Students' Perceptions of Characteristics of Effective College Teachers: A Validity Study of a Teaching Evaluation Form Using a Mixed-Methods Analysis

Anthony J. Onwuegbuzie, Ann E. Witcher, Kathleen M. T. Collins, Janet D. Filer, Cheryl D. Wiedmaier and Chris W. Moore

Additional services and information for *American Educational Research Journal* can be found at:

Email Alerts: http://aerj.aera.net/cgi/alerts

Subscriptions: http://aerj.aera.net/subscriptions

Reprints: http://www.aera.net/reprints

Permissions: http://www.aera.net/permissions

Students' Perceptions of Characteristics of Effective College Teachers: A Validity Study of a Teaching Evaluation Form Using a Mixed-Methods Analysis

Anthony J. Onwuegbuzie
University of South Florida
Ann E. Witcher
University of Central Arkansas
Kathleen M. T. Collins
University of Arkansas, Fayetteville
Janet D. Filer
Cheryl D. Wiedmaier
Chris W. Moore
University of Central Arkansas

This study used a multistage mixed-methods analysis to assess the content-related validity (i.e., item validity, sampling validity) and construct-related validity (i.e., substantive validity, structural validity, outcome validity, generalizability) of a teaching evaluation form (TEF) by examining students' perceptions of characteristics of effective college teachers. Participants were 912 undergraduate and graduate students (10.7% of student body) from various academic majors enrolled at a public university. A sequential mixed-methods analysis led to the development of the CARE-RESPECTED Model of Teaching Evaluation, which represented characteristics that students considered to reflect effective college teaching—comprising four meta-themes (communicator, advocate, responsible, empowering) and nine themes (responsive, enthusiast, student centered, professional, expert, connector, transmitter, ethical, and director). Three of the most prevalent themes were not represented by any of the TEF items; also, endorsement of most themes varied by student attribute (e.g., gender, age), calling into question the content- and construct-related validity of the TEF scores.

KEYWORDS: college teaching, mixed methods, teaching evaluation form, validity

In this era of standards and accountability, institutions of higher learning have increased their use of student rating scales as an evaluative component of the teaching system (Seldin, 1993). Virtually all teachers at most universities and colleges are either required or expected to administer to their students some type of teaching evaluation form (TEF) at one or more points during each course offering (Dommeyer, Baum, Chapman, & Hanna, 2002; Onwuegbuzie, Daniel, & Collins, 2006, in press). Typically, TEFs serve as formative and summative evaluations that are used in an official capacity by administrators and faculty for one or more of the following purposes: (a) to facilitate curricular decisions (i.e., improve teaching effectiveness); (b) to formulate personnel decisions related to tenure, promotion, merit pay, and the like; and (c) as an information source to be used by students as they select future courses and instructors (Gray & Bergmann, 2003; Marsh & Roche, 1993; Seldin, 1993).

TEFs were first administered formally in the 1920s, with students at the University of Washington responding to what is credited as being the first

ANTHONY J. ONWUEGBUZIE is a professor of educational measurement and research in the Department of Educational Measurement and Research, College of Education, University of South Florida, 4202 East Fowler Avenue, EDU 162, Tampa, FL 33620-7750; e-mail: *tonyonwuegbuzie@aol.com*. He specializes in mixed methods, qualitative research, statistics, measurement, educational psychology, and teacher education.

ANN E. WITCHER is a professor in the Department of Middle/Secondary Education and Instructional Technologies, University of Central Arkansas, 104D Mashburn Hall, Conway, AR 72035; e-mail: *annw@uca.edu*. Her specialization area is educational foundations, especially philosophy of education.

KATHLEEN M. T. COLLINS is an associate professor in the Department of Curriculum & Instruction, University of Arkansas, 310 Peabody Hall, Fayetteville, AR 72701; e-mail: *kcollinsknob@cs.com*. Her specializations are special populations, mixed-methods research, and education of postsecondary students.

JANET D. FILER is an assistant professor in the Department of Early Childhood and Special Education, University of Central Arkansas, 136 Mashburn Hall, Conway, AR 72035; e-mail: *janetf@uca.edu*. Her specializations are families, technology, personnel preparation, educational assessment, educational programming, and young children with disabilities and their families.

CHERYL D. WIEDMAIER is an assistant professor in the Department of Middle/Secondary Education and Instructional Technologies, University of Central Arkansas, 104B Mashburn Hall, Conway, AR 72035; e-mail: *cherylw@uca.edu*. Her specializations are distance teaching/learning, instructional technologies, and training/adult education.

CHRIS W. MOORE is pursing a master of arts in teaching degree at the Department of Middle/Secondary Education and Instructional Technologies, University of Central Arkansas, Conway, AR 72035; e-mail: *chmoor@tcworks.net*. Special interests focus on integrating 20 years of information technology experience into the K-12 learning environment and sharing with others the benefits of midcareer conversion to the education profession.

114

TEF (Guthrie, 1954; Kulik, 2001). Ory (2000) described the progression of TEFs as encompassing several distinct periods that marked the perceived need for information by a specific audience (i.e., stakeholder). Specifically, in the 1960s, student campus organizations collected TEF data in an attempt to meet students' demands for accountability and informed course selections. In the 1970s, TEF ratings were used to enhance faculty development. In the 1980s to 1990s, TEFs were used mainly for administrative purposes rather than for student or faculty improvement. In recent years, as a response to the increased focus on improving higher education and requiring institutional accountability, the public, the legal community, and faculty are demanding TEFs with greater trustworthiness and utility (Ory, 2000).

Since its inception, the major objective of the TEF has been to evaluate the quality of faculty teaching by providing information useful to both administrators and faculty (Marsh, 1987; Seldin, 1993). As observed by Seldin (1993), TEFs receive more scrutiny from administrators and faculty than do other measures of teaching effectiveness (e.g., student performance, classroom observations, faculty self-reports).

Used as a summative evaluation measure, TEFs serve as an indicator of accountability by playing a central role in administrative decisions about faculty tenure, promotion, merit pay raises, teaching awards, and selection of full-time and adjunct faculty members to teach specific courses (Kulik, 2001). As a formative evaluation instrument, faculty may use data from TEFs to improve their own levels of instruction and those of their graduate teaching assistants. In turn, TEF data may be used by faculty and graduate teaching assistants to document their teaching when applying for jobs. Furthermore, students can use information from TEFs as one criterion for making decisions about course selection or deciding between multiple sections of the same course taught by different teachers. Also, TEF data regularly are used to facilitate research on teaching and learning (Babad, 2001; Gray & Bergmann, 2003; Kulik, 2001; Marsh, 1987; Marsh & Roche, 1993; Seldin, 1993; Spencer & Schmelkin, 2002).

Although TEF forms might contain one or more open-ended items that allow students to disclose their attitudes toward their instructors' teaching style and efficacy, these instruments typically contain either exclusively or predominantly one or more rating scales containing Likert-type items (Onwuegbuzie et al., 2006, in press). It is responses to these scales that are given the most weight by administrators and other decision makers. In fact, TEFs often are used as the sole measure of teacher effectiveness (Washburn & Thornton, 1996).

Conceptual Framework for Study

Several researchers have investigated the score reliability of TEFs. However, these findings have been mixed (Haskell, 1997), with the majority of studies yielding TEF scores with large reliability coefficients (e.g., Marsh &

Bailey, 1993; Peterson & Kauchak, 1982; Seldin, 1984) and with only a few studies (e.g., Simmons, 1996) reporting inadequate score reliability coefficients. Even if it can be demonstrated that a TEF consistently yields scores with adequate reliability coefficients, it does not imply that these scores will yield valid scores because evidence of score reliability, although essential, is not sufficient for establishing evidence of score validity (Crocker & Algina, 1986; Onwuegbuzie & Daniel, 2002, 2004).

Validity is the extent to which scores generated by an instrument measure the characteristic or variable they are intended to measure for a specific population, whereas *validation* refers to the process of systematically collecting evidence to provide justification for the set of inferences that are intended to be drawn from scores yielded by an instrument (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 1999). In validation studies, traditionally, researchers seek to provide one or more of three types of evidences: *content-related validity* (i.e., the extent to which the items on an instrument represent the content being measured), *criterion-related validity* (i.e., the extent to which scores on an instrument are related to an independent external/criterion variable believed to measure directly the underlying attribute or behavior), and *construct-related validity* (i.e., the extent to which an instrument can be interpreted as a meaningful measure of some characteristic or quality). However, it should be noted that these three elements do not represent three distinct types of validity but rather a unitary concept (AERA, APA, & NCME, 1999).

Onwuegbuzie et al. (in press) have provided a conceptual framework that builds on Messick's (1989, 1995) theory of validity. Specifically, these authors have combined the traditional notion of validity with Messick's conceptualization of validity to yield a reconceptualization of validity that Onwuegbuzie et al. called a meta-validation model, as presented in Figure 1. Although treated as a unitary concept, it can be seen in Figure 1 that content-, criterion-, and construct-related validity can be subdivided into areas of evidence. *All* of these areas of evidence are needed when assessing the score validity of TEFs. Thus, the conceptual framework presented in Figure 1 serves as a schema for the score validation of TEFs.

Criterion-Related Validity

Criterion-related validity comprises concurrent validity (i.e., the extent to which scores on an instrument are related to scores on another, already-established instrument administered approximately simultaneously or to a measurement of some other criterion that is available at the same point in time as the scores on the instrument of interest) and predictive validity (i.e., the extent to which scores on an instrument are related to scores on another, already-established instrument administered in the future or to a measurement of some other criterion that is available at a future point in time as the scores on the instrument of interest). Of the three evidences of validity, criterion-related
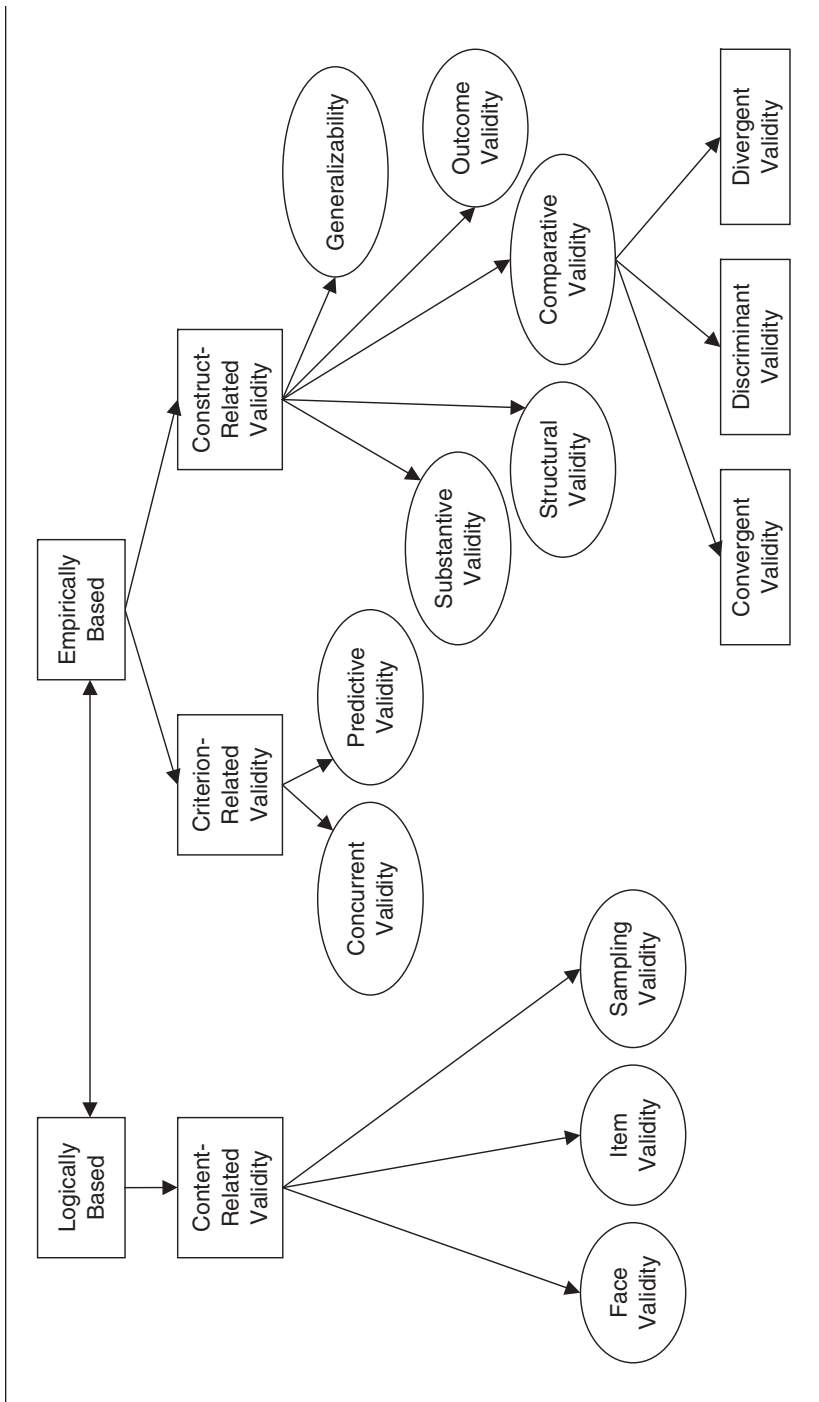
116

Figure 1. **Conceptual framework for score validation of teacher evaluation forms: A metavalidation model.**

validity evidence has been the strongest. In particular, using meta-analysis techniques, P. A. Cohen (1981) reported an average correlation of .43 between student achievement and ratings of the instructor and an average correlation of .47 between student performance and ratings of the course. However, as noted by Onwuegbuzie et al. (in press), it is possible or even likely that the positive relationship between student rating and achievement found in the bulk of the literature represents a "positive manifold" effect, wherein individuals who attain the highest levels of course performance tend to give their instructors credit for their success, whether or not this credit is justified. As such, evidence of criterion-related validity is difficult to establish for TEFs using solely quantitative techniques.

Content-Related Validity

Even if we can accept that sufficient evidence of criterion-related validity has been provided for TEF scores, adequate evidence for content- and construct-related validity has not been presented. With respect to content-related validity, although it can be assumed that TEFs have adequate face validity (i.e., the extent to which the items appear relevant, important, and interesting to the respondent), the same assumption cannot be made for item validity (i.e., the extent to which the specific items represent measurement in the intended content area) or sampling validity (i.e., the extent to which the full set of items sample the total content area). Unfortunately, many institutions do not have a clearly defined target domain of effective instructional characteristics or behaviors (Ory & Ryan, 2001); therefore, the item content selected for the TEFs likely is flawed, thereby threatening both item validity and sampling validity.

Construct-Related Validity

Construct-related validity evidence comprises substantive validity, structural validity, comparative validity, outcome validity, and generalizability (Figure 1). As conceptualized by Messick (1989, 1995), substantive validity assesses evidence regarding the theoretical and empirical analysis of the knowledge, skills, and processes hypothesized to underlie respondents' scores. In the context of student ratings, substantive validity evaluates whether the nature of the student rating process is consistent with the construct being measured (Ory & Ryan, 2001). As described by Ory and Ryan (2001), lack of knowledge of the actual process that students use when responding to TEFs makes it difficult to claim that studies have provided sufficient evidence of substantive validity regarding TEF ratings. Thus, evidence of substantive validity regarding TEF ratings is very much lacking.

Structural validity involves evaluating how well the scoring structure of the instrument corresponds to the construct domain. Evidence of structural validity typically is obtained via exploratory factor analyses, whereby the dimensions of the measure are determined. However, sole use of exploratory

factor analyses culminates in items being included on TEFs, not because they represent characteristics of effective instruction as identified in the literature but because they represent dimensions underlying the instrument, which likely was developed atheoretically. As concluded by Ory and Ryan (2001), this is "somewhat like analyzing student responses to hundreds of math items, grouping the items into response-based clusters, and then identifying the clusters as essential skills necessary to solve math problems" (p. 35). As such, structural validity evidence primarily should involve comparison of items on TEFs to effective attributes identified in the existing literature.

Comparative validity involves *convergent validity* (i.e., scores yielded from the instrument of interest being highly correlated with scores from other instruments that measure the same construct), *discriminant validity* (i.e., scores generated from the instrument of interest being slightly but not significantly related to scores from instruments that measure concepts theoretically and empirically related to but not the same as the construct of interest), and *divergent validity* (i.e., scores yielded from the instrument of interest not being correlated with measures of constructs antithetical to the construct of interest). Several studies have yielded evidence of convergent validity. In particular, TEF scores have been found to be related positively to self-ratings (Blackburn & Clark, 1975; Marsh, Overall, & Kessler, 1979), observer ratings (Feldman, 1989; Murray, 1983), peer ratings (Doyle & Crichton, 1978; Feldman, 1989; Ory, Braskamp, & Pieper, 1980), and alumni ratings (Centra, 1974; Overall & Marsh, 1980). However, scant evidence of discriminant and divergent validity has been provided. For instance, TEF scores have been found to be related to attributes that do not necessarily reflect effective instruction, such as showmanship (Naftulin, Ware, & Donnelly, 1973), body language (Ambady & Rosenthal, 1992), grading leniency (Greenwald & Gillmore, 1997), and vocal pitch and gestures (Williams & Ceci, 1997).

Outcome validity refers to the meaning of scores and the intended and unintended consequences of using the instrument (Messick, 1989, 1995). Outcome validity data appear to provide the weakest evidence of validity because it requires "an appraisal of the value implications of the theory underlying student ratings" (Ory & Ryan, 2001, p. 38). That is, administrators respond to questions such as Does the content of the TEF reflect characteristics of effective instruction that are valued by students?

Finally, generalizability pertains to the extent that meaning and use associated with a set of scores can be generalized to other populations. Unfortunately, researchers have found differences in TEF ratings as a function of several factors, such as academic discipline (Centra & Creech, 1976; Feldman, 1978) and course level (Aleamoni, 1981; Braskamp, Brandenberg, & Ory, 1984). Therefore, it is not clear whether the association documented between TEF ratings and student achievement is invariant across all contexts, thereby making it difficult to make any generalizations about this relationship. Thus, more evidence is needed.

119

Need for Data-Driven TEFs

As can be seen, much more validity evidence is needed regarding TEFs. Unless it is demonstrated that TEFs yield scores that are valid, as contended by Gray and Bergmann (2003), these instruments may be subject to misuse and abuse by administrators, representing "an instrument of unwarranted and unjust termination for large numbers of junior faculty and a source of humiliation for many of their senior colleagues" (p. 44). Theall and Franklin (2001) provided several recommendations for TEFs. In particular, they stated the following: "Include all stakeholders in decisions about the evaluation process by establishing policy process" (p. 52). This recommendation has intuitive appeal. Yet the most important stakeholders—namely, the students themselves—typically are omitted from the process of developing TEFs. Although research has documented an array of variables that are considered characteristics of effective teaching, the bulk of this research base has used measures that were developed from the perspectives of faculty and administrators—not from students' perspectives (Ory & Ryan, 2001). Indeed, as noted by Ory and Ryan (2001), "It is fair to say that many of the forms used today have been developed from other existing forms without much thought to theory or construct domains" (p. 32).

A few researchers have examined students' perceptions of effective college instructors. Specifically, using students' perspectives as their data source, Crumbley, Henry, and Kratchman (2001) reported that undergraduate and graduate students ($n = 530$) identified the following instructor traits that were likely to affect positively students' evaluations of their college instructor: teaching style (88.8%), presentation skills (89.4%), enthusiasm (82.2%), preparation and organization (87.3%), and fairness related to grading (89.8%). Results also indicated that graduate students, in contrast to undergraduate students, placed stronger emphasis on a structured classroom environment. Factors likely to lower students' evaluations were associated with students' perceptions that the content taught was insufficient to achieve the expected grade (46.5%), being asked embarrassing questions by the instructor (41.9%), and if the instructor appeared inexperienced (41%). In addition, factors associated with testing (i.e., administering pop quizzes) and grading (i.e., harsh grading, notable amount of homework) were likely to lower students' evaluations of their instructors. Sheehan (1999) asked undergraduate and graduate psychology students attending a public university in the United States to identify characteristics of effective teaching by responding to a survey instrument. Results of regression analyses indicated that the following variables predicted 69% of the variance in the criterion variable of teacher effectiveness: informative lectures, tests, papers evaluating course content, instructor preparation, interesting lectures, and degree that the course was perceived as challenging.

More recently, Spencer and Schmelkin (2002) found that students representing sophomores, juniors, and seniors attending a private U.S. university perceived effective teaching as characterized by college instructors'

120

personal characteristics: demonstrating concern for students, valuing student opinions, clarity in communication, and openness toward varied opinions. Greimel-Fuhrmann and Geyer's (2003) evaluation of interview data indicated that undergraduate students' perceptions of their instructors and the overall instructional quality of the courses were influenced positively by teachers who provided clear explanations of subject content, who were responsive to students' questions and viewpoints, and who used a creative approach toward instruction beyond the scope of the course textbook. Other factors influencing students' perceptions included teachers demonstrating a sense of humor and maintaining a balanced or fair approach toward classroom discipline. Results of an exploratory factor analysis identified subject-oriented teacher, student-oriented teacher, and classroom management as factors accounting for 69% of the variance in students' global ratings of their instructors (i.e., ". . . is a good teacher" and "I am satisfied with my teacher") and global ratings concerning student acquisition of domain-specific knowledge. Adjectives describing a subject-oriented teacher were (a) provides clear explanations, (b) repeats information, and (c) presents concrete examples. A student-oriented teacher was defined as student friendly, patient, and fair. Classroom management was defined as maintaining consistent discipline and effective time management.

In their study, Okpala and Ellis (2005) examined data obtained from 218 U.S. college students regarding their perceptions of teacher quality components. The following five qualities emerged as key components: caring for students and their learning (89.6%), teaching skills (83.2%), content knowledge (76.8%), dedication to teaching (75.3%), and verbal skills (73.9%).

Several researchers who have attempted to identify characteristics of effective college teachers have addressed college faculty. In particular, in their analysis of the perspectives of faculty ($n = 99$) and students ($n = 231$) regarding characteristics of effective teaching, Schaeffer, Epting, Zinn, and Buskit (2003) found strong similarities between the two groups when participants identified and ranked what they believed to be the most important 10 of 28 qualities representing effective college teaching. Although specific order of qualities differed, both groups agreed on 8 of the top 10 traits: approachable, creative and interesting, encouraging and caring, enthusiastic, flexible and open-minded, knowledgeable, realistic expectations and fair, and respectful.

Kane, Sandretto, and Heath (2004) also attempted to identify the qualities of excellent college teachers. For their study, investigators asked heads of university science departments to nominate lecturers whom they deemed excellent teachers. The criteria for the nominations were based upon both peer and student perceptions of the faculty member's quality of teaching and upon the faculty member's demonstrated interest in exploring her or his own teaching practice. Investigators noted that a number of nomination letters referenced student evaluations. Five themes representing excellence resulted from the analysis of data from the 17 faculty participants. These were *knowledge of subject*, *pedagogical skill* (e.g., clear communicator, one who makes

real-world connections, organized, motivating), *interpersonal relationships* (e.g., respect for and interest in students, empathetic and caring), *research/ teaching nexus* (e.g., integration of research into teaching), and *personality* (e.g., exhibits enthusiasm and passion, has a sense of humor, is approachable, builds honest relationships).

Purpose of the Study

Although the few studies on students' perceptions of effective college instructors have yielded useful information, the researchers did not specify whether the perceptions that emerged were reflected by the TEFs used by the respective institutions. Bearing in mind the important role that TEFs play in colleges, universities, and other institutions of further and higher learning, it is vital that much more validity evidence be collected.

Because the goal of TEFs is to make local decisions (e.g., tenure, promotion, merit pay, teaching awards), it makes sense to collect such validity evidence one institution at a time and then use generalization techniques such as meta-analysis (Glass, 1976, 1977; Glass, McGaw, & Smith, 1981), meta-summaries (Sandelowski & Barroso, 2003), and meta-validation (Onwuegbuzie et al., in press) to paint a holistic picture of the appropriateness and utility of TEFs. With this in mind, the purpose of this study was to conduct a validity study of a TEF by examining students' perceptions of characteristics of effective college teachers. Using mixed-methods techniques, the researchers assessed the content-related validity and construct-related validity pertaining to a TEF. With respect to content-related validity, the item validity and sampling validity pertaining to the selected TEF were examined. With regard to construct-related validity, substantive validity was examined via an assessment of the theoretical analysis of the knowledge, skills, and processes hypothesized to underlie respondents' scores; structural validity was assessed by comparing items on the TEF to effective attributes identified both in the extant literature and by the current sample; outcome validity was evaluated via an appraisal of some of the intended and unintended consequences of using the TEF; and generalizability was evaluated via an examination of the invariance of students' perceptions of characteristics of effective college teachers (e.g., males vs. females, graduate students vs. undergraduate students). Simply put, we examined areas of validity evidence of a TEF that have received scant attention. The following mixed-methods research question was addressed: What is the content-related validity (i.e., item validity, sampling validity) and construct-related validity (i.e., substantive validity, structural validity, outcome validity, generalizability) pertaining to a TEF? Using Newman, Ridenour, Newman, and DeMarco's (2003) typology, the goal of this mixed-methods research study was to have a personal, institutional, and/or organizational impact on future TEFs. The objectives of this mixed-methods inquiry were threefold: (a) exploration, (b) description, and (c) explanation (Johnson &

122

Christensen, 2004). As such, it was hoped that the results of the current investigation would contribute to the extant literature and provide information useful for developing more effective TEFs.

Method

Participants

Participants were 912 college students who were attending a midsize public university in a midsouthern state. The sample size represented 10.66% of the student body at the university where the study took place. These students were enrolled in 68 degree programs (e.g., education, mathematics, history, sociology, dietetics, journalism, nursing, prepharmacy, premedical) that represented all six colleges. The sample was selected purposively utilizing a criterion sampling scheme (Miles & Huberman, 1994; Onwuegbuzie & Collins, in press; Patton, 1990). The majority of the sample was female (74.3%). With respect to ethnicity, the respondents comprised Caucasian American (85.4%), African American (11.0%), Asian American (1.0%), Hispanic (0.4%), Native American (0.9%), and other (1.3%). Ages ranged from 18 to 58 years ($M = 23.00$, $SD = 6.26$). With regard to level of student (i.e., undergraduate vs. graduate), 77.04% represented undergraduate students. A total of 76 students were preservice teachers. Although these demographics do not exactly match the larger population at the university, they appear to be at least somewhat representative. In particular, at the university where the study took place, 61% of the student population is female. With respect to ethnicity, the university population comprises 76% Caucasian American, 16% African American, 1% Asian American, 0.9% Hispanic, 0.86% Native American, and 2.7% unknown; of the total student population, 89% are undergraduates. The sample members had taken an average of 32.24 ($SD = 41.14$) undergraduate or 22.33 ($SD = 31.62$) graduate credit hours, with a mean undergraduate grade point average (GPA) of 2.80 ($SD = 2.29$) and mean graduate GPA of 3.18 ($SD = 1.25$) on a 4-point scale. Finally, the sample members' number of offspring ranged from 0 to 6 ($M = 0.32$, $SD = 0.84$). Because all 912 participants contributed to both the qualitative and quantitative phases of the study, and the qualitative phase preceded the quantitative phases, the mixed-methods sampling design used was a sequential design using identical samples (Collins, Onwuegbuzie, & Jiao, 2006, in press; Onwuegbuzie & Collins, in press).

Setting

The university where the study took place was established in 1907 as a public (state-funded) university. Containing 38 major buildings on its 262-acre campus, this university serves approximately 9,000 students annually (8,555 students were enrolled at the university at the time the study took place), of whom approximately 1,000 are graduate students. The university's departments and programs are organized into six academic colleges and an

123

honors college that offers an array of undergraduate and master's-level programs as well as select doctoral degrees. The university employs more than 350 full-time instructional faculty. It is classified by the Carnegie Foundation as a Masters Colleges and Universities I, and it continues to train a significant percentage of the state's schoolteachers.

Teaching Evaluation Form

At the time of this investigation, the TEF used at the university where the study took place contained two parts. The first part consisted of ten 5-point rating scale items that elicited students' opinions about their learning experiences, the syllabus, course outline, assignments, workload, and difficulty level. The second part contained 5-point Likert-type items, anchored by *strongly agree* and *strongly disagree*, for use by students when requested to critique their instructors with respect to 18 attributes. Thus, the first section of the TEF contained items that primarily elicited students' perceptions of the course, whereas the second section of the TEF contained items that exclusively elicited students' perceptions of their instructor's teaching ability. The TEF is presented in the appendix.

Instruments and Procedure

All participants were administered a questionnaire during class sessions. Participants were recruited via whole classes. The university's "Schedule of Classes" (i.e., sampling frame) was used to identify classes offered within each of the six colleges that represented various class periods (day and evening) throughout the week of data collection. Once classes were identified, instructors/professors were asked if researchers could survey their classes. All instructors/professors agreed. Each data collector read a set of instructions to participants identifying faculty involved in the study, explaining the purpose of the study (to identify students' perceptions of characteristics of effective college teachers), and emphasizing participants' choice in completing the questionnaire. Consent forms and questionnaires were distributed together to all participants. At that point, the data collector asked participants to identify and rank between three and six characteristics they believed effective college instructors possess or demonstrate. Also, students were asked to provide a definition or description for each characteristic. Low rankings denoted the most effective traits. Participants placed completed forms into envelopes provided by the collector. The recruited classes included foundation, core, and survey courses for students pursuing degrees in a variety of disciplines. This instrument also extracted the following demographic information: gender, ethnicity, age, major, year of study, number of credit hours taken, GPA, teacher status, and whether the respondent was a parent of a school-aged child. The instrument, which took between 15 and 30 minutes to complete—a similar time frame to that allotted to students to complete TEFs at many institutions—was administered in classes over a 5-day period. Using Johnson and Turner's (2003) typology, the mixed-methods data collection strategy reflected by the

124

TEF was a mixture of open- and closed-ended items (i.e., Type 2 data collection style).

To maximize its content-related validity, the questionnaire was pilot-tested on 225 students at two universities that were selected via a maximum variation sampling technique (Miles & Huberman, 1994)—one university (n = 110) that was similar in enrollment size and Carnegie foundation classification to the university where the study took place and one Research I university (n = 115). Modifications to the instrument were made during this pilot stage, as needed.

Research design. Using Leech and Onwuegbuzie's (2005, in press-b) typology, the mixed-methods research design used in this investigation could be classified as a fully mixed sequential dominant status design. This design involves mixing qualitative and quantitative approaches within one or more of, or across, the stages of the research process. In this study, the qualitative and quantitative approaches were mixed within the data analysis and data interpretation stages, with the qualitative and quantitative phases occurring sequentially and the qualitative phase given more weight.

Analysis

A sequential mixed-methods analysis (SMMA) (Onwuegbuzie & Teddlie, 2003; Tashakkori & Teddlie, 1998) was undertaken to analyze students' responses. This analysis, incorporating both inductive and deductive reasoning, employed qualitative and quantitative data-analytic techniques in a sequential manner, commencing with qualitative analyses, followed by quantitative analyses that built upon the qualitative analyses. Using Greene, Caracelli, and Graham's (1989) framework, the purpose of the mixed-methods analysis was *development*, whereby the results from one data-analytic method informed the use of the other method. More specifically, the goal of the SMMA was typology development (Caracelli & Greene, 1993).

The SMMA consisted of four stages. The first stage involved a thematic analysis (i.e., exploratory stage) to analyze students' responses regarding their perceptions of characteristics of effective college teachers (Goetz & LeCompte, 1984). The goal of this analytical method was to understand phenomena from the perspective of those being studied (Goetz & LeCompte, 1984). The thematic analysis was generative, inductive, and constructive because it required the inquirer(s) to *bracket* or suspend all preconceptions (i.e., *epoche*) to minimize bias (Moustakas, 1994). Thus, the researchers were careful not to form any a priori hypotheses or expectations with respect to students' perceptions of effective college instructors.

The thematic analysis undertaken in this study involved the methodology of *reduction* (Creswell, 1998). With reduction, the qualitative data "sharpens, sorts, focuses, discards, and organizes data in such a way that 'final' conclusions can be drawn and verified" (Miles & Huberman, 1994, p. 11) while

125

retaining the context in which these data occurred (Onwuegbuzie & Teddlie, 2003). Specifically, a modification of Colaizzi's (1978) analytic methodology was used that contained five procedural steps. These steps were as follows: (a) All the students' words, phrases, and sentences were read to obtain a feeling for them. (b) These students' responses were then *unitized* (Glaser & Strauss, 1967). (c) These units of information then were used as the basis for extracting a list of nonrepetitive, nonoverlapping significant statements (i.e., *horizonalization* of data; Creswell, 1998), with each statement given equal weight. Units were eliminated that contained the same or similar statements such that each unit corresponded to a unique instructional characteristic. (d) Meanings were formulated by elucidating the meaning of each significant statement (i.e., unit). Finally, (e) clusters of themes were organized from the aggregate formulated meanings, with each cluster consisting of units that were deemed similar in content; therefore, each cluster represented a unique emergent theme (i.e., *method of constant comparison*; Glaser & Strauss, 1967; Lincoln & Guba, 1985). Specifically, the analysts compared each subsequent significant statement with previous codes such that similar clusters were labeled with the same code. After all the data had been coded, the codes were grouped by similarity, and a theme was identified and documented based on each grouping (Leech & Onwuegbuzie, in press-a).

These clusters of themes were compared to the original descriptions to verify the clusters (Leech & Onwuegbuzie, in press-a). This was undertaken to ensure that no original descriptions made by the students were unaccounted for by the cluster of themes and that no cluster contained units that were not in the original descriptions. These themes were created a posteriori (Constas, 1992). As such, each significant statement was linked to a formulated meaning and to a theme.

This five-step method of thematic analysis was used to identify a number of themes pertaining to students' perceptions of characteristics of effective college instructors. The locus of typology development was *investigative*, stemming from the intellectual constructions of the researchers (Constas, 1992). The source for naming of categories also was investigative (Constas, 1992). *Double coding* (Miles & Huberman, 1994) was used for categorization verification, which took the form of interrater reliability. Consequently, the verification component of categorization was *empirical* (Constas, 1992). Specifically, three of the researchers independently coded the students' responses and determined the emergent themes. These themes were compared and the rate of agreement determined (i.e., interrater reliability). Because more than two raters were involved, the multirater Kappa measure was used to provide information regarding the degree to which raters achieved the possible agreement beyond any agreement than could be expected to occur merely by chance (Siegel & Castellan, 1988). Because a quantitative technique (i.e., interrater reliability) was employed as a validation technique, in addition to being empirical, the verification component of categorization was *technical* (Constas, 1992). The verification approach was accomplished a posteriori (Constas, 1992). The following

criteria were used to interpret the Kappa coefficient: < .20 = poor agreement, .21-.40 = fair agreement, .41-.60 = moderate agreement, .61-.80 = good agreement, .81-1.00 = very good agreement (Altman, 1991).

An additional method of interrater reliability, namely, peer debriefing, was used to legitimize the data interpretations. Peer debriefing provides a logically based external evaluation of the research process (Glesne & Peshkin, 1992; Lincoln & Guba, 1985; Maxwell, 2005; Merriam, 1988; Newman & Benz, 1998). The ("disinterested") peer selected was a college professor from another institution who had no stake in the findings and interpretations and who served as "devil's advocate" in an attempt to keep the data interpretations as "honest" as possible (Lincoln & Guba, 1985, p. 308).

The second stage of the sequential qualitative–quantitative mixed-methods analysis involved utilizing descriptive statistics (i.e., exploratory stage) to analyze the hierarchical structure of the emergent themes (Onwuegbuzie & Teddlie, 2003). Specifically, each theme was *quantitized* (Tashakkori & Teddlie, 1998). That is, if a student listed a characteristic that was eventually unitized under a particular theme, then a score of 1 would be given to the theme for the student response; a score of 0 would be given otherwise. This dichomotization led to the formation of an *interrespondent matrix* (i.e., Student × Theme Matrix) (Onwuegbuzie, 2003a; Onwuegbuzie & Teddlie, 2003). Both matrices consisted only of 0s and 1s.[1] By calculating the frequency of each theme from the interrespondent matrix, percentages were computed to determine the prevalence rate of each theme.[2]

The third stage of the sequential qualitative–quantitative mixed-methods analysis involved the use of the aforementioned interrespondent matrix to conduct an exploratory factor analysis to determine the underlying structure of these themes (i.e., exploratory stage). More specifically, the interrespondent matrix was converted to a matrix of bivariate associations among the responses pertaining to each of the emergent themes (Thompson, 2004). These bivariate associations represented tetrachoric correlation coefficients because the themes had been quantitized to dichotomous data (i.e., 0 vs. 1), and tetrachoric correlation coefficients are appropriate to use when one is determining the relationship between two (artificial) dichotomous variables.[3,4] Thus, the matrix of tetrachoric correlation coefficients was the basis of the exploratory factor analysis. This factor analysis determined the number of factors underlying the themes. These factors, or latent constructs, yielded *meta-themes* (Onwuegbuzie, 2003a) such that each meta-theme contained one or more of the emergent themes. The *trace*, or proportion of variance explained by each factor after rotation, served as an effect size index for each meta-theme (Onwuegbuzie, 2003a).[5] Furthermore, the combined effect size pertaining to each meta-theme was computed (Onwuegbuzie, 2003a).[6] By determining the hierarchical relationship between the themes, in addition to being empirical and technical, the verification component of categorization was *rational* (Constas, 1992).

The fourth and final stage of the sequential qualitative–quantitative mixed-methods analysis (i.e., confirmatory analyses) involved the determination of

antecedent correlates of the emergent themes that were extracted in Stage 1 and quantitized in Stage 2. This phase utilized the interrespondent matrix to undertake (a) a series of Fisher's Exact tests to determine which demographic variables were related to each of the themes and (b) a canonical correlation analysis to examine the multivariate relationship between the themes and the demographic variables. Specifically, a canonical correlation analysis (Cliff & Krus, 1976; Darlington, Weinberg, & Walberg, 1973; Thompson, 1980, 1984) was used to determine this multivariate relationship. For each statistically significant canonical coefficient, standardized canonical function coefficients and structure coefficients were computed. These coefficients served as *inferential-based effect sizes* (Onwuegbuzie, 2003a).

Onwuegbuzie and Teddlie (2003) identified the following seven stages of the mixed-methods data analysis process: (a) data reduction, (b) data display, (c) data transformation, (d) data correlation, (e) data consolidation, (f) data comparison, and (g) data integration. These authors defined *data reduction* as reducing the dimensionality of the quantitative data (e.g., via descriptive statistics, exploratory factor analysis, cluster analysis) and the qualitative data (e.g., via exploratory thematic analysis, memoing). *Data display* refers to describing visually the qualitative data (e.g., graphs, charts, matrices, checklists, rubrics, networks, and Venn diagrams) and quantitative data (e.g., tables, graphs). This is followed, if needed, by the *data transformation* stage, in which qualitative data are converted into numerical codes that can be analyzed statistically (i.e., *quantitized*; Tashakkori & Teddlie, 1998) and/or quantitative data are converted into narrative codes that can be analyzed qualitatively (i.e., *qualitized*; Tashakkori & Teddlie, 1998). *Data correlation*, the next step, involves qualitative data being correlated with quantitized data or quantitative data being correlated with qualitized data. This is followed by *data consolidation*, whereby both quantitative and qualitative data are combined to create new or consolidated variables, data sets, or codes. The next stage, *data comparison*, involves comparing data from the qualitative and quantitative data sources. *Data integration* is the final stage of the mixed-methods data analysis process, whereby both qualitative and quantitative data are integrated into either a coherent whole or two separate sets (i.e., qualitative and quantitative) of coherent wholes. In implementing the four-stage mixed-methods data analysis framework, the researchers incorporated five of the seven stages of Onwuegbuzie and Teddlie's (2003) model, namely, data reduction, data display, data transformation, data correlation, and data integration.

Using Collins, Onwuegbuzie, and Sutton's (2006) rationale and purpose (RAP) model, the rationale for conducting the mixed-methods study could be classified as (a) participant enrichment, (b) instrument fidelity, and (c) significance enhancement. *Participant enrichment* represents the mixing of quantitative and qualitative approaches for the rationale of optimizing the sample (e.g., increasing the number of participants). *Instrument fidelity* refers to procedures used by the researcher(s) to maximize the utility and/or appropriateness of the instruments used in the study, whether quantitative or qualitative. *Significance enhancement* denotes mixing qualitative and quantitative

128

techniques to maximize the interpretations of data (i.e., quantitative data can be used to enhance qualitative analyses, qualitative data can be used to enhance statistical analyses, or both). With respect to participant enrichment, the present researchers approached instructors/professors before the study began to solicit participation of their students and thus maximize the participation rate. With regard to instrument fidelity, the researchers (a) collected qualitative data (e.g., respondents' perceptions of the questionnaire) and quantitative data (e.g., response rate information, missing data information) before the study began (i.e., pilot phase) and (b) used member checking techniques to assess the appropriateness of the questionnaire and the adequacy of the time allotted to complete it, after the major data collection phases. Finally, with respect to significance enhancement, the researchers used a combination of qualitative and quantitative analyses to get more out of their initial data both during and after the study, thereby enhancing the significance of their findings (Onwuegbuzie & Leech, 2004a). Moreover, the researchers sought to use mixed-methods data-analytic techniques in an attempt to combine descriptive precision (i.e., Stages 1 and 3) with empirical precision (i.e., Stages 2 to 4) (Caracelli & Greene, 1993; Johnson & Onwuegbuzie, 2004; Onwuegbuzie & Leech, 2006). Figure 2 provides a visual representation of how the RAP model was utilized in the current inquiry.

Results

Stage 1 Analysis

Every participant provided at least three characteristics they believed effective college instructors possess or demonstrate. The participants listed a total of 2,991 significant statements describing effective college teachers. This represented a mean of 3.28 significant statements per sample member. Examples of the significant statements and their corresponding formulated meanings and the themes that emerged from the students' responses are presented in Table 1. This table reveals that the following nine themes surfaced from the students' responses: student centered, expert, professional, enthusiast, transmitter, connector, director, ethical, and responsive. The descriptions of each of the nine themes are presented in Table 2. Examples of student centered include "willingness to listen to students," "compassionate," and "caring"; examples of expert include "intelligent," and "knowledgeable"; examples of professional are "reliable," "self-discipline," "diligence," and "responsible"; words that represent enthusiast include "encouragement," "enthusiasm," and "positive attitude"; words that describe transmitter are "good communication," "speaking clearly," and "fluent English"; examples that characterize connector include "open door policy," "available," and "around when students need help"; director includes descriptors such as "flexible," "organized," and "well prepared for class"; ethical is presented by words such as "consistency," "fair evaluator," and "respectful"; finally, examples that depict responsive include "quick turnaround," "understandable," and "informative."
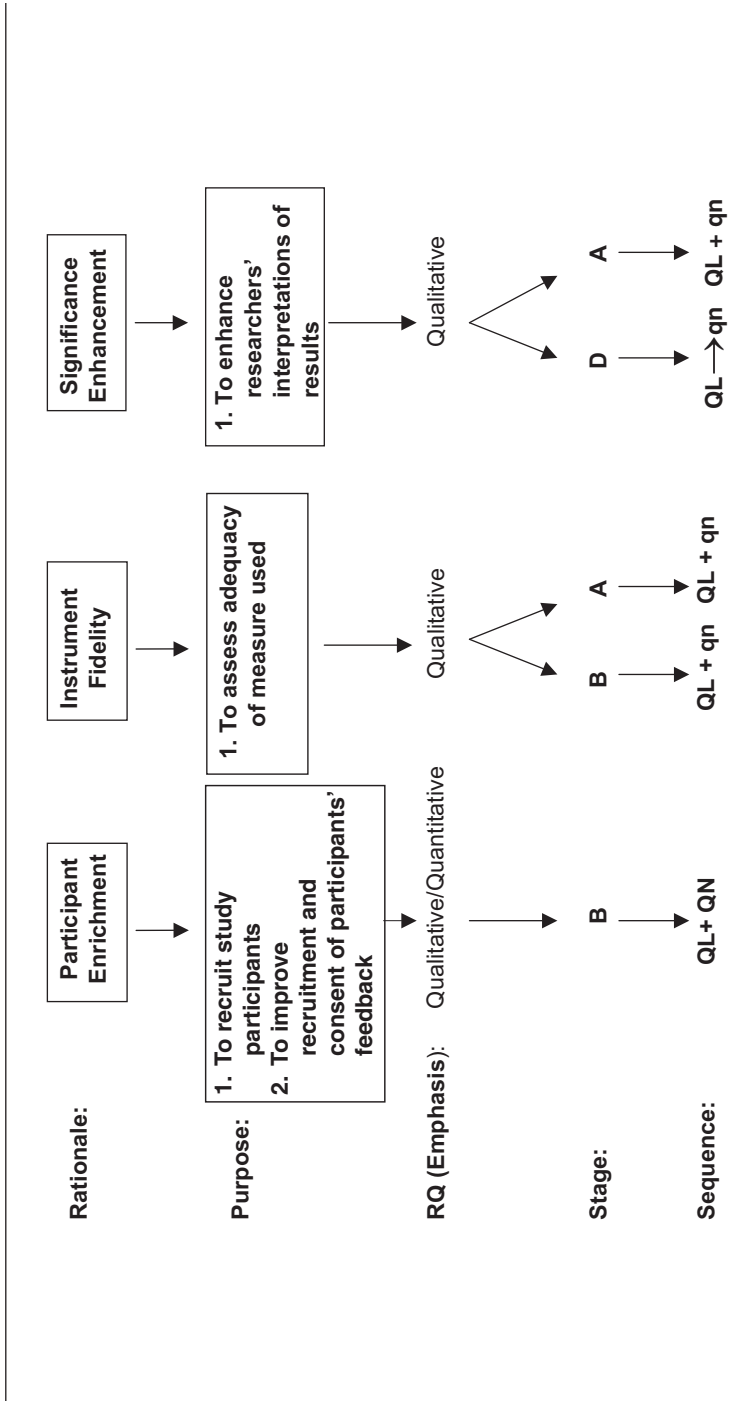
Figure 2. Visual representation of rationale and purpose (RAP) model. RQ = research question; B = before study; D = during study; A = after study; QN/qn = quantitative; QL/ql = qualitative; uppercase = dominant; lowercase = less dominent; → = sequential; + = concurrent.

Table 1
**Stage 1 Analysis: Selected Examples of Significant Statements and
Corresponding Formulated Meanings and Themes Emerging From
Students' Perceptions of Characteristics of Effective College Instructors**

Example of Significant Statement	Formulated Meaning	Theme
"Willing to make time to help if students had problems"	Sensitive to students' needs	Student centered
"Very acquainted with subject matter as well as a holistic knowledge of many other disciplines"	Well informed on course content	Expert
"Has set goals as to what should be accomplished; punctual"	Organized in preparing course	Professional
"A passion for the subject they are teaching"	Animated in delivery of course material	Enthusiast
"Keep students interested during class; good speaking skills"	Clearly conveys course material	Transmitter
"They give office hours where students can reach them and offer additional help"	Available to students	Connector
"Instructor actually know and understand what they are teaching"	Expert in his/her field	Director
"Treating each student the same; give everyone a chance"	Impartial	Ethical
"Teacher lets student know how well he/she has done or can improve"	Provider of student performance	Responsive

The interrater reliability (i.e., multirater Kappa) associated with the three researchers who independently coded the students' responses and determined the emergent themes was 93% ($SE = 0.7$), which can be interpreted as indicating very good agreement. Furthermore, based on the data, the "disinterested" peer agreed with all nine emergent themes. The only discrepancies pertained to the labels given to some of the themes. As a result of these discrepancies,[7] the "disinterested peer" and coders scheduled an additional meeting to agree on more appropriate labels for the themes and meta-themes. This led to the relabeling of some of the themes and meta-themes that were not only more insightful but also evolved into meaningful acronyms—as can be seen in the following sections.

Stage 2 Analysis

The prevalence rates of each theme (Onwuegbuzie, 2003a; Onwuegbuzie & Teddlie, 2003) are presented in Table 3. Interestingly, student centered was the most endorsed theme, with nearly 59% of the sample providing a response

Table 2
**Stage 1 Analysis: Description of Themes Emerging From Students'
Perceptions of the Characteristics of Effective College Instructors**

Theme	Description
Responsive	Provides frequent, timely, and meaningful feedback to students
Enthusiast	Exhibits passion in delivery of curricula, in particular, and representing the field, in general
Student centered	Places students in the center of the learning process, prioritizes instruction in response to student diversity and interests, possesses strong interpersonal skills
Professional	Displays behaviors and dispositions deemed exemplary for the instructor's discipline
Expert	Demonstrates relevant and current content, connects students' prior knowledge and experience with key components of curricula
Connector	Provides multiple opportunities for student and professor interactions within and outside of class
Transmitter	Imparts critical information clearly and accurately, provides relevant examples, integrates varied communication techniques to foster knowledge acquisition
Ethical	Demonstrates consistency in enforcing classroom policies, responds to students' concerns and behaviors, provides equitable opportunities for student interaction
Director	Organizes instructional time efficiently, optimizes resources to create a safe and orderly learning environment

Note. These nine themes were rearranged to produce the acronym RESPECTED.

Table 3
**Stage 2 Analysis: Themes Emerging From Students' Perceptions
of the Characteristics of Effective College Instructors**

Theme	Endorsement Rate (%)
Student centered	58.88
Expert	44.08
Professional	40.79
Enthusiast	29.82
Transmitter	23.46
Connector	23.25
Director	21.82
Ethical	21.60
Responsive	5.04

that fell into this category. The student-centered theme was followed by expert and professional, respectively, both of which secured endorsement rates greater than 40%. Enthusiast, transmitter, connector, director, and ethical each secured an endorsement rate between 20% and 30%. Finally, the responsive theme was the least endorsed, with a prevalence rate of approximately 5%.

132

Stage 3 Analysis

An exploratory factor analysis was used to determine the number of factors underlying the nine themes. This analysis was conducted because it was expected that two or more of these themes would cluster together. Specifically, a maximum likelihood factor analysis was used. This technique, which gives better estimates than does principal factor analysis (Bickel & Doksum, 1977), is perhaps the most common method of factor analysis (Lawley & Maxwell, 1971). As recommended by Kieffer (1999) and Onwuegbuzie and Daniel (2003), the correlation matrix was used to undertake the factor analysis. An orthogonal (i.e., varimax) rotation was employed because of the expected small correlations among the themes. This analysis was used to extract the latent constructs. As conceptualized by Onwuegbuzie (2003a), these factors represented *meta-themes.*

The eigenvalue-greater-than-one rule, also known as K1 (Kaiser, 1958), was used to determine an appropriate number of factors to retain. This technique resulted in four factors (i.e., meta-themes). The "scree" test, which represents a plot of eigenvalues against the factors in descending order (Cattell, 1966; Zwick & Velicer, 1986), also suggested that four factors be retained. This four-factor solution is presented in Table 4. Using a cutoff correlation of .3, recommended by Lambert and Durand (1975) as an acceptable minimum value for pattern/structure coefficients, Table 4 reveals that the following themes had pattern/structure coefficients with large effect sizes on the first factor: student centered and professional*;* the following themes had pattern/ structure coefficients with large effect sizes on the second factor: connector, transmitter, and responsive; the following themes had pattern/structure coefficients with large effect sizes on the third factor: director and ethical; and the following themes had pattern/structure coefficients with large effect sizes on the fourth factor: enthusiast and expert. The first meta-theme (i.e., Factor 1) was labeled *advocate.* The second meta-theme was termed *communicator.* The third meta-theme represented *responsible.* Finally, the fourth meta-theme denoted *empowering.* Interestingly, within the *advocate* meta-theme (i.e., Factor 1), the student-centered and professional themes were negatively related. Also, within the *responsible* meta-theme (i.e., Factor 3), the director and ethical themes were inversely related. The descriptions of each of the four meta-themes are presented in Table 5. The thematic structure is presented in Figure 3. This figure illustrates the relationships among the themes and meta-themes arising from students' perceptions of the characteristics of effective college instructors.

An examination of the *trace* (i.e., the proportion of variance explained, or eigenvalue, after rotation; Hetzel, 1996) revealed that the *advocate* meta-theme (i.e., Factor 1) explained 14.44% of the total variance, the *communicator* meta-theme (i.e., Factor 2) accounted for 13.79% of the variance, the *responsible* meta-theme (i.e., Factor 3) explained 12.86% of the variance, and the *empowering* meta-theme (i.e., Factor 4) accounted for 11.76% of the variance. These four meta-themes combined explained 52.86% of the total variance. Interestingly,

Table 4
Stage 3 Analysis: Summary of Themes and Factor Pattern/Structure Coefficients From Maximum Likelihood (Varimax) Factor Analysis: Four-Factor Solution

Theme	Factor Coefficients[a]				Communality Coefficient
	1	2	3	4	
Student centered	**–.76**	–.31	.01	–.18	.71
Professional	**.75**	–.16	.01	–.01	.59
Connector	–.11	**.64**	.25	.01	.48
Transmitter	.12	**.51**	–.24	.01	.33
Responsive	.01	**.47**	.01	–.32	.10
Director	.16	–.15	**–.72**	–.29	.87
Ethical	.20	–.15	**.72**	–.34	.70
Enthusiast	.01	.01	–.01	**.72**	.52
Expert	.01	–.38	.14	**.55**	.47
Trace	1.24	1.20	1.17	1.15	4.77
% variance explained	14.44	13.79	12.86	11.76	52.86

[a]Coefficients in bold represent pattern/structure coefficients with the largest effect size within each theme using a cutoff value of .3 recommended by Lambert and Durand (1975).

Table 5
Stage 3 Analysis: Description of Meta-Themes Emerging From Factor Analysis

Meta-Themes	Descriptions
Communicator	Serves as a reliable resource for students; effectively guides students' acquisition of knowledge, skills, and dispositions; engages students in the curriculum and monitors their progress by providing formative and summative evaluations
Advocate	Demonstrates behaviors and dispositions that are deemed exemplary for representing the college teaching profession, promotes active learning, exhibits sensitivity to students
Responsible	Seeks to conform to the highest levels of ethical standards associated with the college teaching profession and optimizes the learning experiences of students
Empowering	Stimulates students to acquire the knowledge, skills, and dispositions associated with an academic discipline or field and stimulates students to attain maximally all instructional goals and objectives

Note. These four meta-themes were rearranged to produce the acronym CARE.

this proportion of total variance explained is consistent with that typically explained in factor solutions (Henson, Capraro, & Capraro, 2004; Henson & Roberts, 2006). Furthermore, this total proportion of variance, which provides an effect size index,[8] can be considered large. The effect sizes associated with the
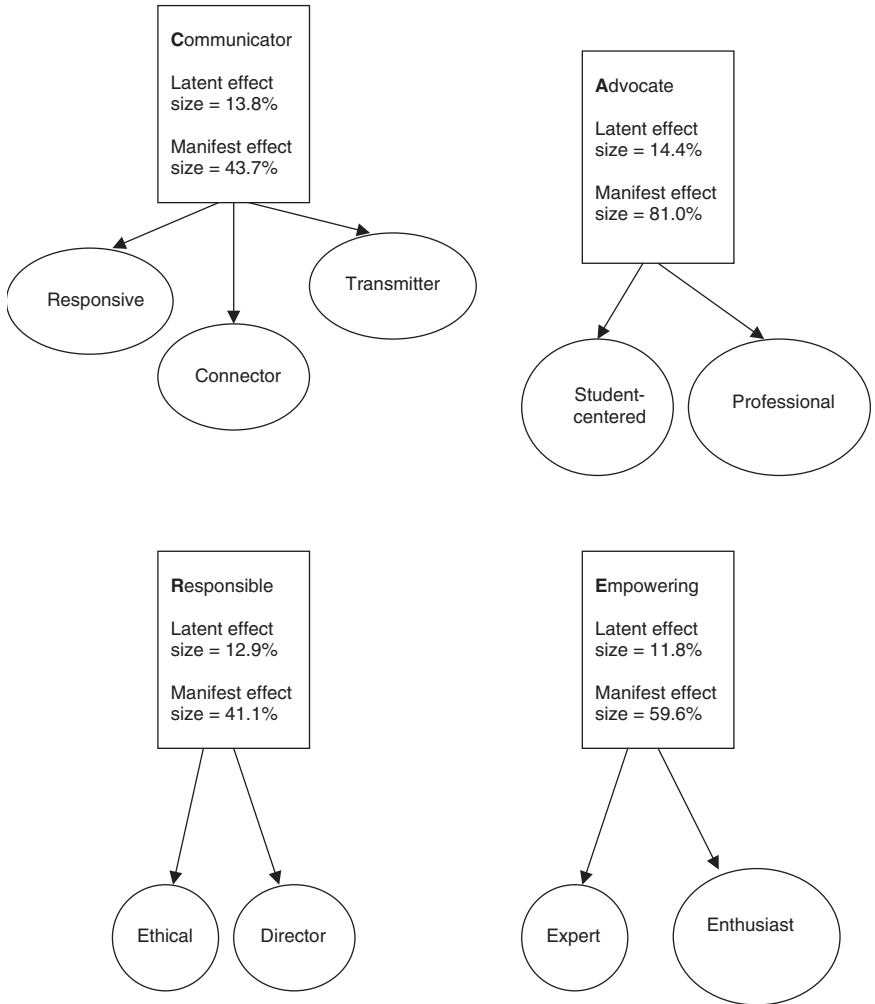
Figure 3. **State 4: Thematic structure pertaining to students' perceptions of the characteristics of effective college instructors: CARE-RESPECTED Model of Effective College Teaching. CARE = communicator, advocate, responsible, empowering; RESPECTED = responsive, enthusiast, student centered, professional, expert, connector, transmitter, ethical, and director.**

four meta-themes (i.e., proportion of characteristics identified per meta-themes)[9] were as follows: *advocate* (81.0%), *communicator* (43.7%), *responsible* (41.1%), and *empowering* (59.6%).

Stage 4 Analysis

A series of Fisher's Exact tests was used to correlate each of the nine themes with each of the following four interval- or ratio-scaled demographic variables: gender, race (Caucasian American vs. minority), level of student (undergraduate vs. graduate), and preservice teacher status (i.e., preservice teacher vs. nonpreservice teacher). Each demographic variable was treated as a family such that the Bonferroni adjustment (i.e., Bonferroni-adjusted α = .05/9 = .0056) was applied for each demographic variable to control for family-wise error. With respect to gender, females (62.3%) tended to place statistically significantly more weight on student centeredness as a measure of instructional effectiveness than did males (49.4%). The effect size associated with this relationship, as measured by Cramer's V, was .12. Furthermore, females were 1.70 times (95% confidence interval [CI] = 1.26, 2.29) more likely than were males to endorse student centeredness. However, gender was not statistically significantly related to any other theme. With respect to race, Caucasian American students (31.6%) were statistically significantly more likely to endorse enthusiastic about teaching as a characteristic of effective instruction than were minority students (19.5%). Cramer's V effect size was .09. More specifically, Caucasian American students were 1.61 times (95% CI = 1.12, 2.32) more likely than were minority students to endorse enthusiasm.

With respect to level of student, graduate students (59.6%) were statistically significantly more likely to deem being an expert in one's field as characteristic of effective instruction than were undergraduate students (39.7%). Cramer's V effect size was .17. Moreover, these graduate students were 2.24 times (95% CI = 1.64, 3.08) more likely than were undergraduates to endorse being an expert. Similarly, graduate students (32.2%) were statistically significantly more likely to consider being a director to exemplify effective instruction than were undergraduate students (18.9%). Cramer's V effect size was .14. These graduate students were 2.03 times (95% CI = 1.44, 2.88) more likely than were undergraduate students to endorse being a director.

With regard to preservice teacher status, preservice teachers (40.8%) were statistically significantly less likely to endorse student centeredness as being indicative of effective instruction than were the other students (60.7%). Cramer's V effect size was .11. Moreover, preservice teachers were 2.24 times (95% CI = 1.39, 3.61) less likely than were other students to endorse student centeredness. Conversely, preservice teachers (44.7%) were statistically significantly more likely to deem being ethical as characterizing effective instruction than were the remaining students (19.5%). Cramer's V effect size was .17. These preservice teachers were 2.29 times (95% CI = 1.72, 3.05) more likely than were other students to endorse ethicalness. Similarly, preservice teachers (23.3%) were statistically significantly more likely to endorse being a director as representing effective instruction than were the other students (6.6%). Cramer's V effect size was .11. These preservice teachers were 4.30 times (95% CI = 1.71, 10.81) more likely than were other students to endorse being a director.

A series of point-biserial correlation coefficients was conducted to correlate each of the nine themes with each of the following four demographic variables: age, GPA, number of credit hours taken, and number of offspring. After applying the Bonferroni adjustment to control for family-wise error, only three associations were statistically significant: (a) Older students were more likely to endorse professionalism as an effective instructional characteristic ($r = .12, p < .001$), (b) students with the most credit hours were more likely to endorse ethicalness ($r = .14, p < .001$), and (c) students with the most credit hours were less likely to endorse being a director ($r = -.09, p < .001$); however, all three correlations were small.

A canonical correlation analysis was undertaken to examine the relationship between the nine themes and the eight demographic variables. The nine themes were treated as the dependent set of variables, whereas the following variables were used as the independent multivariate profile: gender, race, level of student, preservice teacher status, age, GPA, number of credit hours taken, and number of offspring. The number of canonical functions (i.e., factors) that can be generated for a given data set is equal to the number of variables in the smaller of the two variable sets (Thompson, 1980, 1984, 1988, 1990). Because nine themes were correlated with eight independent variables, eight canonical functions were generated.

The canonical analysis revealed that the eight canonical correlations combined were statistically significant ($p < .0001$). Also, when the first canonical root was excluded, the remaining seven canonical roots were statistically significant ($p < .0001$; Canonical $R_{c1} = .31$). Similarly, when the first and second canonical roots were excluded, the remaining six canonical roots were statistically significant ($p < .0001$; Canonical $R_{c1} = .23$). Furthermore, when the first three canonical roots were excluded, the remaining five canonical roots were statistically significant ($p < .001$; Canonical $R_{c1} = .21$). However, when the first four canonical roots were excluded, the remaining four canonical roots were not statistically significant. In fact, removal of subsequent canonical roots did not lead to statistical significance. Together, these results suggested that the first three canonical functions were both statistically significant and practically significant (J. Cohen, 1988), but the remaining five roots were not statistically significant.

Data pertaining to the first canonical root are presented in Table 6. This table provides both standardized function coefficients and structure coefficients. Using a cutoff correlation of .3 (Lambert & Durand, 1975), the standardized canonical function coefficients revealed that student centered, professional, and director made important contributions to the set of themes—with student centered and director being the major contributors. With respect to the demographic set, one's gender, level of student, and preservice teacher status made noteworthy contributions. The structure coefficients pertaining to the first canonical function revealed that student centered, ethical, and director made important contributions (i.e., were practically significant) to the first canonical variate. The square of the structure coefficient indicated that these variables

137

Table 6
Stage 4 Analysis: Canonical Solution for First Function: Relationship Between Nine Themes and Selected Demographic Variables

Theme	Standardization Coefficient	Structure Coefficient	Structure[2] (%)
Theme			
Student centered	.70[a]	.45[a]	20.3
Professional	.37[a]	.18	3.2
Connector	.23	.04	0.2
Transmitter	.27	.16	2.6
Responsive	.16	.09	0.8
Director	.67[a]	.58[a]	33.6
Ethical	−.24	−.45[a]	20.3
Enthusiast	.25	.15	2.3
Expert	.25	.09	0.8
Demographic variable			
Number of credit hours	−.01	−.43[a]	18.5
GPA	.08	.09	0.8
Age	−.09	.05	0.3
Number of offspring	.07	−.01	0.0
Preservice teacher status	.76[a]	.81[a]	65.6
Level of student	.48[a]	.59[a]	34.8
Gender	.32[a]	.30[a]	9.0
Race	.03	.03	0.1

[a]Coefficients with the effect sizes larger than .3 (Lambert & Durand, 1975).

explained 20.3%, 20.3%, and 33.6% of the variance, respectively. With regard to the demographic cluster, preservice teacher status made the strongest contribution, followed by level of student, number of credit hours, and gender. These variables explained 65.6%, 34.8%, 18.5%, and 9.0% of the variance, respectively.

Comparing the standardized and structure coefficients identified professional as a suppressor variable because the standardized coefficient associated with this variable was large, whereas the corresponding structure coefficient was relatively small (Onwuegbuzie & Daniel, 2003). Suppressor variables are variables that assist in the prediction of dependent variables due to their correlation with other independent variables (Tabachnick & Fidell, 2006).

Table 7 presents data pertaining to the second canonical root, containing both standardized function coefficients and structure coefficients. The standardized canonical function coefficients revealed that enthusiast and expert made important contributions to the set of themes—with expert being the major contributor. With respect to the demographic set, one's gender, age, level of student, and number of credit hours made noteworthy contributions. The structure coefficients pertaining to the second canonical function revealed that enthusiast (21.2% explained variance), student centered (11.6% explained variance), and expert (49.0% explained variance) made important contributions.

Table 7

Stage 4 Analysis: Canonical Solution for Second Function: Relationship Between Nine Themes and Selected Demographic Variables

Theme	Standardization Coefficient	Structure Coefficient	Structure^2 (%)
Theme			
Student centered	−.28	−.34[a]	11.6
Professional	.24	.29	8.4
Connector	.09	.07	0.5
Transmitter	−.15	−.18	3.2
Responsive	.21	.17	2.9
Director	.09	.10	1.0
Ethical	−.09	−.05	0.3
Enthusiast	−.52[a]	−.46[a]	21.2
Expert	.74[a]	.70[a]	49.0
Demographic variable			
Number of credit hours	.60[a]	.37[a]	13.7
GPA	−.06	−.03	0.1
Age	−.30[a]	.59[a]	34.8
Number of offspring	.09	.34[a]	11.6
Preservice teacher status	−.07	−.23	5.3
Level of student	.72[a]	.60[a]	36.0
Gender	−.39[a]	.26	6.8
Race	.15	.11	1.2

[a]Coefficients with the effect sizes larger than .3 (Lambert & Durand, 1975).

With regard to the demographic cluster, level of student (36.0% explained variance) made the strongest contribution, followed by age (34.8% explained variance), number of credit hours (13.7% explained variance), and number of offspring (11.6% explained variance). Comparing the standardized and structure coefficients implicated gender as a suppressor variable because the standardized coefficient associated with this variable was large, whereas the corresponding structure coefficient was relatively small.

Table 8 presents data pertaining to the third canonical root, containing both standardized function coefficients and structure coefficients. The standardized canonical function coefficients revealed that enthusiast, student centered, professional, ethical, expert, and director made important contributions to the set of themes—with enthusiast and director being the major contributors. With respect to the demographic set, one's age, race, level of student, and preservice teacher status made similarly noteworthy contributions. The structure coefficients pertaining to the third canonical function revealed that enthusiast (20.3% explained variance), student centered (16.0% explained variance), professional (9.6% explained variance), ethical (10.9% explained variance), expert (10.2% explained variance), and director (16.8% explained variance) made important contributions. With regard to the demographic cluster, race

139

Table 8
Stage 4 Analysis: Canonical Solution for Third Function: Relationship Between Nine Themes and Selected Demographic Variables

Variable	Standardization Coefficient	Structure Coefficient	Structure^2 (%)
Theme			
Student centered	.33[a]	.40[a]	16.0
Professional	.40[a]	.31[a]	9.6
Connector	.16	.20	4.0
Transmitter	−.25	−.21	4.4
Responsive	−.05	.01	0.0
Director	−.48[a]	−.41[a]	16.8
Ethical	−.43[a]	−.33[a]	10.9
Enthusiast	−.47[a]	−.45[a]	20.3
Expert	−.33[a]	−.32[a]	10.2
Demographic variable			
Number of credit hours	.23	.25	6.3
GPA	−.18	−.27	7.3
Age	.52[a]	.32[a]	10.2
Number of offspring	.16	.39[a]	15.2
Preservice teacher status	.46[a]	.16	2.6
Level of student	−.60[a]	−.39[a]	15.2
Gender	.19	.23	5.3
Race	.58[a]	.55[a]	30.3

[a]Coefficients with the effect sizes larger than .3 (Lambert & Durand, 1975).

(30.3% explained variance) made the strongest contribution, followed by level of student (15.2% explained variance), number of offspring (15.2% explained variance), and age (10.2% explained variance). Comparing the standardized and structure coefficients identified preservice teacher status as a suppressor variable because the standardized coefficients associated with this variable were large, whereas the corresponding structure coefficient was relatively small.

In sum, the results of the canonical correlation analysis involving the themes suggest that gender, race, age, level of student, preservice teacher status, number of offspring, and number of credit hours are related in some combination to enthusiast, student centered, professional, ethical, expert, and director. Of the demographic variable set, only GPA did not appear to play a role in the prediction of the themes. On the dependent set, the following three variables consistently were not involved in any of the three multivariate relationships: connector, transmitter, and responsive.

A canonical correlation analysis also was undertaken to examine the relationship between the four meta-themes and the eight demographic variables. The four meta-themes were treated as the dependent set of variables, whereas the eight demographic variables again were utilized as the independent multivariate profile. The canonical analysis revealed that the four canonical

correlations combined were statistically significant ($p < .0001$). When the first canonical root was excluded, the remaining three canonical roots were statistically significant ($p < .0001$; Canonical $R_{c1} = .23$). Similarly, when the first and second canonical roots were excluded, the remaining two canonical roots were statistically significant ($p < .0001$; Canonical $R_{c1} = .21$). However, when the first three canonical roots were excluded, the remaining canonical root was not statistically significant. Together, these results suggested that the first two canonical functions were both statistically significant and practically significant (J. Cohen, 1988), but the remaining two roots were not statistically significant.

Data pertaining to the first canonical root are presented in Table 9. Using Lambert and Durand's (1975) cutoff, the standardized canonical function coefficients revealed that responsible and empowering made important contributions to the set of meta-themes, with empowering slightly being the major contributor. With respect to the demographic set, age, race, level of student, and preservice teacher status made noteworthy contributions, with level of student making by far the largest contribution. The structure coefficients pertaining to the first canonical function revealed that advocate (13.0% explained variance), responsible (37.2% explained variance), and empowering (47.6% explained variance) made important contributions to the first canonical variate. With regard to the demographic cluster, race (24.0% explained variance), level of student (25.0% explained variance), and preservice teacher status (13.7% explained variance) each made important contributions. Comparing the standardized and structure coefficients implicated age as a suppressor variable because the standardized coefficient associated with this variable was large, whereas the corresponding structure coefficient was relatively small.

Data pertaining to the second canonical root are presented in Table 10. Using Lambert and Durand's (1975) cutoff, the standardized canonical function coefficients revealed that communicator, advocate, and responsible made important contributions to the set of meta-themes, with advocate being by far the major contributor. With respect to the demographic set, gender, level of student, and preservice teacher status made noteworthy contributions, with gender making the largest contribution. The structure coefficients pertaining to the first canonical function revealed that advocate (74.0% explained variance) made a significant contribution to the first canonical variate. With regard to the demographic cluster, gender (13.6% explained variance), age (11.6% explained variance), GPA (10.2% explained variance), level of student (27.0% explained variance), and preservice teacher status (14.4% explained variance) each made important contributions. Comparing the standardized and structure coefficients did not reveal any suppressor variables.

In sum, the results of the canonical correlation analysis involving the meta-themes suggest that gender, race, age, GPA, level of student, and preservice teacher status are related in some combination to all four meta-themes: namely, communicator, advocate, responsible, and empowering. Of the demographic variable set, only number of credit hours and number of offspring did not appear to play a role in the prediction of the meta-themes.

141

Table 9
**Stage 4 Analysis: Canonical Solution for First Function: Relationship
Between Four Meta-Themes and Selected Demographic Variables**

Variable	Standardization Coefficient	Structure Coefficient	Structure[2] (%)
Meta-theme			
Communicator	−.02	.08	0.6
Advocate	.29	.36[a]	13.0
Responsible	−.65[a]	−.61[a]	37.2
Empowering	−.72[a]	−.69[a]	47.6
Demographic variable			
Number of credit hours	−.08	.03	0.1
GPA	−.06	−.16	2.6
Age	.56[a]	.15	2.2
Number of offspring	.02	.25	6.3
Preservice teacher status	−.58[a]	.37[a]	13.7
Level of student	.83[a]	.50[a]	25.0
Gender	.13	.09	0.8
Race	.53[a]	.49[a]	24.0

[a]Coefficients with the effect sizes larger than .3 (Lambert & Durand, 1975).

Table 10
**Stage 4 Analysis: Canonical Solution for Second Function: Relation-
ship Between Four Meta-Themes and Selected Demographic Variables**

Variable	Standardization Coefficient	Structure Coefficient	Structure[2] (%)
Meta-theme			
Communicator	.40[a]	.20	4.0
Advocate	.95[a]	.86[a]	74.0
Responsible	.31[a]	.22	4.8
Empowering	.27	.14	2.0
Demographic variable			
Number of credit hours	−.01	−.23	5.2
GPA	.27	.32[a]	10.2
Age	.16	.34[a]	11.6
Number of offspring	.09	.26	6.8
Preservice teacher status	.43[a]	.38[a]	14.4
Level of student	.32[a]	.52[a]	27.0
Gender	.68[a]	.71[a]	13.6
Race	−.14	−.15	2.3

[a]Coefficients with the effect sizes larger than .3 (Lambert & Durand, 1975).

Discussion

The purpose of this study was to conduct a validity study of a TEF by examining students' perceptions of characteristics of effective college teachers, as well as to examine factors that are associated with their perceptions. Participants

were 912 undergraduate and graduate students from various academic majors enrolled in a public university in a midsouthern state. Because the sample represented students at a single university (i.e., threat to population validity and ecological validity) whose perspectives about effective teachers were gathered at a single point in time (i.e., threat to temporal validity), it is not clear the extent to which the present findings are generalizable (i.e., have adequate external validity) to students from other institutions, particularly those from other regions of the United States. In addition, with respect to internal validity, instrumentation was a threat. Specifically, the validity of responses might have been affected by the fact that the students' perceptions were assessed via a relatively brief self-report instrument (Onwuegbuzie, 2003b). However, as stated in Note 1, member checking data revealed that the time allocated for the completion of the survey was more than sufficient for students to express their views of characteristics of effective teachers, which resulted in more than 200 hours of data, in turn yielding nearly 3,000 significant statements.

At the time of the study, the university had 8,555 undergraduate and graduate students enrolled. The sample for this investigation represented 10.7% of the total population and reflected 68 degree programs offered by the university. As such, the findings are representative, at least to some degree, of many students at that institution. In fact, the sample size far exceeded the recommended minimum sample size of 368 for a population size of 9,000 individuals (Krejecie & Morgan, 1970). Notwithstanding, the interpretations that follow pertain *only* to students at the institution where the study took place. Also, the subgroup sizes were large enough to conduct null hypothesis significance tests with very high (i.e., > .95) statistical power (Onwuegbuzie & Leech, 2004b).

Mixed-Methods Validity

Very recently, Onwuegbuzie and Johnson (2006) outlined a new typology of legitimation types in mixed research. This typology contains the following nine legitimation types: sample integration legitimation, insider–outsider legitimation, weakness minimization legitimation, sequential legitimation, conversion legitimation, paradigmatic mixing legitimation, commensurability legitimation, multiple validities legitimation, and political legitimation. Each of these legitimation types is defined in Table 11. The researchers were unable to address sequential legitimation, which is always a threat in sequential mixed-methods designs, because it could not be determined whether the findings would have changed if the quantitative phase had preceded the qualitative phase instead of the QUAL→quan design used in this study. Also, the researchers were unable to address conversion legitimation.

Notwithstanding, the remaining seven legitimation types were addressed. Specifically, sample integration legitimation was optimized by using large and identical samples for both the qualitative and quantitative approaches. This enabled the researchers justifiably to combine the inferences that

Table 11
Typology of Mixed-Methods Legitimation Types

Legitimation Type	Description
Sample integration	The extent to which the relationship between the quantitative and qualitative sampling designs yields quality meta-inferences
Inside–outside	The extent to which the researcher accurately presents and appropriately utilizes the insider's view and the observer's views for purposes such as description and explanation
Weakness minimization	The extent to which the weakness from one approach is compensated by the strengths from the other approach
Sequential	The extent to which one has minimized the potential problem wherein the meta-inferences could be affected by reversing the sequence of the quantitative and qualitative phases
Conversion	The extent to which the quantitizing or qualitizing yields quality meta-inferences
Paradigmatic mixing	The extent to which the researcher's epistemological, ontological, axiological, methodological, and rhetorical beliefs that underlie the quantitative and qualitative approaches are successfully (a) combined or (b) blended into a usable package
Commensurability	The extent to which the meta-inferences made reflect a mixed worldview based on the cognitive process of Gestalt switching and integration
Multiple validities	The extent to which addressing legitimation of the quantitative and qualitative components of the study result from the use of quantitative, qualitative, *and* mixed validity types, yielding high quality meta-inferences
Political	The extent to which the consumers of mixed-methods research value the meta-inferences stemming from *both* the quantitative and qualitative components of a study

emerged from both approaches into meta-inferences (i.e., coherent set inference; Tashakkori & Teddlie, 2003, 2006). Inside–outside legitimation was enhanced by capturing the participants' voices regarding their perceptions of characteristics of effective college instructors (i.e., insiders' views), as well as comparing their perceptions to the TEF items (outsiders' views). Weakness minimization legitimation was improved by combining descriptive precision (i.e., stemming from qualitative analyses) with empirical precision (i.e., stemming from quantitative analyses). Paradigmatic mixing legitimation was enhanced by using a fully mixed-methods research design (Leech & Onwuegbuzie, 2005, in press-b), as well as by undergoing all major steps of the mixed-methods research process (Onwuegbuzie & Leech, 2006). Commensurability legitimation was addressed by using a team of researchers that was diverse with respect to research orientation (e.g., qualitative, quantitative, and mixed-methods research orientations all were represented), college teaching experience (e.g., assistant professor, associate professor, and full professor titles all were represented), and discipline (e.g., special educator, educational foundations specialist, educational assessment, teacher

144

educator, distance-learning specialist, instructional technology specialist, research methodologist). Multiple validities legitimation was enhanced by using the RAP model to optimize participant enrichment, instrument fidelity, and significance enrichment, as well as by using techniques (e.g., interrater reliability, member checking, debriefing) that addressed as many threats to the legitimation of both the qualitative and quantitative findings as possible. Finally, political legitimation was addressed by using rigorous qualitative and quantitative techniques. Nevertheless, despite the extremely rigorous nature of the research design, replications of this inquiry are needed to assess the reliability of the current results. These replications should include the use of other mixed-methods research designs and techniques so that sequential legitimation and conversion legitimation could be addressed.

Stage 1 and Stage 2 Analyses

Using mixed-methods data analysis techniques and a sample size (10.7% of student body) that facilitated generalizations, the perceptions held by college students were found to be multidimensional in nature. Specifically, perceptions were identified that led to the following nine themes: responsive, enthusiast, student centered, professional, expert, connector, transmitter, ethical, and director. These nine themes yield the following acronym: RESPECTED. According to *The American Heritage College Dictionary* (1997, p. 1162), the word *respected* is defined as "the state of being regarded with honor or esteem." Clearly, this is a distinction to which effective teachers aspire. Thus, the acronym RESPECTED is certainly appropriate.

Although the context is primary and secondary schools, the American Association of School Administrators's (AASA's) two-element conceptualization of effective teachers can be used to classify these nine themes. The AASA concluded that characteristics of effective teachers tended to fall into two categories: (a) management and instructional techniques and (b) personal characteristics (Demmon-Berger, 1986). Specifically, the three themes (i.e., student centered, enthusiast, ethical) reflect the category of personal characteristics, whereas the remaining six categories (i.e., expert, professional, transmitter, connector, director, responsive) can be classified as representing management and instructional techniques. Comparing the results of the current study to the AASA's conceptualization revealed that a similarly high proportion of the present sample of college students noted one or more characteristics representing the personal characteristic domain (80.5%), as did those who rated a trait representing management and instructional techniques (88.8%). McNemar's test indicated no statistically significant relationship ($p > .05$) between AASA's two response categories. Specifically, college students who rated a personal characteristic as being evidence of an effective teacher were neither more nor less likely to rate a management and instructional technique. This suggests that personal characteristics and management and instructional techniques appear to represent constructs that are somewhat independent.

The finding that the student-centered theme represented descriptors that received the greatest endorsement is consistent with the results of both Witcher, Onwuegbuzie, and Minor (2001) and Minor, Onwuegbuzie, Witcher, and James (2002), who assessed preservice teachers' perceptions about characteristics of effective teachers in the context of primary and secondary classroom settings. Witcher et al. reported an endorsement rate of 79.5% for the student-centered theme, and Minor et al. documented a 55.2% prevalence rate—both of which represented the highest levels of endorsement in their respective studies. In the present investigation, 58.9% of the sample members provided one or more descriptors that typified a student-centered disposition. All three proportions, which represent very large effect sizes, suggest strongly that student centeredness is considered to be the most important characteristic of effective instruction for teachers at the elementary, secondary, and postsecondary levels. Therefore, as was the case for preservice teachers (Minor et al., 2002; Witcher et al., 2001), college students in the present study, overall, identified the *interpersonal context* as the most important indicator of effective instruction. This study's finding that student centered represented descriptors receiving the strongest student endorsement is consistent with the results of Greimel-Fuhrmann and Geyer's (2003) study that identified a student-oriented teacher (i.e., student friendly, patient, and fair) as an attribute of an effective college teacher. The characteristics of presentation skills, enthusiasm, fairness in grading (Crumbley et al., 2001), and clarity in communication (Spencer & Schmelkin, 2002) are similar to this present study's themes of transmitter, enthusiast, and ethical, respectively.

Witcher et al. (2001) identified the following six characteristics of effective teaching perceived by preservice teachers: student centeredness, enthusiastic about teaching, ethicalness, classroom and behavior management, teaching methodology, and knowledge of subject. Minor et al. (2002), in a follow-up study, replicated these six characteristics and found an additional characteristic, namely, professional. Comparing and contrasting these two sets of findings with the present results reveals several similarities and differences. Specifically, in the current investigation, the following themes from the Witcher et al. and Minor et al. studies were directly replicated: student centered, enthusiast, ethical, and expert (i.e., knowledge of subject area). Also, the professional theme identified in Minor et al.'s inquiry was directly replicated. In addition, the director theme that emerged in the present investigation appears to represent a combination of the classroom and behavior management and teaching methodology themes identified in these previous studies.

Three additional themes emerged in the present study: transmitter (23.46% endorsement rate), responsive (5.04% endorsement rate), and connector (23.25% endorsement rate). These themes have intuitive appeal, bearing in mind the nature of higher education. The emergence of the transmitter and responsive themes likely resulted from the fact that the material covered and homework assigned at the college level can be extremely complex. As such, many students need clear, explicit instructions and detailed feedback. In public schools, classroom teachers are more accessible as teachers are

146

on-site for most, if not all, of the school day. In contrast, college instructors are expected to engage actively in research and service activities that must be undertaken outside their offices. As such, the amount of time that instructors are available for students (i.e., office hours) varies from department to department, college to college, and university to university. In addition, the requirements imposed by administrators for faculty's office hours vary. Some institutions have no office requirements for professors, whereas others expect a minimum of 10 office hours per week. Furthermore, the majority of current undergraduate and graduate students is actively employed while enrolled in college—with a significant proportion working on a full-time basis (Cuccaro-Alamin & Choy, 1998; Horn, 1994). Thus, many students find it difficult to schedule appointments with their instructors during posted office hours. These factors may explain why connector, which includes being accessible, was deemed a characteristic of effective teachers by nearly one fourth of the sample members.

Stage 3 Analysis

Interestingly, all three new emergent themes (i.e., transmitter, responsive, connector) appeared to belong to one factor, namely, the *communicator* meta-theme, indicating that they belong to a set. Consistent with this conclusion, these were the only three themes that were not related to any of the demographic variables. Thus, future research should examine other factors that might predict these three variables. Variables that might be considered include cognitive variables (e.g., study habits), affective variables (e.g., anxiety, self-esteem), and personality variables (e.g., levels of social interdependence, locus of control).

In addition to the *communicator* meta-theme, three other meta-themes emerged: *advocate*, comprising student centered and professional; *responsible*, representing director and ethical; and *empowering*, consisting of expert and enthusiast. The finding within the *advocate* meta-theme that student centered and professional themes were negatively related suggests that college students who were the most likely to endorse being student centered as a characteristic of effective teaching tended to be the least likely to endorse being professional as an effective trait, and vice versa. This result is interesting because it suggests that to some extent, many students view student centeredness and professionalism as lying on opposite ends of the continuum. It is possible that they have experienced teachers who give the impression of being the most professional because they exhibit traits such as efficiency, self-discipline, and responsibility, yet, at the same time, are least likely to display student-centered characteristics such as willingness to listen to students, compassion, and care. This should be the subject of future investigations.

Within the *responsible* meta-theme, the director and ethical themes also were inversely related. In other words, students who deemed ethical to represent characteristics of effective college instructors, at the same time, tended not

to endorse being a director, and vice versa. Indeed, of the sample members who endorsed the ethical theme, 89.3% did not endorse the director theme, yielding an odds ratio of 2.34 (95% CI = 1.53, 3.57). Unfortunately, it is beyond the scope of the present investigation to explain this finding. Thus, follow-up studies using qualitative techniques are needed.

The most compelling finding pertaining to the meta-themes was that student labels represent the acronym CARE. According to *The American Heritage College Dictionary* (1997, p. 212), the following definitions are given for the word *care*: "Close attention," "watchful oversight," "charge or supervision," "attentive assistance or treatment to those in need," "to provide needed assistance or watchful supervision," and "to have a liking or attachment." All of these definitions are particularly pertinent to the field of college teaching. Therefore, the acronym CARE is extremely apt.

Stage 4 Analysis

Themes. The canonical correlation analysis involving the themes revealed that three canonical correlations describe the relationship between students' attributes and their perceptions of characteristics of effective college instructors. The first canonical solution indicated that the traits student centered, professional, director, and ethical are related to the following background variables: gender, level of student, preservice teacher status, and number of credit hours. This suggests that these four themes best distinguish college students' perceptions of effective college teachers as a function of gender, level of student, preservice teacher status, and number of credit hours. That is, these themes combined represent a combination of college students' perceptions (i.e., latent function) that can be predicted by their gender, level of study (i.e., undergraduate vs. graduate), whether they are preservice teachers, and number of credit hours. An inspection of the signs of the coefficients indicates that ethical is inversely related to the remaining themes (i.e., enthusiast, student centered, director). That is, students' attributes that predicted endorsement of the enthusiast, student-centered, and director themes tended to predict nonendorsement of the ethical theme, and vice versa. Interestingly, two themes (i.e., student centered and professional) belonged to the same meta-theme, namely, advocate; whereas the remaining themes, namely, director and ethical, belong to the responsible meta-theme.

The second canonical correlation solution indicated that enthusiast, expert, and student centered composed a set related to the following demographic variables: gender, age, level of student, number of credit hours, and number of offspring. Therefore, these three themes represent a combination of college students' perceptions that can be predicted by their gender, age, level of study, number of credit hours undertaken, and number of offspring. An inspection of the signs of the coefficients indicates that expert is inversely related to enthusiast and student centered. Interestingly, enthusiast and expert represent the empowering meta-theme, whereas student centered represents the advocate meta-theme.

The third canonical correlation solution indicated that enthusiast, student centered, professional, ethical, expert, and director comprised a set related to the following demographic variables: age, race, level of student, preservice teacher status, and number of offspring. Thus, advocate (i.e., student centered, professional), empowering (i.e., enthusiast, expert), and responsible (i.e., ethical, director) represent a combination of college students' perceptions that can be predicted by their age, race, level of student, preservice teacher status, and number of offspring. An inspection of the signs of the coefficients indicates that the two themes that represent the advocate meta-theme are inversely related to the remaining themes that represent this latent variable (i.e., enthusiast, expert, ethical, director).

Meta-themes. The canonical correlation analysis involving the meta-themes revealed that two canonical correlations describe the relationship between students' attributes and the meta-themes that evolved. The first canonical solution indicated that the advocate, responsible, and empowering meta-themes are related to the following background variables: age, race, level of student, and preservice teacher status. This suggests that being an advocate, responsible, and empowering best distinguish college students' perceptions of effective college teachers as a function of age, race, level of student, and preservice teacher status. An inspection of the signs of the coefficients indicates that advocate is inversely related to the remaining meta-themes (i.e., responsible, empowering). That is, students' attributes that predicted endorsement of the responsible and empowering meta-themes tended to predict nonendorsement of the advocate meta-theme, and vice versa. The second canonical correlation solution indicated that communicator, advocate, and responsible as a set are related to the following demographic variables: gender, age, GPA, level of student, and preservice teacher status.

The findings that gender, race, age, level of student, preservice teacher status, number of offspring, and number of credit hours are related in some combination to enthusiast, student centered, professional, ethical, expert, and director and that gender, race, age, GPA, level of student, and preservice teacher status are related in some combination to the four meta-themes suggest that individual differences exist with respect to students' perceptions of the characteristics of effective college teachers. Thus, any instrument that omits items that represent any of the emergent themes or meta-themes may lead to a particular group of students (e.g., graduates, minority students) being "disenfranchised," inasmuch as the instructional attributes that these students perceive play an important role in optimizing their levels of course performance are not available to them for rating. In turn, such an omission would represent a serious threat to the content- and construct-related validity pertaining to the TEF.

Furthermore, the relationships found between the majority of the demographic variables and several themes and meta-themes suggest that when interpreting responses to items contained in TEFs, administrators should consider the demographic profile of the underlying class. Unfortunately, this does not appear to be the current practice. According to Schmelkin, Spencer,

and Gellman (1997), many administrators unwisely aggregate responses for the purpose of summative evaluation and comparison with peers without taking into account the context in which the class was taught. For instance, the finding that female students tend to place more weight on student centeredness than do male students, although replicating the findings of Witcher et al. (2001), suggests that a class with predominantly or exclusively female students—often the case in education courses—might scrutinize the instructor's degree of student centeredness to a greater extent than might a class containing primarily males—often the case in courses involving the hard sciences. Similarly, a class containing mainly Caucasian American students is more likely to assess the instructor's level of enthusiasm than is a class predominantly containing minority students (Minor et al., 2002).

Comparison of Findings With TEF

Of the nine emergent themes, five were represented by items found in the second section of the course/instructor evaluation form (cf. the appendix). These five themes were professional, transmitter, connector, director, and responsive. Specifically, *professional* was represented by the following item: "The instructor is punctual in meeting class and office hour responsibilities." *Transmitter*, the most represented theme, consisted of the following items: (a) "Rate how well the syllabus, course outline, or other overviews provided by the instructor helped you to understand the goals and requirements of this course"; (b) "Rate how well the assignments helped you learn"; (c) "My instructor's spoken English is . . ."; (d) "The instructor communicates the purposes of class sessions and instructional activities"; (e) "The instructor speaks clearly and audibly when presenting information"; (f) "The instructor uses examples and illustrations which help clarify the topic being discussed"; and (g) "The instructor clears up points of confusion." *Accessible* was represented by the following item: "The instructor provides the opportunity for assistance on an individual basis outside of class." *Director* was represented by the following items: (a) "How would you rate the instructor's teaching?" and (b) "The instructor makes effective use of class time." Finally, *responsive* was represented by the following items: (a) "The instructor gives me regular feedback about how well I am doing in the course"; (b) "The instructor returns exams and assignments quickly enough to benefit me"; and (c) "The instructor, when necessary, suggests specific ways I can improve my performance in this course." This instrument, which did not stem from any theoretical framework, was developed by administrators and select faculty, with no input from students.

Four themes were not represented by any of the items in the university evaluation form. These were student centered, expert, enthusiast, and ethical. Disturbingly, student centered, expert, and enthusiast represent three of the most prevalent themes endorsed by the college sample. In an effort to begin the process of generalizing the present findings, the researchers who, between them, have taught at three Research I/Research Extensive and two

150

Research II/Research Intensive institutions, also examined the TEFs used at these sites. It was found that for each of these five institutions, at least three of these themes (i.e., student centered, enthusiast, and ethical) were not represented by any of the items in the corresponding teacher evaluation form. This discrepancy calls into serious question the content-related validity (i.e., item validity, sampling validity) and construct-related validity (i.e., structural validity, outcome validity, generalizability) pertaining to these TEFs.

There appears to be a clear gap between what the developers of TEFs consider to be characteristics of effective instructors and what students deem to be the most important traits. Moreover, this gap suggests that students' criteria for assessing college instructors may not be adequately represented in TEFs; this might adversely affect students' ability to critique their instructors in a comprehensive manner. Thus, even if the scores yielded by this university evaluation form are reliable, the overall score validity of the TEF is in question. In an era in which information gleaned from TEFs is used to make decisions about faculty regarding tenure, promotion, and merit pay issues, this potential threat to validity is disturbing and warrants further research.

Conclusion

Despite the mixed interpretability of TEFs, colleges and universities continue to use students' ratings and interpret students' responses as reliable and valid indices of teaching effectiveness (Seldin, 1999), even though the fact that these TEFs (a) are developed atheoretically and (b) omit what students deem to be the most important characteristics of effective college teachers. Given the likelihood that colleges and universities will continue to use student ratings as an evaluative measure of teaching effectiveness, it is surprising that there has been limited systematic inquiry to examine students' perceptions regarding characteristics of effective college teachers. Thus, the investigators believe that this study has added to the current yet scant body of literature regarding the score validity of TEFs (Onwuegbuzie et al., in press). The current findings cast some serious doubt on the content-related validity (i.e., item validity, sampling validity) and construct-related validity (i.e., substantive validity, structural validity, outcome validity, generalizability) pertaining to the TEF under investigation, as well as possibly on other TEFs across institutions that are designed atheoretically and are not driven by data. This has serious implications for current policies at institutions pertaining to tenure, promotion, merit pay increases for faculty, and other decisions that rely on TEFs.

The next step in the process is to design and score validate an instrument that provides formative and summative information about the efficacy of instruction based upon the various themes and meta-themes making up the CARE-RESPECTED Model of Teaching Evaluation that emerged from this study. The researchers presently are undertaking this task and hope that the outcome will provide a useful data-driven instrument that clearly benefits all stakeholders—college administrators, teachers, and, above all, students.

APPENDIX

Instructor Evaluation Form

MARKING INSTRUCTIONS

- **Use a number 2 pencil only.**
- **Erase changes cleanly and completely.**
- **Do not make any stray marks.**

1. How much do you feel you have learned in this class?
(A) A great deal
(B) More than usual
(C) About the usual amount
(D) Less than usual
(E) Very little

2. How would you rate the instructor's teaching?
(A) Exceptional
(B) Very good
(C) Good
(D) Not very good
(E) Poor

3. How would you rate the course in general?
(A) Exceptional
(B) Very good
(C) Good
(D) Not very good
(E) Poor

4. Rate how well the syllabus, course outline, or other overviews provided by the instructor helped you understand the goals and requirements of this course.
(A) Exceptionally well
(B) Very well
(C) Well
(D) Not very well
(E) Not at all

5. Rate how well the assignments helped you learn.
(A) Exceptionally well
(B) Very well
(C) Well
(D) Not very well
(E) Not at all

6. The workload for this course is
(A) Very light
(B) Light
(C) About average
(D) Heavy
(E) Very heavy

7. The difficulty level of the course activities and materials is
(A) Very easy
(B) Easy
(C) About average
(D) Difficult
(E) Very difficult

8. Of the following, which best describes this course for you?
(A) Major Field (or Graduate Emphasis)
(B) Minor Field
(C) General Education Requirements
(D) Elective
(E) Other

9. Your classification is
(A) Freshman
(B) Sophomore
(C) Junior
(D) Senior
(E) Graduate

10. My instructor's spoken English is
(A) Exceptionally easy to understand
(B) Easy to understand
(C) Understandable
(D) Difficult to understand
(E) Exceptionally difficult to understand

Directions: For each statement completely darken the circle on the answer sheet that corresponds to your rating of the instructor. The description of the rating scale is as follows:

Strongly Agree	**SA**
Agree	**A**
Disagree	**D**
Strongly Disagree	**SD**
I Don't Know/Not Applicable	**NA**

	SA	A	D	SD	NA
11. The instructor communicates the purposes of the class sessions and instructional activities.	O	O	O	O	O
12. The instructor speaks clearly and audibly when presenting information.	O	O	O	O	O
13. The instructor uses examples and illustrations which help clarify the topic being discussed.	O	O	O	O	O
14. The instructor shows meaningful relationships among the topics in this course.	O	O	O	O	O
15. The instructor inspires interest in the subject matter of this course.	O	O	O	O	O
16. The instructor relates course material to life situations when possible.	O	O	O	O	O
17. The instructor asks questions that challenge me to think.	O	O	O	O	O
18. The instructor provides opportunities for me to suggest or discuss issues related to the course.	O	O	O	O	O
19. The instructor develops an atmosphere of respect and trust in the classroom.	O	O	O	O	O
20. The instructor manages classroom discussions so that they are useful.	O	O	O	O	O
21. The instructor clears up points of confusion.	O	O	O	O	O
22. The instructor provides the opportunity for assistance on an individual basis outside of class.	O	O	O	O	O
23. The instructor gives me regular feedback about how well I am doing in the course.	O	O	O	O	O
24. The instructor gives tests and assignments quickly enough to benefit me.	O	O	O	O	O
25. The instructor returns exams and assignments quickly enough to benefit me.	O	O	O	O	O
26. The instructor, when necessary, suggests specific ways I can improve my performance in this course.	O	O	O	O	O
27. The instructor makes effective use of class time.	O	O	O	O	O
28. The instructor is punctual in meeting class and office hour responsibilities.	O	O	O	O	O

DEPARTMENTAL QUESTIONS (Optional)

Notes

This manuscript was adapted from Onwuegbuzie and Johnson (2006). Reprinted with kind permission of the Mid-South Educational Research Association and the editors of *Research in the Schools.* Correspondence should be addressed to Anthony J. Onwuegbuzie, Department of Educational Measurement and Research, College of Education, University of South Florida, 4202 East Fowler Avenue, EDU 162, Tampa, FL 33620-7750; e-mail: tonyonwuegbuzie@aol.com.

[1]This quantitizing of themes led to the computation of what Onwuegbuzie (2003a) called *manifest effect sizes* (i.e., effect sizes pertaining to observable content). Manifest effect sizes are effect sizes that pertain to observable content (Onwuegbuzie & Teddlie, 2003).

153

[2]These prevalence rates provided *frequency effect size* measures (Onwuegbuzie, 2003a). Frequency effect size measures represent the frequency of themes within a sample that can be converted to a percentage (i.e., prevalence rate) (Onwuegbuzie & Teddlie, 2003).

[3]It should be noted that tetrachoric correlation coefficients are based on the assumption that for each manifest dichotomous variable, there is a normally distributed latent continuous variable with zero mean and unit variance. For the present investigation, it was assumed that the extent to which each participant contributed to a theme, as indicated by the order in which the significant statements were presented, represented a normally distributed latent continuous variable. Unfortunately, this assumption could not be tested given only the manifest variable (Nelson, Rehm, Bedirhan, Grant, & Chatterji, 1999). However, this assumption was deemed reasonable given the large sample size (i.e., $n = 912$).

[4]As noted by Bernstein and Teng (1989), dichotomous items are less likely to yield artifacts using factor analytic techniques than are multicategory (Likert-type) items. For more justification about conducting exploratory factor analyses on inter-respondent matrices, see Onwuegbuzie (2003a).

[5]More specifically, the *trace* served as a *latent* effect size for each meta-theme (Onwuegbuzie, 2003a). A latent effect size is an effect size pertaining to nonobservable, underlying aspects of the phenomenon being studied (Onwuegbuzie & Teddlie, 2003).

[6]The combined frequency effect size for themes within each meta-theme represented a manifest effect size (Onwuegbuzie, 2003a).

[7]This additional meeting also was prompted by one of the anonymous reviewers, who questioned some of the labels given to the themes/meta-themes and asked the researchers to derive themes that were more "insightful." Thus, we graciously thank this anonymous reviewer for providing such an important recommendation.

[8]This effect size represents a latent effect size.

[9]These effect sizes represent manifest effect sizes.

References

Aleamoni, L. M. (1981). The use of student evaluations in the improvement of instruction. *NACTA Journal, 20,* 16.

Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman and Hall.

Ambady, N., & Rosenthal, R. (1992). Half a minute. Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology, 64,* 431–441.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing* (Rev. ed.). Washington, DC: American Educational Research Association.

The American Heritage College Dictionary (3rd ed.). (1997). Boston, MA: Houghton Mifflin.

Babad, E. (2001). Students' course selection: Differential considerations for first and last course. *Research in Higher Education, 42,* 469–492.

Bernstein, I. H., & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin, 105,* 467–477.

Bickel, P. J., & Doksum, K. A. (1977). *Mathematical statistics*. San Francisco: Holden-Day.

Blackburn, R. T., & Clark, M. J. (1975). An assessment of faculty performance. Some correlates between administrators, colleagues, students, and self-ratings. *Sociology of Education, 48,* 242–256.

Braskamp, L. A., Brandenberg, D. C., & Ory, J. C. (1984). *Evaluating teaching effectiveness: A practical guide.* Beverly Hills, CA: Sage.

Caracelli, V. W., & Greene, J. C. (1993). Data analysis strategies for mixed-methods evaluation designs. *Educational Evaluation and Policy Analysis, 15,* 195–207.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245–276.

Centra, J. A. (1974). The relationship between student and alumni ratings of teachers. *Educational and Psychological Measurement, 34,* 321–326.

Centra, J. A., & Creech, F. R. (1976). *The relationship between student teachers and course characteristics and student ratings of teacher effectiveness* (Project Report 76-1). Princeton, NJ: Educational Testing Service.

Cliff, N., & Krus, D. J. (1976). Interpretation of canonical analyses: Rotated vs. unrotated solutions. *Psychometrica, 41,* 35–42.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research, 51,* 281–309.

Colaizzi, P. F. (1978). Psychological research as the phenomenologist views it. In R. Vaile & M. King (Eds.), *Existential phenomenological alternatives for psychology* (pp. 48–71). New York: Oxford University Press.

Collins, K. M. T., Onwuegbuzie, A. J., & Jiao, Q. G. (2006). Prevalence of mixed methods sampling designs in social science research. *Evaluation and Research in Education*, *19*(2), 119.

Collins, K. M. T., Onwuegbuzie, A. J., & Jiao, Q. G. (in press). A mixed methods investigation of mixed methods sampling designs in social and health science research. *Journal of Mixed Methods Research*.

Collins, K. M. T., Onwuegbuzie, A. J., & Sutton, I. L. (2006). A model incorporating the rationale and purpose for conducting mixed methods research in special education and beyond. *Learning Disabilities: A Contemporary Journal, 4,* 67–100.

Constas, M. A. (1992). Qualitative data analysis as a public event: The documentation of category development procedures. *American Educational Research Journal, 29,* 253–266.

Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Holt, Rinehart & Winston.

Crumbley, L., Henry, B. K., & Kratchman, S. H. (2001). Students' perceptions of the evaluation of college teaching. *Quality Assurance in Education, 9,* 197–207.

Cuccaro-Alamin, S., & Choy, S. (1998). *Post secondary financing strategies: How undergraduates combine work, borrowing, and attendance* (NCES 98-088). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Darlington, R. B., Weinberg, S. L., & Walberg, H. J. (1973). Canonical variate analysis and related techniques. *Review of Educational Research, 42,* 131–143.

Demmon-Berger, D. (1986). *Effective teaching: Observations from research*. Arlington, VA: American Association of School Administrators. (ERIC Document Reproduction Service No. ED274087)

Dommeyer, C. J., Baum, P., Chapman, K. S., & Hanna, R. W. (2002). Attitudes of business faculty towards two methods of collecting teaching evaluations: Paper vs. online. *Assessment & Evaluation in Higher Education, 27,* 455–462.

Doyle, K. O., & Crichton, L. A. (1978). Student, peer, and self-evaluations of college instruction. *Journal of Educational Psychology, 70,* 815–826.

Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers and courses: What we know and what we don't. *Research in Higher Education, 9,* 199–242.

155

Feldman, K. A. (1989). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education, 30,* 583–645.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Glass, G. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5,* 3–8.

Glass, G. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education, 5,* 351–379.

Glass, G., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Glesne, C., & Peshkin, A. (1992). *Becoming qualitative researchers: An introduction*. White Plains, NY: Longman.

Goetz, J. P., & LeCompte, M. D. (1984). *Ethnography and the qualitative design in educational research*. New York: Academic Press.

Gray, M., & Bergmann, B. R. (2003). Student teaching evaluations: Inaccurate, demeaning, misused. *Academe*, *89*(5), 44–46.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11,* 255–274.

Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist, 52,* 1209–1217.

Greimel-Fuhrmann, B., & Geyer, A. (2003). Students' evaluation of teachers and instructional quality—Analysis of relevant factors based on empirical evaluation. *Assessment and Evaluation in Higher Education, 28,* 229–238.

Guthrie, E. R. (1954). *The evaluation of teaching: A progress report*. Seattle: University of Washington Press.

Haskell, R. E. (1997). Academic freedom, tenure, and student evaluation of faculty: Galloping polls in the 21st century. *Education Policy Analysis Archives, 5*(6). Retrieved July 26, 2006, from http://epaa.asu.edu/epaa/v5n6.html

Henson, R. K., Capraro, R. M., & Capraro, M. M. (2004). Reporting practice and use of exploratory factor analysis in educational research journals: Errors and explanation. *Research in the Schools*, *11*(2), 61–72.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research. *Educational and Psychological Measurement, 66,* 393–416.

Hetzel, R. D. (1996). A primer on factor analysis with comments on patterns of practice and reporting. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 4, pp. 175–206). Greenwich, CT: JAI.

Horn, L. J. (1994). *Undergraduates who work while enrolled in postsecondary education* (NCES 94-311). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Johnson, R. B., & Christensen, L. B. (2004). *Educational research: Quantitative, qualitative, and mixed approaches*. Boston: Allyn & Bacon.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26.

Johnson, R. B., & Turner, L. A. (2003). Data collection strategies in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 297–319). Thousand Oaks, CA: Sage.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23,* 187–200.

Kane, R., Sandretto, S., & Heath, C. (2004). An investigation into excellent tertiary teaching: Emphasizing reflective practice. *Higher Education, 47,* 283–310.

156

Kieffer, K. M. (1999). An introductory primer on the appropriate use of exploratory and confirmatory factor analysis. *Research in the Schools*, *6*(2), 75–92.

Krejecie, R. V., & Morgan, D. W. (1970). Determining sample sizes for research activities. *Educational and Psychological Measurement, 30,* 608.

Kulik, J. A. (2001, Spring). Student ratings: Validity, utility, and controversy. *New Directions for Institutional Research, 109,* 9–25.

Lambert, Z. V., & Durand, R. M. (1975). Some precautions in using canonical analysis. *Journal of Market Research, 12,* 468–475.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. New York: Macmillan.

Leech, N. L., & Onwuegbuzie, A. J. (2005, April). *A typology of mixed methods research designs.* Invited James E. McLean Outstanding Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Leech, N. L., & Onwuegbuzie, A. J. (in press-a). An array of qualitative data analysis tools: A call for qualitative data analysis triangulation. *School Psychology Quarterly*.

Leech, N. L., & Onwuegbuzie, A. J. (in press-b). A typology of mixed methods research designs. *Quality & Quantity: International Journal of Methodology*.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11,* 253–388.

Marsh, H. W., & Bailey, M. (1993). Multidimensional students' evaluations of teaching effectiveness. A profile analysis. *Journal of Higher Education, 64,* 1–18.

Marsh, H. W., Overall, J. U., & Kessler, S. P. (1979). Validity of student evaluations of instructional effectiveness: A comparison of faculty self-evaluations and evaluations by their students. *Journal of Educational Psychology, 71,* 149–160.

Marsh, H. W., & Roche, L. A. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. *American Educational Research Journal, 30,* 217–251.

Maxwell, J. A. (2005). *Qualitative research design: An interactive approach* (2nd. ed.). Thousand Oaks, CA: Sage.

Merriam, S. (1988). *Case study research in education: A qualitative approach*. San Francisco: Jossey-Bass.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). Old Tappan, NJ: Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741–749.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: A sourcebook of new methods*. Thousand Oaks, CA: Sage.

Minor, L. C., Onwuegbuzie, A. J., Witcher, A. E., & James, T. L. (2002). Preservice teachers' educational beliefs and their perceptions of characteristics of effective teachers. *Journal of Educational Research, 96,* 116–127.

Moustakas, C. (1994). *Phenomenological research methods*. Thousands Oaks, CA: Sage.

Murray, H. G. (1983). Low-inference classroom teaching behaviors and student ratings of college teaching effectiveness. *Journal of Educational Psychology, 71,* 856–865.

Naftulin, D. H., Ware, J. E., & Donnelly, F. A. (1973). The Doctor Fox lecture: A paradigm of educational seduction. *Journal of Medical Education, 48,* 630–635.

Nelson, C. B., Rehm, J., Bedirhan, U., Grant, B., & Chatterji, S. (1999). Factor structures for DSM-IV substance disorder criteria endorsed by alcohol, cannabis, cocaine, and opiate users: Results from the WHO reliability and validity study. *Addiction, 94,* 843–855.

Newman, I., & Benz, C. R. (1998). *Qualitative-quantitative research methodology: Exploring the interactive continuum*. Carbondale: Southern Illinois University Press.

Newman, I., Ridenour, C. S., Newman, C., & DeMarco, G. M. P. (2003). A typology of research purposes and its relationship to mixed methods. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 167–188). Thousand Oaks, CA: Sage.

Okpala, C. O., & Ellis, R. (2005). The perceptions of college students on teacher quality: A focus on teacher qualifications. *Education, 126,* 374–378.

Onwuegbuzie, A. J. (2003a). Effect sizes in qualitative research: A prolegomenon. *Quality & Quantity: International Journal of Methodology, 37,* 393–409.

Onwuegbuzie, A. J. (2003b). Expanding the framework of internal and external validity in quantitative research. *Research in the Schools*, *10*(1), 71–90.

Onwuegbuzie, A. J., & Collins, K. M. T. (in press). A typology of mixed methods sampling designs in social science research. *The Qualitative Report*.

Onwuegbuzie, A. J., & Daniel, L. G. (2002). A framework for reporting and interpreting internal consistency reliability estimates. *Measurement and Evaluation in Counseling and Development, 35,* 89–103.

Onwuegbuzie, A. J., & Daniel, L. G. (2003, February 12). Typology of analytical and interpretational errors in quantitative and qualitative educational research. *Current Issues in Education* [Electronic version], *6*(2). Available from http://cie.ed.asu.edu/volume6/number2/

Onwuegbuzie, A. J., & Daniel, L. G. (2004). Reliability generalization: The importance of considering sample specificity, confidence intervals, and subgroup differences. *Research in the Schools*, *11*(1), 61–72.

Onwuegbuzie, A. J., Daniel, L. G., & Collins, K. M. T. (2006). *Student teaching evaluations: Psychometric, methodological, and interpretational issues*. Manuscript submitted for publication.

Onwuegbuzie, A. J., Daniel, L. G., & Collins, K. M. T. (in press). A meta-validation model for assessing the score validity of student teacher evaluations. *Quality and Quantity: International Journal of Methodology.*

Onwuegbuzie, A. J., & Johnson, R. B. (2006). The validity issue in mixed research. *Research in the Schools*, *13*(1), 48–63.

Onwuegbuzie, A. J., & Leech, N. L. (2004a). Enhancing the interpretation of "significant" findings: The role of mixed methods research. *The Qualitative Report, 9,* 770-792. Retrieved July 26, 2006, from http://www.nova.edu/ssss/QR/QR9-4/onwuegbuzie.pdf

Onwuegbuzie, A. J., & Leech, N. L. (2004b). Post-hoc power: A concept whose time has come. *Understanding Statistics, 3,* 201–230.

Onwuegbuzie, A. J., & Leech, N. L. (2006). Linking research questions to mixed methods data analysis procedures. *The Qualitative Report*, *11*, 474-498. Retrieved January 9, 2007, from http://www.nova.edu/ssss/QR/QR11-3/onwuegbuzie.pdf

Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 351-383). Thousand Oaks, CA: Sage.

Ory, J. C. (2000, Fall). Teaching evaluation: Past, present, and future. *New Directions for Teaching and Learning, 83,* 13–18.

Ory, J. C., Braskamp, L. A., & Pieper, D. M. (1980). The congruency of student evaluative information collected by three methods. *Journal of Educational Psychology, 72,* 181–185.

158

Ory, J. C., & Ryan, K. (2001, Spring). How do student ratings measure up to a new validity framework? *New Directions for Institutional Research, 109,* 27–44.

Overall, J. U., & Marsh, H. W. (1980). Students' evaluations of instruction: A longitudinal study of their stability. *Journal of Educational Psychology, 72,* 321–325.

Patton, M. Q. (1990). *Qualitative research and evaluation methods* (2nd ed.). Newbury Park, CA: Sage.

Peterson, K., & Kauchak, D. (1982). *Teacher evaluation: Perspectives, practices, and promises*. Salt Lake City: Utah University Center for Educational Practice.

Sandelowski, M., & Barroso, J. (2003). Creating metasummaries of qualitative findings. *Nursing Research, 52,* 226–233.

Schaeffer, G., Epting, K., Zinn, T., & Buskit, W. (2003). Student and faculty perceptions of effective teaching: A successful replication. *Teaching of Psychology, 30,* 133–136.

Schmelkin, L. P., Spencer, K. J., & Gellman, E. S. (1997). Faculty perspectives on course and teacher evaluations. *Research in Higher Education, 38,* 575–592.

Seldin, P. (1984). *Changing practices in faculty evaluation*. San Francisco: Jossey-Bass.

Seldin, P. (1993). The use and abuse of student ratings of professors. *Chronicle of Higher Education, 39*, A40.

Seldin, P. (1999). Current practices—good and bad—nationally. In P. Seldin (Ed.), *Current practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 1–24). Bolton, MA: Anker.

Sheehan, D. S. (1999). Student evaluation of university teaching. *Journal of Instructional Psychology, 26,* 188–193.

Siegel, S., & Castellan, J. N. (1988). *Nonparametric statistics for the behavioural sciences*. New York: McGraw-Hill.

Simmons, T. L. (1996). Student evaluation of teachers: Professional practice or punitive policy? *JALT Testing & Evaluation N-SIG Newsletter*, *1*(1), 12–16.

Spencer, K. J., & Schmelkin, L. P. (2002). Students' perspectives on teaching and its evaluation. *Assessment & Evaluation in Higher Education, 27,* 397–408.

Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics* (5th ed.). New York: Harper & Row.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches* (Applied Social Research Methods Series, Vol. 46). Thousand Oaks, CA: Sage.

Tashakkori, A., & Teddlie, C. (2003). The past and future of mixed methods research: From data triangulation to mixed model designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 671–701). Thousand Oaks, CA: Sage.

Tashakkori, A., & Teddlie, C. (2006, April). *Validity issues in mixed methods research: Calling for an integrative framework*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Theall, M., & Franklin, J. (2001, Spring). Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instruction? *New Directions for Institutional Research, 109,* 45–56.

Thompson, B. (1980, April). *Canonical correlation: Recent extensions for modelling educational processes*. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Thompson, B. (1984). *Canonical correlation analysis: Uses and interpretations*. Beverly Hills, CA: Sage. (ERIC Document Reproduction Service No. ED199269)

Thompson, B. (1988, April). *Canonical correlation analysis: An explanation with comments on correct practice*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED295957)

Thompson, B. (1990, April). *Variable importance in multiple regression and canonical correlation.* Paper presented at the annual meeting of the American Educational Research Association, Boston. (ERIC Document Reproduction Service No. ED317615)

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* Washington, DC: American Psychological Association.

Washburn, K., & Thornton, J. F. (Ed.). (1996). *Dumbing down: Essays on the strip mining of American culture.* New York: Norton.

Williams, W. M., & Ceci, S. J. (1997). How'm I doing? Problems with student ratings of instructors and courses. *Change, 29*(5), 13–23.

Witcher, A. E., Onwuegbuzie, A. J., & Minor, L. C. (2001). Characteristics of effective teachers: Perceptions of preservice teachers. *Research in the Schools, 8*(2), 45–57.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99,* 432–442.