2011

Community College Student Engagement Patterns: A Typology Revealed Through Exploratory Cluster Analysis

Victor B. Sáenz
University of Texas at Austin, vsaenz@austin.utexas.edu

Deryl K. Hatch
University of Nebraska–Lincoln, derylhatch@unl.edu

Beth E. Bukoski
University of Texas at Austin, bb@austin.utexas.edu

Suyun Kim
University of Texas at Austin

Kye-hyoung Lee
University of Texas at Austin

See next page for additional authors

Authors
Victor B. Sáenz, Deryl K. Hatch, Beth E. Bukoski, Suyun Kim, Kye-hyoung Lee, and Patrick Valdez

Published in *Community College Review* 39:3 (2011), pp 235–267.
doi 10.1177/0091552111416643

Community College Student Engagement Patterns: A Typology Revealed Through Exploratory Cluster Analysis

Victor B. Sáenz, Deryl Hatch, Beth E. Bukoski, Suyun Kim, Kye-hyoung Lee, and Patrick Valdez

University of Texas at Austin

Corresponding author — Victor B. Sáenz, University of Texas at Austin, 1 University Station, D5400, Austin, Texas 78712, USA;
email vsaenz@mail.utexas.edu

Abstract
This study employs survey data from the Center for Community College Student Engagement to examine the similarities and differences that exist across student-level domains in terms of student engagement in community colleges. In total, the sample used in the analysis pools data from 663 community colleges and includes more than 320,000 students. Using data-mining techniques to discover a parsimonious number of natural clusters and, in turn, a *k*-means cluster analysis as a means of revealing a naturally occurring typology of engagement patterns, our findings reveal that support service utilization is the most distinguishing feature of the similarities and dissimilarities across student groups, suggesting areas for further theory development and testing.

Keywords: student engagement, student services, cluster analysis, Community College Survey of Student Engagement, student typologies

Change is ubiquitous in higher education, but in the last 50 years, perhaps no other sector of American higher education has experienced more change and growth within its student population than community colleges. Indeed, the evolving heterogeneous profile of students within this sector has far outpaced the ability of institutions, researchers, and policy makers to keep up, a trend that could undermine attempts to effectively match student services with student needs. In light of this trend, this study investigates the fast-changing nature of the community college student population. In particular, we employ survey data from the Center for Community College Student Engagement (CCCSE) in an exploratory cluster analysis that assesses how student engagement can reveal which groups of students are most similar and dissimilar. In identifying these disparate student groupings, the purpose of our study is to understand how student services provided by different community colleges can be effectively leveraged to maximize the critical outcome of student engagement.

Background

The American Association of Community Colleges (AACC, 2010) reports that there were 1,173 community colleges in the United States enrolling 11.5 million students (including 6.8 million credit-seeking students) in the fall of 2007, which represents 43% of all U.S. undergraduate students enrolled at that time. In addition, community college enrollments have increased 741% since 1963, compared with increases of less than 200% within each of the public and private (nonprofit) 4-year sectors (Provasnik & Planty, 2008). This increase is due in part to mounting demand placed on all educational sectors and, for many students, the choice is simple: the community college or nothing (Cohen & Brawer, 2008). The community college is also a viable option for many students because it requires a lower initial financial investment and leads to increased lifetime earnings (Sanchez & Laanan, 1997). Furthermore, community colleges are unique in terms of their role as an extension of the schools, positioned between secondary education and universities (Palmer, 2000), and in terms of their commitment to providing multiple pathways of access, especially for first-generation, ethnic minority, low-income, and underprepared students (Bragg, Kim, & Barnett, 2006).

Despite enrollment gains among these student groups, more than half of them exhibit characteristics that have been shown to reduce their chances of degree completion or transfer to a 4-year institution, characteristics such as part-time enrollment, full-time work, financial independence from parents, or single parenthood (Hagedorn, 2010). Low-income students and students of color are especially likely to exhibit these characteristics (Lamkin, 2004). To further complicate the situation, community colleges serve a highly heterogeneous population of students across multiple domains, enrolling larger percentages of nontraditional (25 years and older), first-generation, low-income, and ethnic minority students than 4-year colleges and universities (Lamkin, 2004; Provasnik & Planty, 2008). Moreover, these students come with a broad range of academic competencies, from those who are highly prepared to those needing substantial remediation (Bailey & Alfonso, 2005).

As for the critical outcome of student engagement, findings from the most recent CCCSE study (CCCSE, 2010) indicate that students who show the least amount of engagement are at greater risk of dropping out. Given their heterogeneous student populations, community colleges must consider student engagement across several axes. We know that many individual factors can affect engagement levels, including ethnicity, gender, age, whether developmental courses are taken, and the frequency with which students access student services such as tutoring and academic advising (CCCSE, 2010; Harper & Quaye, 2009). There is, however, relatively little research on community college student engagement, in large part due to insufficient national data and the limited extent to which research on 4-year institutions applies to 2-year institutions (Bailey & Alfonso, 2005; Bailey, Calcagno, Jenkins, Kienzl, & Leinbach, 2005). The research literature on student services, meanwhile, is strikingly sparse, especially in terms of how it differentially affects various kinds of community college students. What little work there is tends to emphasize program evaluation and local programmatic initiatives. This study is positioned to help fill this gap.

Literature Review

The researchers' aim is to conduct an exploratory study of engagement patterns across several domains of student and institutional characteristics. Traditionally, the theory section of a study specifies constructs and models underlying the hypotheses that the study ultimately seeks to test (Creswell, 2009). This study, however, explores and describes how students vary across a series of given variables; because it does not predict or account for that variation, a traditional theoretical framework is not entirely necessary. The literature does, however, inform the choice of engagement as the construct of interest as well as the selection of independent variables for describing patterns of variation. Informing this study is a robust literature on student engagement (Kuh, Kinzie, Buckley, Bridges, & Hayek, 2006) and the related theories of involvement (Astin, 2001) and integration (Tinto, 1993).

Rather than testing hypotheses, this exploratory study focuses on typology development, similar to prior work conducted at 4-year institutions (Astin, 2001; Pascarella & Terenzini, 2005) and 2-year institutions (Adelman, 2005; Bahr, 2010; Hagedorn & Prather, 2005). Given the heterogeneity of students in higher education and at 2-year institutions in particular, typology is an important analytical tool that can help institutions understand student behavior and make decisions about how to deploy scarce resources in ways that meet student needs. This section briefly reviews definitions and models of student engagement, the ways engagement relates to student outcomes, and the previous typology work that undergirds this study.

Student Engagement

To understand engagement patterns, it is important to understand what engagement means, especially given the distinctive 2-year institution context with its especially heterogeneous student body. Harper and Quaye (2009) broadly define engagement as "participation in educationally effective practices, both inside and outside the classroom, which leads to a range of measureable outcomes" (p. 3). This definition is certainly useful and has been corroborated by the Community College Survey of Student Engagement (CCSSE) project itself, which also references engagement broadly by calling it the "amount of time and energy that students invest in meaningful educational practices" with (McClenney, 2006, p. 47-48). These broad definitions encompass some of the more specific dimensions of the nomenclature, such as the five benchmarks used by CCSSE: active and collaborative learning, student effort, academic challenge, student–faculty interaction, and support for learners (Marti, 2009; McClenney, 2006).

There is more than one model of engagement, but all models share a common bond concerning the role engagement plays in theories of student development. For Astin (2001), engagement is an environmental factor mediated by student choice. This fits well with his input-environment-outcome (I-E-O) model, which places engagement partially as an environmental (institutional) variable and partly as an outcome (behavioral) variable. That is, students enter an institution predisposed to seek out certain environmental characteristics and then, based on individual choices, students' experiences can vary in many ways. In this way, the notion of engagement—split by Astin into five categories (academic involvement, involvement with faculty, involvement with student peers, involvement in work, and other forms of involvement)—serves as a way to quantify student choices in terms of how and with whom students interact as well as what experiences they do or do not seek. He also found that involvement enhances cognitive and affective development as well as learning, academic performance, and retention.

A more comprehensive model of engagement has been developed by Kuh et al. (2006). They also describe engagement as positioned at the intersection of institutional conditions and student behaviors. Engagement from this perspective implies that institutions can affect the environment by creating conditions and encouraging certain kinds of student behavior. Pascarella and Terenzini (2005) similarly view engagement as a series of academic and social experiences that contribute to student involvement with or integration into the institution. In this way, engagement helps determine college impact. Tinto (1993), however, does not reference engagement specifically, referring instead to academic and social integration. For Tinto, the notion of social and academic integration—the act of students becoming members of college social systems— is vital to student retention. The notion of integration, which is an intermediate measure of student outcomes, has some relationship to student behavior and experiences as well as to student perceptions of and satisfaction with their experiences. However, as Kuh et al. (2006) comment, there is less empirical support for Tinto's model despite its widespread popularity in the literature. For example, Cabrera, Nora, Terenzini, Pascarella, and Hagedorn (1999) found no support for Tinto's assertions that academic preparedness can account for differences in persistence between White and Black students or that persistence requires disconnection from family and community. Though Tinto's model has been found useful across institutional types and ethnically diverse populations, the constructs for academic and social integration in particular have been found to be methodologically flawed (Braxton & Lien, 2000).

Even though these definitions and models of engagement were largely informed by the experiences of students at 4-year institutions, they have some utility for community colleges. However, given the greater student diversity at these institutions, as well as their open-access policies, it should not be assumed that these models apply to community colleges in exactly the same way they apply to 4-year colleges. Although these notions of student engagement, involvement, and integration have served to inspire the present study, this study is not bound by them. Rather, it attempts to fill a crucial gap in understanding the way students at community colleges naturally cluster into groups or typologies of engagement.

Engagement and Outcomes

All of these frameworks share in common the notion that students inhabit the environment of college, that they have various encounters with the environment, and that those experiences can influence many aspects of the students' development and attitudes. Whatever moniker is chosen—whether it is academic and social integration (Tinto, 1993), student involvement (Astin, 1984), or engagement (Harper & Quaye, 2009; Kuh et al., 2006; Marti, 2009)—the act of students accessing the academic, social, and extracurricular activities of an institution has been proven to be very important to their persistence (Astin, 1984; Bean, 2005; Cabrera, Nora, & Castaneda, 1993; Harper & Quaye, 2009; Pascarella & Terenzini, 2005; Tinto, 1993). For the purposes of the present study, we primarily use the term engagement, understanding that it could also be taken to mean involvement and integration.

Student engagement has important implications for many student outcomes, most prominently its positive influence on student learning (Astin, 1984, 2001; Pascarella & Terenzini, 2005). In addition, engagement can aid in student academic and intellectual development and adjustment, and it can also result in increased institutional commitment (Astin, 2001; Cabrera et al., 1999). As positive outcomes such as persistence, learning, and satisfaction are associated with student engagement, it is a useful variable against which similarities and differences across student and institutional characteristics can be measured.

Considering that student engagement is related to student demographic characteristics (Astin, 2001; Tinto, 1993), it is useful to consider engagement across several axes or subgroups. For instance, it plays a role in racial and gender identity development (Harper, Carini, Bridges, & Hayek, 2004), and differences in campus climate can affect how minorities experience engagement (Cabrera et al., 1999). Though engagement patterns tend to be consistent across ethnic groups, there are some variations. African Americans and Asians, for example, are more likely to participate in educational enrichment activities, and African Americans report being more involved in collaborative learning (Kuh et al., 2006). Engagement also differs by gender, with women tending to be more engaged with campus life than men (Harris & Lester, 2009).

Nontraditional status is yet another mediating factor of student engagement and is especially pertinent for 2-year institutions because they tend to have more nontraditional students. Such nontraditional students (e.g., commuter and part-time students as well as returning students who are 25 or older) also face unique challenges; they tend to be more mature but must balance schedules and "multiple life roles" (Silverman, Sarvenaz, & Stiles, 2009, p. 226). At 2-year institutions, nontraditional-aged students also exhibit (in relation to younger students) a higher quality of engagement in their relationships with faculty members, administrative personnel, and other students; this may perhaps be due to the increased sense of purpose they bring to college activities (Gibson & Slate, 2010).

Institutional characteristics also play into student experience, and, though larger institutions often have more resources, size (in terms of enrollment) tends to correlate negatively with student persistence (Bailey et al., 2005). In fact, smaller colleges report higher completion rates, perhaps, due to a more personalized student experience (Bailey et al., 2005). Also of note is urbanicity, which has been reported to predict graduation rates that are approximately 3.5% lower than national rates while more rurally located colleges can expect nearly 4% higher graduation rates. In addition, students at rural colleges are 18% more likely than students nationally to persist toward their goals (Bailey et al., 2005). Another important institutional consideration for community colleges, given the high percentage of part-time students, is that the higher the proportion of part-time students at the institution, the lower the graduation rates are likely to be (Bailey et al., 2005).

A final consideration in engagement is student services (Kuh, Kinzie, Schuh, Whitt, & Associates, 2005). Though there is little solid empirical analysis of the effect of student services (e.g., advising, tutoring, skill labs, and supplemental instruction), students have reported increased grades and high satisfaction with tutoring services (Burnett, 2006); satisfaction in particular usually correlates with increased persistence (Astin, 2001; Cabrera et al., 1993). Supplemental instruction specifically has been found to increase persistence in the targeted class and increase student grades (Zaritsky & Toce, 2006). Learning communities in general, which include tutoring initiatives, also tend to result in retention (Tinto, 1998) and higher engagement (Zhao & Kuh, 2004). Learning communities and accessing academic resources such as libraries also tend to relate to higher engagement (Kuh & Gonyea, 2003; Zhao & Kuh, 2004). In addition, age plays into student service use because nontraditional students are more likely than younger students to participate in individualized academic support services (Kasworm, 1980). Given the scarcity of research on student services and the unique needs of students at community colleges, the present study serves as a launching point for understanding the role of academic services in student engagement.

Most engagement research focuses on 4-year institutions and particularly on fulltime, residential students who are in the traditional college-going age bracket. Meanwhile, we know that most community college students exhibit characteristics that are entirely unique. Nonetheless, it is not advisable to assume that 4-year engagement models do not apply to 2-year institutions and students (Karp, Hughes, & O'Gara, 2008). By using engagement as a foundational concept and examining how engagement actually differs across student characteristics, this study seeks to inform future research on community college student engagement.

Typology

Because of its connection to many positive student outcomes, engagement is a useful variable for the typology work undertaken in this study. Typology work serves an important role in informing and paving the way for future research because it "focus[es] on differences in the way individuals perceive their world or respond to it" (Pascarella & Terenzini, 2005, p. 45). Though typology work can run the risk of oversimplifying constructs or making artificial delineations (Pascarella & Terenzini, 2005), it is still useful for understanding patterns.

In line with the earlier typology work undertaken by researchers such as Astin (1973), Adelman (2005), Hagedorn and Prather (2005), and Bahr (2010), this study seeks to understand community college students through categorization. In 1973, Astin's taxonomy of 4-year college outcomes provided crucial conceptual clarity about cognitive and affective outcomes as well as psychological and behavioral data that are applicable to research on students. In 2001, he used his I-E-O model to work through the domains described in the taxonomy. Astin also included a time element, which this project does not. Our student outcome measure for engagement involves "transactions between the student and environment" and is affective and behavioral in nature (Astin, 2001, p. 10).

Astin's 2001 analysis also employed typology work to better understand student experiences and outcomes. Focusing on 4-year institutions, Astin used eight variables to classify institutions before breaking out specific characteristics for measurement. He first examined environmental variables, clustering such attributes as faculty behavior and values as well as several peer-group characteristics. These clusters were then used in a statistical analysis examining the impact of the institutional environment on student outcomes (such as self-concept, behavior, values, and academic and intellectual development). Astin's earlier 1973 typology of outcomes, which identified seven student types (scholar, social activist, artist, hedonist, leader, status striver, and uncommitted student), influenced his later work (Astin, 1993). Typology, then, can readily serve as the foundation for research on the assessment of college impact and student outcomes within an engagement framework.

Typology work has been conducted at 2-year institutions as well. Interestingly, Adelman (2005) and Hagedorn and Prather (2005) used metaphors in their typologies. Adelman (2005) used a town metaphor, informed by environmental design. He focused on the "settlement behaviors" (p. xiv) of traditional-aged students from an academic perspective, examining when students enrolled in a community college, how long they stayed, and what their academic achievements were before, during, and after their community college term. The results were used to create a series of portraits of student types. Through three portraits, Adelman described six populations—two persistent groups that transferred to 4-year institutions or entered occupations, a group that struggled and stopped, a group that disappeared on entry, a group of temporary transfers who were based in another institutional type (e.g., 4-year college students who took a course or two at a neighboring community college), and a group of reverse transfers (i.e., students who transferred from a 4-year college to a community college)—to emphasize the way in which attendance and attainment patterns play out in community colleges and to revisit the "cooling out" critique of community colleges.

Hagedorn and Prather (2005), however, in studying the Los Angeles Community College District (LACCD), used organizational theory to guide their solar system metaphor. Their cluster analysis revealed seven orbits of students, with students whose lives "gravitated" toward the institution positioned closer to the sun. Moving from closest to furthest, the orbits included those for traditional students, full-voc (i.e., vocationally focused) students, transfer-bound students, transfer-hopeful students, industrious (i.e., underprepared) students, brief-stint students, and unicourse students. This typology, though emanating from one urban district, is useful for understanding the way in which institutions and students relate to one another or gravitate toward or away from one another. It is also useful in problematizing the notion of diversity as it applies to community college students (i.e., diversity should not be limited to immutable characteristics or background descriptors such as gender, ethnicity, or income). Bahr (2010), in his recently published paper using cluster analysis, expanded on this and the wider body of typology work by extending the analysis to first-time students at California community colleges. This study found six clusters of students: transfer, vocational, drop-in, noncredit, experimental, and exploratory. Of note is that two of Bahr's clusters (drop-in and transfer) had strong similarities to, respectively, the unicourse and vocational orbits described by Hagedorn and Prather. The three studies highlighted here all focus on enrollment, transfer, length of stay, and coursetaking patterns. Our study seeks to extend this body of cluster analytic work by using a national data set and by using engagement constructs that have been consistently correlated with student outcomes as the dependent variables.

Method

Cluster analysis is a collective term for several methods of discovering or delineating naturally occurring groups in data sets. It is by its nature a multivariate analysis used in a broad range of applications, from business and social sciences to the physical sciences and engineering (Kaufman & Rouseeuw, 2005; Romesburg, 2004). Beyond merely classifying observations into natural sets, cluster analysis has also been applied as a method to create scientific questions and hypotheses or, under the right circumstances, test such hypotheses (Romesburg, 2004). For our purposes, we do not presume to go quite this far, but we are applying this method as a form of retroduction, using observed evidence to create a research hypothesis that accounts for the observed facts (Sayer, 1992).

As part of a retroductive exploration of a large data set, there are in fact two steps to the cluster method that must be undertaken before results can be analyzed: (a) determining the best, or natural, number of groupings inherent in the data and (b) performing the cluster analysis itself to assign each observation to its best-fit group. The results of both of these steps are described in the Results section. We will first describe the data set and the variables included in the analysis before describing the *k*-means cluster approach, the particular cluster method we implemented.

Data Set

This study employs student survey data from the CCSSE, an instrument administered by the CCCSE at the University of Texas at Austin. The study merges three separate years of cross-sectional survey data—2007, 2008, and 2009—which results in 663 unique institutions. This survey was developed to collect information on community college student engagement, a key indicator of success for community college students. Even though it remains largely untapped as a research tool, CCSSE data are among only a few national data sets available that specifically target 2-year institutions and their students and, as such, its role in institutional research is vital. In this study, we have conducted an exploratory analysis of the similarities and differences that may exist across several domains in terms of student engagement. This is a largely unexplored area of research, but it is greatly needed, given the heterogeneity of 2-year college students and their diverse needs.

In addition, data on institutional characteristics were obtained from the 2008 Integrated Postsecondary Education Data System (IPEDS) and merged into the student- level data. The center provided a randomly selected sample representing 80% (n = 320,338) of the respondents in the center's 3-year cohort of 663 community colleges. Tables 1 and 2 summarize some of the major characteristics of the respondents and institutions in the data set. As CCSSE's sampling method is at the classroom level, there is an overabundance of full-time students, given that they are more likely to be enrolled in a selected course section than their part-time classmates. To correct for this sampling bias, CCCSE calculates an enrollment weight that allows for representative descriptive statistics. CCCSE's sampling method and weight correction provide for a representative sample of each institution's particular student body (Marti, 2009). Note that this weighting is applied in Tables 1 and 2 for descriptive purposes. However, because "one of the thorniest aspects of cluster analysis continues to be the weighting … of variables" (Gnanadesikan, Kettenring, & Tsao, 1995, p. 113), the enrollment weighting variable is not applied in the cluster analysis itself. Weighting, as applied to cluster analysis, is not understood in the same way as correcting for bias of means and frequencies. Rather, weighting has been proposed, for example, as an integrated and iterative process that differentially weights separate variables according to their relative importance to the emergent *k*-means cluster structure (Arabie & Hubert, 1996). It is not at all apparent that a sampling weight as calculated for correcting means and frequencies is appropriate.

Table 1. Demographic Information of Respondents in Data Set for Variables Selected for Cluster Analysis

Student feature	Weighted ***n***	Weighted % (of nonmissing)
Race		
American Indian or other Native American	5,559	1.8
Asian, Asian American or Pacific Islander	15,978	5.1
Native Hawaiian	1,004	0.3
Black or African American, non-Hispanic	37,112	11.9
White, non-Hispanic	202,780	65.1
Hispanic, Latino, Spanish	36,597	11.7
Other	12,606	4.0
First-generation student	88,594	34.7
International student or foreign national	18,885	6.0
Grants and scholarships		
Not a source	167,677	54.0
Minor source	36,754	11.8
Major source	105,941	34.1
Student loans		
Not a source	215,510	69.5
Minor source	25,702	8.3
Major source	68,687	22.2
25 years old or above	125,026	39.9
Seeking a credential	283,436	96.0
Female	188,659	60.0
Enrolled in developmental course(s)	154,810	50.0
Completed 30 or more credit hours	101,914	32.5
Enrolled part-time	186,791	58.3
Student services used—Sometimes or often		
Academic advising/planning	174,762	56.1
Peer or other tutoring	78,868	25.6
Skill labs (writing, math, etc.)	120,436	39.2
Financial aid advising	138,248	45.0
Student organizations	50,094	16.4
Total number of services used—Sometimes or often		
No services	67,537	21.6
1 service	77,677	24.8
2 services	74,855	23.9
3 services	51,705	16.5
4 services	27,624	8.8
5 services	13,882	4.4

Table 2. Frequencies of Selected Institutional Features (n = 663)

Institutional feature	N	%
Location		
Urban	125	18.9
Suburban	142	21.4
Rural	396	59.7
Enrollment size		
Small (<4,500)	331	49.9
Medium (4,500-7,999)	162	24.4
Large (8,000-14,999)	112	16.9
Extra large (15,000+)	58	8.7
Achieving the Dream participant	126	19.0

Variables

The dependent variables in the study were selected as a way of collectively providing a fine-grained measure of student engagement as conceptualized by research undergirding the CCSSE. Each item on the instrument was crafted by CCCSE researchers on the basis of theoretical and empirical evidence in the literature (CCCSE, 2011). CCCSE researchers developed benchmarks that community colleges could use at the local level as a way to inform decision making by administrators in discrete areas of practice. Even though the benchmarks were informed by an in-depth factor analysis study (Marti, 2009), the instrument "was intended to be a holistic measure of engagement, with no latent factor structure" (Angell, 2009, p. 565). The limitations and cautions of using benchmarks for theoretical inquiry have been well documented (Angell, 2009; Roman, Taylor, & Hahs-Vaughn, 2010), but researchers can take advantage of the original latent constructs (which are themselves subject, of course, to further investigation) and the large amount of data made available through nationwide administration over several years as long as we are thoughtful about how to approach the data.

Marti (2009) provided a comparison of what he called the model of best fit (MBF)— that is, the latent constructs—and the model of effective educational practices (MEEP)—that is, what came to be adopted as CCSSE benchmarks. A later validation study (McClenney & Marti, 2006) demonstrated how both the MBF factors and MEEP benchmarks perform as proxies for a wide range of desirable student outcomes such as grade point average (GPA), credit hours completed, first-to-second term persistence, and degree completion. Of the nine MBF constructs, it was found that not all of them were consistently correlated with student outcomes, and those that did not were set aside in the construction of benchmarks. However, at least three clearly were, namely, *collaborative learning, class assignments*, and *mental activities*. (These clusters, incidentally, provided the bulk of items used in the benchmarks that are also most correlated with student outcomes.) For this study, we selected the 13 individual survey items from these three latent constructs. They each use a 4-level ordinal response scale as answers to prompts about behaviors, experiences, requirements of the college classroom, and learning activities. This equivalently scaled coding scheme is important because *k*-means algorithms (and most clustering algorithms in general) assume equivalent distances for all measures, even if through standardization. Table 3 lists the 13 dependent variables used in this multivariate analysis.

Whereas the *k*-means cluster analysis uses these 13 measures of student engagement to form groups, the data set includes more than 250 other variables, allowing a broad range of independent variables (listed in Table 1) for investigating the composition of those groups. For this study, we investigated the demographic features of the groups, which included race, gender, age, parents' education level, enrollment status, international status, student loans, grants and scholarships, academic goals, developmental classes, credit hours completed, and the frequency of use of student services (including advising, tutoring, skill labs, financial aid advising, and student organizations).

Analytic Method: k-Means Cluster Analysis

Cluster analysis is performed to identify relatively homogeneous groups of cases in data (Kaufman & Rouseeuw, 2005; SPSS Inc., 2009). In other words, the method provides the researcher "with clusters that are as different from each other as possible, with the members within each cluster as similar to each other as possible" (Ammon, Bowman, & Mourad, 2008, p. 34). Clusters are determined by computing distances between cases in multidimensional space.

Cluster analysis methods come in two general approaches: hierarchical and partitional. Hierarchical approaches—both agglomerative (bottom-up) and divisive (top-down) types—are amenable to plotting linkages as dendrogram trees, allowing for visual inspection and comparison with theoretical divisions or practical divisions informed by logical deduction. Unfortunately, these methods are only feasible for studies with up to a few hundred observations because the graphs quickly become unwieldy to interpret and the computations become too impractical to perform with commonly available technology. Partitional clustering, such as the *k*-means method, have, in turn, the advantage of being able to handle very large numbers of observations but require an assigned number of clusters at the outset (Duran & Odell, 1974; Hartigan, 1985; Manton, Lowrimore, Yashin, & Kovton, 2005).

As the purpose of our research was to investigate what characteristics of students lead them into academic engagement, we implemented the partitional *k*-means cluster analysis so that it could cluster students into *k* homogeneous groups based on their pattern of academic engagement. *k*-means clustering requires that a researcher determine, a priori, the number of clusters in the data and then proceed by iteratively calculating the mathematical center of each cluster (in as many dimensions as there are variables) using randomly or purposively selected seed values. This method is commonly used to automatically divide a data set into *k* groups and can easily handle large sample

Table 3. Dependent Variables for Multivariate Analysis, by Construct, Coded on Four-Level Ordinal Scale of Frequency

Model of best fit construct	Variable	Prompt
		How often have you …[a]
Collaborative learning	CLASSGRP	worked with other students on projects during class?
	OCCGRP	worked with classmates outside of class to prepare class assignments?
	COMMPROJ	participated in a community-based project as a part of a regular course?
	TUTOR	tutored or taught other students (paid or voluntary)?
	REWROPAP	prepared two or more drafts of a paper or assignment before turning it in?
Class assignments	INTEGRAT	worked on a paper or project that required integrating ideas or information from various sources?
	CLPRESEN	made a class presentation?
	WORKHARD	worked harder than you thought you could to meet an instructor's standards or expectations?
		How much has your coursework at this college emphasized … [b]
Mental activities	ANALYZE	analyzing the basic elements of an idea, experience, or theory?
	SYNTHESZ	synthesizing and organizing ideas, information, or experiences in new ways?
	EVALUATE	making judgments about the value or soundness of information, arguments, or methods?
	APPLYING	applying theories or concepts to practical problems or in new situations?
	PERFORM	using information you have read or heard to perform a new skill?

All variables are coded on a 4-level ordinal response scale. See the 2009 Community College Survey of Student Engagement codebook (CCSSE, n.d.) for full details.

a. The frequency scale is as follows: 1 = *never*, 2 = *sometimes*, 3 = *often*, 4 = *very often*.

b. The extent scale is as follows: 1 = *very little*, 2 = *some*, 3 = *quite a bit*, 4 = *very much*.

sizes (Ammon et al., 2008). We used randomly selected seed values and set the program to 15 iterations of centering, a decision that will be further explored in our Results section.

As was mentioned above, scaling of variables is an important consideration in k-means clustering because variables measured with different scales may lead to misinterpreting the result (Kaufman & Rouseeuw, 2005). Thus, it is crucial to standardize variables before performing the *k*-means cluster analysis. In this respect, the equivalent scale variables we use fit to the *k*-means clustering.

The *F* statistic in *k*-means one-way analysis of variance (ANOVA) is calculated "to help identify the variables that drive the clustering" (Ammon et al., 2008, p. 35). The relative size of the *F* statistic provides information about each variable's contribution to the separation of the groups (SPSS Inc., 2009). In other words, a large *F* value indicates that the variable is an important factor. However, although the size of the *F* statistic is important in assessing the contribution, the significance level does not provide any useful information (Ammon et al., 2008). Additional between-group (post hoc) analyses were not computed, although the authors encourage future iterations of this research to explore these between-group differences more closely.

Determining the Number of Clusters

Before running the *k*-means analysis, we were faced with the dilemma of determining how many natural clusters there are so that the algorithm can assign each observation to its appropriate group. One of the assumptions of *k*-means cluster analysis is that researchers need to select an optimal and parsimonious number of clusters so that they account for all relevant variables (SPSS Inc., 2009). In cluster analysis, the challenge of determining the appropriate number of *k* clusters is distinct from the actual clustering task at hand. Indeed, "the problem of determining the 'true' number of clusters has been called the fundamental problem of cluster validity" (Hardy, 1996, p. 83). Whether one selects a hierarchical clustering method, which reveals various levels of agglomerations simultaneously, or a partitional clustering method, which requires a preestablished number of groupings, the problem remains: How many clusters?

Many methods have been proposed over time to determine the appropriate number of clusters but there is little consensus as to which is most efficient or accurate, let alone how this determination might be made. It is advisable to use several methods to make an informed but nonetheless subjective decision based on theory, preliminary investigation of the data, or both (Fraley & Raftery, 1998; Milligan & Cooper, 1985). With little theoretical guidance as to how student characteristics and behaviors ultimately combine into typical patterns of community college student engagement, our first task was to investigate the data to provide clues about how many groupings there are. We dealt with these results first as a key to interpreting the results of the cluster analysis task at hand. Finally, we carried this data-mining exercise further to see how demographic and institutional factors differ among these engagement clusters.

Results

The results of this study are divided in three areas: (a) determining the number of clusters, (b) inspecting how the clusters differ in terms of the multiple dependent variables, and (c) inspecting the clusters in terms of other relevant descriptive or behavioral factors that shed light on theoretical issues of community college student engagement.

Determining 15 Clusters

Existing theories concerning student-level factors that influence student engagement offer little guidance in selecting a likely number of clusters because studies regarding demographic differences in engagement are inconsistent (Pike & Kuh, 2005) and because engagement theory in general tells us that student experiences and institutional conditions, rather than demographic variables, are the critical factors (Kuh, 2004; Kuh et al., 2006). These concerns are further complicated by the heterogeneity of community college students and institutions. We are left with an inspection of the data set itself to reveal a sensible number of clusters. We considered the results of several methods both sequentially and collectively. In all, we used four approaches: (a) a rule of thumb, (b) the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), (c) the clustergram approach, and (d) the elbow method. Each is described below.

A rule of thumb. We began the analysis by employing a rule of thumb suggested by Mardia, Kent, and Bibby (1979). This rule of thumb stipulates that the number of clusters k is approximately the square root of n / 2. However, in the case of the CCSSE data set, this would be roughly 400 clusters, clearly not a workable number to make sense of patterns of student engagement. We thus turned to the AIC and BIC.

AIC and BIC. The SPSS two-step clustering component includes the option to calculate the AIC or BIC for a range of possibilities and then automatically estimate the number of clusters in a data set (SPSS Inc., 2001). However, this approach works best with continuous variables and where clusters do not overlap extensively (Bacher, Wenzig, & Vogler, 2004), conditions that unfortunately do not describe the characteristics of the CCSSE data set. Rather, our 13 dependent variables are not continuous but use a 4-level ordinal response scale, and there are 13 separate measures of student engagement from tens of thousands of respondents, which make for understandably ill-defined borders between clusters. As a consequence, this method calculated only two overlapping clusters and proved to be impractical. As the CCSSE data set is so large, and because the degree of separation between potential clusters for those students within one or two standard deviations from the mean on the various measures makes for a crowded center, a visual inspection of the data was necessary. To this end, we employed two more methods: the clustergram method and the elbow method.

Clustergrams. The clustergram is a diagnostic plot proposed by Schonlau (2002, 2004) that offers an alternative to dendrograms for visualizing how clusters are formed and how individual observations are assigned among clusters as the number of clusters increases. As it is amenable to nonhierarchical clustering, such as k-means clustering, it can handle very large data sets. One important advantage of clustergrams over dendrograms is that the size of clusters can be seen by the relative number of observations in each one and by

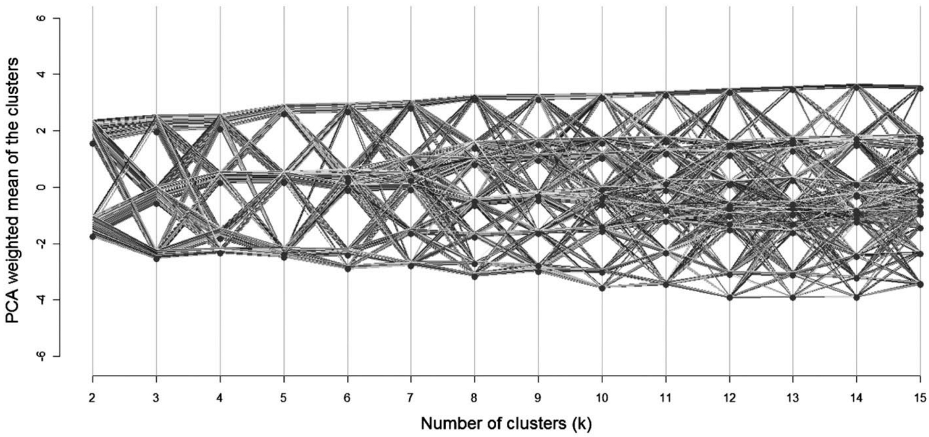

Figure 1. Clustergram plot showing group centers for 2 to 15 ***k***-means clusters The cluster centers were determined through principal components analysis (PCA), employed as an integral part of the cluster analysis methodology as a way to mean center the data for each attribute in multivariate statistics (Shaw, 2003).

how close cluster centers are relative to each other, an indication of whether increasing divisions of clusters add anything to a descriptive typology of the data set. We used an R-code program for clustergrams, which was developed by Galili (2010) on the basis of Schonlau's (2002, 2004) work and which implements R's default *k*-means algorithm as used in its cluster package. As this is a random, nondeterministic, algorithm (Hartigan & Wong, 1979), the value of the seed numbers influences the outcome, necessitating several runs to see which cluster assignment patterns occur (Steinley, 2006). Simply put, because the number of clusters was not predetermined, the researchers ran the algorithm eight times with a random starting point to discern the trend of the number of clusters.[1]

At least three pertinent observations can be made from the resulting plots, seen in Figure 1. First, cluster assignment is very fluid; many individual observations change cluster assignments throughout the range of 2 to 15 clusters, suggesting that engagement patterns are quite similar for all students with no clear-cut distinctions—though this may be an artifact of a 4-level ordinal scale. Second, by the time the data set is divided into eight or nine clusters, there are three prominent strands: a dense middle strand and two smaller strands above and below the middle strand. Third, even at higher numbers of clusters, their centers nonetheless tend to group together in three to four major clumps, though this latter trend is not consistent.

On the one hand, this analysis would agree with the SPSS 2-step method that suggested there are only a few parsimonious clusters. On the other hand, the possibility that there may in fact be many discernible clusters in the dense center—even if their means are not distant from each other—suggested that

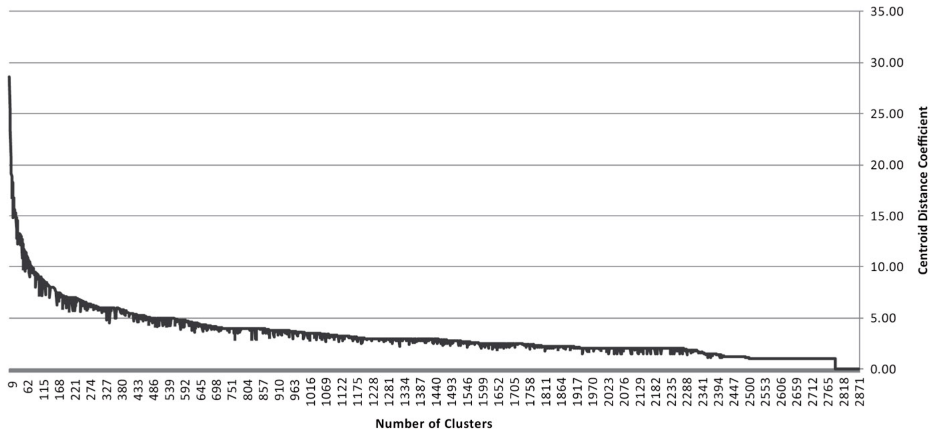

Figure 2. Elbow method plot: Centroid distance coefficient versus number of hierarchical centroid clusters.

the data were worth inspecting in terms of how noteworthy those small distances may be. To test the possibility of potentially larger numbers of clusters, we used the popular elbow method.

Elbow method. The so-called elbow method is based on the notion that the number of clusters is parsimoniously maximized at the point where adding more clusters does little to reduce the amount of variation between clusters (Thorndike, 1953). This curve of diminishing returns can be inspected on various kinds of diagrams, usually plotting the increasing number of clusters on the x-axis and plotting either the percentage of variance explained or the agglomeration coefficient as provided by hierarchical clustering procedures in many software packages on the y-axis. The agglomeration coefficient is, among other things, an index of how similar two clusters are at the point they are merged. Thus, the value of the coefficient is high at the first stages of agglomeration because like observations naturally fit. The point at which the change in coefficients levels out—the point of diminishing returns or the elbow in the plot—suggests roughly the number of natural clusters, though it is not always clearly discernible (Ketchen & Shook, 1996). As seen in Figure 2, as the number of clusters increases, the value of the coefficient drops, reflecting increasing dissimilarities among merging clusters. For the CCSSE data set, there were too many observations for the SPSS hierarchical cluster algorithms to handle, making it necessary to take a random sample of observations to calculate coefficients. With a 1% sample (n = 2,872), the elbow suggests that there may be as many 100 natural clusters. However, the greatest drop in coefficient values occurs within the first 15 to 30 clusters (where the coefficient value is 16.75 and 12.22, respectively), suggesting perhaps a reasonable number to work with in an initial exploration of whether there are distinguishing features among students associated with different engagement patterns.

Whereas the clustergram method suggested as few as three or four consistent clusters, it was likewise clear that, within the very crowded center region, there may be discernible trends between closely grouped clusters that were worth inspecting. The elbow method suggested as much—that at least 15 and perhaps as many 100 clusters were sufficiently distinguishable. For the purpose of this study, we selected 15 clusters as a reasonable and manageable number to work with.

Patterns of Engagement Between Clusters

With the number of optimal clusters set at 15, we performed a *k*-means procedure in SPSS allowing for cluster assignment of each observation. The cluster assignment ranged from n = 10,932 to n = 28,485. (See Appendix A for the mean values of the dependent variables for each group and the between-group variance F for each variable.) A 5-step grayscale coloring scheme (at break points of 0.5 on the range of 1.0 to 4.0) was added to the table to highlight engagement patterns. In addition, the clusters were arranged from lower to higher engagement levels by an ad hoc method of summing the means for each column and ordering them accordingly. To group the 15 clusters into three main groupings—as suggested by the clustergram method described above—a separate *k*-means procedure was performed for a 3-cluster solution and the assignments of both methods cross-tabulated to see where they most consistently overlapped. The degree of overlap for this 15 × 3 crosstab table led to the categorization of *low engagers* and *high engagers* on both extremes, with a broad set of *diverse engagers* in between, as labeled in Appendix A. This pattern of strong trending groups on both extremes along with a dense center is in accordance with the clustergram analysis discussed above.

The results can be read both horizontally and vertically to detect trends. Reading across columns, as has been noted, there are two extremes: certain kinds of students who are either consistently detached from their school environment (e.g., those in Cluster 1, for whom mean scores on the items listed in Table 3 are somewhere between *never or very little* and *sometimes or some*) or those who are highly involved on all fronts (e.g., students in Cluster 15, who, on average, are *very often* involved in the collaborative learning or classroom assignments that are associated with engagement and whose coursework *very much* emphasizes mental activities). In the middle, there is a wide variety of engagement.

Scanning the table from top to bottom, we can make observations about the MBF constructs too. For *collaborative learning*, it can be seen that, for all clusters, group work both inside (CLASSGRP) and outside (OCCGRP) the classroom happens more regularly than community-based projects (COMMPROJ) and tutoring and teaching (TUTOR) among peers. With the limited time available to many community college students, it is not surprising that community-based projects are consistently rare, except for two or three clusters. However, it is discouraging that tutoring and teaching among peers is likewise so consistently neglected. Interestingly, these latter two variables happen to be the ones with the largest mean values of all variables for Cluster 10 and Cluster

12, calling attention to what might be the characteristics of the individuals in these groups that lead them to have such a divergent pattern of engagement (discussed in next section).

For *class assignments*, the pattern of how often students are preparing drafts of papers (REWROPAP), integrating ideas from many sources (INTEGRAT), and making presentations (CLPRESEN) increases quite regularly from left to right in the table, except for some lower-than-average showings in a couple of instances among the *diverse engagers*. For example, the students of Cluster 13—among the most highly engaged overall—only have a mean rate of 1.70 for presenting in class (between *never* and *sometimes*), as do the students in Cluster 9 for both presenting and creating several drafts of papers.

As far as the construct of *mental activities*, Cluster 9 again presents an interesting case. Whereas these students have, on average, weak scores on *collaborative learning* and *class assignments*, they are regularly and highly engaged due to the emphasis placed by their colleges on analysis, synthesis, and evaluation of ideas, information, and experiences as well as on performing and applying that knowledge (variables ANALYZE through PERFORM). A similar group of students are those in Cluster 3. Though not claiming, on average, such high levels of engagement in mental activities as their peers in Cluster 9, the students of Cluster 3 still cite their colleges as emphasizing these activities *quite a bit*, whereas they score poorly on *collaborative learning* and *class assignments*.

In all, the engagement factors with the highest rate of between-group variance were REWROPAP, focusing on how often students prepared two or more drafts of a paper or assignment before turning it in, $F(14, 296{,}184) = 30{,}342$, $p < .001$, and APPLYING, which focused on how much coursework emphasized applying theories or concepts to practical problems or in new situations, $F(14, 296{,}184) = 28{,}720$, $p < .001$. The lowest, but always significant statistically, was WORKHARD, which offered a measure of how often students worked harder than they thought they could to meet an instructor's standards or expectations, $F(14, 296{,}184) = 6{,}890$, $p < .001$.

Differences in Demographic and Behavioral Factors Between Clusters

With cluster assignments established, based on students' patterns of engagement, and a review of those patterns out of the way, a natural follow-up question is how these groups differ in terms of critical demographic characteristics and behaviors that are described in the literature. Also, more to the point, which characteristics or behaviors most distinguish the clusters? The table in Appendix B includes several of these characteristics and behaviors, ordered from lower to higher levels of between-group variance. (Similar to data in Appendix A, values in Appendix B are shaded to visually highlight patterns, this time at break points of 20% on a scale from 0% to 100%.) Given the sheer size of the data set and the numbers of students within each cluster (between 10,932 and 28,485), it is overly reductive to propose typical student profiles of engagement for each one. That is to say, cluster analysis is often used for identifying a typical profile; however, the reader will notice in Appendix B by scanning across a row of a given characteristic that the composition remains consistent across clusters, thus the typical demographic profile of students for each cluster looks very similar to any other cluster profile (though it may be more feasible to distinguish different demographic profiles with 100 or more clusters). Nevertheless, some trends are worth noting, especially those that show which characteristics vary most widely between clusters.

Of all the demographic variables, racial identification is the one with the least amount of variation across clusters, with an F statistic of only 26.2, which is very small considering the number of observations and degrees of freedom involved, $F(14, 289{,}091) = 26.2$, $p < .001$. Many of the other selected breakout variables presented relatively little variance between clusters, including first-generation status, citizenship status, the relative importance of student loans as a revenue source, age, and even gender to some extent, $F(14, 291{,}549) = 251.1$, $p < .001$. However, between-group variance for gender does show a definite pattern of more highly engaged groups trending to more dominant female composition: The average proportion of females among the *high engagers* (67% and 62% in Clusters 14 and 15, respectively) was higher than the average proportion of females among the *low engagers* (54%). It would appear that a distinguishing feature of Clusters 10 and 12, noted above as having higher-than-average scores in completing projects together and tutoring each other (in comparison with nearly all other clusters), is that more than 80% are full-time students.

The variables that likewise showed more pronounced patterns of between-group variance, and that also varied along with levels of engagement, include grants and scholarships, enrollment in developmental courses, number of credits earned, enrollment status, and the number and types of student services used. The use of student services was coded on a binary scale of either 0 (*do not know*/*rarely*) or 1 (*sometimes*/*often*). Whereas the use of individual student services has different levels of between-group variance—tutoring services more prominently distinguish groups than does financial aid advising, for example—the variable that most clearly distinguishes clusters is the number of services used, $F(14, 290{,}873) = 1{,}832.7$, $p < .001$. Scanning this section of the table from left to right, it is apparent that from *low engagers* through to *high engagers*, the number of services used has a definite and direct correlation with increasing levels of student engagement. Specifically, 37% of the least engaged students used no services at all, and 48% of them used at most two services sometimes or often, whereas a full 60% of the most engaged students used three, four, or five services sometimes or often. The distribution of student service utilization in between suggests that when a quarter of the students use, on average, at least three services regularly, their levels of engagement are correspondingly above average.

Discussion

Students in today's community colleges have more heterogeneous backgrounds than their counterparts in 4-year institutions in terms of age, race and ethnicity, levels of academic preparation, education aspirations, and academic goals (Ammon et al., 2008; Hagedorn, 2010; Schuetz, 2002). To address the varied and diverse needs of students with such a heterogeneous set of characteristics, our study sought to examine patterns of student engagement and of service utilization in greater detail. In conducting this exploratory cluster analysis, our study shed light on how student services provided by institutions could be effectively leveraged to maximize overall student engagement.

To begin, we carefully examined the notion of engagement within the context of community college students. Harper and Quaye (2009) broadly define engagement as participating in educationally effective practices both inside and outside the classroom, and the Center for Community College Student Engagement (2010) also addresses engagement in similarly broad terms (McClenney, 2006, p. 18). Kuh et al. (2006) point to a more comprehensive model of engagement focused on the intersection of institutional conditions and student behaviors. With these operational definitions in mind, we delved into our exploratory cluster analysis to examine distinct typologies of student behavior with an eye toward understanding the forms of service utilization that might yield optimal engagement results for students.

Student typologies can be instrumental for research on the assessment of college impact and student outcomes within an engagement framework (Astin, 1973, 1993; Pascarella & Terenzini, 2005). This prior typology research provided an important precedent for conducting similar work with community college students. Our study in particular sought to allow groups (k = 15) to arise naturally from the data rather than imposing artificial partitions. The elbow method suggested optimal parsimonious clusters set at 15, but as many 100 clusters were distinguishable within our analysis. This underscores the vast diversity of community college student experiences related to their engagement and service utilization.

Though cluster analysis often helps to create a typical demographic profile, our results yielded clusters not readily distinguished by demographics. Instead, our clusters were most distinguishable according to students' use of services. In short, the clusters did not arise from who the students are but from the activities students choose to engage in. Our results identified three distinct patterns of engagement among 15 clusters: low, diverse, and high engagers. When comparing the extremes, we found that certain clusters of students were either consistently detached from their school environments or highly involved on all fronts. More specifically, students who were well prepared for assignments and who reported that coursework emphasized performing and applying knowledge ranked highest on engagement. This implies that student engagement can be encouraged not only by individual student effort but also by well-designed curricular and pedagogical practices as well as by invasive student service practices.

Our findings also indicated that female students are more engaged than their male counterparts. Between-group variance for gender demonstrated a definite pattern of more highly engaged groups trending toward mostly female composition, a finding perhaps related to the fact that females are more likely to cooperate with their peers or tutors in completing projects. This finding indicates that cooperation and collaboration among students is strongly related to a higher level of engagement.

In examining the clusters of low engagers to high engagers, the number of student services used has a definite and direct correlation with an increasing amount of student engagement. The distribution of student service utilization between clusters suggests that when one quarter of the students utilize, on average, at least three services regularly, their levels of engagement are correspondingly above average. Thus, our findings suggest that support service utilization is the most distinguishing feature of the similarities and dissimilarities across cluster groups, suggesting areas for further theory development and testing. Quite simply, the more that institutions can encourage students to seek out and utilize support services, the more likely their overall engagement will increase, resulting in increased positive outcomes.

Whether this relationship is causal or correlative cannot be determined by this analysis, and our analysis does not directly allow us to make inferences about which services are most closely associated with higher levels of engagement (though the different levels of between-group variation points us in a direction to look). However, what we do see from these results is that the utilization of student services, whether measured in terms of the types of services used or how many, is among the features that most clearly demarcate levels of student engagement regardless of other characteristics, including ethnicity, parents' education, gender, or even enrollment status, all of which are shown in the literature to mark important differences. This finding raises numerous questions—some already mentioned—as to which services are most critical and in what combination, not to mention how much the utilization of services depends on student behavior and how much depends on institutional practices.

Regardless of this analytic distinction, we do know that community colleges can take practical steps in facilitating greater service utilization that will ultimately yield positive results for some students more than others. Thus, the overarching implication of our study is that community colleges can be proactive in crafting academic and social environments that create optimal conditions for engagement by encouraging the use of more student support services among students. In addition, given the heterogeneity of students at 2-year institutions, typology is an important analytical tool that can be used by institutions to understand student behavior and make decisions about how to deploy scarce resources in ways that meet the diverse needs of students. Cluster analysis using CCSSE data afforded us an opportunity to make more sense of a complex community college student population in new and compelling ways. Future research using such data should consider the utility of typologies as a viable means of uncovering unique patterns of behavior and service utilization related to community college student engagement.

Limitations

There are several limitations to this study that deserve mention. Some of the more prominent limitations stem from the nature of the data set, which yielded many of its strengths—namely, the depth and the breadth that such a large data set affords. With tens of thousands of observations from a wide variety of institutions all taken from a single administration, the cross-sectional data both reveal and conceal phenomena simultaneously. Moving forward to follow-up research, it would be very useful to concentrate on programs and practices used at community colleges that are designed to foster student engagement and success. It is also recommended that researchers examine how, why, where, and when students access these services. This level of analysis was not possible with the data set employed in the study because the CCCSE does not reveal identifying information that would allow for this level of institutional disaggregation (beyond size categories and the urban, suburban, or rural location of colleges, which may admittedly be good starting points). In the same vein, the exploratory (descriptive) nature of this study affords a high-level view of community college student engagement nationwide as a way of suggesting a direction for theoretical development or an agenda for future research; however, this retroductive stance limits us to mostly speculation, even though the analysis was guided by an underlying conceptual framework.

The question of how many clusters naturally comprise the data set in terms of student engagement patterns also remains open. Though we followed a rigorous set of procedures to arrive at the number of 15, it is likely (as noted above) that many more cluster groupings could be explored. Each cluster numbered in the many thousands of respondents, and some of the other demographic and behavioral characteristics of the students may very well be more prominently distinct in smaller clusters or within subsets along institutional- and student-level lines. We concede that there could easily be many more "viable" clusters beyond the 15 that we chose. This is one of the primary questions that future research could more closely examine. An attendant limitation is that between-group analyses comparing individual clusters were not performed. Although *F* statistics were calculated for our *k*-means to identify variables that drove the clustering, additional post hoc analyses that could have further explored these between-group differences were not computed. We encourage future iterations of this research to examine these differences more closely.

Finally, other limitations of the current study concern the validity of engagement constructs as proxies for desirable student outcomes and the validity of the CCSSE items and factors used to operationalize those constructs. In the case of the former, student engagement in its several forms has a relatively strong body of literature to support its validity and reliability as a critical factor in academic endeavors that are closely associated with achievement, persistence, and completion. In the case of the latter, however, the validity of CCSSE constructs and items have been subject to closer scrutiny because existing validation studies, though promising, have not yet formed a substantial enough body of research to understand fully their implications for theoretical inquiry. CCSSE benchmarks, especially, are problematic because they were not designed for theoretical research but, rather, as tools for approaching institutional and system factors at a local, disaggregated level (Marti, 2004; Roman et al., 2010). Although this study used the socalled MBF factors that were statistically derived and that are empirically correlated with many student outcomes (McClenney & Marti, 2006), we are unaware of any other studies that have employed these factors in assessing the validity of the CCSSE. Instead, validation research of the CCSSE has tended to focus on the MEEP (e.g., Angell, 2009). The MBF constructs, though limited in their own regard, have the potential advantage of addressing this issue, subject to further research.

Appendix A

Results of k-Means Cluster Analysis With Mean Values of Dependent Variables by Cluster and Between-Group Variance by Variable

	Low engagers		Diverse engagers											High engagers		
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	Cluster 11	Cluster 12	Cluster 13	Cluster 14	Cluster 15	
Variables	*N* = 26,157	*N* = 26,043	*N* = 26,344	*N* = 27,061	*N* = 18,013	*N* = 14,833	*N* = 28,485	*N* = 22,336	*N* = 16,588	*N* = 11,179	*N* = 22,713	*N* = 11,846	*N* = 16,621	*N* = 17,048	*N* = 10,932	*F*
Collaborative learning																
CLASSGRP	1.84	2.37	1.72	2.13	3.08	3.04	2.13	2.67	2.4	2.71	3.08	2.94	2.89	2.94	3.57	10,899
OCCGRP	1.32	1.76	1.39	1.47	1.51	3.26	1.49	1.92	1.81	2.32	2.47	2.84	2.21	2.25	3.49	17,332
COMMPROJ	1.08	1.24	1.13	1.13	1.16	1.24	1.15	1.22	1.27	3.06	1.28	1.59	1.28	1.34	2.84	19,031
TUTOR	1.14	1.27	1.20	1.17	1.26	1.48	1.20	1.25	1.34	1.85	1.29	3.24	1.37	1.38	2.59	14,895
Class assignments																
REWROPAP	1.48	2.21	1.64	3.33	1.69	2.14	3.32	1.90	1.70	2.79	3.56	2.75	3.40	3.35	3.38	30,342
INTEGRAT	1.75	2.73	2.13	2.97	2.01	2.5	3.14	3.17	2.27	2.94	3.55	3.1	3.56	3.61	3.66	16,908
CLPRESEN	1.37	2.57	1.54	1.69	1.57	1.84	1.87	2.79	1.73	2.58	3.01	2.34	1.70	3.42	3.22	21,374
Mental activities																
ANALYZE	1.92	2.12	2.93	2.44	2.65	2.64	3.09	3.04	3.73	2.78	2.86	3.19	3.74	3.73	3.72	18,645
SYNTHESZ	1.67	1.95	2.71	2.24	2.41	2.46	3.03	2.93	3.68	2.71	2.77	3.07	3.75	3.75	3.72	25,360
EVALUATE	1.56	1.91	2.59	2.04	2.11	2.26	2.9	2.88	3.54	2.57	2.54	2.91	3.62	3.63	3.61	21,641
APPLYING	1.61	1.85	2.62	1.99	2.64	2.41	2.97	2.95	3.76	2.67	2.59	3.09	3.74	3.74	3.74	28,720
PERFORM	1.80	1.98	2.51	2.13	3.08	2.64	3.01	2.94	3.75	2.86	2.78	3.15	3.73	3.75	3.78	19,470
WORKHARD	1.83	2.18	2.12	2.43	2.42	2.55	2.73	2.42	2.71	2.7	2.94	2.89	3.25	3.22	3.5	6,890

Note: See Table 3 for a description of the variables. A 5-step grayscale coloring scheme (at break points of 0.5 on the range of 1.0 to 4.0) highlights engagement patterns—from least engaged (no shading) to most engaged (the darkest shading).

Appendix B

Means of Selected Independent Variables by Cluster and Between-Group Variance by Variable

	Low engagers		Diverse engagers											High engagers		
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	Cluster 11	Cluster 12	Cluster 13	Cluster 14	Cluster 15	
Variables	*N* = 26,157	*N* = 26,043	*N* = 26,344	*N* = 27,061	*N* = 18,013	*N* = 14,833	*N* = 28,485	*N* = 22,336	*N* = 16,588	*N* = 11,179	*N* = 22,713	*N* = 11,846	*N* = 16,621	*N* = 17,048	*N* = 10,932	*F*
Race																26.2
Native American	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02	.03	
Asian American, Pacific Islander	.05	.06	.05	.05	.04	.06	.05	.05	.05	.09	.05	.08	.05	.05	.06	
Native Hawaiian	0	0	0	0	0	0	0	0	0	.01	0	0	0	0	.01	
Black, non-Hispanic	.10	.10	.10	.11	.10	.11	.12	.08	.13	.13	.12	.11	.14	.14	.15	
White, non-Hispanic	.69	.67	.70	.67	.71	.67	.65	.70	.65	.58	.63	.61	.62	.60	.57	
Hispanic, Latino, Spanish	.10	.11	.10	.12	.10	.11	.13	.11	.11	.11	.13	.12	.13	.14	.13	
Other race	.04	.04	.04	.04	.03	.04	.04	.04	.04	.05	.05	.05	.04	.05	.05	
First generation	.34	.32	.31	.36	.34	.33	.35	.28	.32	.33	.35	.29	.34	.35	.32	39.3
International	.04	.06	.05	.06	.05	.06	.06	.05	.06	.10	.06	.09	.07	.07	.08	67.7
Student loans																109
Not a source	.73	.69	.71	.69	.69	.65	.68	.69	.67	.60	.66	.65	.65	.64	.59	
Minor source	.08	.10	.08	.09	.08	.10	.08	.08	.08	.13	.90	.11	.07	.08	.09	
Major source	.19	.21	.21	.22	.22	.25	.24	.23	.26	.27	.25	.25	.28	.27	.32	

(continued)

Appendix B (continued)

	Low engagers		Diverse engagers										High engagers		
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	Cluster 11	Cluster 12	Cluster 13	Cluster 14	Cluster 15	
Variables	*N* = 26,157	*N* = 26,043	*N* = 26,344	*N* = 27,061	*N* = 18,013	*N* = 14,833	*N* = 28,485	*N* = 22,336	*N* = 16,588	*N* = 11,179	*N* = 22,713	*N* = 11,846	*N* = 16,621	*N* = 17,048	*N* = 10,932	*F*
Below 25	.34	.24	.34	.30	.36	.29	.35	.25	.41	.28	.30	.31	.41	.37	.36	229.9
Seeking a credential	.93	.98	.96	.98	.96	.98	.99	.99	.97	.98	.99	.98	.99	.99	.99	248.1
Female	.54	.54	.54	.61	.51	.57	.61	.54	.62	.54	.65	.51	.69	.67	.62	251.1
Grants, scholarships																258.2
Not a source	.59	.49	.56	.5	.54	.47	.48	.50	.51	.40	.44	.40	.44	.43	.37	
Minor source	.11	.15	.12	.12	.13	.15	.12	.14	.11	.18	.14	.16	.11	.12	.14	
Major source	.3	.36	.32	.38	.33	.39	.40	.36	.38	.41	.42	.44	.44	.45	.49	
In developmental classes	.41	.50	.43	.56	.45	.50	.56	.46	.42	.60	.60	.50	.55	.56	.56	327.4
30+ credits completed	.25	.35	.30	.23	.31	.36	.27	.42	.39	.43	.37	.47	.35	.45	.53	586.2
Part-time	.47	.25	.40	.30	.37	.27	.29	.22	.33	.19	.21	.18	.24	.20	.14	838.7
Services used																
Advising	.39	.54	.50	.55	.52	.60	.62	.61	.56	.68	.68	.67	.68	.72	.74	747.2
Tutoring	.14	.22	.18	.22	.19	.34	.26	.22	.25	.39	.32	.51	.35	.34	.52	1,001.5
Skill labs	.25	.36	.31	.41	.36	.44	.45	.38	.38	.51	.5	.52	.51	.52	.59	665.6
Financial aid advising	.34	.43	.40	.47	.43	.48	.52	.46	.48	.54	.54	.54	.58	.59	.62	434.0
Student organization	.08	.17	.11	.14	.13	.21	.16	.18	.16	.37	.23	.34	.21	.25	.45	995.7

(continued)

Appendix B (continued)

	Low engagers		Diverse engagers										High engagers		
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	Cluster 11	Cluster 12	Cluster 13	Cluster 14	Cluster 15	
Variables	*N* = 26,157	*N* = 26,043	*N* = 26,344	*N* = 27,061	*N* = 18,013	*N* = 14,833	*N* = 28,485	*N* = 22,336	*N* = 16,588	*N* = 11,179	*N* = 22,713	*N* = 11,846	*N* = 16,621	*N* = 17,048	*N* = 10,932	*F*
Number of services used																1,832.7
No services	.37	.22	.27	.19	.23	.16	.15	.18	.19	.12	.11	.10	.10	.10	.08	
1 service	.28	.27	.29	.26	.27	.23	.24	.26	.26	.19	.21	.17	.20	.19	.13	
2 services	.20	.24	.24	.27	.25	.25	.27	.27	.26	.22	.26	.22	.26	.25	.19	
3 services	.09	.15	.13	.18	.15	.20	.20	.17	.17	.21	.22	.23	.23	.23	.22	
4 services	.04	.07	.05	.08	.06	.11	.10	.08	.08	.15	.13	.17	.14	.16	.19	
5 services	.02	.04	.02	.03	.03	.05	.04	.04	.04	.11	.06	.11	.06	.08	.19	

Note: Values are shaded to visually highlight patterns at break points of 20% on a scale from 0% to 100%, reflecting (from lightest to darkest) the intensity of engagement.

Acknowledgments — The authors wish to thank Kay McClenney, Jeff Crumpley, and Michael Bohlig with the Center for Community College Student Engagement for their support in allowing the authors access to data from the Community College Survey of Student Engagement (CCSSE). The authors also thank them for their continued encouragement and helpful feedback throughout the writing of this article.

Declaration of Conflicting Interests — The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding — The authors received no financial support for the research, authorship, and/or publication of this article.

Note

1. As the R code used does not handle missing values, missing values were estimated using linear interpolation and rounding to whole digits. Plots from all eight iterations are available from the author.

References

Adelman, C. (2005). *Moving into town—and moving on: The community college in the lives of traditional-age students*. Washington, DC: U.S. Department of Education.

American Association of Community Colleges. (2010). *2010 fact sheet*. http://www.aacc.nche.edu/AboutCC/Documents/factsheet2010.pdf

Ammon, B. V., Bowman, J., & Mourad, R. (2008). Who are our students? Cluster analysis as a tool for understanding community college student populations. *Journal of Applied Research in the Community College*, *16*(1), 32-44.

Angell, L. R. (2009). Construct validity of the Community College Survey of Student Engagement (CCSSE). *Community College Journal of Research and Practice*, *33*, 564-570. doi 10.1080/10668920801901217

Arabie, P., & Hubert, L. J. (1996). An overview of combinatorial data analysis. In P. Arabie, L. J. Hubert, & G. de Soete (Eds.), *Clustering and classification* (pp. 5-63). River Edge, NJ: World Scientific.

Astin, A. W. (1973). Measurement and determinants of the outputs of higher education. In L. Solomon & P. Taubman (Eds.), *Does college matter? Some evidence on the impacts of higher education* (pp. 107-127). New York, NY: Academic Press.

Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel*, *25*, 297-308. Astin, A. W. (1993). An empirical typology of college students. *Journal of College Student Development*, *34*, 36-46.

Astin, A. W. (2001). *What matters in college? Four critical years revisited*. San Francisco, CA: Jossey-Bass.

Bacher, J., Wenzig, K., & Vogler, M. (2004). *SPSS Two step cluster: A first evaluation*. http://www.soziologie.wiso.uni-erlangen.de/publikationen/a-u-d-papiere/a_04-02.pdf

Bahr, P. (2010). The bird's eye view of community colleges: A behavioral typology of first-time students based on cluster analytic classification. *Research in Higher Education*, *51*, 724-749. doi 10.1007/s11162-010-9180-5

Bailey, T., & Alfonso, M. (2005). *Paths to persistence: An analysis of research on program effectiveness at community colleges* (New Agenda Series, Vol. 6, No. 1). http://www.luminafoundation.org/publications/PathstoPersistence.pdf

Bailey, T., Calcagno, J. C., Jenkins, D., Kienzl, G., & Leinbach, T. (2005). *The effects of institutional factors on the success of community college students* (CCRC Brief No. 24). http://ccrc.tc.columbia.edu/Publication.asp?UID=208

Bean, J. P. (2005). Nine themes of college student retention. In A. Seidman (Ed.), *College student retention: Formula for student success* (pp. 215-244). Washington, DC: ACE and Praeger.

Bragg, D. D., Kim, E., & Barnett, E. A. (2006). Creating access and success: Academic pathways reaching underserved students. In D. D. Bragg & E. A. Barnett (Eds.), *Academic pathways to and from the community college* (New Directions for Community Colleges, No. 135, pp. 5-19). San Francisco, CA: Jossey-Bass. doi 10.1002/cc243

Braxton, J. M., & Lien, L. (2000). The viability of academic integration as a central construct in Tinto's interactionalist theory of college student departure. In J. Braxton (Ed.), *Reworking the student departure puzzle* (pp. 11-28). Nashville, TN: Vanderbilt University Press.

Burnett, S. (2006, December 18). Student engagement: More counseling, tutoring and better communication with professors are considered key to improving the success of community college students. *Community College Week*, 6-8.

Cabrera, A. F., Nora, A., & Castaneda, M. B. (1993). College persistence. *Journal of Higher Education*, *64*, 123-139.

Cabrera, A. F., Nora, A., Terenzini, P. T., Pascarella, E., & Hagedorn, L. S. (1999). Campus racial climate and the adjustment of students to college. *Journal of Higher Education*, *70*, 134-160.

Center for Community College Student Engagement. (2010). *The heart of student success: Teaching, learning, and college completion*. http://www.ccsse.org/publications/national_report_2010/36379tw/CCCSE_2010_national_report.pdf

Center for Community College Student Engagement. (2011). *The community college student report*. http://www.ccsse.org/aboutsurvey/biblio/page1.cfm

Cohen, A. M., & Brawer, F. B. (2008). *The American community college* (5th ed.). San Francisco, CA: Jossey-Bass.

Community College Survey of Student Engagement. (n.d.). *CCSSE 2009 codebook*. http://www.ccsse.org/ExampleReport/Codebook/CCSSE%202009%20Codebook.pdf

Creswell, J. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: SAGE.

Duran, B. S., & Odell, P. L. (1974). *Cluster analysis*. New York, NY: Springer-Verlag.

Fraley, C., & Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal*, *41*, 578-588. doi 10.1093/comjnl/ 41.8.578

Galili, T. (2010, June 15). Clustergram: Visualization and diagnostics for cluster analysis (R code) [Web log post]. http://www.r-statistics.com/2010/06/clustergram-visualization-and-diagnostics-for-cluster-analysis-r-code/

Gibson, A. M., & Slate, J. R. (2010). Student engagement at two-year institutions: Age and generational status differences. *Community College Journal of Research & Practice*, *34*, 371-385.

Gnanadesikan, R., Kettenring, J. R., & Tsao, S. L. (1995). Weighting and selection of variables for cluster analysis. *Journal of Classification*, *12*, 113-136. doi 10.1007/BF01202271

Hagedorn, L. S. (2010). The pursuit of student success: The directions and challenges facing community colleges. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 25, pp. 181-218). New York, NY: Springer.

Hagedorn, L. S., & Prather, G. (2005, May-June). *The community college solar system: If university students are from Venus community college students must be from Mars*. Paper presented at the Annual Forum of the Association of Institutional Research, San Diego, CA.

Hardy, A. (1996). On the number of clusters. *Computational Statistics & Data Analysis*, *23*(1), 83-96.

Harper, S. R., Carini, R. M., Bridges, B. K., & Hayek, J. C. (2004). Gender differences in student engagement among African American undergraduates at historically Black colleges and universities. *Journal of College Student Development*, *45*, 271-284.

Harper, S. R., & Quaye, S. J. (2009). Beyond sameness, with engagement and outcomes for all. In S. R. Harper & S. J. Quaye (Eds.), *Student engagement in higher education: Theoretical perspectives and practical approaches for diverse populations* (pp. 1-16). New York, NY: Routledge.

Harris, F., III, & Lester, J. (2009). Gender-specific approaches to enhancing identity development among undergraduate men and women. In S. R. Harper & S. J. Quaye (Eds.), *Student engagement in higher education: Theoretical perspectives and practical approaches for diverse populations* (pp. 99-116). New York, NY: Routledge.

Hartigan, J. A. (1985). Statistical theory in clustering. *Journal of Classification*, *2*, 63-76. doi 10.1007/BF01908064

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A *k*-means clustering algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *28*, 100-108.

Karp, M. M., Hughes, K. L., & O'Gara, L. (2008). *An exploration of Tinto's integration framework for community college students* (CCRC Working Paper No. 12). http://ccrc.tc.columbia.edu/Publication.asp?UID=615

Kasworm, C. E. (1980). Student services for the older undergraduate student. *Journal of College Student Personnel*, *21*, 163-169.

Kaufman, L., & Rouseeuw, P. J. (2005). *Finding group in data: An introduction to cluster analysis*. Hoboken, NJ: Wiley-Interscience.

Ketchen, D. J., & Shook, C. L. (1996). The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal*, *17*, 441-458.

Kuh, G. D. (2004, September). *Student engagement: Understanding it; measuring it; and grasping the implications*. Invited presentation to University of Illinois Higher Education Colloquium, Urbana, IL.

Kuh, G. D., & Gonyea, R. M. (2003). The role of the academic library in promoting student engagement in learning. *College & Research Libraries*, *64*, 256-282.

Kuh, G. D., Kinzie, J., Buckley, J. A., Bridges, B. K., & Hayek, J. C. (2006). *What matters to student success: A review of the literature commissioned report for the National Symposium on Postsecondary Student Success: Spearheading a dialog on student success*. Washington, DC: National Postsecondary Education Cooperative.

Kuh, G. D., Kinzie, J., Schuh, J. H., & Whitt, E. J., & Associates. (2005). *Student success in college: Creating conditions that matter*. San Francisco, CA: Jossey-Bass.

Lamkin, M. D. (2004). To achieve the dream, first look at the facts. *Change*, *36*(6), 12-15.

Manton, K. G., Lowrimore, G., Yashin, A., & Kovton, M. (2005). Cluster analysis: Overview. In *Encyclopedia of Statistics in Behavioral Science* (Vol. 1, pp. 305-315). West Sussex, UK: Wiley. doi 10.1002/0470013192.bsa101

Mardia, K., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. San Diego, CA: Academic Press.

Marti, C. N. (2004). *Overview of the CCSSE instrument and psychometric properties*. http://www.pvc.maricopa.edu/ccsse/docs/psychometrics.pdf

Marti, C. N. (2009). Dimensions of student engagement in American community colleges: Using the Community College Student Report in research and practice. *Community College Journal of Research and Practice*, *33*, 1-24. doi 10.1080/10668920701366867

McClenney, K. (2006). Benchmarking effective educational practice. In J. A. Seybert (Ed.), *Benchmarking: An essential tool for assessment, improvement, and accountability* (New Directions for Community Colleges, No. 134, pp. 47-55). San Francisco, CA: Jossey-Bass doi 10.1002/cc.236

McClenney, K. M., & Marti, C. N. (2006). *Exploring relationships between student engagement and student outcomes in community colleges: Report on validation research*. Austin, TX: Community College Survey of Student Engagement. http://www.ccsse.org/publications/CCSSE%20Working%20Paper%20on%20Validation%20Research%20December%202006.pdf

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, *50*, 159-179. doi 10.1007/BF02294245

Palmer, J. C. (2000). Demographics, state education reform policies, and the enduring community college role as an extension of the schools. In J. C. Palmer (Ed.), *How community colleges can create productive collaborations with local schools* (New Directions for Community Colleges, No. 111, pp. 93-103). San Francisco, CA: Jossey-Bass.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: Vol. 2. A third decade of research*. San Francisco, CA: Jossey-Bass.

Pike, G. R., & Kuh, G. D. (2005). A typology of student engagement for American colleges and universities. *Research in Higher Education*, *46*, 185-209. doi 10.1007/s11162-004-1599-0

Provasnik, S., & Planty, M. (2008). *Community colleges: Special supplement to the condition of education 2008* (NCES 2008-033). http://nces.ed.gov/pubs2008/2008033.pdf

Roman, M. A., Taylor, R. T., & Hahs-Vaughn, D. (2010). The retention index of the Community College Survey of Student Engagement (CCSSE): How meaningful is it? *Community College Journal of Research and Practice*, *34*, 386-401. doi 10.1080/10668920701382484

Romesburg, H. C. (2004). *Cluster analysis for researchers*. Raleigh, NC: Lulu Press.

Sanchez, J. R., & Laanan, F. S. (1997). The economic returns of a community college education. *Community College Review*, *25*(3), 73-87.

Sayer, A. (1992). *Method in social science: A realist approach* (2nd ed.). London, UK: Routledge.

Schonlau, M. (2002). The clustergram: A graph for visualizing hierarchical and nonhierarchical cluster analyses. *The Stata Journal*, *2*, 391-402.

Schonlau, M. (2004). Visualizing non-hierarchical and hierarchical cluster analyses with clustergrams. *Computational Statistics*, *19*, 95-111. doi 10.1007/BF02915278

Schuetz, P. (2002). *Emerging challenges for community colleges*. Los Angeles, CA: ERIC Clearinghouse for Community Colleges. (ERIC Document Reproduction Service No. ED477829)

Shaw, P. J. A. (2003). *Multivariate statistics for the environmental sciences*. Hoboken, NJ: Wiley.

Silverman, S. C., Sarvenaz, A., & Stiles, M. R. (2009). Meeting the needs of commuter, parttime, transfer, and returning students. In S. R. Harper & S. J. Quaye (Eds.), *Student engagement in higher education: Theoretical perspectives and practical approaches for diverse populations* (pp. 223-242). New York, NY: Routledge.

SPSS Inc. (2001). *The SPSS Two step cluster component: A scalable component enabling more efficient customer segmentation*. http://www.spss.ch/upload/1122644952_The%20SPSS%20TwoStep%20Cluster%20Component.pdf

SPSS Inc. (2009). *PASW Statistics 18.0 for Windows user's guide*. Chicago, IL: Author.

Steinley, D. (2006). *k*-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, *59*, 1-34. doi 10.1348/000711005X48266

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, *18*, 267-276. doi 10.1007/BF02289263

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). Chicago, IL: University of Chicago Press.

Tinto, V. (1998). Colleges as communities: Taking research on student persistence seriously. *Review of Higher Education*, *21*, 167-177.

Zaritsky, J. S., & Toce, A. (2006). Supplemental instruction at a community college: The four pillars. In M. E. Stone & G. Jacobs (Eds.), *Supplemental instruction: New visions for empowering student learning* (New Directions for Teaching and Learning, No. 106, pp. 23-31). San Francisco, CA: Jossey-Bass. doi 10.1002/tl.230

Zhao, C. M., & Kuh, G. D. (2004). Adding value: Learning communities and student engagement. *Research in Higher Education*, *45*, 115-138.

About the authors

Victor B. Sáenz is an assistant professor in the Higher Education Administration Program in the Department of Educational Administration at The University of Texas at Austin.

Deryl Hatch was a doctoral student at the University of Texas at Austin. (He is currently Assistant Professor at the University of Nebraska–Lincoln.)

Beth E. Bukoski is a doctoral student at the University of Texas at Austin.

Suyun Kim is a doctoral student at the University of Texas at Austin.

Kye-hyoung Lee is a doctoral student at the University of Texas at Austin.

Patrick Valdez is a doctoral student at the University of Texas at Austin.