TESE
Wednesday Sep 22 2004 10:08 AM
Allen Press • DTPro System
GALLEY 92
tese 27_408 Mp_92
File # 08em

Teacher Education and Special Education
2004, Volume 27, No. 4, 000–000

Identifying Attributes of High Quality Special Education Teachers

Elaine Carlson, Hyunshik Lee, & Karen Schroll Westat

Abstract: *Using a nationally representative sample of 1,475 special education teachers, researchers used factor analysis to test five teacher quality factors with empirical and theoretical grounding in general education research: experience, credentials, self-efficacy, professional activities, and selected classroom practices. These factors were combined to create an aggregate teacher quality measure. All five factors emerged as viable components of an aggregate teacher quality measure, although some modest factor loadings suggested a need for further research into the precise nature of these teacher quality dimensions and the best ways to measure them.*

In this age of educational accountability, it is critical that we understand the results of our efforts to educate teachers, both in their initial preparation and in on-going professional development. Evaluating teacher preparation programs requires an objective, comprehensive measure of teacher quality, something currently missing in both special and general education.

The proximal outcomes of preservice programs or on-going professional development may include number of graduates, job placement results, or mastery of specific knowledge and skills. However, the distal outcomes of our efforts are much more ambitious. We aim to prepare high quality teachers. If we cannot define a high quality teacher or measure teacher quality, we cannot adequately evaluate the true effectiveness of our teacher preparation programs. Until we have well-developed outcome measures and carefully designed evaluations, preservice and inservice preparation programs will be subject to unsubstantiated criticism from audiences both within and outside the education community.

Previous research shows that the quality of children's teachers is important in influencing academic achievement. Sanders and Rivers (1996) found that, on average, the least effective teachers in one district produced annual gains of roughly 14 percentile points among low-achieving students, while the most effective teachers produced gains of 53 percentile points. They concluded that students with similar initial achievement levels have "vastly different academic outcomes as a result of the sequence of teachers to which they are assigned" (p. 6). Similar results have been documented in Dallas and Boston (Bain et al., as cited in Haycock, 1998; Jordan, Mendro, & Weerasinghe, 1997).

While these studies of student achievement support the hypothesis that classroom teachers are critically important to children's

This project is supported by the U.S. Department of Education, Office of Special Education Programs (OSEP). Opinions expressed herein are those of the authors and do not necessarily represent the position of the U.S. Department of Education. The authors would like to acknowledge the contributions of Dr. Patricia Gonzalez. Her insight and guidance throughout the study were invaluable.

Correspondence concerning this article should be address to Dr. Elaine Carlson, 1650 Research Boulevard, Rockville, MD 20850.

educational attainment, they leave many questions unanswered. For example, they do not indicate what teaching practices, attitudes, or attributes account for differences in student outcomes. Certainly, teacher quality is multifaceted, and researchers need frameworks for capturing this complexity that they can draw on when designing their research.

In general education, teachers who demonstrate the greatest gains in student achievement seem to have certain attributes. Specifically, they seem to be more academically capable than other teachers, and this capability is demonstrated through two different types of studies. They score higher on a variety of tests, such as teacher licensure exams and tests of verbal ability (Rice, 2003; Wayne & Youngs, 2003). They attend more selective colleges, although it is unclear whether this reflects the tested ability of the teachers (e.g., SAT and ACT scores used in college admissions) or the quality of the undergraduate education they receive (Rice, 2003; Wayne & Youngs, 2003).

General education teachers who secure the greatest gains in student achievement also exhibit certain beliefs about students and the learning process. They report higher levels of self-efficacy, meaning they have faith in their capability to succeed in specific instructional endeavors (Ashton & Webb, 1986; Brownell & Pajares, 1999; Moore & Esselman, 1992; Ross, 1992; Midgley, Feldlaufer, & Eccles, 1989). Additionally, these teachers exhibit beliefs that are specific to the content they teach. For instance, teachers who demonstrate the best student gains on tests of mathematics achievement tend to view making connections between concepts as an important underlying principle of effective mathematics instruction (Muijs & Reynolds, 2002).

What general education teachers do also seems to play a strong role in the amount students learn. For example, the use of specific classroom practices accounts for differences in student achievement. Results from process-product research asserted that the more time students spent actively engaged in tasks they completed with high rates of success, the more they learned (Fisher et al. in Sindelar, Smith, Harriman, Hale, & Wilson, 1986). Sindelar, et al. (1986) found that time spent in teacher-directed reading instruction was the single best predictor of gains in reaching achievement for students with disabilities. More recently, classroom observation research has demonstrated linkages between student achievement and generic instructional strategies, such as those gleaned from the process-product research, as well as more specific instructional strategies in reading, such as teaching phonemic awareness and decoding, developing vocabulary, and engaging students in meaningful interactions about text (Haager, Gersten, Baker, & Graves, 2003).

Many studies document a positive relationship between teachers' years of experience and student achievement (Biniaminov & Glasman, 1983; Ferguson, 1991; Greenwald, Hedges, & Laine, 1996; Lopez, 1995; Murnane, 1981). However, some researchers hesitate to draw conclusions about the association between teachers' experience and student achievement because the data on years of experience can be difficult to interpret. For example, if teachers who leave the profession are less skilled than those who stay, measures of experience would reflect those differences as well as the knowledge and skills accumulated through on-the-job training (Wayne & Youngs, 2003).

Research on the importance of teacher certification is ambiguous. Several studies support a relationship between secondary mathematics achievement and teacher certification in math (Goldhaber & Brewer, 2000; Hawk, Coble, & Swanson, 1985). However, research is less clear in documenting an association between student achievement and elementary certification or secondary certification in subjects other than math (Rice, 2003).

In her important theoretical work, Kennedy (1992) identified five dimensions of teacher quality: credentials; tested ability; demographic representation, which refers to the mix of educators working in individual schools; professionalism, meaning the extent to which teachers are given real responsibility for their work and are able to make sound professional decisions; and classroom teaching practices. Many of Kennedy's teacher quality dimensions are supported by empirical research on student achievement (such as

TESE, **Volume 27, No. 4**
Fall 2004

credentials, tested ability, and classroom teaching practices, as noted above). Kennedy asserts that we cannot assume teaching practices will improve simply by ensuring high quality in credentials, tested ability, and/or professionalism. Instead, she insists that policymakers and researchers must separately address each dimension of quality.

Purpose

The purpose of this study was to develop a model of teacher quality in special education using data from a national data set and identify components of that model that were consistent with research on teacher quality in general education. There were several reasons to believe that teacher quality in special education might differ from teacher quality in general education. First, special education teachers often work collaboratively with general education teachers, so their influence on student achievement may be intermingled with the influence of their general education colleagues. Second, special education teachers may require skills and knowledge that differ from general educators'. The clearest examples pertain to discrete skills, like Braille or sign language, but may certainly encompass broader skill sets such as behavior management.

Methods

In this study, the authors used confirmatory factor analysis to test whether the data reported by special education teachers supported previous theoretical and empirical work on teacher quality in general education and to derive an aggregate teacher quality measure. Factor analysis is a method for defining or testing broad constructs, such as intelligence or achievement, and identifying the important dimensions within those constructs. As such, factor analysis was used in this study to determine the important facets of teacher quality that can be assessed through large data sets.

Large-scale data sets, like the one used here, play an important role in social science research. They can identify relationships among variables and explore constructs, like teacher quality, that are not well defined. They can also test the validity of emerging theories by controlling for a wide range of confounding variables.

The Study of Personnel Needs in Special Education (SPeNSE), funded by the U.S. Department of Education's Office of Special Education Programs (OSEP), was designed to describe personnel who serve students with disabilities and factors associated with workforce quality. It included computer-assisted telephone interviews with a nationally representative sample of local administrators (n = 358) and service providers (n = 8,061), including elementary and secondary special and general education teachers, preschool special education teachers, speech-language pathologists, and special education paraprofessionals. This article includes results only for 1,475 special education teachers (preschool through secondary school), who had full responses for the variables used in the analysis.

Sample Design and Weighting

SPeNSE used a two-phase sample design because no national sampling frame was available with suitable lists of special education teachers. In the first phase of the sample, that frame was created by contacting selected local education agencies (LEAs), state schools for students with visual or hearing impairments, and intermediate education units (IEUs) and asking them to submit lists of all their special education teachers.[1]

Samples of LEAs and IEUs were selected from the November 5, 1998 version of Quality Education Data's (QED) National Education Database.[2] The sample of LEAs was stratified by geographic region and district size (i.e., total student enrollment). IEUs that did not employ staff who provide direct services to students with disabilities were deleted from the frame, and the IEU sample was stratified by geographic region only. All state schools (76) were included in the first-phase sample because there were so few of them.

The second-phase was a stratified simple

[1] In some states, intermediate education units are responsible for providing services to member districts, such as special education services, vocational education services, and professional development.

[2] QED's National Education Database is a commercially available data set with basic demographic information on schools and districts nationwide.

random sample of service providers from rosters of personnel that were obtained from 370 participating LEAs, IEUs, and state schools. The stratification was done by personnel type to facilitate separate analysis by the type of service provider. Table 1 shows the total number of eligible teachers sampled and the number and percentage of respondents, by type of teacher in special education.

The special education teachers were selected with different probabilities through the two-phase sample design, and such differences were reflected in weighting so that national estimates could be generated from SPeNSE data. Survey weighting produces a weight for each respondent. It reflects the overall sampling process encompassing both sample selection and the nonrespondent determination process. The final weight, therefore, consists of two components: the sampling weight, which is determined by the sample design, and the nonresponse adjustment weight, which is based on the researcher's choice of the adjustment procedure.

Our weighting was done in phases reflecting the two-phase sample design. In the first phase, the weighting for the sampling units was based on the recruitment results. The sampling weights were first calculated for the recruited respondents for sample selection. In the second phase, their sampling weights were adjusted for recruitment nonrespondents.[3]

The service provider base weight, which was the multiple of its first-phase weight and the second-phase sampling weight, was adjusted to compensate for nonrespondent teachers within each job assignment.[4] This elaborate weighting was necessary not only to account for different sampling probabilities but also to counter the effects of the low rate of district participation, which could introduce nonresponse bias. The LEA participation rate at the recruitment stage was quite low; the overall rate was 42%. The IEU participation rate of 44% was also low. Only the state school sample met the expected participation rate of 70%. To ensure that the weighting was effective in eliminating nonresponse bias, Westat conducted an extensive nonresponse study, which did not suggest any systematic nonresponse bias. The nonresponse study was based on an independent sample of 23 previously nonparticipating districts and 202 special education teachers (Carlson & Lee, 2003).

Data Collection Procedures

Sampled teachers received letters explaining the study and requesting their participation. The letters also indicated that all teachers sampled for the study would be entered in a drawing for one $2,000 gift certification to Circuit City, ten $250 gift certificates to Amazon.com, and fifty $10 gift certificates to Starbucks. If individuals did not have access to those retailers, alternative gift certificates were provided.

Data collection was done by computer-assisted telephone interview (CATI) from May 2000 through November 2000. CATI is a highly-structured form of telephone interview that typically involves complex skip patterns; in other words, CATI allows interviewers to skip questions that do not apply to the respondent. An interviewer would have difficulty reliably following the skip patterns without a computer that automatically displays the appropriate questions based on previous responses.

The SPeNSE instruments were developed specifically for this study (available at www.spense.org). However, those instruments used many items from the Schools and Staffing Survey (U.S. Department of Education, 2000) and other previous or ongoing studies of school personnel.

In data collection, SPeNSE devoted particular attention to five instructional areas: teaching reading, managing behavior, facilitating secondary transition, teaching English language learners (ELLs), and promoting in-

[3] After analyzing the response pattern using Chi-squared Automatic Interaction Detection (CHAID), first-phase nonresponse adjustment weighting was conducted within each of the 36 design strata (24 district size-region cells in the LEA sample, 6 regions in each of the IEU and state school samples). The CHAID analysis enabled us to account for the differential nonresponse tendency of the nonrespondents, thus reducing nonresponse bias. After careful nonresponse weight adjustment, the weights for the LEA and IEU samples were further adjusted by poststratification weighting using teacher population data for the poststrata from QED. This was designed to reduce the nonresponse bias as well as to enhance the efficiency of estimation of population parameters.

[4] CHAID analysis was performed to construct weighting cells for the weighting adjustment using design variables (region and district size) and some auxiliary variables from the QED.

TESE, **Volume 27, No. 4**
Fall 2004

Table 1. Number of Eligible Teachers Sampled and Response Rates, by Type of Teacher

Type of teacher	Sample size[a]	Response rate	
		Number	Percent
Special education teachers who serve primarily children with disabilities ages 3-5	1,171	881	75.2
Special education teachers who serve primarily students with visual or hearing impairments	1,679	1,054	62.8
Special education teachers who serve primarily students with emotional disturbance	1,190	859	72.2
Special education teachers who are not included in the previous three categories	3,688	2,633	71.4
Total	7,668	5,427	70.8

[a] Excludes those sampled individuals found ineligible in the verification process or the screening portion of the interview and individuals who died or became incapacitated between the time when the sampling frame was developed and the data collection period.

clusion.[5] Two of the instructional areas, teaching ELLs and facilitating secondary transition, were excluded from this analysis because the items were inappropriate for many of the respondents owing to the types of students they taught. Teachers' responses on the frequency with which they used various classroom practices were combined into scales for teaching reading (alpha = .87 for children ages 5–11 and .83 for 6–21), managing behavior (alpha = .86), and promoting inclusion (alpha = .85), where alpha denotes the Cronbach's alpha.

Item nonresponse on the service provider questionnaire was very low. Consequently, researchers did not impute any item responses on this instrument.

Data Analysis Methods

In this analysis, a two-level confirmatory factor analysis was performed using LISREL (LInear Structural RELations) software (Jöreskog and Sörbom, 1996). This kind of tool is useful for finding the linearly-related structure of correlated variables by which a coherent summary measure with a meaningful interpretation can be defined. In our case, we used the tool to define a teacher quality measure from various teacher characteristics. Using results from empirical studies and Kennedy's theoretically framework, a large set of variables, believed to be related to teacher quality, were grouped into several first-order latent teacher quality factors, such as experience and credentials. These first-order factors were further summarized into a single second-order factor to create aggregate teacher quality scores. Since the analysis was confirmatory, the factor model was specified by the analysts.

For factor loadings that range from −1 to 1, the size indicates the relative importance of each variable among those variables that define the factor. The factor loadings are the correlations between the variables and the factor. Their squares tell how much variance is explained by the factor. For example, if a factor loading of a variable is 0.5, then 25% of its variance is explained by the factor. Factor loadings of 0.0–0.3 were considered low; 0.3–0.5, medium; 0.5–0.7, high; and 0.7–1.0, very high.

In the second-order factor analysis, the first-order factors (experience, credentials, self-efficacy, professional activities, and selected classroom practices) were combined to generate a broad teacher-quality factor. This allowed the authors to explore the relative importance of the five factors.

Results

Researchers tested five factors using the SPeNSE data on special education teachers:

[5] In each of these professional areas, service providers were asked the extent to which they used various best practices identified by experts in the field. For example, 12 instructional practices were listed for teaching reading, and respondents were asked, for each of the 12, whether they use that approach *not at all, to a small extent, to a moderate extent,* or *to a great extent.* Scale scores were created by combining responses to those items that were highly correlated.

Table 2. Factor Loadings of the First-Order Factors[a]

Variable description	Factor loadings					
	Exper	Cred	Self-Ef	Prof	Class prac	Error variance
Years teaching	0.986					0.03
Years teaching special education	0.908					0.18
Level of certification		0.560				0.69
Highest degree earned		0.367				0.88
Number of fields in which certified		0.181				0.97
Self-efficacy score[b]			0.510			0.73
General self-assessment of performance as a teacher			0.682			0.53
CEC skills score[c]			0.768			0.40
Number of journals read regularly				0.301		0.90
Number of association memberships				0.333		0.89
Times per month asked advice from colleagues				0.331		0.89
Frequency with which teachers reported using identified best practices to teach Reading					0.504	0.75
Extent to which teachers individualized reading instruction					0.163	0.90
Frequency with which teachers reported using identified best practices to manage behavior					0.295	0.92
Frequency with which teachers reported using identified best practices to promote inclusion					0.523	0.72
Plans to remain teaching special education[d]						
Distance teacher relocated to accept job[d]						
Extent to which teachers know the cultures of the students in their school[d]						

[a] All variables with significant factor loadings in LISREL were retained.
[b] Derived from Gibson & Dembo (1984) scale.
[c] Score derived from teachers' self-assessment on a subset of skills in the Council for Exceptional Children's Standards for Entry into Practice
[d] This indicates variables with insignificant factor loadings in the specified model.

experience, credentials, self-efficacy, professional activities, and selected classroom practices. Tables 2 and 3 show the factor loadings for the model.

Factor 1: Experience

This factor included two variables—years teaching and years teaching special education. The factor loadings for the two experience variables are close to 1, which is very high. This means that the factor explains most of the variance in the two component variables.

Factor 2: Credentials

This factor included three variables: level of certification (none, emergency, certified out of field, fully certified for position); number of fields in which teachers were certified; and highest degree earned. In defining the credential factor, level of certification was most important. The variable that measured the number of fields in which teachers were certified was least important, with its variance largely unexplained. This provides evidence for the hypothesis that the match between area of certification and job assignment is an important one. That is, a teacher's credential is most valuable if it matches the field in which the teacher works. Additional areas of certification add very little to this dimension of teacher quality.

Factor 3: Self-Efficacy

This factor included three variables. The first was a scale on special education teachers'

TESE, **Volume 27, No. 4**
Fall 2004

Table 3. Factor Loadings of the Second-Order (Teacher Quality) Factor

Variable	Factor loading	Error variance
Experience	0.400	0.84
Credentials	0.414	0.83
Self-efficacy	0.874	0.24
Professional Activities	0.924	0.15
Selected Classroom Practices	0.441	0.81

Note. Model fit statistics for the two-level factor analysis: Root mean square of approximation = .041; Comparative Fit Index = .965; Goodness of Fit Index = .967.

perceptions of their skill in completing a variety of tasks related to their work, such as using appropriate instructional techniques, managing behavior, monitoring student progress and adjusting instruction accordingly, and working with parents. These were a subset of the CEC Standards for Entry into Practice. The second was teachers' assessment of their own job performance. Respondents were asked "*How would you characterize your overall performance as a teacher?*" The third summarized several items designed to measure teacher beliefs (e.g., "*If you try hard you can get through to even the most difficult student*"). In Table 2, it is labeled *self-efficacy score.* The factor loadings for all three self-efficacy variables were reasonably high. This suggests that a single latent construct is at work in shaping teachers' responses to all three component variables.

Factor 4: Professional Activities

This factor included three variables: the number of professional journals teachers read regularly, the number of professional associations to which they belonged, and the number of times per month that colleagues asked them for professional advice. The three variables have moderate and more or less equal factor loadings; their variances are largely unexplained by the professional activities factor. This means that the variables that comprise the professional activities factor are not very coherent. Nevertheless, the factor emerged as a strong predictor of teacher quality, which indicates that the variables used in defining this factor are complimentary.

Factor 5: Selected Classroom Practices

This factor included four variables. Three of them were scale scores for the frequency with which special education teachers reported using specified best practices in teaching reading, managing behavior, and promoting inclusion. The fourth was a variable on the extent to which teachers individualized reading instruction. The reading scale and the inclusion scale have reasonable factor loadings. The other variables, although significant, have moderate or small factor loadings. This suggests that the factor is incoherently defined, and some important variables needed to define a stronger factor are missing.

Second Order Factor Analysis: An Aggregate Teacher-Quality Measure

In the aggregate teacher-quality measure, the professional activities factor was the most important factor, followed by the self-efficacy factor. The other three were almost equal, with moderate factor loadings. The results suggest that each of the five teacher-quality factors is an important component of an aggregate teacher-quality measure and should be considered in future research on teacher quality in special education.

It is important to note that the first order factor loadings for the professional activities factor were quite low, and the variance for its component variables (the number of professional journals teacher read; the number of professional associations to which they belonged, and the number of times per month that colleagues asked them for professional advice) was largely unexplained by the factor. So while it appears that teachers with higher quality scores seek to improve their knowledge and skills through journals and association memberships, the individual variables that define the professional activities factor are not very coherent. It is possible

that there are two or more similar but distinct constructs at work or that the one construct (professional activities) is poorly measured in the study. Nonetheless, the component variables complement each other, so the factor emerged as a strong contributor as a whole to the teacher quality measure.

The SPeNSE teacher quality model was later validated using data from the Special Education Elementary Longitudinal Study (SEELS) (Blackorby, Lee, & Carlson, 2004). SEELS is a study of elementary and middle school students with disabilities, which includes both teacher data and achievement data for a nationally representative sample of students with disabilities. Researchers used a variety of outcome measures to model academic performance as a function of various inputs, including disability, demographic characteristics, functional skills, and teacher quality. While SEELS did not include all of the variables in the SPeNSE teacher quality model, the SPeNSE model was replicated as closely as possible, with a correlation of .97 between the SPeNSE and SEELS teacher quality scores. Outcome measures included standard scores on four Woodcock Johnson scales: letter-word identification, passage comprehension, calculation, and mathematical problem solving, as well as a mean of these four achievement scores. Two models were specified for each outcome variable, one that included the teacher quality score and one that did not. In each case, the amount of variation explained by the model improved when the teacher quality score was included, albeit modestly. Blackorby and colleagues (2004) characterized the changes as educationally meaningful but still smaller than average effect sizes in intervention studies, with effect sizes of .3 to .5, depending on the outcome measure. As a comparison, in meta-analyses of special education interventions, Kavale and Forness (2000) found effect sizes of .52 for computer-assisted instruction, .84 for direct instruction, and 1.62 for mnemonic strategies.

Limitations

There were several limitations to the study. First, all of the items in the teacher quality model were based on self-report. Therefore, caution should be used in interpreting results, particularly with regard to use of specific classroom practices, where self-report may be particularly suspect.

Although SPeNSE interviews included a few items on teachers' test participation and performance, specifically tests for certification or licensure, an insufficient number of special education teachers took those tests to include the items in the factor analysis. Furthermore, because tests for certification have become more prevalent in recent years, those who took them had significantly fewer years of teaching experience than those who did not. This precluded entering teachers' years of experience and test performance (as defined) in the same model, and we decided to drop the latter. Consequently, we cannot speak to verbal ability, specifically, or tested ability, more generally, as a component of special education teacher quality.

Discussion

This analysis found that high quality special education teachers share many attributes with their general education colleagues. Many general education studies document a positive relationship between experience and student achievement. In this study of special education teachers, experience again emerged as an important teacher-quality factor, although no achievement measure was used. As in the general education research, teacher attitudes and beliefs, such as self-efficacy, proved important for special education teachers. Specific classroom practices play a role in explaining how much general education students learn and, likewise, classroom practices appear to be an important part of teacher quality in special education.

Previous research suggests that general education teachers with the greatest gains in student achievement are more academically capable than other teachers. As noted earlier, we were unable to address the role of academic competence or tested ability in this analysis of special education teachers.

Kennedy (1992) made the case for five separate dimensions of teacher quality. While there is substantial overlap in the dimensions she described and those tested here, this analysis provided evidence for a single teacher

TESE, **Volume 27, No. 4**
Fall 2004

quality construct with multiple component parts. There is likely a role for both a broad, aggregate measure of teacher quality, like the one described here, and finer measures of teacher attributes associated with quality.

A broad teacher quality construct is important for a number of reasons. First, it is possible that even when the separate dimensions of teacher quality (either Kennedy's or others) may be insignificant or weak predictors of student achievement, an aggregate measure of teacher quality for the same teachers may account for substantially more variance in student achievement. Therefore, using such an aggregate measure would allow researchers to document the effects of teacher quality on student achievement. Second, an aggregate measure of teacher quality may be preferable for evaluating the long-term outcomes of personnel policies, teacher preparation programs, and professional development programs because of its breadth, assuming the ultimate goal of such policies and programs is to prepare high quality teachers, not simply to prepare teachers with specific credentials, attitudes, or skills. The aggregate measure would also allow researchers to examine the role of various inputs and processes in explaining differences in teacher quality, such as personnel policies or working conditions.

In contrast, specific dimensions of teacher quality, such as classroom practices or self-efficacy, may be better than a broad measure of teacher quality as an independent outcome in evaluating specific interventions because they may have a more direct relationship to the program's purpose and may be more sensitive to change. For example, a district-wide professional development program on effective instructional strategies in reading may improve classroom teaching practices sooner than it improves the broader teacher-quality measure, since some aspects of the aggregate teacher quality measure (like experience or credentials) do not change quickly. Used in combination, the aggregate teacher quality measure and its components may help enlarge our knowledge base on how to hire, prepare, and retain high quality teachers through improved policy and program evaluations.

Clearly, more research is needed to better define and measure the different facets of teacher quality and to determine which measures have the most potential for achieving different goals. This study provides a first step in developing and testing a theory of teacher quality in special education. To move this work ahead, it may be important to learn whether teacher attitudes linked to student achievement are relatively stable aspects of a teacher's personality or whether they can be taught during preservice preparation and, if so, how that is best accomplished. An orientation toward life-long learning and professional identity may be another important area of research relevant to teacher quality. In particular, the dimension of professional activities requires further exploration to determine its composition and its relationship to teacher quality and student achievement. Additional research on the validity of self-reported classroom practices and the role of academic competence would also contribute considerably to the measurement of special education teacher quality.

References

Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement.* New York: Longman.

Biniaminov, I., & Glasman, N. S. (1983, Summer). School determinants of student achievement in secondary education. *American Educational Research Journal, 20,* 251–268.

Blackorby, J., Lee, H., & Carlson, E. (2004). Teacher quality as a predictor of achievement for students with disabilities. Manuscript in preparation.

Brownell, M. T., & Pajares, F. (1999, Summer). Teacher efficacy and perceived success in mainstreaming students with learning and behavior problems. *Teacher Education and Special Education, 22,* 154–164.

Carlson, E., & Lee, H. (2003). *Study of Personnel Needs in Special Education, Paperwork Substudy: Nonresponse Component.* Rockville, MD: Westat.

Ferguson, R. F. (1991, Summer). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation, 28,* 464–498.

Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76,* 569–582.

Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis, 23,* 79–86.

Greenwald, R. L., Hedges, V., & Lane, H. (1996).

The effect of school resources on student achievement. *Review of Educational Research, 36,* 42–48.

Haager, D., Gersten, R., Baker, S., & Graves, A. (2003). The English-Language Learner Classroom Observation instrument for beginning readers. In S. Vaughn and K. Briggs (Eds.), *Reading in the classroom: Systems for the observation of teaching and learning* (p. 111–144). Baltimore: Paul H. Brooks Publishing.

Hawk, P. P., Coble, C. R., Swanson, M. (1985). Certification: Does it matter. *Journal of Teacher Education, 36,* 13–15.

Haycock, K. (1998). Good teaching matters a lot. *Thinking K-16, 2.* Washington, DC: The Education Trust. Retrieved June 25, 2001. Available: http://www.cftl.org/ documents/k16.pdf

Jordan, H., Mendro, R. L., & Weerasinghe, D. (1997, July). *Teacher effects on longitudinal student achievement: A report on research in progress.* Paper presented at the CREATE Annual Meeting, Indianapolis, IN.

Jöreskog, K. G., and Sörbom, D. (1996). *LISREL 8: User' reference guide.* Chicago: Scientific Software International.

Kavale, K. A., & Forness, S. R. (2000). Policy decisions in special education: The role of meta-analysis. In R. Gersten, E. P. Schiller, & S. Vaughn (Eds.), *Contemporary Special Education Research: Synthesis of the Knowledge Base on Critical Instructional Issues.* Mahwah, NJ: Lawrence Erlbaum Associates.

Kennedy, M. M. (1992). The problems of improving teacher quality while balancing supply and demand. In E. E. Boe & D. M. Gilford (Eds.), *Teacher Supply, Demand, and Quality: Policy Issues, Models, and Data bases: Proceedings of a Conference* (pp 65–108). Washington, DC: National Academy Press.

Lopez, O. (1995). *The effect of the relationship between classroom student diversity and teacher capacity on student performance.* (ERIC Document Reproduction Service No. ED386423)

Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Change in teacher efficacy and student self- and task-related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology, 81,* 247–258.

Moore, W. P., & Esselman, M. E. (1992, April). *Teacher efficacy, empowerment, and a focused instructional climate: Does student achievement benefit?* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Muijs, D., & Reynolds, D. (2002). Teachers' beliefs and behaviors: What really matters? *Journal of Classroom Interaction, 37*(2), 3–15.

Murnane, R. J. (1981). Inner-city schools can make a difference. In Educational Testing Service (Ed.), *The minority student in public schools: Fostering academic excellence* (pp. 31–41). Princeton, NJ: Educational Testing Service, Office of Minority Education. (ERIC Document Reproduction Service No. ED243981)

Rice, J. K. (2003). Teacher *Quality: Understanding the Effects of Teacher Attributes.* Washington, DC: Economic Policy Institute.

Ross, J. A. (1992, Winter). Teacher efficacy and the effect of coaching on student achievement. *Canadian Journal of Education, 17,* 51–65.

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement.* Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Sindelar, P. T., Smith, M. A., Harriman, N. E., Hale, R. L., & Wilson, R. J. (1986). Teacher effectiveness in special education. *Journal of Special Education, 20*(2), 195–207.

U.S. Department of Education, National Center for Education Statistics (2000). The Schools and Staffing Survey. Washinton, DC.: Retrieved May 2000. Available at http://www.nces.ed.gov/surveys/SASS.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73* (1), 89–122.

Elaine Carlson, Hyunshik Lee, and Karen Schroll Westat, Rockville, Maryland.