What Matters for Excellence in PhD Programs?: Latent Constructs of Doctoral Program Quality Used by Early Career Social Scientists

Emory Morrison
Elizabeth Rudd
William Zumeta
Maresi Nerad

The Journal of Higher Education, Volume 82, Number 5, September/October 2011, pp. 535-563 (Article)

Published by The Ohio State University Press
DOI: 10.1353/jhe.2011.0029

Emory Morrison
Elizabeth Rudd
William Zumeta
Maresi Nerad

What Matters for Excellence in PhD Programs? Latent Constructs of Doctoral Program Quality Used by Early Career Social Scientists

What do recent doctoral graduates have in mind when they evaluate their program's quality? Do graduates consider the quality of *academic* or other job training, mentoring, and/or socialization when evaluating their programs? Do these doctorate-holders consider similar factors when making their assessments, or does each individual rely on their own idiosyncratic considerations? Do the types of considerations that inform evaluations vary across the many status characteristics that differentiate doctorate-holders (such as gender, field of study, and status of graduate training program)? In this paper, we address these questions. Understanding how doctoral degree recipients approach the evaluation of their programs is an important first step in integrating their evaluations (and thus their perspectives which we think are valuable) into the more general assessment of doctoral degree granting programs.

The authors are grateful for the helpful comments and reviews of this work provided by Frederick Thompson (Willamette University) and Rachelle Brooks (Northwestern University) as well as those from two anonymous reviewers. The reviewers bear no responsibility for the any errors or oversights that remain. We are also grateful for the technical assistance provided by Joseph Picciano. Our work has been partially sponsored by the Ford Foundation. A previous version of this paper was presented at the 2009 annual meeting of the Association for Studies in Higher Education.

Emory Morrison is an Assistant Professor of Sociology at Mississippi State University. Elizabeth Rudd is a research scientist at the University of Washington, Seattle. Maresi Nerad is an Associate Professor in the College of Education and Director of the Center for Innovation and Research in Graduate Education (CIRGE) at the University of Washington, Seattle. William Zumeta is a Professor, Daniel J. Evans School of Public Affairs and College of Education, University of Washington.

The Journal of Higher Education, Vol. 82, No. 5 (September/October 2011)

In our view, graduates who experienced the program relatively recently yet have worked long enough to reflect on how program elements affected their career competencies are a highly credible group of program raters.[1] They provide a view of PhD program quality rooted in their experiences of those programs as students and graduates. Their perspectives can help us identify dimensions of program quality not readily accessible through traditional reputational measures. This information may be useful to faculty and graduate deans interested in improving the educational experiences of current students and the career-relevance of PhD education. It may also identify dimensions of program quality that prospective students could use as a guide for assessing alternative graduate programs.

Traditionally, excellence in PhD programs has been seen as largely coincident with the scholarly reputation of program faculty (Brooks, 2005; Cartter, 1966). Yet, no matter how capable the faculty in research, to consistently produce well prepared graduates, it seems logical that a doctoral program also needs to be well designed, its requirements made clear to students, and its faculty attentive to teaching and students' needs. Such dimensions of quality may be assessed by surveying students or recent graduates about their doctoral program. We view quality in doctoral education then as multi-dimensional, including both the scholarly attributes of the faculty and also these dimensions related to educational effectiveness. In this paper, we explore the latter dimensions in some depth and also relate alumni perceptions of quality to the standard reputational measure of scholarly quality of a doctoral program. The heart of our analysis is our examination of the relationship between alumni raters' global assessment of the overall quality of their doctoral program and their assessments of quality on 26 specific dimensions (reduced to seven factors). We are then able to estimate models of how these alumni construct their assessments of program quality and relate these latent constructs to several demographic variables, discipline within the social sciences, and the scholarly reputation of their doctoral program's faculty.

The empirical basis for our analysis is a sample of more than 2,000 social science PhDs in six disciplines who graduated between July 1, 1995 and June 30, 1999 from whom survey data were collected in 2005 and 2006. Respondents assessed the overall quality of their PhD program and evaluated more than two-dozen specific quality-related dimensions (Nerad, Rudd, Morrison, & Picciano, 2007).

We first review studies of doctoral program quality, including findings on the relationship of faculty scholarly prestige to student or alumni evaluation of educational effectiveness. We also propose sev-

eral evaluator characteristics that might influence a respondent's latent construct of program quality. Then, we describe our dataset, a national survey called *Social Science PhDs–Five+ Years Out* (*SS5*). This is followed by details of our analysis and results. We conclude with implications for knowledge and policy.

Concepts of PhD Program Quality

The longest-standing and most widely used indicators of PhD program quality are reputational measures of the scholarly quality of a program's graduate faculty. The most comprehensive assessments—conducted by the National Research Council (NRC) in 1981, 1993, and 2007—use the approach pioneered by Alan Cartter's (1966) study (see Brooks 2005; Ostriker & Kuh 2003; Roose & Andersen 1970). Cartter staunchly defended the validity of reputational measures, arguing that quality is inherently a subjective concept and reputation would be closely correlated with more "objective" measures. Sponsored by the American Council on Education, Cartter's study included 106 universities and sampled 4,000 faculty, who ranked PhD-granting programs on the "quality of the graduate faculty" and separately rated the program's "educational effectiveness." In practice these measures, with correlation coefficients above 0.9, are redundant (Clark, Hartnett, & Baird, 1976; Jones, Lindzey, & Coggeshall, 1982; Roose & Anderson, 1970).

While recognizing the utility of more direct indicators of educational effectiveness, the National Research Council's assessments of PhD programs relied on faculty surveys and produced reputational measures of quality. Surveying students and alumni was deemed infeasible given the scale of the studies (Ostriker & Kuh, 2003). The 1981 assessment of research-doctorate programs encompassed 32 disciplines, more than 200 institutions, 2,699 programs, and surveyed more than 5,000 graduate faculty. In 1993, the NRC covered 3,600 programs in 270 institutions and 41 fields and surveyed more than 8,000 graduate faculty (Goldberger, Maher, & Flattau, 1995). The most recent study involved 226 institutions, more than 5,000 programs, and 76 fields. This assessment surveyed students in selected fields and gathered data about various dimensions of programs, but its main measure of overall quality was still derived (in a complex way) from a faculty survey that in the end weights faculty research activity highly.[2] The final report on this study is expected to appear by the end of 2009.[3]

Existing evidence suggests that, unlike faculty rater views, student and alumni views of educational effectiveness rely on program dimensions that are not necessarily correlated with faculty scholarly repu-

tation. *PhDs—Ten Years Later*, a national survey in 1996–97 of PhD cohorts of 1982–1985, found that graduates' assessment of program quality was not correlated with the NRC's 1982 measures of faculty reputation (Nerad & Cerny, 1999a). In a sample of students in chemistry, psychology, and history from 25 institutions, Clark et al. (1976) found that student ratings of faculty teaching effectiveness were not closely correlated with faculty research productivity (chapter 5). In a sample of recent social science PhDs, graduates of more prestigious programs were not more likely to be satisfied with their training during graduate school in writing and publishing (CIRGE, n.d.).

The National Association of Graduate-Professional Students (NAGPS) surveyed more than 32,000 current PhD students and alumni in 2000 about elements of their PhD program such as training for teaching assistants and time to degree. Results show that reputational measures of scholarly quality may diverge from student and alumni perceptions of educational effectiveness.[4]

If students and alumni evaluate educational effectiveness as a dimension distinct from faculty reputation for scholarly quality, then what elements of their PhD programs are important to them when evaluating program quality? Further, how might evaluator characteristics influence the relative importance of various elements?

Studies of job requirements and graduate student experiences suggest that several evaluator characteristics could be associated with needing or wanting different kinds of things from a PhD program, which, in turn, might influence criteria used to evaluate program quality. We look specifically at career goals, current job sector, age at PhD, gender, prestige of the PhD-granting program, and PhD discipline.[5]

Career goals at the time of degree and current employment sector might influence concepts of program quality because different kinds of jobs require different kinds of skills. For instance, compared to faculty, PhD social scientists working in business, government, or non-profit sectors more often reported that teamwork, communication, and management skills were "very important" in their jobs (Nerad, Rudd, Morrison, & Picciano, 2007).

Younger PhD students might have different priorities than older students, for whom mentoring and guidance might be less important but career preparation and support in the job search more urgent. Since women often experience PhD programs less positively than men do (Fox, 2001; Nerad & Cerny, 1999b; Nerad & Stewart, 1991), possibly women are less likely than men to feel that they really "belong" in doctoral programs and PhD-level careers. If so, women might place a higher value on efforts to explicitly socialize students. In the social sci-

ences today, however, men and women are about equally likely to complete PhD programs (Sowell, 2008) and several disciplines award more PhDs annually to women than to men.

The academic prestige of one's PhD-granting department might be associated with notions of PhD program quality because graduates of higher-ranking programs more strongly internalize the abstract academic values that legitimate prestige rankings (Morrison, Rudd, Picciano, & Nerad, 2010). Graduates of lower-ranking programs might be influenced more strongly by other elements of the PhD program—for instance, support and guidance in meeting program requirements—in evaluating overall quality.

Finally, disciplinary differences might exist. Disciplines such as communication and geography with a tradition of applied work might value skills needed outside academia more highly. Holders of doctorates in fields such as history and anthropology in which students are mostly academically oriented might hold abstract academic values more strongly but also place a higher value on career preparation and help with the job search because of perennially weak academic job markets.

To answer the research questions posed at the beginning of this paper, we undertake an inductive, exploratory analysis relating respondent ratings of "overall program quality" to patterns of their responses on 26 other items evaluating the quality of separate elements of the respondent's PhD program, guided by the following specific questions:

1. What do recent doctoral graduates have in mind when they evaluate their program's quality? In terms of our analysis, how do domain specific assessments influence global assessments of overall program quality?
2. Are career goals, employment sector, age at PhD, gender, prestige of PhD-granting department and discipline associated with different sets of domain specific criteria—which we will call "latent constructs"—for the global evaluation of overall PhD program quality by respondents?

Data and Methods

Data

Social Science PhDs—Five+ Years Out (SS5) was a national survey of recent recipients of doctoral degrees from U.S. programs in anthropology, communication, geography, history, political science, and sociology conducted by the Center for Innovation and Research in Graduate Education (CIRGE) at the University of Washington. Respondents, who

earned their PhD between July 1995 and June 1999, provided information in 2005–2006 on career and family spanning the time from starting graduate school to 6–10 years post-PhD and assessed their graduate school experiences—including answering questions about the perceived quality of various elements of their PhD program, training in specific skills, and mentoring by their dissertation advisor.

Sixty-five U.S. institutions—diverse with respect to geography, control, and NRC scholarly prestige rating—were selected to participate in the SS5. CIRGE located reliable contact information for 6,670 doctorate-holders from these universities, 45% of whom responded. The National Opinion Research Council (NORC) conducted a non-response analysis to determine if the non-respondents differed from the respondents in any important way. Differences on the Survey of Earned Doctorates[6] between those who responded to the SS5 and those who did not respond to the SS5 were not noteworthy, with one exception. Eighty-three percent of the SS5 respondents indicated having definite academic plans at the time of the PhD award, whereas 78% of the non-respondents reported having definite academic plans.

Of the 3,025 SS5 respondents, 2,192 completed the graduate school evaluation inventory we employ and were thus available to be analyzed. These respondents accounted for 15% of all U.S. PhDs awarded in the study fields and time frame. The 833 SS5 respondents who were not part of the sample for the following analyses were dropped either because they completed a different version of the survey or because they did not finish the survey (503 respondents) which often required an investment of an hour. Those individuals who were retained from the SS5 for our analysis were more likely to be male (52.3% vs. 47.7%), to be under 30 years of age (21.8% vs. 16.0%) at the time of PhD award, and to be a ladder faculty member at the time of the survey (64.1% vs. 60.8%) than those who were dropped. Those dropped from the sample and those retained were similarly distributed across doctoral fields, across the reputational rankings of their PhD programs, and across their career goals while in graduate school. The non-response analysis was done for both eligible non-respondents (graduates of the 65 participating institutions) and the population of PhD graduates (n = 15,677) in sampled years from sampled disciplines with essentially the same results.

Survey Items Used in the Analysis

Analyses in this paper use 27 items from three questionnaire inventories that elicited assessments of particular aspects of respondents' doctoral education. These items and exact wording of the inventories are listed in Table 1. The third column of Table 1 indicates the inventory

in which each item is embedded. Respondents rated 10 program elements and "overall program quality" as "excellent," "adequate," "poor," or "N/A." They rated their satisfaction with the quality of mentoring by their dissertation advisor in six areas using the scale "very satisfied," "somewhat satisfied," "somewhat unsatisfied," and "very unsatisfied." Respondents also rated the quality of training (formal or informal) they received in particular competencies during their PhD studies.

The first column of Table 1 displays the percentage of respondents who rated the attribute as "excellent" (or "very satisfied" for items about mentoring) in their doctoral program. We focus on ratings at the highest end of the scale because it is widely believed that excellence, rather than adequacy or something similar, should be the *sine qua non* of PhD programs and because SS5 responses tended to be skewed upward.[7] In aggregate, respondents rated their PhD programs most favorably on the analytical items: "thinking critically" (77% "excellent"), academic rigor" (65% "excellent"), and "analyzing and synthesizing data" (60%"excellent"). The items least likely to be rated excellent were "managing people and budgets," "working collaboratively," "writing proposals for funding," and "non-academic career preparation." On four of six of the mentoring items, about half the respondents were "very satisfied," but only one quarter were very satisfied with "help with publishing." Nearly half—48%—rated "overall program quality" as "excellent" (top row of data).

Additional variables used here include reputation for scholarly quality of the PhD-granting program, respondent's current employment sector, career goal at PhD, age in years at PhD, gender, and PhD discipline. Scholarly quality rankings used are from the NRC's 1993 assessment of research doctorate programs (Goldberger et al., 1995) for all disciplines except communication (which was not included in the NRC study). Programs with faculty that received an average rating above 4.0 in the NRC study—on a one to five point scale—were categorized as having an "excellent" scholarly reputation. For PhD programs in communication, data from the 1996 Speech Communication Association (currently the National Communication Association) study of doctorate program reputation was used to assess scholarly reputation (National Communication Association, 1996). Overall, just under one third (32.8%) of our respondents had graduated from one of these elite programs (second row of data in Table 1).

Employment sector was categorized as: ladder faculty (which includes tenured or tenure-track faculty), non-ladder faculty, non-faculty academic position, and business/government/non-profit sector. Career goals included: professorship, non-faculty work in the academic sector

TABLE 1

Evaluations of Program Quality and Reputational Rank—Percent Excellent and Concordant with Rating of Overall Program Quality

	% Excellent	Concordance	Item List[a]
Overall program quality	48.1%		1
Reputational rank	32.8%	56.8%	N/A
Academic rigor	65.6%	75.4%	1
Academic career preparation	31.0%	73.8%	1
Clear program requirements	53.1%	73.0%	1
Support and guidance during dissertation	41.8%	72.2%	1
Feedback on student progress	30.9%	71.4%	1
Socializing students into an academic community	33.0%	71.0%	1
Preparation for qualifying exam	36.0%	70.6%	1
Presentation skills	33.7%	69.2%	3
Managing people and budgets	3.1%	66.7%	3
Working with people from diverse educational and social backgrounds	25.9%	66.7%	3
Working in interdisciplinary context	31.2%	66.0%	3
Writing and publishing reports and articles	29.0%	65.9%	2
Quality of advice from your dissertation chair in developing your dissertation topic	53.9%	65.8%	2
Your dissertation chair's support of your career decisions	49.9%	65.7%	3
Writing proposals for funding	14.6%	65.2%	2
Financial support	33.4%	64.6%	1
Quality of guidance from your dissertation chair in completing your PhD	54.1%	64.4%	3
Quality of help from your dissertation chair in publishing	26.2%	63.9%	3
Working collaboratively, in a team	14.3%	61.4%	2
Thinking critically	77.0%	60.9%	1
Having a diverse student population	25.6%	60.6%	1
Research design (experiments, surveys, etc.)	35.3%	60.5%	2
Analyzing or synthesizing data	60.6%	58.3%	2
Your dissertation chair's support of your job search	41.3%	57.4%	3
Non-academic career preparation	3.7%	55.5%	1
Overall quality of mentoring you received from your dissertation chair	47.1%	53.0%	3

[a] Item lists 1 and 2 on a scale of "Excellent, Adequate, Poor, N/A"; list 3 scale is "Very satisfied, Somewhat satisfied, Somewhat unsatisfied, Very unsatisfied." Exact question wording: (1) *How would you evaluate your doctoral program on each of the following? Please evaluate your perception of your program for the years that you were in the program.* (2) *For each activity, please indicate the quality of your doctoral training (either formal or informal):* (3) *As you look back on your doctoral studies, to what extent were you satisfied with the following types of support from your dissertation chair or advisor?*

Source. CIRGE, *Social Science PhDs—Five+ Years Out.*

(e.g., academic administration, researcher), non-academic career of any kind, and "other" (e.g., "start my own business" and "no formulated goal"). Age in years at PhD award is divided into three categories: younger than 30; between 30 and 40; and age 40 or older. PhD discipline is based on self-report and was confirmed by comparisons with data from the *Survey of Earned Doctorates* done by the National Opinion Research Center for *Social Science PhDs—Five+ Years Out* (Picciano et al., 2007).

Methods

We investigate whether value sets that inform graduates' assessments of "overall program quality" vary systematically across characteristics of the evaluator. For example, are some items—e.g., perception of excellence in socialization into the academic community—predictive of perception of overall program quality for some respondents, but not for others? If so, are those individuals for whom socialization matters different from others with respect to some characteristic—e.g., gender or prestige of the graduate training institution?

In order to address these types of questions, we first reduced the 26 quality-related attributes into a manageable set of dimensions using first exploratory and then confirmatory factor analysis. The exploratory factor analysis (EFA) suggested that a model with seven factors fits the data best.[8] We did not, however, rely on the EFA to specify our factors because the EFA is designed to arrive at factors that are not correlated with one another. We, on the other hand, assume that the latent factors driving the correlation structure of our observed variables ought to be correlated. Both idiosyncratic characteristics of each respondent (such as tendency to rely on certain portions of the Likert scale, even personality, mood, etc.) and his or her global perception of the overall quality of their graduate training program ought to influence such factors in similar directions. We therefore built a confirmatory factor model (CFA) that incorporated these assumptions. For the sake of parsimony and ease of interpretation, we constructed our seven factor CFA by specifying that items load onto one and only one latent dimension, while each latent dimension correlates freely with the other six latent dimensions. We built our CFA in an iterative process evaluating improvement to model fit through a series of small adjustments to how items are specified to load onto factors.[9]

We next applied latent class regression analysis (LCRA) to identify how different patterns of scores on these dimensions (factors) differentially predict respondent-perceived excellence in overall program quality. In other words, we seek to identify whether different "quality

constructs" (i.e., latent classes) are evident in the respondents' thinking. The LCRA also allows us to identify whether the different quality constructs are associated with particular groups—groups defined by scholarly quality of PhD-granting program, PhD discipline, gender, etc.

Latent class regression analysis (LCRA) is an extension of latent class analysis (LCA), a method of assigning cases to different classes that enables application of statistical tests to determine the number and definition of classes that best fit the data.[10] LCA assigns cases to different classes as a function of the distribution of scores on a specified set of observed variables. In LCA, the observed, endogenous variables are assumed to be related to one another as a function of being co-determined by an unobserved latent variable (the latent class). With LCA we can use formal hypothesis testing to determine how many classes best fit the data and to estimate the contribution of a particular observed endogenous variable—here the quality attribute dimensions derived from the factor analysis—to class assignment. In addition, LCA allows the researcher to test whether exogenous covariates (here discipline, gender, etc.) predict membership in a given class. LCRA extends LCA by estimating latent classes as a function of how the specified set of observed endogenous variables (i.e., the factors) predicts a single dependent variable (the respondent's assessment of overall program quality).[11]

To identify the number of latent classes that best fits the data, several models are estimated and compared to each other using maximum likelihood estimation. Analysis begins with the simplest, one-class model. Subsequent models vary the allowed number of classes and whether or not the parameter estimates are permitted to be different across classes[12] in order to identify the model with the best fit. Using the Bayesian Information Criteria (BIC), the best fitting model can then be selected for interpretation (Raftery, 1995).[13]

Findings

First, we find that a program's reputation for scholarly quality clearly does not predict well respondents' evaluation of their own PhD program as "excellent" (Table 1, second row of data at the top). Each respondent's evaluation of each item shown is either concordant or discordant with their evaluation of "overall quality" (column 2). Items are concordant with overall quality when a respondent evaluates both as excellent or evaluates neither as excellent. The second row of data shows that about one-third of respondents come from programs with excellent scholarly reputations according to the National Research Council ratings. The concordance measure of just 56.8% on this item indicates

that respondents from programs with an excellent reputation rate their doctoral program's overall quality as excellent only slightly more often than respondents from programs not regarded as excellent in scholarly reputation. Many of the items from the three survey inventories tapping into different domains of graduate program quality are much more strongly related to respondent-perceived overall program quality. "Academic rigor," "Academic career preparation," and "Clear program requirements" are items that are most strongly related to perceived "Overall program quality." On the other hand, "Analyzing and synthesizing data," Mentoring with respect to "job search," "Non-academic career preparation," and "Overall quality of mentoring" are the items that are least associated with perceived "Overall program quality."

Data Reduction

Exploratory and confirmatory discrete factor analyses reduced the 26 survey items related to quality to seven underlying dimensions. Table 2 displays the pattern of associations that emerged. We term the seven factors *mentoring*, *abstract academic qualities*, *research skills*, *support with PhD program requirements*, *non-academic career, diversity,* and *belonging.* Correlations among the seven latent dimensions are positive and tend to be strong. Additionally, all seven dimensions are positively correlated with all 26 items from the questionnaire. Thus, observing any attribute rated as excellent increases the odds that any other attribute will be rated excellent.

Latent Class Regression Analysis

The dependent variable in this analysis is the respondent's rating of overall program quality, which is observed as a dichotomy: excellent or not excellent. The independent variables are the seven dimensions of program quality, i.e., the factors (see Table 2). Respondents' scores on each factor were calculated according to the confirmatory discrete factor analysis model described above. Covariates that are hypothesized to predict assignment to latent classes are respondent characteristics, including: gender, scholarly prestige of PhD program, job sector at the time of the survey, career goal at PhD, age at PhD, and PhD discipline.

We began with a baseline model, the most parsimonious LCRA specification possible with the set of variables described above—the single class model. This baseline model is compared to models with two or more latent classes with all coefficients of the effects on overall program quality free to vary across classes,[14] i.e., these are multi-class unconstrained models in this exploratory phase. Coefficients that are not significantly different across classes in the unconstrained models are

TABLE 2

Confirmatory Discrete Factor Analysis: Inter-factor Correlations, Component Loadings, and Correlations between Factors and Non-Component Items

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Latent Factors							
(1) Mentoring							
(2) Abstract academic	0.33						
(3) Research skills	0.41	0.65					
(4) PhD program requirements	0.57	0.60	0.58				
(5) Non-academic career	0.37	0.46	0.70	0.53			
(6) Diversity	0.23	0.33	0.27	0.33	0.54		
(7) Belonging	0.32	0.49	0.61	0.57	0.55	0.30	
(1) Mentoring							
Mentoring overall	***0.90***	0.30	0.37	0.51	0.33	0.20	0.28
Mentoring completing PhD	***0.75***	0.25	0.31	0.43	0.27	0.17	0.24
Mentoring dissertation topic	***0.68***	0.22	0.28	0.39	0.25	0.15	0.21
Mentoring career decisions	***0.64***	0.21	0.26	0.36	0.23	0.15	0.20
Mentoring job search	***0.60***	0.20	0.24	0.34	0.22	0.14	0.19
Mentoring publishing	***0.49***	0.16	0.20	0.28	0.18	0.11	0.16
(2) Abstract Academic Qualities							
Academic rigor	0.25	***0.75***	0.49	0.45	0.34	0.25	0.37
Thinking critically	0.20	***0.60***	0.39	0.36	0.27	0.20	0.29
(3) Research Skills							
Academic career preparation	0.24	0.37	***0.57***	0.33	0.40	0.15	0.35
Analyzing or synthesizing data	0.21	0.33	***0.50***	0.29	0.35	0.13	0.31
Writing and publishing	0.20	0.32	***0.49***	0.28	0.35	0.13	0.30
Research design	0.20	0.31	***0.48***	0.27	0.33	0.13	0.29
(4) Support for PhD Program Requirements							
Support and guidance during dissertation	0.36	0.38	0.37	***0.64***	0.34	0.21	0.36
Feedback on student progress	0.33	0.35	0.34	***0.59***	0.31	0.19	0.34
Preparation for qualifying exam	0.29	0.30	0.29	***0.51***	0.27	0.17	0.29
Clear program requirements	0.28	0.29	0.28	***0.49***	0.26	0.16	0.28
(5) Non-Academic Career Skills							
Presentation skills	0.21	0.26	0.40	0.30	***0.57***	0.31	0.31
Working collaboratively	0.16	0.19	0.30	0.22	***0.43***	0.23	0.23
Writing proposals for funding	0.12	0.15	0.23	0.18	***0.33***	0.18	0.18
Non-academic career prep.	0.08	0.10	0.16	0.12	***0.23***	0.12	0.13
Managing people and budgets	0.08	0.10	0.15	0.11	***0.21***	0.12	0.12

TABLE 2 (*Continued*)

Confirmatory Discrete Factor Analysis: Inter-factor Correlations, Component Loadings, and Correlations between Factors and Non-Component Items

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(6) Diversity							
Working within diversity	0.17	0.25	0.20	0.25	0.41	***0.76***	0.23
Working in interdisciplinary context	0.13	0.19	0.15	0.19	0.31	***0.58***	0.17
Having a diverse student pop.	0.09	0.12	0.10	0.12	0.20	***0.38***	0.11
(7) Belonging							
Socializing into an academic community	0.23	0.35	0.44	0.41	0.40	0.21	***0.72***
Financial support	0.11	0.16	0.21	0.19	0.18	0.10	***0.34***

then constrained to be equal across classes in the multi-class models in order to improve parsimony. A comparison of the BIC statistics across all models enables us to adjudicate the model that best balances parsimony with goodness-of-fit to the data (Raftery, 1995).

The one-class model explains 55% of the variation in respondent-perceived overall program quality. The BIC score of 1636.13 provides a baseline model fit score against which alternative models may be compared to evaluate model preference. The addition of a second latent class increases the explained variance by 27 percentage points but also requires 26 additional parameters to be estimated. A three-class model adds another 26 parameters, and explains no additional variation over the two-class model, and a four-class model performed even worse.

The rise in the BIC statistic associated with the two-class model suggests that the increase in explained variance is not a good trade-off relative to the reduction in parsimony when comparing the one-class and the two-class models. This led us to consider specifications of the two-class model in which we constrain parameters across classes to be equal and thus improve parsimony while potentially maintaining levels of explained variance. The coefficients from Model 2 (as seen in Table 4) suggest that the specification of the effect of abstract academic should not be constrained to be equal across classes. Table 3 summarizes the model comparisons among 27 different specifications of the LCRA that we used to identify the number of latent classes and mix of constrained and unconstrained coefficients that best fit the data. The specification of each model is described in Table 3 under the column heading "Notes."

TABLE 3

Model Fit Comparisons for Latent Class Regression Analysis: Perceptions of Excellence in Overall Program Quality Regressed on Seven Dimensions of Program Quality with Respondent Characteristics Predicting Latent Class Membership

Model	Classes	Parameters	L^2	R^2	BIC[a]	Notes[b]
1	1	8	1574.61	0.55	1636.13	No Constraints – One-Class Baseline Model
2	2	34	1457.27	0.82	1718.71	No Constraints – Two-Class Model
3	3	60	1382.11	0.82	1843.47	No Constraints – Three-Class Model
4	4	86	1347.79	0.81	2009.08	No Constraints – Four-Class Model
5	2	28	1463.55	0.81	1678.85	Model 2 + Equity Constraints on All IVs But Abstract Academic Quality
6	2	29	1462.21	0.81	1685.21	Model 5 + Free Mentoring
7	2	29	1463.56	0.81	1686.56	Model 5 + Free Research Skills
8	2	29	1462.30	0.81	1685.30	Model 5 + Free Support in Meeting Ph.D. Program Requirements
9	2	29	1463.00	0.81	1685.99	Model 5 + Free Non-Academic Career Skills
10	2	29	1463.50	0.81	1686.50	Model 5 + Free Diversity
11	2	29	1463.55	0.81	1686.54	Model 5 + Free Belonging
12	3	48	1435.39	0.82	1804.58	Model 5 + Additional Latent Class
13	3	54	1421.78	0.79	1837.00	Model 12 + Free Parameters For Additional Latent Class
14	2	13	1470.44	0.81	1579.81	Model 5 + All Effects of Predictors of Latent Class—Except Doctoral Program Rank—Constrained to Zero
15	2	14	1472.47	0.81	1585.66	Model 14 + Free Gender
16	2	15	1474.59	0.81	1593.80	Model 14 + Free Age
17	2	16	1474.81	0.81	1600.66	Model 14 + Free Career Goal at Time of Doctorate
18	2	17	1471.22	0.81	1604.77	Model 14 + Free Job at Time of Survey
19	2	18	1471.05	0.81	1612.29	Model 14 + Free Doctoral Field

TABLE 3 (*Continued*)

Model Fit Comparisons for Latent Class Regression Analysis: Perceptions of Excellence in Overall Program Quality Regressed on Seven Dimensions of Program Quality with Respondent Characteristics Predicting Latent Class Membership

Model	Classes	Parameters	L^2	R^2	BIC[a]	Notes[b]
20	2	14	1471.57	0.81	1582.03	Model 18 + Job at Time of Survey Respecified as Binary: faculty/non-faculty
21	2	15	1470.23	0.81	1588.34	Model 20 + Free Gender
22	2	15	1474.18	0.81	1589.57	Model 20 + Free Age Respecified as Binary: young/not-young
24	2	19	1465.90	0.81	1614.83	Model 20 + Free Doctoral Field
26	3	20	1458.82	0.84	1615.45	Model 20 + Additional Latent Class
27	3	26	1486.77	0.84	1689.55	Model 26 + Free Parameters for Additional Latent Class

*$p < 0.05$ and **$p < 0.01$

[a] Smaller BIC statistics indicate models that better balance fit to data and parsimony.

[b] In a model with "No Constraints" all parameters are said to be "free" that is to be computed through maximum likelihood estimation. A parameter is "constrained" when set to a specific value prior to model estimation. "Equity constraints" on coefficients for independent variables (IVs) indicate that across latent classes coefficients are set to be equal (i.e., class does not moderate the effect). "Constrained to zero" indicate that the coefficient is set to equal zero, i.e., no effect.

Each of the models is nested and most, as shown in the table, involve a minor modification of a previously estimated model.

The model that best balances fit and parsimony according to the BIC statistic is Model 14, a two-class model in which the effects of all independent variables—except the effects of abstract academic qualities—are constrained to be equal across classes and the effects of all covariates on class membership—except the scholarly prestige of the respondent's PhD program—are constrained to equal zero. The BIC statistic for Model 14 is 1579.81, lower than for the other 26 models. Compared with Model 20, however, the difference in the BIC statistics is only 2.2—just slightly over the threshold of 2.0 that suggests "weak" empirical support in favor of a particular model in such a comparison[15] (Raftery, 1995, p. 141). Since the comparison of BIC statistics between Model 14 and 20 is not conclusive, and since Model 20 includes an additional estimated parameter of theoretical interest that obtains significance (and no additional parameters that fail to obtain significance), we select Model 20 as our preferred model and base our interpretation on this specification. Substantively, all findings in Model 14 and Model 20 are identical, except for the estimation of the effect of being in a ladder faculty position, the one parameter that is freed in Model 20.

Findings from Model 20 are presented in Tables 4 and 5. Table 4 presents the estimated effects of each of the seven dimensions of program quality (i.e., the factors) on perceptions of overall program quality for both of the latent classes in Model 20 (far right-hand panel). Table 5 presents the estimated effects of individual level characteristics on latent class assignment. Tables 4 and 5 also present the same information for Model 2, Model 5, and Model 14 since these three models were integral steps in the path to discovery of Model 20, and the estimated coefficients from these models illuminate why Model 20 is the preferred model.[16]

According to Model 20, there are clearly two different classes of respondents who utilize different value sets or quality constructs in developing their perceptions of overall program excellence. The model estimates that 54% of social scientists belong to Class 1 and that the defining element for this group is to very strongly rely on their assessment of abstract academic qualities (academic rigor and critical thinking) when assessing overall quality of their graduate program (Table 4). For these doctorate-holders, other dimensions of graduate program quality also inform their assessment but not nearly as strongly, therefore we refer to this as the "academics dominant" perspective on PhD program quality. The remaining 46% of social science PhD holders (Class 2) seem to draw on a multi-dimensional conceptual scheme in assessing quality.

TABLE 4

Estimated Effects of Perception of Dimensions of Program Quality on Perceptions of Overall Program Excellence[a, b,c]

	Model 2		Model 5		Model 14		Model 20	
	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2
Mentoring	0.62	-0.27	0.34	0.34	0.41	0.41	0.37	0.37
	(1.86)	*(0.51)*	*(1.25)*	*(1.25)*	*(1.49)*	*(1.49)*	*(1.33)*	*(1.33)*
	1.98							
Abstract academic	41.82**	2.12**	43.34**	2.37**	48.26**	2.31**	42.79**	2.34**
	(3.08)	*(3.78)*	*(2.83)*	*(5.11)*	*(2.71)*	*(5.12)*	*(2.81)*	*(5.04)*
	8.55**		**7.16****		**6.70****		**7.09****	
Research skills	0.75	2.58*	1.36*	1.36*	1.43*	1.43*	1.46*	1.46*
	(0.94)	*(2.21)*	*(2.24)*	*(2.24)*	*(2.36)*	*(2.36)*	*(2.39)*	*(2.39)*
	1.63							
PhD program req.	3.19**	9.49**	4.22**	4.22**	4.23**	4.23**	4.43**	4.43**
	(3.75)	*(2.80)*	*(4.38)*	*(4.38)*	*(4.39)*	*(4.39)*	*(4.28)*	*(4.28)*
	3.08							
Non-academic career	-1.12	-3.74*	-1.70**	-1.70**	-1.72**	-1.72**	-1.78**	-1.78**
	(1.58)	*(2.53)*	*(2.78)*	*(2.78)*	*(2.82)*	*(2.82)*	*(2.89)*	*(2.89)*
	2.47							
Diversity	0.77	1.50*	1.02**	1.02**	1.02**	1.02**	1.04**	1.04**
	(1.73)	*(2.14)*	*(2.82)*	*(2.82)*	*(2.84)*	*(2.84)*	*(2.88)*	*(2.88)*
	0.76							
Belonging	2.79**	2.00	2.83**	2.83**	2.81**	2.81**	2.77**	2.77**
	(5.89)	*(1.27)*	*(6.25)*	*(6.25)*	*(6.22)*	*(6.22)*	*(6.14)*	*(6.14)*
	0.23							

TABLE 4 (*Continued*)

Estimated Effects of Perception of Dimensions of Program Quality on Perceptions of Overall Program Excellence[a, b,c]

	Model 2		Model 5		Model 14		Model 20	
	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2
% Overall program quality "excellent"	33.9%	69.4%	32.6%	66.9%	32.8%	65.4%	32.5%	66.9%
Intercept	-1.92**	-0.36	-2.48**	-0.31	-2.53**	-0.38*	-2.59**	-0.32
	(3.70)	*(1.51)*	*(4.57)*	*(1.57)*	*(4.59)*	*(2.03)*	*(4.48)*	*(1.64)*
	8.21**		**19.85****		**18.50****		**19.14****	
R^2 individual class	71.8	72.6	71.9	69.6	72.2	70.3	72.0	70.1
R^2 model with 2 classes	0.82		0.81		0.81		0.81	
L^2	1457.27		1463.55		1470.44		1471.57	
BIC	1718.71		1678.85		1579.81		1582.03	

[a] Four two-class LCRA specifications predicting Overall Program Quality "Excellent" based on perceived quality of program dimensions and predicting class assignment based on evaluator characteristics.

[b] *T*-test statistics are in parentheses.

[c] Wald statistic computed for assumption that the coefficient is equal in both classes presented in bold and underlined.

*$p < 0.05$ ** $p < 0.01$

TABLE 5
Estimated Effects of Evaluator Characteristics on Assignment to Class 1 (Academics Dominant)[a, b, c]

	Model 2		Model 5		Model 14		Model 20	
Gender	**0.98**		**1.38**					
Female	-0.05	(0.99)	-0.06	(1.17)				
Faculty Scholarly Reputation	**31.42****		**31.22****		**31.39****		**29.80****	
High	0.38**	(4.71)	0.40**	(4.94)	0.36**	(4.74)	0.36**	(4.68)
Medium	-0.01	(0.16)	-0.04	(0.54)	0.00	(0.19)	-0.01	(0.19)
Low	-0.36**	(5.20)	-0.36**	(4.98)	-0.36**	(5.20)	-0.35**	(5.03)
Most Recent Job	**5.13**		**3.98**				**5.95***	
Ladder faculty	-0.19	(1.86)	-0.17	(1.60)			-0.13*	(2.44)
Non-tenure-track faculty	0.00	(0.00)	-0.03	(0.16)				
Academic other	0.07	(0.43)	0.06	(0.38)				
BGN[d]	0.14	(1.01)	0.12	(0.89)				
Not specified	-0.02	(0.07)	0.01	(0.03)				
Career Goal	**0.18**		**0.40**					
Professor	-0.04	(0.41)	-0.06	(0.62)				
Academic research/other	0.00	(0.03)	0.00	(0.01)				
BGN	0.05	(0.27)	0.05	(0.28)				
Other	-0.01	(0.07)	0.01	(0.10)				
Age at PhD	**1.61**		**2.60**					
Under 30	0.03	(0.33)	0.04	(0.41)				
30–40	0.07	(1.10)	0.10	(1.42)				
Over 40	-0.10	(1.10)	-0.13	(1.37)				

TABLE 5 (*Continued*)

Estimated Effects of Evaluator Characteristics on Assignment to Class 1 (Academics Dominant)[a, b, c]

	Model 2		Model 5		Model 14	Model 20
PhD Field	**6.50**		**6.32**			
Anthropology	0.10	(0.72)	0.09	(0.66)		
Communication	-0.11	(0.85)	-0.08	(0.64)		
Geography	0.23	(1.02)	0.21	(0.93)		
History	0.09	(0.93)	0.10	(1.03)		
Political science	-0.11	(1.07)	-0.10	(0.96)		
Sociology	-0.20	(1.78)	-0.22	(1.84)		
L^2	1457.27		1463.55		1470.44	1471.57
BIC	1718.71		1678.85		1579.81	1582.03

[a] Four two-class LCRA specifications predicting Overall Program Quality "Excellent" based on perceived quality of program dimensions and predicting class assignment based on evaluator characteristics. [b] *T*-test statistics are in parentheses.

[c] Wald statistics are in bold and underlined (computed assuming the category has no effect on class assignment).

[d] "BGN" stands for business, government, or non-profit sector.

*$p < 0.05$ **$p < 0.01$

For Class 2, abstract academic qualities is just one of several dimensions that more equally contribute to the assessment of overall program quality. For Class 2 respondents, support in meeting graduate program requirements and provision of indications of belonging in the academic community (the factor including socialization and financial support)[17] are both stronger predictors of the assessment of overall program quality than is assessment of abstract academic qualities. Net of other dimensions, mentoring by the dissertation adviser does not seem to matter for either group's assessment of overall program quality. For both classes, the net effect of the positive assessment of non-academic skill training is negative on one's assessment of overall program quality. To sum up, the substantive distinction between the two groups lies in the importance of the assessment of abstract academic qualities in determining the evaluation of overall program quality.

Table 5 shows that certain respondent characteristics help to predict whether a graduate is likely to have an academics-dominant or a multidimensional perspective on overall program quality. As shown in Model 20, those from elite PhD programs in terms of scholarly reputation have a much higher likelihood of relying dominantly on their assessment of abstract academic qualities (academic rigor and critical thinking skills). One's job at the time of the survey also has an effect, albeit a much smaller one, on which value set is employed in evaluating overall program quality. However, the effect is in the opposite direction from that expected. Ladder faculty are more likely to use a multi-dimensional framework in evaluating overall program quality. Those not in ladder faculty positions are more likely to evaluate overall program quality based dominantly on the assessed abstract academic quality of their graduate program. This counter-intuitive finding might hide differences among faculty. The current study does not differentiate between types of ladder faculty positions (B. Clark, 1987). We do not know how the finding of differences between ladder faculty and others would be influenced by controlling for sector of employment—institutional type, stature, etc.—within the academic labor market.

The data provide no evidence that respondents differ in the way they assess program quality by gender, by age at PhD, by original career goal, or by social science discipline of PhD. In sum, the LCRA finds a great deal of homogeneity in the value sets that underlie respondents' evaluations of overall program excellence. Two such value sets are sufficient to explain 82 percent of the variation in perceptions of overall program excellence. The value sets differ in the dominance of the effect of perceptions of excellence in abstract academic qualities (academic rigor and training in critical thinking) in accounting for perceptions of

excellence in overall program quality. For just over half of these social science doctorate holders, the perception of excellence in abstract academic qualities dominantly drives the perception of overall program excellence. For the other nearly half, perceptions of excellence in abstract academic qualities are roughly as important as other factors. Those trained at elite doctoral institutions are much more likely to be in the first group. Those trained at less prestigious institutions are more likely to have a multi-dimensional approach to evaluating overall program quality.

Conclusions and Implications

These findings clearly indicate that alumni assessments of program quality tap a different dimension than peer (faculty) ratings of the scholarly quality of doctoral programs. Because recent alumni evaluate their education in light of career experiences, they offer important and underutilized perspectives on PhD program quality. In contrast to peer raters of faculty scholarly reputation, alumni have recent, direct experience of their PhD program and can evaluate how well it prepared them for their careers. To investigate this important dimension of PhD program quality in the social sciences this research used alumni perceptions of the quality of 26 separate elements of their own PhD program to examine the relationship between (a) these evaluations and the scholarly reputation of the PhD programs, (b) patterns in respondent evaluations of specific program elements and their global evaluation of program quality—patterns termed here "latent constructs" of quality, and (c) respondent characteristics and latent constructs of PhD program quality.

We found that the scholarly reputation of program faculty (i.e., NRC rating) was unrelated to alumni perceptions of overall program quality. Graduates of programs considered excellent according to the NRC's (1995) reputational assessment were only slightly more likely than others to return a rating of excellent for the "overall quality" of their PhD program (see Table 1).

We then turned to the content of alumni quality constructs. Following data reduction through factor analysis, latent class regression analysis revealed homogeneity in commitment to the value of abstract academic qualities (academic rigor and critical thinking). Yet, the LCRA also revealed distinct value sets associated with two classes of respondents. For a slight majority of respondents, who tended to come from doctoral programs with excellent reputations for scholarly quality according to the NRC, abstract academic qualities were strongly dominant in determining assessments of overall program quality. However, another group

(nearly as large) saw other elements—especially support in meeting program requirements and fostering a sense of belonging—as at least as important for evaluating overall program quality. For both groups training in research skills and diversity attributes of the doctoral program also contributed somewhat to their assessments of overall program quality. Surprisingly, the perceived quality of non-academic skills training and career preparation was negatively associated with "excellent" ratings of overall program quality.

Contrary to expectations, these distinct value sets were not associated with respondent characteristics such as career goal at PhD award or having a PhD in a field with a relatively large non-academic labor market. Further, we found no evidence that respondent demographic characteristics differentiate the value sets used in assessing overall program quality. Men and women and older and younger graduates, for instance, were equally committed to abstract academic values and equally likely to assess the quality of their PhD program in terms of socialization efforts and diversity attributes. There were no significant disciplinary differences. The primary factor that affected the value set respondents applied in assessing overall program quality is a feature of the academic value system itself: the reputation for scholarly quality of the alumni evaluator's PhD program.

Generalizability of our findings about the meaning of PhD program quality in the social sciences may be limited in certain ways. In particular, our data set covered only those who actually earned the doctorate whereas attrition rates among PhD students can be as high as 50 percent. Non-completers would offer another important perspective on the meanings of PhD program quality (e.g., Lovitts, 2001). Similarly, those who did not respond to the SS5+ survey and those outside the sampling frame could conceivably assess their graduate programs differently than the subjects in our study. However, such possible differences do not invalidate the patterns we found in our data, and we can be confident that we can generalize our findings to a substantial proportion of those who completed doctorates in the social sciences.

We cannot know the extent to which the patterns that we found for degree completers from the mid-to-late 1990s apply to more recent graduates, or will apply to contemporary graduate students who are nearing completion of the doctorate. Our methodology of using recent graduates' perspectives to evaluate doctoral training in light of their subsequent experience precludes the perspectives of those currently in graduate training. However, our findings are still relevant to contemporary graduate education. First, these findings help frame issues and raise considerations for those concerned about the quality of contemporary

graduate education. For example, our findings encourage the consideration in subsequent inquiries of how recent efforts at improving professional training have influenced contemporary students' general esteem of their graduate training. Second, these findings can serve as a benchmark to compare change over time in the ways in which social science doctorate-holders construct program quality.

Findings also could be different in disciplines outside the social sciences. We suspect that a class of alumni with an "academics dominant" perspective on PhD program quality would be found in all fields and disciplines; however, assessment of alumni perceptions of quality in natural sciences, humanities and professional fields would need to be based on field appropriate skills inventories.

Implications

The insights of our study into PhD program quality assessments by recent alumni have implications for current and future efforts to improve the educational experiences of doctoral students and ensure the relevance of doctoral education for graduates' careers. First, what are implications of the dimensions used by all graduates in assessing program quality—the dimensions shared by the two latent classes? Members of both classes place a high priority on abstract academic qualities such as academic rigor and critical thinking. Program elements related to these academic qualities should clearly have high priority in any doctoral program. But we found that other elements that may be influenced by faculty and administrators also affect alumni evaluations of overall program quality. These elements include support in meeting program requirements, fostering a sense of belonging, training in research skills, and diversity attributes.

As previously noted, we found an inverse relationship between graduate-perceived quality of training in non-academic skills and ratings of overall program quality. For decades now graduate education scholars, policy makers, and funders have been arguing that doctoral students need targeted training in communication skills, teamwork, and managing people and budgets (Clark et al., 1976; COSEPUP, 1995; Nerad, 2004; Nyquist, 2002). Social science PhD holders are obviously not convinced. Our findings suggest that advocates have more work to do to communicate the value of their proposals.

The high value that graduates place on abstract academic qualities irrespective of career goal and career type, together with the finding of a negative association between training in non-academic skills and career preparation and one's global assessment of doctoral program

quality, may suggest that for social science PhD graduates one of the primary values of doctoral education is the conferring of an academic identity. Even among—or perhaps especially among—graduates who work outside of academia, the value of the PhD may reside as much in its identity and status conferring function as in the fact that graduates have gained marketable skills and a credential with economic value in the labor market. If so, this might explain why alumni who perceived their programs as offering high quality training in non-academic skills and excellent non-academic career preparation also tended to view their programs as less than excellent. A focus on non-academic skills and careers may undermine the academic identity conferring function of the PhD program in the minds of students and alumni.

If this is the case, then efforts to promote professional skills training among social science PhD students must be mindful of the potential identity and status-conferring functions of graduate school. Such efforts might find greater resonance among students if they can be presented as an integral part of gaining the academic identity that students value so highly. Concretely this might mean, for instance, holding such trainings within the PhD department and/or sponsored and co-taught by respected professors.

Our finding of no association between demographic factors and the two quality constructs suggests that all kinds of doctoral students—and potentially prospective students—are equally likely to place a high value on abstract academic qualities such as academic rigor and critical thinking and are equally likely to value other program elements, such as support in meeting program requirements and fostering a sense of belonging, on which programs can take actions to improve. This could have implications for efforts to diversify doctoral programs. For instance, the easy assumption that women might value diversity attributes more than men is contradicted by our findings.

Conclusion

This study demonstrates clearly that the scholarly reputation of a PhD program is not the same as its alumni perceived program quality. Our finding of two quality constructs, one in which abstract academic qualities dominate that is more likely to be used by graduates of elite PhD programs and a second multi-dimensional quality construct more often used by graduates of lower-status programs is intriguing but difficult to interpret without a different line of empirical analysis. What we can say is that (a) these distinct value sets are not associated with factors such as gender, age, career goal and social science discipline and (b) members of both classes place a high priority on abstract academic qualities

such as academic rigor and critical thinking. Program elements enhancing academic quality should clearly have high priority in efforts to improve doctoral program quality. But other elements should also receive attention, including support in meeting program requirements, fostering a sense of belonging, training in research skills, and diversity attributes. We believe these findings provide both credible knowledge about important aspects of how doctoral programs work (or do not) and useful insights about how they might be improved.

Notes

[1] We do not claim that alumni-judged quality of program should eclipse scholarly standing of the faculty as the primary basis for assessing overall doctoral program quality, but only that it is an important dimension of quality in its own right that has heretofore received too little attention. Ideally, comprehensive assessments of quality would also include extensive empirical data about program outcomes.

[2] The new NRC study utilizes both "direct weights" provided by samples of faculty raters of some 20 elements related to quality in each discipline and faculty assessments of overall quality of each program in the discipline. Separate groups of raters were employed for each task. Regressions of the ratings on the elements produce "indirect" weightings of the various program elements. By each method, indicators of faculty scholarly activity dominate the weightings (National Research Council, 2009).

[3] The *US News and World Report*'s ranking of PhD programs in arts and sciences relies purely on reputational measures and is less comprehensive than the NRC research doctoral program assessments (Morse, Flannigan, & Yerkie, 2005).

[4] Generalizability is limited by possible self-selection bias among respondents and correlations of program scholarly reputation with student satisfaction are not available on the NAGPS web site.

[5] With our data, we cannot investigate the impact of race/ethnicity due to small numbers of non-whites among social science PhD recipients during the years under study.

[6] This is the National Research Council's survey of all who receive research doctorates from U.S. institutions each year.

[7] One might wonder why we chose not to use all the information provided by the several response options on each item. To expand a bit on what is said above, on only four of the 27 items was "poor" the most common response and these were the same four items that had very low proportions indicating that the quality of training was "excellent" so little would be gained by complicating the analysis and presentation considerably. Three of these also had fairly high proportions that gave no response on the item which we think implies that their program did not address the particular dimension and was certainly not "excellent" on it. Finally, since the items about mentoring used different wording and provided more response options, it would have been quite difficult to achieve basic equivalence in any other way. (We equated "very satisfied" on the mentoring items with "excellent" responses on the other items.)

[8] In the exploratory factor analysis, the number of factors was selected by comparing model fit statistics across models that varied according to the number of specified factors. The model fit statistics for the seven-factor model suggested that this model ought to be preferred. Model fit statistics and factor loading for the exploratory factor analysis are available from the authors upon request.

[9] See Vermunt and Magidson (2004) for a detailed explanation of the confirmatory (and exploratory) factor analysis techniques that we used.

[10] For example, the number of distinguishable "quality constructs" used by our survey respondents in forming their assessment of the overall quality of their doctoral program.

[11] Using maximum likelihood, LCRA simultaneously estimates (a) within each of the latent classes, a coefficient and a *t*-test statistic for the net effect of each independent variable on the dependent variable; (b) a Wald-test statistic the size and significance of which provides a test of the null hypothesis that the effects of a given variable on the dependent variable are equal across latent classes; (c) a likelihood for each case of membership in each of the latent classes; (d) for each of the covariates modeled, a coefficient and *t*-test statistic that tests the hypothesis that the covariate predicts membership in a latent class; (e) a goodness-of-fit statistic for the model—the likelihood ratio statistic (L^2)—that reflects the odds of finding the observed distribution in the data given the model parameter estimates; and (f) explained variance statistics (R^2) capturing how much variance is explained within and across classes (Magidson & Vermunt, 2004).

[12] In a LCRA with multiple classes, parameters that estimate the effect of an independent variable on a dependent variable may be unique for each class. However, a parameter may be constrained to be the same across classes—suggesting that for more than one group the effect of "x" on "y" is identical—which therefore limits the number of parameters to be estimated and thus gains parsimony. We constrained parameters to be the same only where parameters had been found to be not significantly different.

[13] The BIC statistic is a function of the model likelihood statistic (how likely we are to observe the data as it is distributed if the estimated parameters represented the "true" model) and the number of parameters estimated in the model. Thus, the BIC statistic balances parsimony with model fit. In evaluating models, the smaller the BIC statistic the more empirical evidence in favor of the model (Raftery 1995).

[14] A coefficient that is "free to vary across classes" is estimated to be unique for each class. The effect of "x" on "y" for those in one class may be estimated to be different than the effect of "x" on "y" for another class. If the coefficient is estimated to be the same across classes it is said to be "constrained" as explained in the text.

[15] BIC comparisons of Model 14 with the other 25 models each produce differences greater than or equal to six—the threshold for "strong" empirical support in favor of a particular model.

[16] In Model 2 (see Table 4) six of the seven Wald statistics for the hypothesis that effects are unique for each class are not significant, and thus these effects are constrained to be equal in Model 5. In Model 5 (see Table 5), five of the six Wald statistics for the hypotheses that the exogenous covariates have an effect on class assignment are not significant. Thus, these coefficients are constrained to zero in Model 20.

[17] We note that "belonging" is not as well measured as the other factors insofar as it consists of only two components that do not link particularly well together as indicated by the factor loadings shown in Table 2. Thus, it is possible that, could it be better measured, this construct might exert an even greater proportionate influence on "overall program quality" than we observe in our estimates.

References

Brooks, R. L. (2005). Measuring university quality. *The Review of Higher Education, 29,* 1–21.

Cartter, A. M. (1966). *An assessment of quality in graduate education*. Washington, DC: American Council on Education.

CIRGE (Center for Innovation and Research in Graduate Education). (n.d.). Data reports. Seattle, WA: Author.

Clark, B. (1987). *The academic life: Small worlds, different worlds.* Princeton, NJ: The Carnegie Foundation for the Advancement of Teaching.

Clark, M. J. (1976). The meaning of quality in graduate and professional education. In J. Katz & R. T. Hartnett (Eds.), *Scholars in the making: The development of graduate and professional students* (pp. 85–102). Cambridge, MA: Ballinger.

Clark, M. J., Hartnett, R. T., & Baird, L. L. (1976). *Assessing dimensions of quality in doctoral education: A technical report of a national study in three fields.* Princeton, NJ: Educational Testing Service.

COSEPUP (Committee on Science, Engineering, and Public Policy). (1995). *Reshaping the graduate education of scientists and engineers*. Washington, DC: National Academies Press.

Fox, M. F. (2001). Women, science, and academia: Graduate education and careers. *Gender & Society, 15,* 654–666.

Goldberger, M. L., Maher, B. A., & Flattau, P. E. (Eds.). (1995). *Research-doctorate programs in the United States: Continuity and change.* Washington, DC: National Academies Press.

Golde, C. M., & Dore, T. M. (2001). *At cross purposes: What the experiences of doctoral students reveal about doctoral education*. Philadelphia, PA: The Pew Charitable Trusts. Retrieved from http://www.PhD-survey.org

Jones, L.V., Lindzey, G., & Coggeshall, P.E. (1982). *An assessment of research-doctorate programs in the United States: Social & behavioral sciences.* Washington, DC: National Academies Press.

Lovitts, B. (2001). *Leaving the ivory tower: The causes and consequences of departure from doctoral study.* Lanham, MD: Rowman & Littlefield.

Magidson, J., & Vermunt, J. K. (2004). Latent class models. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 175–198). Thousand Oaks, CA: Sage Publications.

Morrison, E., Rudd, E., Picciano, E., & Nerad, M. (2011). Are you satisfied? PhD education and faculty taste for prestige—limits of the prestige value system. *Research in Higher Education, 52*(1), 24–46.

Morse, R. J., Flannigan, S., & Yerkie, M. (2005). America's best colleges. *U.S. News and World Report*, *139*(7), 78.

NAGPS (National Association of Graduate and Professional Students). (n.d.). *2000 national doctoral program survey executive summary.* Retrieved from http://survey.nagps.org/about/execsummary.php

National Communication Association. (1996). A study of the reputations of doctoral programs in communication. Retrieved from http://www.natcom.org/uploadedFiles/Content/Education/A%20Study%20of%20the%20Reputations%20of%20Doctoral%20Programs%20in%20Communication%201996.pdf

National Research Council (2009). *A guide to the methodology of the national research council assessment of doctorate programs*. Washington, DC: National Academies Press.

Nerad, M. (2004). The PhD in the US: Criticisms, facts, and remedies. *Higher Education Policy, 17,* 183–199.

Nerad, M., Aanerud, R., & Cerny, J. (2004). "So you want to be a professor!" Lessons from the PhDs—Ten Years Later study. In D. H. Wulff, A. E. Austin, & Associates, (Eds.), *Paths to the professoriate: Strategies for enriching the preparation of future faculty* (pp. 137–158). San Francisco: Jossey-Bass.

Nerad, M., & Cerny, J. (1999a). PhDs—ten years later: Final results. Paper presented at the CGS Annual Meeting, Washington DC.

Nerad, M., & Cerny, J. (1999b). Widening the circle: Another look at women graduate students. *CGS Communicator, 32*(6).

Nerad, M., Rudd, E., Morrison, E., & Picciano, J. (2007). *Social science PhDs—Five+ years out, A national survey of PhDs in six fields highlights report.* Seattle, WA: Center for Innovation and Research in Graduate Education. Available at www.cirge.washington.edu

Nerad, M., & Stewart, C. L. (1991, May). Assessing doctoral student experience: Gender and departmental culture. Paper presented at the 31st Annual Conference of the Association for Institutional Research, San Francisco, CA.

Nyquist, J. D. (2002). The PhD: A tapestry of change for the 21st century. *Change,* November/December.

Ostriker, J. P., & Kuh, C. V. (Eds.) (assisted by Voytuk, J. A.). (2003). *Assessing research-doctorate programs: A methodology study.* Committee to Examine the Methodology for the Assessment of Research-Doctorate Programs. National Research Council. Washington, DC: National Academies Press.

Picciano, J., Rudd, E., Morrison, E., & Nerad, M. (2007). *Social science PhDs—Five+ years out: Survey methods.* CIRGE Report 2007-01. Seattle, WA: Center for Innovation and Research in Graduate Education. Available at www.cirge.washington.edu

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25,* 111–163.

Roose, K. D., &. Andersen, C. J. (1970). *A rating of graduate programs*. American Council on Education: Washington, DC.

Sowell, R. (2008). PhD completion and attrition: Analysis of baseline data. Washington, DC: Council of Graduate Schools.

Vermunt, J. K., & Magidson, J. (2004). Factor analysis with categorical indicators: A comparison between traditional and latent class approaches. In A. Van der Ark, M. A. Croon, & K. Sijtsma (Eds.), *New developments in categorical data analysis for the social and behavioral sciences* (pp. 41–63). Mahwah, NJ: Erlbaum.