one of the programs. As Table 2 indicates, there were significant differences on CC5 (Designing and adapting assessments, curriculum & instruction), on CC2 (Managing the classroom environment), and on CC3 and CC4.

--Insert Table 2--

Finally, after conducting a bi-factor analysis, CC-specific secondary factor scores were generated for each program to be used as ANOVA measures. The results indicate significant differences in two individual CCs: CC1 (Demonstrating mastery of and pedagogical expertise in content taught), and CC7 (Supporting academic language development and English Language Acquisition).

Conclusions:

A variety of methods and approaches were used in this research assessing the dimensionality of CC instrument which feature potential multidimensionality of the competencies hypothesized to be necessary to be practiced in a teacher preparation program. The purpose of the study was to have a better understanding to the degree that instrument present multidimensionality as intended and examining the effect of different ways of measuring the CCs for comparison of teacher preparation programs by using the data from two samples which are hypothesized to come from same population.

On the basis of a purely exploratory approach an argument can be advanced for collapsing CCs into an overall composite; on the basis of a confirmatory approach an argument can be advanced for reporting 8 dimensions; and our bi-factor approach can be seen as a compromise between these first two approaches. In answering second research question we see the consequence of decisions made about how to represent the dimensional structure of the instrument such that each different outcome measure lead different results. The overall examination showed the need for new insights useful when considering future use and development of the CCs instrument as well as revision of theory that underlies the instrument.

Appendices

Appendix A. References

Blank, R., Porter, A., & Smithson, J. (2001). New Tools for Analyzing Teaching, Curriculum and Standards in Mathematics & Science. Washington, DC: Council of Chief State School Officers.

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons: New York.

Koedel, C., Parsons, E., Podgursky, M., & Ehlert, M. (2012*). Teacher preparation programs and teacher quality: are there real differences across programs*? National Center for Analysis for Longitudinal Data in Education Research, Working Paper 79.

Pianta, R.C, La Paro, K.M., & Hamre, B. (2007). *Classroom Assessment Scoring System (CLASS) Manual, K-3*. Baltimore, MD: Paul H. Brookes Publishing.

Pianta, R.C., Hamre, B.K., Haynes, N.J., Mintz, S.L., and La Paro, K.M. (2007). *CLASS Classroom Assessment Scoring System Manual: Middle/Secondary Version (Pilot).* University of Virginia: CASTL.

Reise, S.P., Morizot, J., & Hays, R.E (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16 (1), 19-31.

Porter, A. (2002). Measuring the Content of Instruction: Uses in Research and Practice, *Educational Researcher* October 2002 31: 3-14.

Wisconsin Center for Education Research. (2004). *Analyzing Instructional Content.* WCER Research Highlights, vol 16, no 1, pp. 4-5. University of Wisconsin-Madison. Online at: http://www.wcer.wisc.edu/publications/highlights/

Appendix B. Tables and Figures

Table 1. *Fit statistics for Graduate and Candidate Survey Responses*

	Model	$\chi 2$	Df	p-value	GFI	RMSEA	CFI	AIC
Graduate Survey	Model 1	1115.1	918	<.001	0.796	0.033	0.965	1349.1
	Model 2	1888.1	946	<.001	0.660	0.071	0.833	2066.1
	Model 3	406.45	406	0.48	0.884	0.002	1.000	586.5
	Model 4	968.45	434	<.001	0.698	0.079	0.859	1092.4
Candidate Survey	Model 1	971.9	601	<.001	0.794	0.048	0.930	1175.9
	Model 2	1743.5	629	<.001	0.598	0.094	0.790	1891.5
	Model 3	410.6	271	.001	0.862	0.051	0.960	570.6
	Model 4	881.04	299	<.001	0.722	0.099	0.834	985.0

Note: Model 1 (8 factors, all items); Model 2 (1 factor, all items); Model 3 (8 factor, subset of items); Model 4 (1 factor, subset of items)

Table 2. *Comparison of Programs with Respect to different outcomes of CCs*

Composite CCs			**Confirmatory CCs**			**Bifactor CCs**		
CCs	(Bonferroni)	(B& H correction)	CCs	(Bonferroni)	(B&H correction)	CCs	(Bonferroni)	(B & H correction)
CC1	A vs. B (d=0.14)	B vs. O(d=-0.7)	CC2	C vs. D alt(d=-1.03)	H vs. D alt(d=-0.81)	CC1	N vs. H (d=0.71)	B vs. H (d=0.62)
	C vs. B (d=0.10)	C vs. D Alt (d=-0.6)		P vs. A (d= -0.31)	E vs. D alt (d=-0.77)		A vs. N (d=-0.58)	E alt vs. H (d=0.71)
	A vs. C (d=0.24)	G vs.B(d=0.94)		E alt vs. C (d=0.77)	C vs. D alt (d=-1.03)			E Alt vs. C (d=0.64)
		A vs. B (d=0.65)			M vs. D alt(d=-1.21)			C vs. H (d=-0.55)
		C vs. B(d=0.81)			J vs. D alt (d=-1.11)			A vs. E Alt (d=-0.58)
		C vs. G(d=-0.81)			E alt vs. C (d=0.77)			D vs. H (d=0.79)
		A vs. C(d=0.53)			G vs. P (d=1.10)			
		G vs. C(d=0.71)			G vs. R (d=0.37)			
CC2	C vs. D Alt (d=0.32)	H-D Alt (d=-0.65)	CC5	C vs. D alt (d=-0.98)	B vs. D alt (d= -0.86)	CC7	E alt vs. N (d=-0.07)	E alt vs. H (d=0.06)
	C vs. E Alt (d=0.21)	C vs. D (d=-0.15)			L vs. D alt (d=-0.744)		A vs. N (d-0.07)	E alt vs. B(d=0.07)
		J vs. D alt(d=-0.8)			M vs. D alt (d=-1.1)			L vs. E Alt(d=-0.07)
		E alt vs. H (d=0.43)			G vs. M (d=1.06)			A vs. E Alt (d=-0.08)
		K vs. H (d=0.52)						
		G vs. H (d=0.63)						
CC3	NO	P vs. H (d=-0.14)						
		G vs. M (d=0.69)						
CC5	D Alt vs. C(d=0.16)	B vs. O (d=-0.73)						
	D Alt vs. L(d=0.21)	G vs. D alt (d=-0.53)						
		C vs. D Alt (d=-0.81)						
		P vs. O (d=-0.70)						
		J vs. D alt (d=-0.74)						
		C vs. H (d=-0.49)						
		D vs. J (d=0.82)						
		L vs. D Alt (d=-0.72)						