

USING LATENT PROFILE METHODOLOGY TO OBTAIN A MIDWESTERN  
COMMUNITY COLLEGE STUDENT TYPOLOGY

A DISSERTATION  
SUBMITTED TO THE GRADUATE SCHOOL  
IN PARTIAL FULFILLMENT OF THE REQUIRMENTS  
FOR THE DEGREE  
DOCTOR OF PHILOSOPHY

BY

KAREN R. HUM

DISSERTATION ADVISOR: DR. W. HOLMES FINCH

BALL STATE UNIVERSITY

MUNCIE, INDIANA

July 2016

Copyright © July, 2016

Karen R. Hum

All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the author.

## ABSTRACT

**DISSERTATION:** Using Latent Profile Methodology to Obtain a Midwestern Community

College Student Typology

**STUDENT:** Karen R. Hum

**DEGREE:** Doctor of Philosophy

**COLLEGE:** Teachers College

**DATE:** July 2016

**PAGES:** 103

The accurate classification of college students is critical if educators are to better understand their student populations. Doing so will enable institutions to target specific student groups with interventions geared toward improving persistence, retention, and overall achievement, which ultimately could result in more effective and efficient approaches. The current study examined the development of student typologies in higher education using student enrollment behavior. The investigation looked at whether a typological model could be shared between institutions, or if regional policies, processes, and student demographics dictated a locally grown solution. Using Latent Profile Analysis (LPA), the results found that each institution would be best served to establish its own model. Additional findings supported the use of behavioral variables, but suggested that the inclusion of gender and race had an impact on a student's group classification.

## DEDICATION

To my husband, children, and mother, who have served as tireless supporters and motivators through this extensive journey. I love and thank you with all my heart.

## ACKNOWLEDGEMENTS

I would first like to thank my family for being my most ardent cheerleaders. You succeeded in keeping me going when I was ready to give up. Also, a big thank-you to Dr. Holmes Finch and Dr. Jerrell Cassady for their knowledge and patience while guiding me through the dissertation process. Lastly, I want to express my gratitude to my entire committee—Dr. Finch, Dr. Cassady, Dr. Sharon Paulson, and Dr. Michelle Glowacki-Dudka—for their time and expertise.

## TABLE OF CONTENTS

ABSTRACT .....	III
DEDICATION .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
ACKNOWLEDGEMENTS .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
LIST OF TABLES .....	X
LIST OF FIGURES .....	XI
CHAPTER I INTRODUCTION .....	1
TYPOLOGIES IN HIGHER EDUCATION .....	1
COMMUNITY COLLEGE STUDENT TYPOLOGIES .....	4
IMPLICATIONS OF STUDENT TYPOLOGIES .....	6
TYPOLOGICAL METHODOLOGIES .....	8
CONCLUSION .....	12
PURPOSE OF THE STUDY .....	12
SIGNIFICANCE OF THE PRESENT STUDY .....	14
CHAPTER II REVIEW OF THE LITERATURE .....	16
ESTABLISHING STUDENT TYPOLOGIES IN HIGHER EDUCATION .....	16
TYPOLOGICAL RESEARCH IN EDUCATION .....	18
<i>Attributes and Institutions</i> .....	18
<i>Historical Effects</i> .....	22
<i>Attitudes, Performance, and Engagement Patterns</i> .....	23
<i>Comparison of Classification Schemes</i> .....	34
ANALYTIC STRATEGIES FOR STUDENT TYPOLOGIES .....	38

<i>Data Clustering</i> .....	38
<i>Demographic v. Behavioral Variables</i> .....	44
<i>Bahr's Methodology</i> .....	46
APPLICATIONS OF STUDENT TYPOLOGIES .....	47
SUMMARY .....	49
CHAPTER III METHODOLOGY .....	51
PARTICIPANTS AND PROCEDURE.....	51
STATISTICAL ANALYSES .....	52
<i>LPA and CLPA Analyses Overview</i> .....	53
CHAPTER IV RESULTS.....	58
LATENT PROFILE ANALYSIS (LPA) VALIDATION OF BAHR'S CLUSTER ANALYSIS .....	58
<i>Model Comparison</i> .....	58
CROSS-VALIDATION OF INITIAL LPA RESULTS .....	63
<i>Model Comparison</i> .....	63
CROSS-VALIDATION OF THE 6-CLASS LPA MODEL.....	71
<i>Model Comparison</i> .....	71
CHAPTER V DISCUSSION .....	76
LPA CROSS-VALIDATION OF BAHR'S CLUSTER ANALYSIS .....	76
CROSS-VALIDATION OF INITIAL LPA RESULTS .....	78
CROSS-VALIDATION OF THE 6-CLASS LPA MODEL.....	80
GENDER AND RACE AS PREDICTORS OF LPA CLASS MEMBERSHIP.....	80
LIMITATIONS AND FUTURE DIRECTIONS .....	82

CONCLUSION..... 83

REFERENCES ..... 85

APPENDIX A BAHR VALIDATION LPA SYNTAX..... 100

APPENDIX B BAHR VALIDATION CLPA SYNTAX..... 101

APPENDIX C LPA VALIDATION SYNTAX ..... 103

## LIST OF TABLES

Table 1. Comparison of Methodologies for Typological Studies Having Conflicting Demographic Outcomes.....	45
Table 2. Analyzed Measures.....	52
Table 3. Posterior Classification Probabilities for a 5-Class Cross-Validation Model.....	59
Table 4. Variable Means for Bahr's 5-Class Cluster Model.....	60
Table 5. Latent Class Means and Standard Deviations for a 5-Class LPA Cross-Validation Model.....	61
Table 6. Frequencies in Gender and Race for Bahr's 5-Class Cluster Model.....	62
Table 7. Frequencies in Gender and Race for a 5-Class LPA Cross-Validation Solution.....	62
Table 8. Fit Statistics for Competing Models.....	64
Table 9. Posterior Classification Probabilities for a 6-Class Covariate Model.....	66
Table 10. Standardized Latent Class Means and 95% Confidence Intervals for a 6-Class Covariate Model.....	67
Table 11. Frequencies in Gender and Race for a 6-Class Covariate Model.....	68
Table 12. Logistic Regression and Odds-Ratio Results for a 6-Class Covariate Model.....	70
Table 13. Posterior Classification Probabilities for a 6-Class Cross-Validation Model.....	72
Table 14. Standardized Latent Class Means and 95% Confidence Intervals for a 6-Class Cross-Validation Model.....	73
Table 15. Frequencies in Gender and Race for a 6-Class Cross-Validation Model.....	74
Table 16. Logistic Regression and Odds-Ratio Results for a 6-Class Cross-Validation Model..	75

## LIST OF FIGURES

Figure 4.1. Latent class means for a 6-class covariate model.....	66
Figure 4.2. Latent class means for a 6-class cross-validation model.....	73

## CHAPTER I

### INTRODUCTION

Since the 1960s, researchers have systematically examined student attitudes, values, and experiences for a better understanding of personal and social factors that are likely to affect student behaviors. Guided by theoretical models of student subtypes, studies have examined the relationship between student attributes and institutions. Over time, numerous researchers, including Adelman (2005), Bahr (2010), Clark and Trow (1966), Horowitz (1987), Katchadourian and Boli (1985), and Zhao, Gonyea, and Kuh (2003), have developed empirically oriented typologies based on students' personality, interests, values, and behavior. By utilizing this classification technique, educators are able to make better sense of the populations they are trying to serve, allowing them to target interventions and approaches geared to improving persistence and retention.

#### **Typologies in Higher Education**

While the practice of classifying students has been common, coming to a consensus on appropriate typologies that can meet the goals for higher education has not been a simple task. Some typologies are more descriptive in nature, focusing on student characteristics, while others relate to student outcomes, or concentrate on motivational and behavioral characteristics, such as student engagement. Certain student classifications, such as Perry's (1970) levels of intellectual and ethical development, can also be characterized as developmental or hierarchical, where students are expected to move upward from stage to stage, with the implication that each successive stage is considered a higher level than the preceding stage. The presumption is that higher stages are preferred over lower stages (Astin, 1993).

In higher education, the typological approach is a method of organizing and viewing cohorts of college students within and across time periods (Kuh, Hu, & Vesper, 2000) to identify homogeneously distinct groups of students as a way to better understand and predict outcomes and behavior (Kuh, 1990; Kuh et al., 2000). Although the focus of student typologies is not on the process or implications of student advancement, typologies can help to reveal the similarities and differences of college students over time (Bahr, 2010).

Early typological research (Clark & Trow, 1966; Horowitz, 1987) was often based more on theory than on empirical evidence such as that seen in recent studies (Bahr, 2010; Kuh, Hu, & Vesper, 2000). However, a truly comprehensive college student typology depends on a combined use of both conceptual and observed evidence (Krathwohl, 1998) as this allows educators to better understand students so as to improve educational outcomes. The science of behaviorism is primarily concerned with observable actions rather than unobservable events such as thoughts or feelings, as the behavioral school of thought holds that behaviors can be described scientifically (Skinner, 1945, 1953; Watson, 1925; Zuckerman, 1994). Identifying student behavior patterns not only assists educators to better understand their students, it can also provide educators with opportunities to address and possibly alter negative behavior, which is not an option with physical characteristics alone (Bahr, 2010). Results comparing differences in engagement measures have consistently demonstrated a predictable relationship between persistence pathways and levels of engagement. For example, Marti (2008) found that students who follow less efficient pathways consistently exhibit lower levels of engagement.

Although nominal, qualitative variables are instrumental to the development of student typologies, taxonomic designs can be quantitative in nature (Luan, 2006). Yet, most have been based on categorical information such as gender, race, and major field of study (Ammon,

Bowman, & Mourad, 2008; Astin, 1993; Clark & Trow, 1966; Horowitz, 1987; Katchadourian & Boli, 1985; Keniston, 1973; Mauss, 1967; Newcomb, Koenig, Flacks, & Warwick, 1967; Perry, 1970). Typologies used to characterize college students' academic performance are not as common. From a quantitative data perspective, however, taxonomies can be applied to continuous variables or traits, such as standardized test scores and grade-point averages (Clagett, 1995; Horn, 2009; Hu & McCormick, 2011).

Bandura's social cognitive theory (Bandura, 1986) suggests that human activity is a function of behavior and person variables—cognitive processes such as attitude, motivation, and goals—as well as the environment. Compared to declared educational goals, however, student behaviors, such as social integration and campus engagement (Astin, 1993; Tinto, 1975) and progression based on course-taking activity (Bahr, 2010, 2011), have been shown to be more reliable measures of whether or not students are actually achieving their desired objectives (Perry, 2005). Pascarella and Terenzini (1991) and Zhao, Gonyea, and Kuh (2003) also noted that student behavior is a much better predictor of optimal college outcomes than is student background information. Still, many educators focus more on student demographics such as gender or race than on student behavior when reporting achievement outcomes.

Classifying students by behaviors rather than by their stated objectives allows educators to sort students into distinct categories based on what students do, supplemented by existing measures such as demographics, major, and GPA to describe who they are (Bahr, 2010). Typological research results have supported the importance of student experience and effort to learning outcomes (Kuh, Hu, & Vesper, 2000). Utilizing large-scale databases such as the National Survey of Student Engagement (2004, 2005), researchers have been able to show some consistency in groups across student typologies. For example, students found to exhibit high

levels of involvement in educational activities have been identified as scholars, intellectuals, or maximizers, whereas those demonstrating low engagement have been classified as disengaged (Hu & McCormick, 2011; Kuh et al., 2000; Zhao, Gonyea, & Kuh, 2003).

Typologies can be designed to capture student individuality by incorporating details regarding their values, attitudes, and behavior. The underlying assumption in typological research is that student experiences and outcomes can be organized in meaningful ways to reflect similarities within and differences across groups (Hu & Li, 2011). Doing so will result in categories that are more descriptive than hierarchical, as one is not necessarily better than another. Although each group is distinct, a linear, developmental process should not be presumed, as students frequently alternate among categories (Bahr, 2010).

Researchers should also be careful to validate typologies by examining the relationships among student types and outcomes such as academic achievement, learning, and persistence (Bahr, 2010). To this end, earlier studies have used direct assessment, grade point averages, and retention (Hu & McCormick, 2011) as well as self-reported student gains (Hu & McCormick, 2011; Kuh et al., 2000).

### **Community College Student Typologies**

Student typologies generally have been utilized for placement and interventions based on student status, and personal and physical characteristics such as gender, age, and race (Astin, 1993). The emphasis, however, has primarily been on students located at four-year institutions (Bahr, 2010). Policy makers often fail to fully note the community college's role in facilitating student career advancement due to a focus on conventional completion measures such as degree or credential attainment and upward transfer to four-year institutions (Bahr, 2010, 2011). Commonly referred to as less traditional, community college student populations generally differ

from the standard four-year institution student profiles in a variety of ways. Students are often disadvantaged, attend part-time, and are less academically accomplished. Yet, a review of community college research (Bailey & Alfonso, 2005) found the quality and quantity of research on community college institutional practice to be inadequate.

Consequently, there is growing awareness of the need to identify and differentiate among various types of community college students to determine who is enrolling, and how and to what end they are using the community college, such as for a terminal degree, eventual transfer to a four-year institution, or job training (Ammon, Bowman, & Mourad, 2008; Attinasi, Stahl, & Okun, 1982; Bahr, 2010, 2011; Clagett, 1995; Hagedorn & Prather, 2005; Horn, 2009; Luan 2006; Marti, 2008; Mauss, 1967). It is crucial for community college stakeholders to have this information available in order to properly address the needs of their student population, especially when facing increased expectations and declining resources. A community college should have a clear understanding of its student population subtypes to ensure students are being served effectively and efficiently (Astin, 1993; Saenz, Hatch, Bukoski, Kim, Lee, & Valdez, 2011; Santibáñez, Gonzalez, Morrison, & Carroll, 2007; Schuck & Zeckhauser, 2008).

According to Bahr (2011), students use the community college in a wide variety of ways to achieve an equally wide variety of ends, some of which align closely with the institutional goals, priorities, and performance indicators, while others do not. Thus, a typology of community college students based on their use of the institution has the potential to be of great informational and interpretive value to community college stakeholders.

To this end, Bahr (2010, 2011) developed a community college student typology based on course-taking behavior and enrollment patterns. Using cluster analysis of longitudinal data, he identified five categories of behavior, and subsequently determined how well each group in

the cluster solution predicted students' long-term academic outcomes as measured by credential completion and upward transfer. Bahr's results suggested that students in the transfer, vocational, and exploratory clusters demonstrated higher probabilities of persistence and course success rates.

### **Implications of Student Typologies**

Interest in college student classification has been driven by a heightened focus on achieving greater student outcomes using increasingly limited resources (Bahr, 2011; Goldrick-Rab, 2010). Typological research findings can be used to guide institutional improvement efforts via the development and delivery of targeted and cost-effective student interventions to maximize outcomes (Bahr, 2011). Thus, an institution's clear and comprehensive understanding of its student population can facilitate policymakers, administrators, and other stakeholders in optimizing student benefits (Astin, 1993; Bahr, 2011; Borden, 1995; Hom, 2009; Santibáñez, Gonzalez, Morrison, & Carroll, 2007). By comparing student typology information with institutional data as it relates to the types of programs and practices available, educators can help to ensure adequate fit between interventions and individuals (Bahr, 2010). Having an improved understanding of its students will also enable higher educators to better determine how and to what extent the institution is helping or hindering its students in their academic efforts.

Additionally, increased emphasis on institutional accountability (Bahr, Hom, & Perry, 2004; Dowd & Tong, 2007; Layzell, 1999) and accurately assessing student outcomes (Dellow & Romano, 2002; Dowd, 2003; Gillmore & Hoffman, 1997) has drawn attention to the need to better distinguish student types, objectives, and needs (Shulock & Moore, 2007). Clarity regarding variation in student composition, both across and within institutions, provides a framework on which to build awareness and understanding of institutional performance (Bahr,

2011; Bahr, Hom, & Perry, 2004, 2005; Dellow & Romano, 2002; Hoachlander, Sikora, & Horn, 2003). For example, driven by efforts to measure college-level rates of upward transfer (Bahr, Hom, & Perry, 2005), institutions have struggled to clearly identify transfer-seeking students (Bradburn & Hurst, 2001; Townsend, 2002; Wassmer, Moore, & Shulock, 2004). Yet, previous attempts to define transfer-seeking students have been based primarily on anecdotal information (Hom, 2009; Townsend, 2002) rather than empirical data. To assure precision in measurement, accountability indices such as transfer rates and degree completion should be carefully operationalized (Hagedorn & Kress, 2008).

In order to design and implement programs to enhance student experiences and improve student learning, educators first need to better understand college students, what they do, and what they gain—or hope to gain—from their college experience. Determining student engagement, or how college students spend their time, is a viable means to understanding college student success. A student's peers can significantly influence student engagement—what that student chooses to do or not to do—and student engagement has been shown to be directly related to student outcomes (Bahr, 2010). Typologies can be employed to identify students who are more likely to engage in educationally oriented activities, such as intellectuals (Katchadourian & Boli, 1985), scientists and individualists (Ku, Hu, & Vesper, 2000), and maximizers (Zhao, Gonyea, & Kuh, 2003). Conversely, they can also be used to determine which students show a higher probability of participating in activities associated with less desirable outcomes, such as recreators and socializers (Hu & McCormick, 2011), unconventional (Zhao et al., 2003), and disengaged (Hu & McCormick, 2011; Kuh et al., 2000; Luan, Zhao, & Hayek, 2009; Zhao et al., 2003). By using these student typologies, a community college could also determine the proportion of students who demonstrate behaviors and outcomes that align

with its institutional mission (Zhao et al., 2003), and even predict how students are likely to respond to interventions (Bahr, 2011).

Community colleges attract students of exceptionally diverse backgrounds, pursuing a broad range of academic objectives with widely varying levels of academic preparation (Adelman, 2005; Bahr, 2011; Goldrick-Rab, 2010; Hagedorn & Prather, 2005; Hoachlander, Sikora, & Horn, 2003; Kim, 2002; Kim, Sax, Lee, & Hagedorn, 2010; Laanan, 2000; VanDerLinden, 2002; Voorhees & Zhou, 2000). This expanse of student characteristics and ambitions highlights the democratizing function of community colleges in the United States (Dowd, 2003) and the growing role of community colleges as their local institutions of higher education (Noftsinger & Newbold, 2007; Shaw & Jacobs, 2003; Wang, 2004). This extreme variation in student population, however, can complicate the effective delivery of student services and support (Hagedorn, 2010; Shannon & Smith, 2006), while also impeding accurate measurement of institutional performance (Bahr, 2011; Bahr, Hom, & Perry, 2004, 2005; Bailey, Calcagno, Jenkins, Leinbach, & Kienzl, 2006; Dowd, 2007; Lane, 2003; Seybert, 2002).

### **Typological Methodologies**

A variety of approaches have been used to determine student typologies, including cluster analysis, factor analysis, path analysis, latent trajectory analysis, and latent class analysis. Of these, the two primary methods shown to be valuable in this line of inquiry are cluster analysis and latent class analysis. Cluster analysis is clearly the more common approach, but not necessarily the most appropriate. To have a better understanding of the extant literature, it is important to be clear on the benefits and limitations of each method as applied to student classification.

Typology is defined as the science and practice of classification (Bahr, 2010). Although the use of typologies to organize information is not limited to higher education, researchers have long used typologies to study college students. Typology is often equated with taxonomy, with both having been described as the two essential approaches to the classification process (Lewis-Beck, 1994). Typology, however, has been considered more conceptual and deductive, and taxonomy empirical and inductive (Bailey, 1994). Taxonomies differ from typologies in that they classify items on the basis of observable and measurable characteristics. Astin (1993) noted that “it is virtually impossible to carry on a meaningful conversation about American college students without invoking taxonomic language” (p. 36).

Taxonomy has been regarded more as a scientific system of classification (Sokal & Sneath, 1963) bound by specific rules regarding nomenclature, hierarchical classification, and interrelationships among groups (Fenske, Keller, & Irwin, 1999). Although these limitations have sometimes inhibited the use of taxonomy in fields related to social sciences and education, taxonomic methods such as cluster analysis have been successfully employed in a variety of disciplines (Lorr, 1983; Mezzich & Solomon, 1980). While typologies have fewer restrictions when classifying observations with no obvious commonalities (Fenske et al., 1999), drawbacks to their use can include categories that are neither exhaustive nor mutually exclusive, are descriptive rather than explanatory or predictive, and are based on subjective criteria (Bailey, 1994).

The terms typology and taxonomy suggest that individuals can be sorted into distinct and independent categories, or clusters. Data clustering is the process of dividing data elements into classes or clusters so that items in the same class are as similar as possible, while items in different classes are as dissimilar as possible (Kaymak & Setnes, 2000). The two fundamental

questions to be addressed in any clustering scenario are how many clusters are actually present in the data, and how real or good is the clustering itself (Sousa & Kaymak, 2002).

In hard clustering, data are divided into distinct clusters, where each data element belongs to exactly one cluster. In fuzzy, or soft, clustering, data elements can belong to more than one cluster, and associated with each element is a set of membership levels, which indicate the strength of the association between that data element and a particular cluster (Kaymak & Setnes, 2000). Fuzzy clustering is a process of designating these membership levels, and then using them to assign data elements to one or more clusters (Sousa & Kaymak, 2002). Fuzzy clustering recognizes that no group is truly distinct and isolated from another, as there tends to be an overlap among types.

An increasingly popular alternative to cluster analysis in determining student typologies is latent class analysis (Dugan, 2011). Latent class analysis (LCA) is a statistical method that can identify underlying groups based on a set of observed response variables (Finch & Bronk, 2011). The function of LCA is to find subtypes of related cases (latent classes) from multivariate categorical data. When the indicators are continuous, however, latent profile analysis (LPA) is preferred instead (Vermunt & Magidson, 2002).

LCA is similar to cluster analysis in that it is used to discover groups or types of cases based on observed data and to assign cases to groups. LCA is considered an exploratory modeling technique (Hojtink, 2001) to be used anterior to determining the number of latent classes underlying data (Laudy, Boom, & Hoijtink, 2005). Conversely, confirmatory latent class analysis (CLCA) allows a researcher to hypothesize as to the number and type of latent classes in a particular set of data prior to analysis (Finch & Bronk, 2011).

In purpose, LCA is closely related to CA as both are used to discover groups or types of cases based on observed data. However, LCA can also be used to determine if sub-classes exist in a cluster solution that are not apparent through cluster analysis (Magidson & Vermunt, 2002). An important difference between standard cluster analysis techniques and LCA is that LCA is a model-based approach, such that a statistical latent class (LC) model is theorized for the population from which the data sample is obtained (Lazarsfeld, & Henry, 1968). With non-hierarchical techniques such as k-means clustering, the assignment of objects to clusters is based on specified criteria (e.g. student demographics or types of behaviors). This process typically involves minimizing within-cluster and maximizing between-cluster variation (Lorr, 1983). Advantages of using a statistical model such as LCA include the approach involves rigorous statistical tests and the choice of the cluster criterion is less random (Magidson & Vermunt, 2002).

In LC modeling, the criteria used to make decisions regarding the number of classes are considered more formal, as a number of statistics are available that can assist with choosing one model over another, such as the BIC statistic (Samuelsen & Raczynski, 2013). This is not the case with k-means, however, as the researcher must determine the number of clusters without the benefit of formal diagnostic statistics (Magidson & Vermunt, 2002). While k-means uses an ad hoc approach for classification, the LC method is probability-based, in that cases are classified into clusters using model-based posterior probabilities estimated by maximum likelihood (ML) methods, which also yield estimates for misclassification rates (Muthén & Muthén, 2007).

LCA's model-based clustering allows working with variables of mixed-measurement levels to be relatively simple as compared to k-means, in which variables must be standardized beforehand to ensure equal variance so as to avoid clusters dominated by variables with the most

variation (Magidson & Vermunt, 2002). Conversely, standardization of variables in LC clustering is not needed. Discriminant analysis is commonly used following a k-means clustering to describe differences between the clusters on one or more possible covariates. In contrast, the LC model can be broadened easily to include demographic and other exogenous variables, allowing for the classification and cluster description to be performed at the same time (Magidson & Vermunt, 2002).

### **Conclusion**

Understanding its student population can enable an educational institution to better serve those students' needs, and developing a typology of student types is a good way to classify students so as to target interventions based on those different types. It seems, however, that most of the empirical data available to promote the use of student typologies relates to four-year institutions. Because community college students are generally less traditional, it is highly unlikely that typologies developed for four-year students will prove adequate. Clearly, continued typological research into students of two-year institutions is needed.

Bahr's publishing of his process (2011) for determining student classes suggests that his methodology is transferrable across institutions. Due to differences in student behavior and institutional processes, however, use of his model should be approached carefully. Student classes determined via inaccurate or inappropriate means could result in ineffective interventions and wasted resources.

### **Purpose of the Study**

The main purpose of the current study is to validate Bahr's 2011 cluster analysis methodology and findings of community college students' course-taking behavior to ascertain if comparable results can be obtained when utilizing a different regional sample and alternative

statistical analyses. An ancillary, applied focus of the study is to determine a more sound method for classifying students for the purpose of targeted interventions. This study is interested in the following questions:

1. When validating Bahr's 2011 methodology using Confirmatory Latent Profile Analysis and constraining for Bahr's cluster solution, will similar classes be identified in a Midwestern sample of first-time community college students as determined by model fit, variable means, and comparison of group size?
2. Using Latent Profile Analysis to compare models comprised of Bahr's eight variables based on student enrollment behavior in a sample of Midwestern community college students not limited to first-time attendees, which model will be determined to be the best-fitting?
3. Can the best-fitting model confirmed in Question 2 be validated via Confirmatory Latent Profile Analysis when using a new sample of Midwestern community college students not limited to first-time attendees?
4. Does including a student's gender and race impact a student's latent class assignment?

It is hypothesized that the use of a different regional sample of first-time students with Bahr's methodology will result in divergent outcomes with respect to the nature and number of typological categories. This expectation is based on the limitation of Bahr's research to California Community College Students, as the sample to be used in this study will be Midwestern community college students, and his use of cluster analysis, which is a non-statistical method of classification that can be somewhat limiting in scope. Also, Bahr's student typology was based on student enrollment behaviors only without consideration of any confounding variables that might have influenced the results. It is anticipated that variations in

student demographics will influence the outcome with respect to the makeup of the resulting typological categories.

### **Significance of the Present Study**

The accurate classification of college students is critical if educators are to better understand their student populations. Doing so will enable institutions to target specific student groups with interventions geared toward improving persistence, retention, and overall achievement, which ultimately could result in more effective and efficient approaches. In this era of increased emphasis on institutional accountability, educators are continually being asked to improve student outcomes using fewer resources. Having a clear and comprehensive understanding of its student population will enable an institution—as well as its policymakers and other stakeholders—to maximize student benefit while more accurately measuring student outcomes. Community college student populations are generally more diverse, which underscores the need to have a clear understanding of student subtypes to ensure students are being served both effectively and efficiently. In order to design and implement programs to enhance college student experiences and improve student learning, educators first need to better understand what students do, and what they gain—or hope to gain—from their college experience. Determining student engagement, or how college students spend their time, is crucial to facilitating college student success.

Typological research findings also can be used to guide institutional improvement efforts. Increased clarity regarding variation in student composition, both across and within institutions, can provide a framework on which to build awareness and understanding of institutional performance. By comparing student typology information with institutional data as they relate to the types of programs and practices available, educators can help to ensure adequate fit between

interventions and individuals. Basically, having an improved understanding of its students will allow a community college to better determine how and to what extent the institution is helping or hindering its students in their academic efforts.

Bahr's development of a community college student typology used a sample of first-time California Community College students, enrolled fall 2001 through summer 2008. Bahr has made available a detailed accounting of his methodology so that other researchers might be able to reproduce his findings. To date, however, the literature do not show that this has been done. Replicating Bahr's analysis using a different regional sample of both first-time and non-first-time students will establish whether his published methodology is generalizable across both different community colleges and geographical regions. Additionally, validating his methodology using latent profile analysis will determine if sub-classes exist in his cluster solution that are not apparent through k-means.

Replication is the key to the support of any worthwhile theory. However, it can also involve applying the theory to new situations in an attempt to determine the generalizability of the outcomes to different age groups, locations, races, or cultures. Replication, therefore, is important for a number of reasons, including (a) assurance that results are valid and reliable; (b) determination of generalizability or the role of extraneous variables; (c) application of results to real-world situations; and (d) inspiration toward new research. At the same time, using a common set of indicators will help to advance not only local understanding, but also the field in identifying relevant variables that schools should systematically track and document in their systems.

## CHAPTER II

### REVIEW OF THE LITERATURE

This chapter provides a review of the literature as it pertains to the current study. The chapter first defines typologies and describes their use in higher education, with a look at the statistical analyses commonly used, as well as a consideration of the importance of careful variable selection. The relevance of gender, race, and SES to student typology outcomes is subsequently addressed. The three different categories of student typologies are then outlined, using chronologically organized empirical examples as support for each. This is followed by a historical comparison of study outcomes for the different classification schemes and a closer look at the benefits of Bahr's methodology, as well as a discussion of how student typologies have been applied in the field.

#### **Establishing Student Typologies in Higher Education**

Typology has been defined as the science and practice of classification (Bahr, 2010). Although the use of typologies to organize information is not limited to higher education, researchers have long used typologies to study college students. Typology is often equated with taxonomy, with both having been described as the two essential approaches to the classification process (Lewis-Beck, 1994). Typology, however, has been considered more conceptual and deductive, and taxonomy empirical and inductive (Bailey, 1994). Taxonomies differ from typologies in that they classify items on the basis of observable and measurable characteristics. Astin (1993) noted that "it is virtually impossible to carry on a meaningful conversation about American college students without invoking taxonomic language" (p. 36).

Taxonomy has been regarded more as a scientific system of classification (Sokal & Sneath, 1963) bound by specific rules regarding nomenclature, hierarchical classification, and

interrelationships among groups (Fenske, Keller, & Irwin, 1999). Although these limitations have sometimes inhibited the use of taxonomy in fields related to social sciences and education, taxonomic methods such as cluster analysis have been successfully employed in a variety of disciplines (Lorr, 1983; Mezzich & Solomon, 1980). While typologies have fewer restrictions when classifying observations with no obvious commonalities (Fenske et al., 1999), drawbacks to their use can include categories that are neither exhaustive nor mutually exclusive, are descriptive rather than explanatory or predictive, and are based on subjective criteria (Bailey, 1994).

Researchers have systematically examined student attitudes, values, and experiences for a better understanding of personal and social factors that are likely to affect student behaviors. Guided by theoretical models of student subtypes, studies have examined the relationship between student attributes and institutions. Over time, numerous researchers, including Adelman (2005), Bahr (2010), Clark and Trow (1966), Horowitz (1987), Katchadourian and Boli (1985), and Zhao, Gonyea, and Kuh (2003), have developed empirically oriented typologies based on students' personality, interests, values, and behavior. By utilizing this classification technique, educators are able to make better sense of the populations they are trying to serve, allowing them to target interventions and approaches geared toward improving persistence and retention.

Coming to a consensus on appropriate typologies, however, has not been a simple task. Some typologies are more dispositional in nature, focusing on student descriptions and traits, while others relate to student outcomes, or concentrate on motivational and behavioral characteristics, such as student engagement. Additionally, the methods used to determine student typologies have varied as well.

## **Typological Research in Education**

Since 1960, three lines of inquiry have examined student attitudes, values, and experiences in two- and four-year institutions to better understand the factors most likely to affect student behavior (Luo & Jamieson-Drake, 2005): (a) attributes and institutions; (b) historical effects; and (c) attitudes, performance, and engagement patterns. The first two strategies have used theoretical models of student subtypes to guide the development of typological categories. Type one has focused on variations in student attributes (e.g. interest in school) with respect to institutional characteristics such as social structure. Type two has investigated student traits directly focused on historical effects of the purpose and function of higher education—essentially, what the institution has to offer. The third approach has operated more from a data-driven, empirically derived orientation to typologies based on students' attitudes, performances, and engagement patterns, including student progression activity. For all three approaches, empirical data from four-year institutions far outweighs that of two-year. Additionally, the analyses and subsequent outcomes of the various studies have been mixed, if not conflicting, suggesting a need for further research.

### **Attributes and Institutions**

One of the most frequently cited student typologies is Clark and Trow's (1966) qualitative analysis of student data collected in the late 1950s and early 1960s at the University of California. To determine how larger social structures impacted a student's life and relationships, two primary "orientations" were used to categorize students into subsets: identification with the college and involvement with ideas (i.e., intellectual inquiry). These categories were then combined to create four distinct student subcultures: (a) collegiate, (b) vocational, (c) academic, and (d) nonconformist. Students in the collegiate group showed high

identification with the college and low involvement with ideas. Students in the vocational group displayed low identification with the college and low involvement with ideas, taking part in few college activities, as they viewed college as a stepping stone to a good job. Students in the academic group scored high on both dimensions, while the nonconformists, who ranked low in institutional identification, were highly interested in intellectual matters and issues related to art, literature, and politics.

Similarly, in an examination of student self-report data at Bennington College, Newcomb, Koenig, Flacks, and Warwick (1967) identified two categories—individualism and intellectualism—that led to their identification of six dominant student subgroups: (a) creative individualists, (b) wild ones, (c) scholars, (d) the social group, (e) leaders, and (f) political activists. With strong beliefs in principles and commitment to creative pursuits, creative individualists ranked high on both individualism and intellectuality. Students in the wild group displayed high individualism but low intellectuality, caring more for wild parties than academic work. Scholars, focusing more on academics, scored high on intellectualism but low on individualism, while the social group ranked low on both dimensions, as their interest was solely in social life and having fun. Leaders participated actively in the student government and were popular among major student groups, while political activists were interested in public affairs, civil rights, campus politics, and social conditions.

Continuing the intellectual theme, Perry's (1970) findings included ethical development rather than individualism, identifying college students' tendencies and directions in growth when moving from less to more complex reasoning. Due to individual differences, however, these paths were not considered definite, nor did they advance at identical rates. Perry proposed that college students progress through four major stages of intellectual and ethical development: (a)

dualism, (b) multiplicity, (c) relativism, and (d) commitment. These four stages were then subdivided into nine “positions.” Dualism was defined as received knowledge, suggesting that there are both right and wrong answers, whereas multiplicity was subjective knowledge. Relativism prescribed that individuals should learn to evaluate solutions, while commitment held that individuals integrate what they learn from others with personal experience and reflection.

Katchadourian and Boli (1985) examined a cohort of Stanford students from their freshman year to graduation and classified them on a two-dimensional scheme that also included intellectualism, but paired it instead with careerism. This interaction resulted in four types of students: (a) careerists, (b) intellectuals, (c) strivers, and (d) unconnected. By definition, careerists viewed higher education mainly as getting training for a career, while intellectuals aspired to enjoy the learning process while extending their knowledge. Strivers attempted to excel in their studies and, at the same time, prepare for a successful career of their choice, while the unconnected scored low on both dimensions and failed to engage in college.

In 1980, Pascarella and Terenzini developed a theoretical model to predict student dropout, which resulted in two distinct typologies: persisters and voluntary dropouts. Using Tinto’s Student Integration Model (1975) as a basis for their research, the authors employed a combination of factor, discriminant, and classification analyses of four-year student response data to demonstrate the effects of academic and social integration on a student’s level of commitment.

Tinto’s model holds that academic and social integration consist of several basic components, which consist of a student’s commitments to an institution as well as goals associated with graduation and career (Tinto, 1975). According to the model, as the level of institutional and goal commitment increase, the likelihood of the student persisting at the

institution will also increase. Pascarella and Terenzini (1980) noted that their findings generally supported “the predictive validity of the major dimensions of the Tinto model” (p. 72), and suggested that student-faculty relationships, as measured by interactions with faculty and faculty concern for student development and teaching scales, contributed highly to student persistence.

In 2002, VanDerLinden used cluster analysis to investigate response data from 100,000 community college students across 300 institutions. The study identified six clusters based on the students’ stated reasons as to why they attended community college: (a) skill upgrader (upgrading skills for career advancement); (b) career prep; (c) life changer (recently encountered a major life change); (d) personal enrichment/transfer; (e) transfer only; and (f) no definite purpose for enrolling. The results found the majority of study participants to be enrolled for career preparation (29%), followed closely by personal enrichment/transfer (21%) and transfer only (21%). Only 2% of the respondents were classified as having no definite purpose for enrolling.

Like VanDerLinden (2002), Ammon, Bowman, and Mourad (2008) used cluster analysis of community college students to identify three similar categories: (a) skill upgraders, (b) career advancers, and (c) transfer students. Each of these categories was subsequently organized into sub-clusters. The skill-upgraders category, comprised of students who enrolled to increase their skill and knowledge, was subdivided into personal enrichment and nontraditional student clusters. The career advancer group was shown to be interested more in education used for employment purposes, and included the clusters (a) certificate for work, (b) skills for work, (c) degree for the computer field, and (d) degree for other fields: culinary arts, applied technologies, or math, natural and behavioral sciences fields. Lastly, the transfer student clusters included (a)

degree to transfer to four year; (b) non-degree to transfer to four year; and (c) dual enroll to transfer to high school or community college.

Although he also used cluster analysis of community college students, Horn's (2009) groupings were based on three different attributes: (a) student intent to complete or transfer to a four-year institution; (b) attendance intensity; and (c) enrollment in a formal program of study for degree seekers. The purpose of creating the taxonomy was to investigate its relationship with three-year outcomes related to enrollment status and degree attainment. Using longitudinal data, the researchers identified three clusters of student direction: (a) strongly, (b) moderately, and (c) not directed. Students within each cluster were then further divided into tracks based on their program of study (e.g. associate's degree or vocational certificate). Subsequent logistic regression analysis showed that strongly directed students had higher odds of retention and completion.

### **Historical Effects**

Other research into student typologies has focused more on characterizing student traits through the functions of colleges and the purposes of higher education. Examining the missions of colleges and universities, Keniston (1973) identified seven types of students: (a) gentleman-in-waiting, (b) apprentice, (c) big man on campus, (d) professionalist, (e) underachiever, (f) activist, and (g) disaffiliate. According to Keniston, colleges in early times largely served to educate children from the privileged class, so students at the time were considered gentlemen-in-waiting, as they primarily went to college to prepare themselves for their birthright as part of the upper-class life. In the early twentieth century, the function of education began to change from vocational training to the teaching of such soft skills as the ability to be likable and persuasive. Thus, a student with especially strong social prowess began to emerge as the big man on campus.

Later, the development of modern technology gave rise to new types of students: the professionalist, who had high professional expertise and was capable of handling technical problems, as well as the activist, the disaffiliate, and the underachiever. Engaging in student demonstration, the activists protested against the university or society to press for reform or improvement. The disaffiliates were politically inactive but culturally alienated students, whereas the underachievers were students who accepted their own deficiencies.

Horowitz (1987) looked at Yale students and their subcultures from a historical perspective, classifying them into four major groups: (a) college men, (b) outsiders, (c) rebels, and (d) new outsiders. Outsiders were serious about academic work, whereas college men/women were more interested in the social life in college. New outsiders, serious about academic work, were more concerned about job opportunities after college, while rebels were concerned about social issues on and off campus.

### **Attitudes, Performance, and Engagement Patterns**

A third line of research into student typologies examines college student attitudes, performance, and engagement patterns in college activities. Houle (1961) developed a typology that described the orientations of adult education participants. Based on audiotaped interviews with 22 continuing education participants in the Chicago area, Houle proposed a typology to characterize their motivational influence: (a) goal-, (b) learning-, or (c) activity-oriented. Goal-oriented learners used education to accomplish specific objectives, while the activity-oriented students found meaning in the activity of learning itself. Learning-oriented students were simply interested in obtaining knowledge.

Armand Mauss (1967) also used self-report data to adapt Clark and Trow's (1966) typology, incorporating sociological factors into the educational performance of junior college,

rather than university, students. Differences in student classifications were said to be based primarily on students' value commitments to the "adult world" and their pursuit of intellectual ideas. These classifications included the (a) academic type, (b) vocational type, (c) incipient rebel, and (d) perpetual teenager. Mauss found that the academic and vocational types generally tended to have stronger self-concepts and identified with the adult community. However, the academic type was involved with ideas, or idea-oriented, while the vocational type was not. The incipient rebel was also involved with ideas but did not identify with the adult community, whereas the perpetual teenager did neither.

Similar to Houle and Mauss, Attinasi, Stahl, and Okun (1982) analyzed interview data to categorize community college students based on their motivational orientations toward enrollment in individual courses. Students were compared based on how they defined (a) their major criterion of success; (b) the aspect of education they most emphasized; and (c) their willingness to endure a delay in application of learning. Findings showed that most students could be classified as either (a) requirement meeters (task-oriented with a goal to earn credit hours and/or a good grade but spending as little time as possible on school-related activities); (b) knowledge seekers (genuinely interested in the subject matter); or (c) specific information users (focused more on obtaining practical knowledge or skills).

Based on ratings of student performance on a wide range of dimensions, Taber and Hackman (1976) used factor analysis to identify twelve distinct patterns of Yale student performance classified into two groups: success types and nonsuccess types. The success types included (a) leaders, (b) scholars, (c) careerists, (d) grinds, (e) artists, (f) athletes, and (g) socializers. Conversely, the nonsuccess types consisted of (a) extreme grinds and (b) disliked, (c) alienated, (d) unqualified, and (e) directionless students.

Taber and Hackman (1976) found that leaders ranked highest in organizational participation and played leading roles in student organizations and activities, while the scholars ranked highest in intellectual performance. Noted for their remarkable mathematical proficiency, the grinds attached great emphasis to academic performance. The artists were characterized by exceptionally high levels of artistic performance, whereas the athletes were differentiated by their exceptionally high levels of athletic performance. Socializers were rated relatively high on interpersonal sociability but low on academic dimensions. Among the nonsuccess types of students, the extreme grinds failed to balance academic work with nonacademic aspects of college life, while the disliked students scored low on all personal and interpersonal behaviors. The alienated students had high ratings on artistic performance, but they had not yet developed career plans, were low in self-direction, and were unhappy with their college experience. With poor academic performance, the unqualified students were viewed as unlikely candidates for any advanced study. The directionless students, chiefly interested in socializing, identified least with college.

Also using factor analysis, Astin (1993) examined student behavioral characteristics at the time of college entry to identify seven types of students comparable to those found by Taber and Hackman (1976): (a) scholar, (b) leader, (c) hedonist, (d) status striver, (e) social activist, (f) artist, and (g) uncommitted. Scholars displayed a high degree of academic and intellectual self-esteem with high aspirations for academic success, while the social activists showed an elevated level of activity, assertiveness, and social involvement. Artists scored high on self-ratings of artistic ability and values, whereas hedonists were largely defined by their party behaviors. Status strivers were committed to career and financial success, and leaders had high self-ratings on popularity, social self-confidence, and leadership ability, viewing themselves as popular and

outgoing. The uncommitted, however, exhibited more disengaged behaviors, including changing major or career choice, dropping out of college, or transferring to another college before graduating.

Clagett (1995) adopted a more outcomes-based approach, noting that community college assessment measures focusing exclusively on graduation rates are misleading. He suggested that a student outcomes typology should (a) be accepted as legitimate by all parties to whom the college is accountable; (b) take into account the full range of student goals; and (c) acknowledge student enrollment behavior patterns. To that end, seven outcomes-based categories were developed: (a) award and transfer, (b) transfer/no award, (c) award/no transfer, (d) sophomore status in good standing, (e) persisters, (f) special motive, and (g) other exiters. Award and transfer students earned a degree or certificate, and then transferred to a four-year institution. The transfer/no award students transferred to a four-year institution without earning a degree or certificate. Award/no transfer students were those who had earned a degree or certificate, but for whom there was no record of transfer. Students in the sophomore status in good standing category had not graduated but had earned at least 30 credits with a cumulative GPA of 2.0. The persisters category identified students who were still in the college but did not fit any of the other categories, while the other exiters left the college without graduating or earning 30 credits, and with no record of transfer. The special motive students were those who indicated short-term, non-degree goals of personal enrichment or job skill upgrading.

Kuh (1995) grouped undergraduate students based on interview responses and used qualitative analysis and cross-tabulation to determine the relationship between the students' out-of-class experiences and five outcome domains. These included (a) interpersonal competence, (b) practical competence, (c) cognitive complexity, (d) knowledge and academic skills, and (e)

humanitarianism. The eight types of experiences, or antecedents, were defined as (a) leadership, (b) peers, (c) academic, (d) faculty, (e) work, (f) travel, (g) ethos, and (h) other. Kuh's results showed that leadership and work experiences highly contributed to practical competence, while gains in cognitive complexity were more often associated with academic activities, other experiences, and institutional ethos. In addition, peers were found to be the most important influence in the areas of humanitarianism, interpersonal competence, and cognitive complexity.

As a response to Kuh's (1995) recommendation for an increase in research on the interaction between student behaviors and outcomes-based characteristics, Dugan (2011) employed latent class analysis to develop a typology reflective of how patterns of college student involvement in clubs and organizations related to leadership development. Eight unique categories were identified: (a) affinity group affiliates, (b) identity and expression leaders, (c) cultural collegiate, (d) academic careerists, (e) recreational academics, (f) athletes, (g) social recreators, and (h) social collegiates.

Dugan's (2011) affinity group affiliates were connected to a limited number of associations, while the identity and expression leader members were involved in an average of six different types of group experiences, including those that were leadership-oriented. Students described as cultural collegiates reported the highest levels of involvement and showed the greatest probability of leadership activity. The academic careerists largely focused on group activities related to specific disciplines or career interests. Students in the recreational academics category were also more likely to participate in groups related to their academic interests, but this included intramurals as well. Group involvement for the athletes was primarily limited to sports and intramurals, whereas social recreators encompassed not only students involved in intramurals and sports, but also sororities and fraternities. Lastly, the social collegiates, who

more closely resembled classifications described by other researchers (Astin, 1993; Clark & Trow, 1966; Kuh, Hu, & Vesper, 2000; Taber and Hackman, 1976), maintained levels of group involvement that were high across most campus activities.

In 2000, Kuh, Hu, and Vesper developed an activities-based typology by classifying four-year students based on the quality of their educational efforts, which were then linked to self-reported college gains. Combining factor and cluster analysis to link student types to patterns of behaviors and self-reported outcomes, the researchers identified ten major groups:

(a) disengaged, (b) recreator, (c) socializer, (d) collegiate, (e) scientist, (f) individualist, (g) artist, (h) grind, (i) intellectual, and (j) conventional. Students in the disengaged group scored below average on all college activities, whereas recreators devoted a considerable amount of effort to sports and exercise but were below average in most other activities. Socializers had substantial social interaction with their peers, while collegiates distinguished themselves by their active involvement in co-curricular activities. Scientists scored extremely high on science and quantitative activities, whereas individualists interacted often with peers and participated in artistic activities. Artists scored high on artistic activities and faculty interaction, and the grinds characterized themselves by a high level of academic effort. Intellectuals were engaged in all types of college activities, while conventionals displayed a mixed pattern of involvement.

Also using factor analysis, Zhao, Gonyea, and Kuh (2003) assayed National Survey of Student Engagement (NSSE) data to determine five “psychographic” categories of student engagement: (a) student-faculty interactions; (b) experiences with diversity; (c) academic effort; (d) out-of-class experiences; and (e) integrative activities. Subsequent cluster analysis resulted in eight groups of students: (a) unconventional, (b) collegiate, (c) vocational, (d) conventional, (e) grinds, (f) academics, (g) maximizers, and (h) disengaged. While most of the groups shared

characteristics already identified through other researchers' typologies (Astin, 1993; Dugan, 2011; Kuh, Hu, & Vesper, 2000; Mauss, 1967; Taber and Hackman, 1976), the unconventional category was unique and included students of below average social and academic activity who often interacted with people of diverse backgrounds.

Luan, Zhao, and Hayek (2009) employed data mining followed by cluster analysis of the NSSE dataset to identify eight student-engagement typologies: (a) high interaction, (b) traditional-learning focused, (c) homework-emphasized, (d) diverse and spread, (e) meeting service needs, (f) disengaged, (g) collegiate, and (h) easy-pass. The high-interaction, traditional-learning-focused, and collegiate students were found to be more traditional, enrolling full-time and having few off-campus responsibilities. Conversely, the diverse-and-spread, meeting-service-needs, disengaged, and easy-pass students were generally part-time and non-traditional students, working professionals, and minorities. The collegiate student was shown to be "conventional," while the high-interaction student was described as a "doer or busy bee" and the traditional-learning focused student was a "learner and teacher's pet." The homework-emphasized student was an "avid reader," the diverse and spread student was considered a "driller," and the meeting service needs students were described as "contented." The disengaged students, however, were found to be "over-challenged," whereas easy-pass students were "insufficiently challenged."

Hu and McCormick (2011) expanded upon earlier analyses (Kuh, Hu, & Vesper, 2000; Zhao, Gonyea, & Kuh, 2003) to further refine the engagement-based student typology. Also utilizing NSSE data, the authors examined the utility of this typology in understanding learning outcomes and student persistence. Using cluster analysis to demonstrate conceptually related student behaviors, experiences, and perceptions, seven groups of student types—(a) academics,

(b) unconventional, (c) disengaged, (d) collegiate, (e) maximizers, (f) grinds, and (g) conventionals—were identified based on the five different NSSE benchmarks (2004, 2005): (a) level of academic challenge, (b) active and collaborative learning, (c) student-faculty interaction, (d) enriching educational experiences, and (e) supportive campus environment. Relationships between these resultant student types and outcome variables such as grade point average (GPA), self-reported gains, and persistence from first to second year of college were subsequently determined via linear and logistic regression. Findings suggested that certain groups of students (e.g. Academics and Maximizers) are more successful in college because they spend time on such educationally purposeful activities as academics. In addition, students who were more engaged in campus and academic activities were shown to be more likely to persist to a second year at the same institution. Stressing that student engagement is important to student learning and development, the results suggested the diagnostic potential of an engagement-based typology to identify students at risk for limited learning and attrition.

Focusing more on the impact of academic process on student success, Boughan (2000) utilized structural equation modeling (SEM), as well as path and cluster analysis, to identify ten distinct achievement-based categories of students: (a) extra effort, (b) supported scholars, (c) collegians, (d) true grit, (e) pragmatists, (f) full-time struggle, (g) part-time struggle, (h) vanishers, (i) unprepared, and (j) casuals. Extra effort, supported scholars, and collegians were described as high achievers with elevated levels of course performance and study load scores. True grit students were considered to be of high medium achievement, as they experienced problems with remedial and credit courses while still managing to achieve. Average achievement clusters included the pragmatists, who were more oriented to occupational courses and job-related goals, and full-time strugglers, who typically entered college unprepared and

demonstrated only moderate drive and persistence. The low-achievement clusters were found to include the part-time strugglers, vanishers, unprepared, and casuals. The students in these groups tended to have high dropout and stopout rates, with low levels of achievement.

Adelman (2005) developed a student typology for community college learners based on the number of college credits earned, which resulted in three clusters: (a) homeowners, (b) tenants, and (c) visitors. Students in the homeowners category were described as credential-oriented (vocational), while those in the tenant group were shown to be using the community college as a means to transfer to a four-year institution. Visitors were those students earning less than 30 credits and typically terminating without a credential.

To develop his five-category typology of student engagement, Cox (2011) used qualitative data to examine faculty-student interactions outside the classroom, which resulted in (a) disengagement, (b) incidental contact, (c) functional interaction, (d) personal interaction, and (e) mentoring. Disengagement was defined as an absence of interaction between faculty members and students outside the classroom, whereas personal interaction suggested faculty members and students engaged in purposeful interactions revolving around shared interests. Incidental contact was considered unintentional and arose when a student and faculty member interacted just by being in the same place at the same time. Functional interactions were characterized as more formal, as they related to student and faculty academic relationships, such as those that took place during office hours. Mentoring occurred when a student and faculty member worked so closely together that personal and functional interactions began to converge.

Knight (2013) used cluster analysis to measure outcomes-based learning for engineering students based on National Academy of Engineering strategies for the “engineers of 2020” (E2020). This resulted in seven distinct typologies: (a) E2020 engineers, (b) theory-focused

engineers, (c) connecting engineers, (d) E2020-lite engineers, (e) non-reflective engineers, (f) E2020 deficient engineers, and (g) professionally oriented engineers. E2020 engineers reported high proficiencies on all outcomes, whereas E2020 deficient reported the lowest proficiencies overall. The E2020-lite students reported well-rounded outcomes, but were still less than the E2020 group. The connecting category contrasted with the E2020 group in fundamental skills, although they were very similar otherwise. The non-reflective engineers were also similar to the E2020 students, but scored markedly lower on the Reflective Behavior Practice scale. The professionally oriented group was most similar to the E2020 students, but reported lower proficiencies on dimensions of interdisciplinary competence and design/contextual awareness.

Typologies focusing on students' actions with respect to progression through an institution also can enable researchers to predict student outcomes. Concentrating on what a student does rather than who that individual is has been shown to be the strongest predictor of educational gains (Kuh, Kinzie, Schuh, & White, 2005; Pascarella & Terenzini, 2005).

Robinson (2004) used path analysis of longitudinal data to merge indicators of undergraduate enrollment, progress, and completion into six categories: (a) commencing student; (b) continuing with no repeats; (c) unit of study repeats following failure; (d) stopout (temporary); (e) transfer (enrollment in another degree course at the same university); and (f) no enrollment at the university. According to the study, a benefit of this procedure is its adaptability and resistance to changes in attendance from full- to part-time.

In their study of community college student progression, Hagedorn and Prather (2005) chose four variables from the literature to inform their cluster analysis: (a) lowest level of math enrollment; (b) stated intention to transfer; (c) average number of units per semester; and

(d) total number of semesters enrolled. The math cluster variable, based on Adelman's (1999) finding that mathematics was one of the most important predictors of community college success, was defined as the first or lowest level of mathematics in which a student enrolled.

Hagedorn and Prather (2005) subsequently used a solar system metaphor to characterize the resultant groupings, suggesting that the further a group of students was from the sun (full-time attendance), the less likely those students were to achieve their goals. Planet one was comprised of traditional students, planet two included full-vocs (vocational), and planet three, transfer-bound. Students aligned with planet four were identified as transfer-hopefuls, while those in planets five through seven were described as industrious, brief-stint, and uni-course, respectively.

Taking a data-mining approach, Luan (2006) explored three indicators of community college student behavior—course volume, unit load, and course withdrawals, as demonstrated by an adjustment factor—to identify six clusters of learners: (a) careful nibblers, (b) confident unit loaders, (c) well-adjusted course packers, (d) overly burdened, (e) total withdraw, and (f) unit maximizers. Careful nibblers were found to be low on all three indices, while the confident unit loaders were high on unit load only. The well-adjusted course packers were shown to be high on course volume but low on unit load and adjustment factor. The overly burdened students were high on both course volume and adjustment factor, whereas the unit maximizers were highest on unit load. The total withdraw students, however, dropped all courses.

Marti (2008) modeled enrollment patterns of community college students to determine five distinct persistence pathways: (a) full-time, long-term; (b) two years and out; (c) long-term decliners; (d) part-time, long-term; and (e) one term and out. Using latent trajectory analysis to examine diversity in persistence pathways related to student characteristics and behaviors, Marti

found group membership to be predictive of mean differences in such factors as class assignments, collaborative learning, information technology, and student services.

Bahr (2010, 2011) developed a community college student typology based on course-taking behavior and enrollment patterns. Using cluster analysis of longitudinal data, he identified five categories of behavior: (a) Completion Likely, (b) Career Technical Education, (c) Completion Unlikely, (d) Skills Builder, and (e) Noncredit. Bahr subsequently determined how well each group in the cluster solution predicted students' long-term academic outcomes as measured by credential completion and upward transfer. The results suggested that students in the transfer, vocational, and exploratory clusters demonstrated higher probabilities of persistence and course success rates. However, experimental students, averaging just two semesters of enrollment, experienced elevated rates of failure. Bahr later replicated his cluster analysis solution using a second sample of community college students, demonstrating that the identified student types were consistent across cohorts.

### **Comparison of Classification Schemes**

Across typological studies, students have been grouped by similarities in preferences, predispositions, and values (Clark & Trow, 1966; Horowitz, 1987; Katchadourian & Boli, 1985; Keniston, 1973; Mauss, 1967; Newcomb, Koenig, Flacks, & Warwick, 1967; Perry, 1970), and attitudes, behaviors, and outcomes (Adelman, 2005; Ammon, Bowman, & Mourad, 2008; Astin, 1993; Attinasi, Stahl, & Okun, 1982; Bahr, 2010, 2011; Boughan, 2000; Clagett, 1995; Cox, 2011; Dugan, 2011; Hagedorn & Prather, 2005; Horn, 2009; Houle, 1961; Hu & McCormick, 2011; Knight, 2013; Kuh, 1995; Kuh, Hu, & Vesper, 2000; Luan, 2006; Luan, Zhao, & Hayek, 2009; Marti, 2008; Pascarella & Terenzini, 1980; Robinson, 2004; Taber & Hackman, 1976; VanDerLinden, 2002; Zhao, Gonyea, & Kuh, 2003). Recent typological studies based on student

activities and behaviors were found to be consistent with earlier research showing that students learn and develop from their engagement in educationally oriented, or purposeful, activities (Bahr, 2010, 2011; Hu & McCormick, 2011; Kuh, Hu, & Vesper, 2000; Pascarella & Terenzini, 1991, 2005). Self-reported gains, GPAs, and persistence have been shown to be preferred outcomes associated with student engagement (Hu & McCormick, 2011; Kuh et al., 2000).

The collegiates group described in more recent typologies (Hu & McCormick, 2011; Kuh et al., 2000; Luan, Zhao, & Hayek, 2009; Zhao, Gonyea, & Zu, 2003) was similar to Clark and Trow's (1966), whereas the intellectuals in Kuh et al. (2000), and maximizers in Hu and McCormick (2011) and Zhao et al. (2003) mirror Clark and Trow's (1966) academics. Zhao, Gonyea, and Kuh's (2003) and Hu and McCormick's (2011) academic group was comparable to Astin's (1993) scholar, Katchadourian and Boli's (1985) intellectual striver, Taber and Hackman's (1976) scholar, and Clark and Trow's (1966) academic. Zhao et al.'s (2003) results, which included a vocational group similar to Clark and Trow's (1966), was comparable to Taber and Hackman's (1976) and Katchadourian and Boli's (1985) careerist.

Several typologies, including Taber and Hackman (1976), Kuh, Hu, and Vesper (2000), and Hu and McCormick (2011), included a group of students similar to Clark and Trow's (1966) grinds, who appeared to place more value on academic work. In addition to grinds, Hu and McCormick (2011), Kuh et al. (2000), and Zhao et al. (2003) all identified collegiate, conventional, and disengaged groups. The disengaged group of students was shown to demonstrate substantial decreases in learning outcomes, academic achievement, and persistence, as evidenced by below-average scores in all college activities (Hu & McCormick, 2011). Recreators (Kuh et al., 2000; Taber & Hackman, 1976), said to comprise a tenth of all undergraduates, and the disengaged were identified as a major concern due to an increased focus

on student learning outcomes and low graduation rates. Despite an average overall effort based primarily on elevated sports and exercise factor scores, recreators did not report comparable college gains (Kuh, Hu, & Vesper, 2000).

Additionally, student engagement in a wide range of activities, as demonstrated by intellectuals (Astin, 1993; Kuh et al., 2000; Taber & Hackman, 1976) and maximizers (Astin, 1993; Hu & McCormick, 2011; Taber & Hackman, 1976; Zhao, Gonyea, & Zu, 2003), was shown to be more beneficial for students than emphasizing individual activities such as social interaction with peers (socializers; Kuh et al., 2000), sports and recreation (recreators; Kuh et al., 2000; Taber & Hackman, 1976), the arts (artists; Astin, 1993; Kuh et al., 2000; Taber & Hackman, 1976), or academics (grinds; Kuh et al., 2000; Taber & Hackman, 1976; Zhao et al., 2003).

Zhao, Gonyea, and Kuh's (2003) and Hu and McCormick's (2011) findings also complemented Kuh, Hu, and Vesper's (2000) results by not only identifying similar types of student groups, but by connecting these groups to student outcomes as well. The unconventional group, which contained a large percentage of part-time students, was almost at the opposite end of the engagement spectrum, with below-average social and academic activity (Zhao et al., 2003; Hu & McCormick, 2011). Nine percent of students in Zhao, Gonyea, and Kuh's (2003) study and 10 percent from Hu and McCormick's (2011) were labeled as maximizers, a group shown to be highly engaged in all measured outcomes.

Consistent with Pascarella and Terenzini's (1991, 2005) observation that equal numbers of most student types are found in four-year colleges, recent typological research has indicated that, generally speaking, student types are not directly related to institutional characteristics (Hu, Katherine, & Kuh, 2011). Zhao, Gonyea, and Kuh (2003), however, found the majority of

students at liberal arts colleges to be collegiates and maximizers, while Kuh, Hu, and Vesper (2000) determined that individualists were overrepresented at highly selective colleges. Small colleges also showed slightly larger proportions of intellectuals and artists, with greater numbers of scientists found at research-oriented universities (Kuh, Hu, & Vesper, 2000).

Student background information can help to explain some of the differences across student types. Zhao, Gonyea, and Kuh (2003) determined maximizers to generally be traditional-age, full-time students who were more satisfied overall with their college experience than were other groups. Hu and McCormick (2011) and Zhao et al. (2011) found the unconventional group, which included students of below-average social and academic activity, frequently interacted with individuals of diverse perspectives and backgrounds. Kuh, Hu, and Vesper (2000) suggested that student group membership may be influenced by a combination of psychosocial developmental stage and amount of college experience.

Student typologies have also shown group membership to be stable over the course of college enrollment. Kuh et al. (2000) found that for first-year students, the largest groups were the conventionals, disengaged, and socializers. The number of students in these groups, however, appeared to drop sharply from freshman to sophomore year, suggesting that students in these classifications may not persist into their second year. Another possibility is that, having successfully managed the psychosocial challenges common to the first year of college, these students modified their pattern of engagement such that they identified with a different student group or groups during subsequent years (Kuh et al., 2000).

Peers have been shown to influence most aspects of college student development (Astin, 1993; Pascarella & Terenzini, 1991, 2005). A student's friends and affinity groups can be highly influential with respect to the quantity and quality of time that the student uses to study or

socialize (Clark & Trow, 1966; Kuh & Whitt, 1988). Flacks and Thomas (2007) noted that, although college attendance affected students' values and orientations in many ways, this was greatly impacted by a student's social environment. The attitudes and behaviors that are the basis for a particular student type can provide insight into those individuals with whom a student is spending time and explain peer effects on college outcomes (Kuh, Hu, & Vesper, 2000; Kuh & Whitt, 1988).

### **Analytic Strategies for Student Typologies**

A variety of approaches have been used to determine student typologies, including cluster analysis, factor analysis, path analysis, latent trajectory analysis, and latent class analysis. Of these, the two primary methods shown to be valuable in this line of inquiry are cluster analysis and latent class analysis. Cluster analysis is clearly the more common approach, but not necessarily the most appropriate. To have a better understanding of the extant literature, it is important to be clear on the benefits and limitations of each method as they apply to student classification.

### **Data Clustering**

The terms typology and taxonomy suggest that individuals can be sorted into distinct and independent categories, or clusters. The objective of clustering is to find structure in data (Pakhira, Bandyopadhyay, & Maulik, 2004). Data clustering is the process of dividing data elements into classes or clusters so that items in the same class are as similar as possible, while items in different classes are as dissimilar as possible (Kaymak & Setnes, 2000). Depending on the nature of the data and the purpose of the clustering, various measures of similarity may be used to assign items to classes, where the similarity measure controls how the clusters are formed (Kaymak & Setnes, 2000). Ultimately, however, the two fundamental questions to be addressed

in any clustering scenario are how many clusters are actually present in the data, and how real or good is the clustering itself (Sousa & Kaymak, 2002).

**Cluster analysis.** Cluster analysis is a common method of classifying students (Adelman, 2005; Ammon, Bowman, & Mourad, 2008; Bahr, 2010, 2011; Bahr, Bielby, & House, 2011; Boughan, 2000; Hagedorn & Prather, 2005; Horn, 2009; Hu & McCormick, 2011; Knight, 2013; Ku, Hu, & Vesper, 2000; Luan, 2006; Luan, Zhao, & Hayek, 2009; Marti, 2008; VanDerLinden, 2002). It involves data mining techniques designed to identify groups of similar observations in otherwise undifferentiated data (Rapkin & Luke, 1993). The method is intended to sort individuals into subsets, or clusters that share similar characteristics while differing in significant ways from other subsets (Borden, 2005; Jain & Dubes, 1988; Lorr, 1983; Punj & Stewart, 1983).

The goal of cluster analysis is to find the combination of observations that maximizes both within-group homogeneity and between-group heterogeneity (Bahr, Bielby, & House, 2011; Borden, 2005; Cormack, 1971; Everitt, Landau, Leese, & Stahl, 2011; Tryon, 1939, 1958). The extent to which student observations attributed to a given cluster share similar characteristics, based on the variables included in the cluster analysis, is referred to as within-group homogeneity. Between-group heterogeneity, however, describes the degree to which each cluster differs from other clusters as they relate to the variables included in the analysis. Basically, an optimal cluster solution groups together those students who are most alike on the variables of interest while creating clusters that are most different from one another on the variables of interest (Bahr, Bielby, & House, 2011).

With respect to community college students, clusters may be formed on the basis of such measures as student aspirations (VanDerLinden, 2002), student course-taking and enrollment

behaviors (Adelman, 2005; Bahr, 2010, 2011; Bahr et al., 2011; Hagedorn & Prather, 2005; Horn, 2009; Luan, 2006), student demographics (Ammon, Bowman, & Mourad, 2008), or a combination of these or other characteristics (Boughan, 2000). In large student data sets, the identification of clusters can also assist in determining trends in behaviors or characteristics (Mezzich & Solomon, 1980). Because cluster analysis is a mechanical rather than statistical procedure (Aldenderfer & Blashfield, 1984), however, it is prudent for researchers to include conceptual evidence when reviewing empirically produced groups of students in order to determine whether the resulting clusters are appropriate.

A number of student classification studies have used cluster analysis to categorize community college students (Adelman, 2005; Ammon, Bowman, & Mourad, 2008; Bahr, 2010, 2011; Hagedorn & Prather, 2005; Horn, 2009; Luan, 2006; Marti, 2008; VanDerLinden, 2002). Some of this research, however, was rather limited in scope. For example, several studies were based on first-time students (Bahr, 2010, 2011; Boughan, 2000; Horn, 2009; Kuh, Hu, & Vesper, 2000; Marti, 2008; Robinson, 2004), or those who had not previously enrolled in college. While the intent is to prevent duplication of students, this method can cause certain student populations (e.g. returning or career and technical/noncredit) to appear underrepresented in the results.

Although Hagedorn and Prather's (2005) study was longitudinal in nature and analyzed a sample of 15,296 students across nine college campuses, it was restricted to a single geographic district and incorporated only four variables: self-reported goal of upward transfer, first or lowest math course enrollment, average unit load per semester, and total number of enrolled semesters. In particular, one of these variables—self-reported goal (Ammon, Bowman, & Mourad, 2008; Astin, 1993; Boughan, 2000; Luo & Jamieson-Drake, 2005; Pascarella & Terenzini, 1980)—has been determined by researchers to be an unreliable measure of student intent (Adelman, 2005;

Hom, 2009). Similarly, VanDerLinden's (2002) and Kuh's (1995) studies focused on a self-reported measure of students' reasons for enrolling in college, which has been shown to carry the same reliability concerns as self-reported goal (Bahr, 2010).

***K-means cluster analysis.*** One form of cluster analysis, k-means, is frequently used (Ammon, Bowman, & Mourad, 2008; Bahr, 2010, 2011; Boughan, 2000; Hu & McCormick, 2012; Knight, 2013; Luan, 2006; Luan, Zhao, & Hayek, 2009; Zhao, Gonyea, & Kuh, 2003) due to its simple process of classifying a given data set through a fixed number of clusters (k) determined a priori (Likas, Vlassis, & Verbeek, 2003). The goal of the analysis is to define k centers, one for each cluster. An advantage of k-means is that it is a fast, robust, and efficient procedure that is relatively easy to comprehend. Disadvantages include the required a priori specification of the number of cluster centers, and an inability to accommodate non-linear data or outliers. A key limitation of k-means is its use of a cluster model in which clusters are expected to be of similar size.

***Fuzzy clustering.*** In hard clustering, data are divided into distinct clusters, where each data element belongs to exactly one cluster. In fuzzy, or soft, clustering, data elements can belong to more than one cluster, and associated with each element is a set of membership levels, which indicate the strength of the association between that data element and a particular cluster (Kaymak & Setnes, 2000). Fuzzy clustering is a process of designating these membership levels, and then using them to assign data elements to one or more clusters (Sousa & Kaymak, 2002). Fuzzy clustering recognizes that no group is truly distinct and isolated from another, as there tends to be an overlap among types.

***Latent class analysis.*** Latent class analysis (LCA) is an increasingly popular alternative to cluster analysis in determining student typologies (Dugan, 2011). LCA is a statistical method

for identifying unmeasured class membership among subjects using categorical or continuous observed variables (Finch & Bronk, 2011). LCA allows for model comparison, with the benefit of fit statistics to determine which model is the better fit for the data. For example, LCA could be used to categorize students based on their drinking behaviors (observations) into different types of drinkers (latent classes), which could lead to finding categories such as abstainers, social drinkers, and alcohol abusers. Models could then be created to predict why one would fall into a particular class membership (e.g. which people become alcohol abusers and why) and to explore the consequences of such class memberships (e.g. if being an alcohol abuser predicts other variables).

LCA can identify underlying groups based on a set of observed response variables (Finch & Bronk, 2011). The function of LCA is to find subtypes of related cases (latent classes) from multivariate categorical data. When the indicators are continuous, however, latent profile analysis (LPA) is preferred instead (Vermunt & Magidson, 2002). LPA is simply LCA for continuous data. LCA is considered an exploratory modeling technique (Hojtink, 2001) to be used anterior to determining the number of latent classes underlying data (Laudy, Boom, & Hoijtink, 2005). Conversely, confirmatory latent class analysis (CLCA), in which constraints are placed on model parameters, allows a researcher to hypothesize as to the number and type of latent classes in a particular set of data prior to analysis (Finch & Bronk, 2011).

Dugan's (2008, 2011) use of LCA to determine latent factors in student involvement helped to provide a framework for recognizing relationships among and between types of co-curricular group involvement experiences. The results suggested that existing survey research data could be collapsed to align with latent factors or classes and then used in further analyses of student needs, trends, and outcomes.

**Comparison of methods.** In purpose, latent class analysis (LCA) is closely related to cluster analysis as both are used to discover groups or types of cases based on observed data. However, LCA can also be used to determine if sub-classes exist in a cluster solution that are not apparent through cluster analysis (Magidson & Vermunt, 2002). An important difference between standard cluster analysis techniques and LCA is that LCA is a model-based approach, such that a statistical latent class (LC) model is theorized for the population from which the data sample is obtained (Lazarsfeld, & Henry, 1968). With non-hierarchical techniques such as k-means clustering, the assignment of objects to clusters is based on specified criteria (e.g. student demographics or types of behaviors). This process typically involves minimizing within-cluster and maximizing between-cluster variation (Lorr, 1983). An advantage of using a statistical model such as LCA is that the approach includes rigorous statistical tests and the choice of the cluster criterion is less random (Magidson & Vermunt, 2002).

In LC modeling, the criteria used to make decisions regarding the number of classes are considered more formal, as a number of statistics are available that can assist with choosing one model over another, such as the BIC statistic (Samuelsen & Raczynski, 2013). This is not the case with k-means, however, as the researcher must determine the number of clusters without the benefit of formal diagnostic statistics (Magidson & Vermunt, 2002). While k-means uses an ad hoc approach for classification, the LC method is probability-based, in that cases are classified into clusters using model-based posterior probabilities estimated by maximum likelihood (ML) methods, which also yield estimates for misclassification rates (Muthén & Muthén, 2007).

LCA's model-based clustering allows working with variables of mixed-measurement levels to be relatively simple as compared to k-means, in which variables must be standardized beforehand to ensure equal variance so as to avoid clusters dominated by variables with the most

variation (Magidson & Vermunt, 2002). In LC clustering, however, standardization of variables is not needed. Discriminant analysis is commonly used following a k-means clustering to describe differences between the clusters on one or more possible covariates. In contrast, the LC model can be broadened easily to include demographic and other exogenous variables, allowing for the classification and cluster description to be performed at the same time (Magidson & Vermunt, 2002).

### **Demographic v. Behavioral Variables**

The issue of variable selection in the classification process is extremely important (Borden, 1995, 2005). Because groupings are designed to minimize differences within clusters and maximize differences between clusters on all variables included in the analysis (Boughan, 2000; Rapkin & Luke, 1993), cluster formation is highly sensitive to variable selection (Fowlkes, Gnanadesikan, & Kettenring, 1988). Thus, researchers should take care to focus on selecting variables germane to their research questions (Punj & Stewart, 1983; Rapkin & Luke, 1993), as the inclusion of unrelated variables reduces the ability of cluster analytic algorithms to return ideal solutions (Milligan, 1980).

Student typographical research outcomes related to the demographic variables gender, race, and socioeconomic status (SES) are mixed (Table 1). Several researchers have shown no statistically significant impact of gender or race (Horn, 2009; Kuh, 1995; Kuh, Hu, & Vesper, 2000; Luan, 2006) or SES (Kuh et al., 2000) on a student's inclusion in a particular typology. Ammon, Bowman, and Mourad (2008) determined that lower  $F$  values for GPA, gender, race, new-student status, and full-time enrollment indicated that these background variables played less of a role in differentiating clusters, suggesting that academic variables were more important in categorizing students. Ammon et al.'s (2008) analysis included a number of informative

variables, such as employment status, course unit load, program of study, and non-degree status. However, their research was limited to two small cross-sectional samples ( $n_1 = 403$ ;  $n_2 = 435$ ) from a single community college, and, akin to Hagedorn and Prather (2005) and VanDerLinden (2002), it incorporated variables of questionable utility, such as minority status, sex, age, and self-reported goal. Ammon et al.'s (2008) study, which was similar to VanDerLinden's (2002) focus on educational goals, had issues generalizing across cohorts of students and proved to be sensitive to confounding variables.

Table 1. Comparison of Methodologies for Typological Studies Having Conflicting Demographic Outcomes

Study	Sample Origin	Analysis Used	Type of Student	Type of Data
Boughan, 2000	2-year college	Factor, Path, Cluster	First-time	Process, Transcript
Dugan, 2011, 2013	4-year college	Latent class	All	Self-report
Hagedorn & Prather, 2005	2-year college	Cluster	All	Self-report Transcript
Hu & McCormick, 2012	4-year college	Cluster	First-time	Self-report
Knight, 2013	4-year college	Cluster	All	Self-report
Luan, Zhao, & Hayek, 2009	4-year college	Cluster	All	Process, Transcript
Mauss, 1967	2-year college	Cluster	All	Self-report
Pascarella & Terenzini, 1980	4-year college	Factor	All	Self-report
VanDerLinden, 2002	2-year college	Cluster	All	Self-report

Luan (2006) found students' "biological markers" (p. 13) were less informative than their academic behaviors as performance measures. Regardless of gender or ethnicity, all students stood an equal chance to be in any of the learner typologies, as student behavior determined a student's classification (Luan, 2006). Hagedorn and Prather (2005) concurred, noting that academic background, course intensity, and transfer intentions appear to be stronger descriptors of students than are gender and ethnic demographics.

Outcomes from other studies, however, help to illustrate a lack of consensus on the impact of including demographic variables. Some researchers (Mauss, 1967; Pascarella & Terenzini, 1980) have determined only gender to have an effect, while others (Dugan, 2011, 2013; Hagedorn & Prather, 2005; Hu & McCormick, 2012; Knight, 2013; Luan, Zhao, & Hayek, 2009; VanDerLinden, 2002) have found both gender and race to be significant. Boughan (2000) showed just SES and race to be moderating variables. As seen by the lack of trends in Table 2.1, however, the reason for these differences is not readily apparent.

### **Bahr's Methodology**

Compared to the other published typological methodologies, Bahr's (2010, 2011) offers a number of advantages. Unlike many of the other researchers (Dugan, 2011, 2013; Hagedorn & Prather, 2005; Hu & McCormick, 2012; Knight, 2013; Mauss, 1967; Pascarella & Terenzini, 1980; VanDerLinden, 2002), Bahr used student progression rather than self-report data. He also excluded demographic characteristics from the cluster analytic process, which increased the external validity of his results. By focusing on course-taking behavior and enrollment patterns instead of demographic variables to determine his typology, Bahr's outcomes were intended to not only be generalizable to other samples at the same institution, but to other institutions as well. To this end, Bahr's published methodology (2011) was designed to simplify the process of developing student typologies at other colleges. Bahr's seven-year longitudinal study involved an extremely large sample of students ( $N = 165,921$ ), and unlike VanDerLinden's (2002) and Adelman's (2005) focus on for-credit only, he included both for-credit and non-credit students. Bahr also tested the efficacy of his cluster solution with respect to predicting students' long-term academic outcomes—specifically, credential completion and upward transfer—which, although

integral to the process, is often neglected during the development of college student typologies (Kuh et al., 2000).

Bahr noted that interest in the classification of community college students is motivated by a number of important objectives beyond the broad scientific goal of advancing knowledge and understanding of community colleges. For example, in measuring rate of transfer, not all students who enroll in a community college intend to transfer (Hom, 2009), so the calculation of transfer rate should not include all enrolled students in the denominator. However, past efforts to define transfer-seeking have been based primarily on anecdotal evidence and supposition (Hom, 2009; Townsend, 2002). Thus, the development of a comprehensive classification scheme of community college students is an important step toward determining empirically based operational definitions of student classifications such as transfer-seeking and degree-seeking. Similarly, decisions regarding the operationalization of institutional accountability measures, including transfer rate, degree completion rate, etc., should be equally informed (Hagedorn & Kress, 2008).

### **Applications of Student Typologies**

Determining a student typology is only the first step—appropriately applying the typology in a way to enhance student interventions is the ultimate goal. The development of student typologies can help colleges to better understand the range of students they are serving. Group membership can be used to predict successes such as persistence and graduation (Ammon, Bowman, & Mourad, 2008; Pascarella & Terenzini, 1980) as well as potential failures, and including student types in exploratory regression models can improve accuracy and explanatory power when compared to models that pool different subgroups together (Zhao, Gonyea, & Kuh, 2003).

Student engagement also has been shown to be important to learning and development. Luan (2006) found GPA and other outcome data to be powerful predictors of persistence when analyzed by a typology of student behaviors. Hu and McCormick (2012) showed that distinct patterns of engagement corresponded to different patterns of learning and development in the first year of college, and different rates of persistence to the second year, even after controlling for demographic variation. Their results demonstrated dramatic differences in predictions of persistence by student type, suggesting the diagnostic potential of an engagement-based typology to identify students at risk in the areas of learning and attrition.

Looking at student engagement through the process of dropping out longitudinally rather than via cross-section is also important (Tinto, 1975). Robinson's (2004) pathway technique is described as a means of categorizing the progress of students over time, while allowing easy comparisons between institutions, fields of study, or cohorts within the same field of study. As a longitudinal technique, it can help to provide a fuller understanding of the process of dropping out, which can then be used in the validation of longitudinal retention models.

By determining how reported learning outcomes may vary across different groups of students, clusters can be used to improve instructor accountability and inform institutional decision-making in such areas as minority under-representation (Knight, 2013). Because predictions based on typologies may be more robust, policy makers and practitioners can more effectively and efficiently direct institutional resources to students who are more likely to benefit (Katchadourian & Bolin, 1985). Frequently revising a typology is discouraged, however. Once the learner types within a typology have been identified, institutions should continue to use these patterns to try to segment, or classify, incoming students as a way for advisors to monitor change (Luan, 2006) and for administrators to use data to support targeted interventions (Knight, 2013).

## Summary

This chapter provided a review of the literature as it pertains to the current study. The chapter first defined typologies and described their use in higher education, with a look at the statistical analyses commonly used, as well as a consideration of the importance of careful variable selection. The relevance of gender, race, and SES to student typology outcomes was subsequently addressed. The three different categories of student typologies were then outlined, using chronologically organized empirical examples as support for each. This was followed by a historical comparison of study outcomes for the different classification schemes and a closer look at the benefits of Bahr's methodology, as well as a discussion of how student typologies have been applied in the field.

Due to the considerable amount of student typological research available, resulting in many different types of student typologies, determining which direction is best for an institution to follow can prove difficult. A common thread throughout the literature, however, is that the development of a student typology can be highly subjective. Typologies based on student behaviors such as course enrollment and campus engagement seem to offer more predictive power. But even though established typologies may suggest a common trait, such as a behavior that regularly defines a particular type of student, all typologies are essentially local and relational (Luan, 2006). Thus, it is important for each institution to develop its own typology based on its own student population.

Bahr's publishing of his process (2011) for determining student classes suggests that his methodology is transferrable across institutions. Due to differences in student behavior and institutional processes, however, use of his model should be approached carefully. Student

classes determined via inaccurate or inappropriate means could result in ineffective interventions and wasted resources.

Validating Bahr's analysis using a different regional sample of both first-time and non-first-time students will show whether his published methodology is generalizable across both different community colleges and geographical regions. Additionally, using latent profile analysis will determine if sub-classes exist in the population that are not apparent through k-means.

## CHAPTER III

### METHODOLOGY

#### **Participants and Procedure**

The participants in the study were part of a two-year (four-semester) cohort of community college students from a Midwestern two-year institution. The sample included full-time, first-time students, which excluded students who enrolled in college at any point prior to cohort, and full-time non-first-time attendees. The sampled student population found more than 129,000 students, including part-time and dual-credit, enrolled statewide. Of this group, 36% were considered minority, 53% female, and 37% aged 25 or older. Full-time enrollments totaled nearly 53,000, of which 32% were classified as minority, 55% female, and 40% aged 25 or older.

Student demographic and enrollment data were obtained from the college's statewide Banner (Oracle) database using a Discoverer Plus query. In keeping with Bahr's methodology, students were removed from the analyses in the event of missing data, and the remaining students were then randomly assigned to three samples: one with first-time enrollees only (n=10,892) and two with all student types (n=13,822 and n=13,814). For each sample, the student enrollment data were transformed into eight continuous variables as defined by Bahr (2011; Table 2), and prior to analysis, the continuous variables in each dataset were standardized.

Table 2. Analyzed Measures

Variable	Abbreviation	Definition
Course success	CS	Percentage of courses in which student earned a grade of C or better
Mean credits attempted	MCA	Mean number of credits attempted during fall and spring terms
Non-credit courses attempted	NCA	Count of non-credit courses attempted
Non-transferrable degree credits attempted	NTDCA	Credits attempted in all non-transferrable degree courses
Terms enrolled	TE	Count of terms in which a student was enrolled (fall, spring, summer)
Transferrable English credits attempted	TECA	Credits attempted in all transferrable English courses
Transferrable math credits attempted	TMCA	Credits attempted in all transferrable math courses
Transferrable other credits attempted	TOCA	Credits attempted in all transferrable other courses

### Statistical Analyses

The current study seeks to address four questions: 1) When validating Bahr's 2011 methodology using Confirmatory Latent Profile Analysis and constraining for Bahr's cluster solution, will similar classes be identified in a Midwestern sample of first-time community college students as determined by model fit, variable means, and comparison of group size? 2) Using Latent Profile Analysis to compare models comprised of Bahr's eight variables based on student enrollment behavior in a sample of Midwestern community college students not limited to first-time attendees, which model will be determined to be the best-fitting? 3) Can the best-fitting model confirmed in Question 2 be validated via Confirmatory Latent Profile Analysis when using a new sample of Midwestern community college students not limited to first-time attendees? 4) Does including a student's gender and race impact a student's latent class assignment?

## **LPA and CLPA Analyses Overview**

Latent profile analysis (LPA) is an increasingly popular alternative to cluster analysis in determining student typologies (Dugan, 2011). LPA is a statistical method for identifying unmeasured class membership among subjects using continuous observed variables (Finch & Bronk, 2011). However, LPA may also serve simply as a data-reduction tool. LPA allows for model comparison, with the benefit of fit statistics to determine which model is the better fit for the data (Nylund, Asparouhov, & Muthén, 2007).

LPA is considered an exploratory analysis such that there is often no well-developed theory regarding the number and type of latent groups to be found in the population sample (Laudy, Boom, & Hoijtink, 2005). Thus, LPA should be used anterior to determining the number of latent classes underlying data (Laudy, Boom, & Hoijtink, 2005). Conversely, confirmatory latent profile analysis (CLPA), in which constraints are placed on model parameters, allows a researcher to hypothesize as to the number and type of latent classes in a particular set of data prior to analysis (Finch & Bronk, 2011).

LPA supposes a simple parametric model and uses observed data to estimate two parameter values for each model: (1) conditional response probabilities (the probability for each combination of latent class, item, or variable, and response level for the item or variable, that a randomly selected member of that class will make that response to that item/variable); and (2) the probability of being in a specific latent class (McCutcheon, 2002). Parameters are estimated by maximum likelihood (MLE), and the MLE estimates are those most likely to account for the observed results (Muthén, 2001). Basically, to find the set of parameter values that optimize some criterion, the primary method is to statistically assess latent class models with two, three, or more latent classes, and to statistically assess the fit of each one to the data (Nylund,

Asparouhov, & Muthén, 2007). In general, as the number of classes becomes fewer, a model will fit the data worse (Nylund, Asparouhov, & Muthén, 2007).

In LP modeling, the criteria used to make decisions regarding the number of classes are considered more formal, as a number of statistics are available that can assist with choosing one model over another, such as the BIC statistic (Samuelsen & Raczynski, 2013). This is not the case with k-means, however, as the researcher must determine the number of clusters without the benefit of formal diagnostic statistics (Magidson & Vermunt, 2002). While k-means uses an ad hoc approach for classification, the LP method is probability-based, in that cases are classified into clusters using model-based posterior probabilities estimated by maximum likelihood (ML) methods, which also yield estimates for misclassification rates (Muthén & Muthén, 2007).

LPA's model-based clustering allows working with variables of mixed-measurement levels to be relatively simple as compared to k-means, in which variables must be standardized beforehand to ensure equal variance so as to avoid clusters dominated by variables with the most variation (Magidson & Vermunt, 2002). Discriminant analysis is commonly used following a k-means clustering to describe differences between the clusters on one or more possible covariates. In contrast, the LPA model can be broadened easily to include demographic and other exogenous variables as predictors of the latent classes, thereby allowing for the classification and cluster description to be performed at the same time (Magidson & Vermunt, 2002). It is important to note, however, that when covariates for latent classes are included, the latent class results are altered (Vermunt & Magidson, 2002).

**Confirmatory latent profile analysis (CLPA).** Midwestern student enrollment behavior data as defined by Bahr's variables in Table 2 initially was to be analyzed with confirmatory latent profile analysis (CLPA) using the *Mplus*, Version 5 (Muthén & Muthén, 2008) statistical

package. The intent was to answer whether the same typological categories could be identified when validating Bahr's 2011 methodology using Midwestern samples of first-time and non-first-time community college students.

Early attempts were made to conduct confirmatory latent profile analysis (CLPA) to validate Bahr's 5-class solution. However, despite a series of different approaches, the CLPA model (Appendix B), which required 32 constraints, would not converge. The initial model requested 500-50 starts, 50-20-50-20 LRT starts, 500 LRT bootstrap starts, and 500 iterations. As each model failed to converge, the number of starts was increased, with a final 5,000-500 starts, 500-200-500-200 LRT starts, 1,000 LRT bootstrap starts, and 2,000 iterations. To further promote model convergence, the number of constraints were limited as well (e.g. including only one constraint rather than all 32), but these solutions were equally unsuccessful. Thus, it was determined that CLPA would not be an option for this study.

**Latent profile analysis (LPA).** Midwestern student enrollment behavior data was analyzed with latent profile analysis (LPA) using the *Mplus*, Version 5 (Muthén & Muthén, 2008) statistical package to determine what classes existed, and if underlying, latent dimensions were evident as well. As each latent profile model was identified, students were classified to the most likely latent class by means of recruitment probabilities. From the recruitment probabilities, the posterior classification probability of a student's membership in each latent class was calculated. Students were then assigned to the latent class with the highest a posteriori probability.

**Model fit.** Comparative fit statistics were used to determine which model (i.e. number of latent classes and covariate structure) best fit the data. These fit statistics included the bootstrap likelihood ratio test (BLRT), the sample-size adjusted Bayesian information criterion (aBIC),

and the Lo-Mendell-Rubin Adjusted Likelihood Ratio Test (LMR). In an examination of the performance of likelihood-based tests used in latent class analysis, Nylund et al. (2007) found the BLRT, aBIC, and LMR to be preferred when determining the number of classes in mixture modeling.

The BLRT uses a bootstrap resampling method to approximate the  $p$ -value of the generalized likelihood ratio test, comparing the likelihood values for mixtures with differing numbers of classes by resampling under the null hypothesis of no difference (Nylund et al., 2007; McLachlan and Peel, 2000). In other words, instead of assuming that the difference distribution follows a known distribution (e.g., the chi-square distribution), the BLRT empirically estimates the difference distribution. The BLRT test compares the fit of a  $k$  class solution to that of a  $k-1$  solution, e.g. 6-class ( $k$ ) versus 5-class ( $k-1$ ), with a statistically significant result implying better fit for the  $k$  model. The Lo-Mendell-Rubin Adjusted Likelihood Ratio Test or LMR (Lo, Mendell, & Rubin, 2001) is similar to the BLRT in that it compares whether a  $k$  class solution fits better than a  $k-1$  solution, was also used as an indicator of model fit.

Because the sample size adjusted Bayesian information criterion (aBIC) has been shown to be superior to other information criteria (Nyland et al., 2007), such as the Akaike (AIC) and standard Bayesian (BIC), it was included as a measure of model fit (Table 4.6). The aBIC applies a penalty to the likelihood ratio statistic for an increased number of model parameters, such that lower values indicate better fit.

***Posterior Classification Probabilities.*** Group classification based on most likely class membership provides another approach for assessing model fit. These classification diagnostics are based on the estimated posterior class probabilities of group membership. Posterior class

probabilities are the model-estimated values for each individual's probabilities of being in each of the latent classes based on the maximum likelihood parameter estimates and the individual's observed responses on the indicator variables (Masyn, 2013). Posterior classification probabilities range between 0 and 1, and gauge uncertainty in assigning individuals to a specific class (Bauer & Shanahan, 2007). Thus, individuals will have a high probability of being assigned to the latent class most associated with the parameters reflective of their values on the variables used to fit the model (Vichi & Opitz, 2012). In other words, each student is placed in the latent class for which the student has the highest probability of belonging to. A perfectly fitting model would have 1s on the diagonal and 0s elsewhere. Thus, the higher the values, or classification probabilities, on the diagonal, the greater the certainty for the model in terms of class membership.

## CHAPTER IV

## RESULTS

**Latent Profile Analysis (LPA) Validation of Bahr's Cluster Analysis****Model Comparison**

A latent profile analysis (LPA; Mplus program in Appendix A) was conducted to validate Bahr's 5-class cluster solution (Bahr, 2001). The LPA sample was comprised of standardized enrollment data for 10,892 students, and, in keeping with Bahr's methodology, only first-time attendees were included. Because the goal of the initial analysis was to determine whether Bahr's proposed 5-class model using California Community College enrollment data would be supported with a sample of Midwestern community college students, only a 5-class LPA was conducted.

**Model Fit.** In terms of model fit of the 5-class LPA, the bootstrap likelihood ratio test (BLRT) was significant at  $p = 0.0000$ , indicating that the 5 class solution did not fit the data. The results suggested that the 6-class solution fits the data better than does the 5-class, which is contrary to Bahr's finding of 5 clusters.

Looking at the posterior classification probabilities, the value on the diagonal for Latent Class 3 was 0.964, suggesting that some uncertainty did exist regarding students' assignment to this class. Considering the overall 5-class LPA cross-validation solution, however, the classification probabilities on the diagonal (Table 3) ranged from 0.964 to 1.000, suggesting the model was reasonably confident in its ability to group students based on their enrollment behavior.

Table 3. Posterior Classification Probabilities for a 5-Class Cross-Validation Model

Latent Class	Most Likely Class Membership				
	Class 1	Class 2	Class 3	Class 4	Class 5
1	1.000	0.000	0.000	0.000	0.000
2	0.000	0.998	0.001	0.001	0.000
3	0.000	0.000	0.964	0.036	0.000
4	0.000	0.000	0.028	0.972	0.000
5	0.000	0.000	0.000	0.000	1.000

**Latent Class Comparison.** The means from the LPA results were first used to characterize the latent classes identified by the data, followed by a comparison to Bahr's outcomes to determine how well the results corresponded to one another. When the unstandardized LPA means (Table 5) were compared against Bahr's (Table 4), however, they were found to be markedly divergent, suggesting that the pattern of means seen with Bahr's solution was not identified through LPA.

**Variable Means.** Despite a few similarities between the results found in the current study and those reported by Bahr, Bahr's classes were generally not evident in the LPA results. The average number of Terms Enrolled (TE) ranged from 2 to 10 in Bahr's cluster analysis, but only 2 to 4 in the LPA. Likewise, Bahr's mean for Math Credits Attempted (MCA) ranged from 2 to 11; whereas, the LPA range was 0 to 3. Further, a number of Bahr's means were significantly discrepant from the highest value for the same variable in the LPA. Transferrable Math Credits Attempted (TMCA) was 14.5 versus the LPA's 6, while Transferrable English Credits Attempted (TECA) equaled 11, as opposed to 4 in the LPA. More salient disparities involved Non-Transferrable Degree Credits Attempted (NTDCA) and Noncredit Courses Attempted (NCA), which were 43 versus 14 and 25 versus 2, respectively.

A prominent example of the divergence between the solutions for the LPA and Bahr's cluster analysis involved Bahr's Career and Technical Education (CTE) cluster, which differed greatly from the most comparable class found in the LPA. Bahr identified his CTE cluster as having an increased number of terms enrolled and non-transferrable credits, with a high level of course success. However, the LPA class most resembling Bahr's CTE cluster did not align with Bahr's results. In Bahr's solution, the mean for Terms Enrolled (TE) was 9 versus 4 for the LPA. The mean for Non-Transferrable Degree Credits Attempted (NTDCA) was also much higher in Bahr's analysis than in the LPA, at 43 versus 10, respectively. Similarly, at 25, the mean for Transferrable Other Credits Attempted (TOCA) was greater in Bahr's solution than in the LPA, which was only 8.

Table 4. Variable Means for Bahr's 5-Class Cluster Model

Variable	Class 1 Completion Likely n=55,215 (32%)	Class 2 CTE n=4,165 (3%)	Class 3 Completion Unlikely n=41,613 (30%)	Class 4 Skills Builder n=44,976 (32%)	Class 5 Noncredit n=1,977 (3%)
Terms Enrolled (TE)	10	9	2	2	13
Course Success (CS)	0.72	0.80	0.26	0.94	0.96
Mean Credits Attempted (MCA)	11	10	7	3	2
Non-transferrable Degree Credits Attempted (NTDCA)	2.5	43	1	1	2
Noncredit Courses Attempted (NCA)	3	2	0	1	25
Transferrable Math Credits Attempted (TMCA)	14.5	4	2	0	3
Transferrable English Credits Attempted (TECA)	11	4	2	0	3
Transferrable Other Credits Attempted (TOCA)	58.5	25	9	3	9

Table 5. Latent Class Means and Standard Deviations for a 5-Class LPA Cross-Validation Model

Variable (SD)	Class 1 n=76 (1%)	Class 2 n=677 (5%)	Class 3 n=3,453 (32%)	Class 4 n=4,429 (41%)	Class 5 n=2,257 (21%)
Terms Enrolled (TE; 1.943)	4	3	2	4	4
Course Success (CS; 0.395)	0.68	0.13	0.16	0.85	0.90
Mean Credits Attempted (MCA; 0.834)	3	0	3	3	3
Non-transferrable Degree Credits Attempted (NTDCA; 10.135)	14	6	1	10	12
Noncredit Courses Attempted (NCA; 0.224)	2	0	0	0	0
Transferrable Math Credits Attempted (TMCA; 2.747)	2	0	0	0	6
Transferrable English Credits Attempted (TECA; 3.235)	2	0	3	3	4
Transferrable Other Credits Attempted (TOCA; 10.378)	7	3	8	8	15

Bahr's Completion Likely cluster also showed marked differences in select variable means when set against the LPA class with the most similar characteristics. The mean for terms enrolled in Bahr's analysis was 10 versus 4 in the LPA, while the mean for Transferrable Other Credits Attempted (TOCA) was 58.5 for Bahr, but only 15 in the LPA. The main difference seen with Bahr's Skills Builder cluster involved Course Success (CS). In Bahr's solution, the mean for this variable was 0.94, but in the LPA class with the most similar characteristics, the value was much lower, at 0.13.

With respect to Bahr's Noncredit cluster, a comparable class in the LPA did not seem to exist. At 2, only one LPA class had a Noncredit Courses Attempted (NCA) mean greater than 0; however, the mean for this variable in Bahr's solution equaled 25. Other notable differences could be seen with Terms Enrolled (TE), Course Success (CS), and Non-Transferrable Degree Credits Attempted (NTDCA). Comparing Bahr's solution to the LPA, the means for these variables were 13 versus 4, 0.96 versus 0.68, and 2 versus 14, respectively.

**Latent Class Size.** A review of latent class sizes (Tables 6 and 7) also revealed a number of differences. Most notably, Bahr’s Career and Technical Education (CTE) cluster accounted for only 3% of his overall sample. Whereas, the latent class with the most similar characteristics in the LPA solution was 41% of the sample. Also, at 32%, Bahr’s Completion Likely group was much larger than the LPA solution’s most comparable latent class of 21%, and Bahr’s Skills Builder cluster was 32% of his sample versus the LPA’s 5%.

Table 6. Frequencies in Gender and Race for Bahr’s 5-Class Cluster Model

	Class 1 Completion Likely (32%)	Class 2 CTE (3%)	Class 3 Completion Unlikely (30%)	Class 4 Skills Builder (32%)	Class 5 Noncredit (3%)
Male	45%	57%	52%	48%	39%
Female	55%	43%	48%	52%	61%
Minority	58%	61%	63%	64%	73%
Nonminority	42%	39%	37%	36%	27%

Table 7. Frequencies in Gender and Race for a 5-Class LPA Cross-Validation Solution

	Class 1 (1%)	Class 2 (5%)	Class 3 (32%)	Class 4 (41%)	Class 5 (21%)
Male	38% (29)	44% (295)	46% (1,604)	38% (1,703)	38% (855)
Female	62% (47)	56% (382)	54% (1,849)	62% (2,726)	62% (1,402)
Minority	71% (54)	71% (483)	70% (2,423)	74% (3,287)	64% (1,434)
Nonminority	29% (22)	29% (194)	30% (1,030)	26% (1,142)	36% (823)

**Demographic Composition.** Latent class variations in demographic composition were seen as well (Tables 6 and 7). The majority of Bahr’s Completion Unlikely cluster was male at 52%, while the most equivalent LPA latent class was 54% female. Similarly, Bahr’s Career and Technical Education (CTE) cluster was chiefly male at 57%, whereas the seemingly CTE-oriented latent class in the LPA was 62% female. In the area of race, Bahr’s Career and

Technical Education (CTE) cluster was 61% minority. The most comparable latent class in the LPA solution was also primarily minority, but the difference was more pronounced at 74%.

**Summary.** The 5-class LPA did not appear to support Bahr's 5-class solution. Marked differences in class assignment and variable means, as well as in demographic composition, suggested that the 5-class model proposed by Bahr may not be present in Indiana, with respect to classification of students by their enrollment behavior.

### **Cross-Validation of Initial LPA Results**

#### **Model Comparison**

Based on the results of the previous analysis, indicating that Bahr's 5-cluster model may not provide optimal fit to the Indiana data, a series of latent profile analysis (LPA) models (Appendix C) were run first without the covariates gender and race, and then with in a sample of 13,822 students. The goal was to identify the number of latent classes in the sample based on the pattern of means of Bahr's eight variables (2011), as well as to determine the latent class to which an individual belonged. Due to the mixed results seen in the literature (Dugan, 2011, 2013; Hagedorn & Prather, 2005; Hu & McCormick, 2012; Knight, 2013; Luan, Zhao, & Hayek, 2009; Mauss, 1967; Pascarella & Terenzini; VanDerLinden, 2002), an additional objective was to explore whether a relationship existed between gender, race, and class assignment. It should be noted that inclusion of covariates in a model may potentially change the actual latent class solutions themselves. Unlike with the Bahr cross-validation analysis, however, the sample used for these solutions was comprised of all student types, rather than just first-time only.

**Model Fit.** Comparative fit statistics (Table 8) were used to determine which model (i.e. number of latent classes and covariate structure) best fit the data. These fit statistics included the

bootstrap likelihood ratio test (BLRT), the sample-size adjusted Bayesian information criterion (aBIC), and the Lo-Mendell-Rubin Adjusted Likelihood Ratio Test (LMR).

Table 8. Fit Statistics for Competing Models

Models	Fit Statistic		
	aBIC	BLRT p-value	LMR p-value
Non-covariate* Models			
4-Class	181754.607	= 0.0000	= 0.0000
5-Class	176207.625	= 0.0000	= 0.0000
6-Class	175981.993	= 0.0000	= 0.4310
7-Class	178288.398	= 1.0000	= 0.8673
Covariate* Models			
4-Class	142096.112	= 0.0000	= 0.7603
5-Class	127942.753	= 0.0000	= 0.7800
6-Class	107961.878	= 0.0000	= 0.2299
7-Class	112961.530	= 1.0000	= 0.7601
Cross-Validation Covariate* Models			
5-Class	194572.180	=0.0000	=0.6432
6-Class	108775.193	=0.0000	=0.7513
7-Class	113823.201	=1.0000	=0.7612

*Note.* LMR = Lo-Mendell-Rubin; BLRT = bootstrap likelihood ratio test; aBIC = adjusted Bayesian information criterion. \*Covariates = gender and race.

In the covariate and non-covariate models, the BLRT for the 4-, 5-, and 6-class solutions was statistically significant at  $p = 0.0000$ , but not for the 7-class (Table 8). This finding suggests that for each group, the fit for the 5-class solution was better than that of the 4-class, and the 6-class was better than the 5-class, but the 7-class was not better than the 6-class. Based on this result, the 6-class model was determined to be optimal for the both the covariate and non-covariate solutions.

In the non-covariate models, the LMR was statistically significant at  $p = 0.0000$  for both the 4- and 5-class models, implying better fit for the 5-class model than the 4-class, as well as better fit for the 4-class than a 3-class. However, the LMR statistic was not significant for the 6-

and 7-class models, suggesting that the 6-class model ( $p = 0.4310$ ) did not fit better than the 5-class, and the 7-class ( $p = 0.8673$ ) did not fit better than the 6. The LMR statistic was not shown to be statistically significant for any of the covariate models, although the result for the 6-class model was smallest at  $p = 0.2299$ , suggesting that the 5-class non-covariate model was optimal.

The aBIC estimate for the 6-class non-covariate model was smallest among the models excluding the covariates, at 175981.99, implying a better fit to the observed data than the other latent class models excluding the covariates. However, the inclusion of the covariates in the 6-class model lowered the aBIC value to 107961.878. Indeed, the aBIC value for the 6-class covariate model was the lowest across all of the models assessed here, indicating that it yielded the best fit to the data. In short, because the aBIC for the 6-class model in each group was lowest and the BLRT values significant, the results suggested that each 6-class model demonstrated the best fit to the data. Contrasting the 6-class covariate and non-covariate models, the fit of the 6-class covariate model (Table 8) indicated that the addition of the variables gender and race to the model improved the model's ability to predict the data. Specifically, aBIC for the covariate model was smaller at 107961.878 versus 175981.993 for the non-covariate.

***Posterior Classification Probabilities.*** The classification probabilities for the 6-class covariate model (Table 9) ranged from 0.964 (Latent Class 4) to 0.999 (Latent Class 2). The high probability that students were assigned to the appropriate class was an additional affirmation of the fit of the model to the data.

Table 9. Posterior Classification Probabilities for a 6-Class Covariate Model

Latent Class	Most Likely Class Membership					
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
1	0.976	0.000	0.000	0.023	0.000	0.000
2	0.001	0.999	0.000	0.000	0.000	0.000
3	0.004	0.000	0.996	0.000	0.000	0.000
4	0.036	0.000	0.000	0.964	0.000	0.000
5	0.002	0.000	0.000	0.000	0.998	0.000
6	0.013	0.000	0.000	0.004	0.000	0.982

**Characterization of Latent Classes.** Although the 6-class covariate outcome did not affirm Bahr's 5-class, differences in enrollment behavior as demonstrated by variable means (Figure 1; Table 10) did suggest a clear delineation in student type. Following is a description of each latent class.

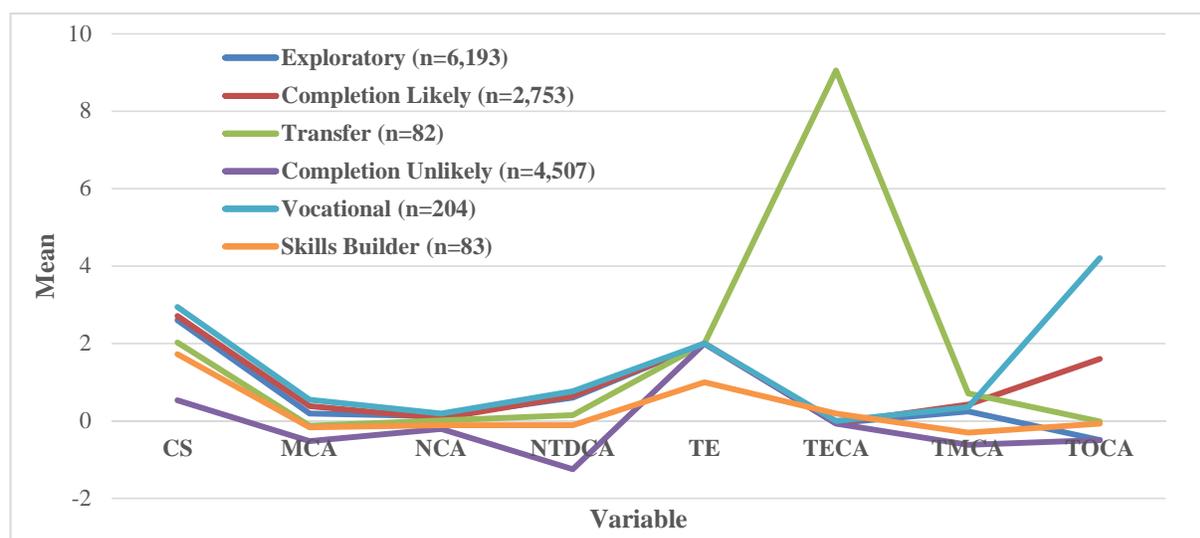


Figure 4.1. Latent class means for a 6-class covariate model.

Table 10. Standardized Latent Class Means and 95% Confidence Intervals for a 6-Class Covariate Model

Variable (SD)	Class 1 Exploratory n=6,193 (44%)	Class 2 Completion Likely n=2,753 (20%)	Class 3 Transfer n=82 (.5%)	Class 4 Completion Unlikely n=4,507 (33%)	Class 5 Vocational n=204 (2%)	Class 6 Skills Builder n=83 (.5 %)
Terms Enrolled (TE; 0.000)	2.000 (2.000/2.000)	2.000 (2.000/2.000)	2.000 (2.000/2.000)	2.000 (2.000/2.000)	2.000 (2.000/2.000)	1.002 (1.002/1.002)
Course Success (CS; 1.053)	2.592 (2.560/2.623)	2.702 (2.662/2.741)	2.025 (1.726/2.324)	0.532 (0.502/0.561)	2.942 (2.801/3.083)	1.728 (1.435/2.022)
Mean Credits Attempted (MCA; 0.741)	0.195 (0.170/0.221)	0.381 (0.344/0.418)	-0.120 (-0.336/0.096)	-0.517 (-0.542/-0.492)	0.538 (0.416/0.659)	-0.165 (-0.380/0.051)
Non-transferrable Degree Credits Attempted (NTDCA; 0.056)	0.603 (0.589/0.617)	0.643 (0.623/0.663)	0.150 (-0.057/0.356)	-1.248 (-1.267/-1.229)	0.759 (0.697/0.822)	-0.108 (-0.336/0.120)
Noncredit Courses Attempted (NCA; 0.955)	0.122 (0.103/0.141)	0.068 (0.044/0.093)	0.014 (-0.151/0.180)	-0.214 (-0.256/-0.172)	0.183 (0.106/0.259)	-0.106 (-0.343/0.131)
Transferrable Math Credits Attempted (TMCA; 0.650)	0.247 (0.220/0.275)	0.422 (0.381/0.462)	0.712 (0.406/1.018)	-0.618 (-0.630/-0.607)	0.354 (0.204/0.503)	-0.300 (-0.455/-0.144)
Transferrable English Credits Attempted (TECA; 0.263)	-0.051 (-0.061/-0.041)	-0.053 (-0.076/-0.029)	9.050 (9.050/9.050)	-0.066 (-0.075/-0.056)	-0.006 (-0.006/-0.006)	0.185 (0.185/0.185)
Transferrable Other Credits Attempted (TOCA; 0.001)	-0.493 (-0.494/-0.492)	1.602 (1.599/1.605)	-0.013 (-0.013/-0.013)	-0.494 (-0.495/-0.493)	4.204 (4.204/4.204)	-0.065 (-0.065/-0.065)

***Latent Class 1.*** With 6,193 participants, Latent Class 1 was markedly larger than the rest of the classes, representing 45% of the student sample (Table 10). Measured against those in other classes, students in this Exploratory class were higher in course success (CS), but lower in mean credits attempted (MCA), as well as in all transferrable courses attempted (TMCA, TECA, TOCA). Additionally, these students were also second-highest in noncredit courses attempted (NCA). With respect to demographics (Table 11), gender in this class leaned heavily toward

females rather than males, with 3,908 versus 2,285, respectively. In the area of race, however, the group was almost evenly split at 3,160 minority students versus 3,033 nonminority.

Table 11. Frequencies in Gender and Race for a 6-Class Covariate Model

	Class 1 Exploratory (n=6,193)	Class 2 Completion Likely (n=2,753)	Class 3 Transfer (n=82)	Class 4 Completion Unlikely (n=4,507)	Class 5 Vocational (n=204)	Class 6 Skills Builder (n=83)
Male	2,285	941	33	2,054	116	34
Female	3,908	1,812	49	2,453	88	49
Minority	3,160	1,398	36	1,956	84	46
Nonminority	3,033	1,355	46	2,551	120	37

**Latent Class 2.** Comparatively, Latent Class 2 was smaller than Class 1 at 2,753 participants, or 20% of the sample (Table 10). However, students in this Completion Likely class had the second-highest level of course success (CS) and the second-highest level of mean credits attempted (MCA), non-transferrable degree credits attempted (NTDCA), and transferrable math (TMCA) and non-math or English (TOCA) credits attempted. Although the number of females (1,812) was nearly twice that of males (941), minority student status was closely divided at 1,398 minority and 1,355 nonminority (Table 11).

**Latent Class 3.** Latent Class 3, which was smallest with only 82 participants, or <1% of the sample, appeared to be more Transfer-oriented. Relative to the other classes, students in this class were high in course success (CS), but low in mean credits attempted (MCA). These students also had the highest level of transferrable math (TMCA) and English (TECA) credits attempted (Table 10). As before (Table 11), the number of female students (49) surpassed the number of males (33); however, the total nonminority students (46) in this class was greater than that of minority (36).

***Latent Class 4.*** Latent Class 4, the Completion Unlikely class, was the second largest, with 4,507 participants, or 33% of the sample. Yet, with the exception of number of terms enrolled (TE), these students were lowest across all variables (Table 10). Similar to Latent Class 3 (Table 11), this class was comprised of a greater number of nonminority students (2,551) than minority students (1,956), but had a higher number of females (2,453) than males (2,054).

***Latent Class 5.*** Latent Class 5 was relatively small, with only 204 participants, or 2% of the sample (Table 10). Students in this Vocational class had the highest levels of course success (CS), mean credits attempted (MCA), non-transferrable degree credits attempted (NTDCA), and transferrable non-math or English credits attempted (TOCA). These students also had the highest number of noncredit courses attempted (NCA). Students in this class (Table 11) also were predominantly male (116) as opposed to female (88), and nonminority (120) rather than minority (88).

***Latent Class 6.*** With only 83 participants, or <1% of the sample, Latent Class 6 was comparable in size to Latent Class 3 (Table 10). However, students in this Skills Builder class demonstrated the lowest level of terms enrolled (TE), which at 1.002 was nearly half that of the other classes. Comparatively, these students were average in course success (CS) and transferrable English credits attempted (TECA), but low across the remaining variables. Converse to Latent Class 5 (Table 11), students in this class were largely minority (46) females (49), as opposed to nonminority (37) males (34).

**Logistic Regression.** Using the Skills Builder class (Latent Class 6) as the reference category, logistic regression results reflecting the impact of gender and race on class membership (Table 12) indicated a statistically significant relationship between latent class membership and the student's gender and race. In the Completion Unlikely class (Latent Class 4; OR 0.626; 95%

CI: 0.404-0.970;  $p = 0.036$ ), the odds of being male ( $n = 2,054$ ) were 0.626 times that of being female ( $n = 2,453$ ) when compared to the Skills Builder class. Whereas, in the Vocational class (Latent Class 5; OR 0.575; 95% CI: 0.343-0.963;  $p = 0.035$ ), the odds of being female ( $n = 88$ ) were only 0.575 times that of being male ( $n = 116$ ). With respect to race, however, membership in the Vocational class (Latent Class 5; OR 1.881, 95% CI: 0.537-1.285;  $p = 0.017$ ) alone showed a significant impact, with the odds of being nonminority ( $n = 120$ ) nearly twice that of being minority ( $n = 84$ ).

Table 12. Logistic Regression and Odds-Ratio Results for a 6-Class Covariate Model

	Est.	SE	Est./SE	p-Value	OR	95% CI
Latent Class 1 (Exploratory)						
Gender	-0.185	0.223	-0.833	0.405	0.831	-0.622/0.251
Race	-0.171	0.225	-0.759	0.448	0.843	-0.612/0.270
Latent Class 2 (Completion Likely)						
Gender	-0.195	0.224	-0.868	0.385	0.823	-0.635/0.245
Race	-0.285	0.227	-1.256	0.209	0.752	-0.730/0.160
Latent Class 3 (Transfer)						
Gender	-0.470	0.314	-1.498	0.134	0.625	-1.085/0.145
Race	-0.035	0.317	-0.110	0.912	0.966	-0.657/0.587
Latent Class 4 (Completion Unlikely)						
Gender	-0.469	0.223	-2.098	0.036	0.626	-0.906/-0.031
Race	0.183	0.226	0.812	0.417	1.201	-0.259/0.625
Latent Class 5 (Vocational)						
Gender	-0.554	0.263	-2.104	0.035	0.575	-1.069/-0.038
Race	0.632	0.264	2.388	0.017	1.881	0.113/1.150

*Note:* Reference category = Latent Class 6 (Skills Builder)

**Summary.** Of the sample, the Exploratory class (Latent Class 1) was largely female and appeared, although accomplished, more focused on determining an academic direction than in immediate completion or transfer. The Completion Likely students (Latent Class 2) were also predominantly female and demonstrated the highest level of successful but non-transferrable completion. With their high level of course success and increased number of transfer credits,

students in Latent Class 3, the Transfer class, seemed better positioned for transition to a different institution. Whereas, students in Latent Class 4, comprised primarily of nonminority females, seemed unlikely to complete due to their elevated enrollment but low level of achievement. The largely nonminority male students in the Vocational class (Latent Class 5), while successful, did not demonstrate an enrollment pattern suggestive of completion or transfer and had the highest number of noncredit courses attempted. Moreover, the primarily minority female students in Latent Class 6 (Skills Builder) appeared more remediation-oriented.

Looking at the logistic regression results, in the Completion Unlikely class, males were found to be more likely to complete their degrees than females. However, in the Vocational group, nonminority males were much more prominent, suggesting that nonminority males were more interested in improving jobs skills than in pursuing a degree.

### **Cross-Validation of the 6-Class LPA Model**

#### **Model Comparison**

To validate the whether the 6-class covariate model's structure was indeed optimum, a second LPA (Appendix C) was conducted using a new sample of 13,814 students. Similar to the previous 6-class LPA, the sample was comprised of all student types, not just first-time attendees.

**Model Fit.** As demonstrated by a significant BLRT value ( $p = 0.0000$ ; Table 8), the model continued to be a good fit to the data. To confirm that the 6-class model was ideal, a 7-class model was run using the same data, with the BLRT results suggesting poor model fit ( $p = 1.0000$ ). Additionally, the posterior classification probabilities (Table 13) were comparable to those for the initial exploratory LPA sample, with values ranging from 0.944 (Latent Class 1) to 1.000 (Latent Class 6).

Table 13. Posterior Classification Probabilities for a 6-Class Cross-Validation Model

Latent Class	Most Likely Class Membership					
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
1	0.944	0.000	0.019	0.036	0.000	0.000
2	0.004	0.996	0.000	0.000	0.000	0.000
3	0.000	0.000	0.967	0.033	0.000	0.000
4	0.000	0.000	0.023	0.977	0.000	0.000
5	0.002	0.000	0.000	0.000	0.998	0.000
6	0.000	0.000	0.000	0.000	0.000	1.000

**Latent Class Comparison.** A review of the latent class counts and means for the cross-validation analysis (Figure 2; Table 14) found the 6-class covariate solution to be extremely similar across the two samples. Latent Class 1 in the cross-validation LPA, which was most similar to Latent Class 6 (Skills Builder) in the initial analysis, demonstrated a higher mean in Noncredit Courses Attempted (0.015 versus -0.106). Similarly, the means for Mean Credits Attempted (MCA) and Transferrable Other Credits Attempted (TOCA) in Latent Class 2 (Transfer) of the cross-validation analysis (0.057 and 0.067, respectively) were greater than that of the most comparable class in the initial LPA, Latent Class 3 (-0.120 and -0.013, respectively). Additionally, the mean for Transferrable English Credits Attempted (TECA) in Latent Class 5 (Vocational) of the cross-validation LPA (0.029) was larger than that of its counterpart in the initial analysis, Latent Class 5 (-0.006).

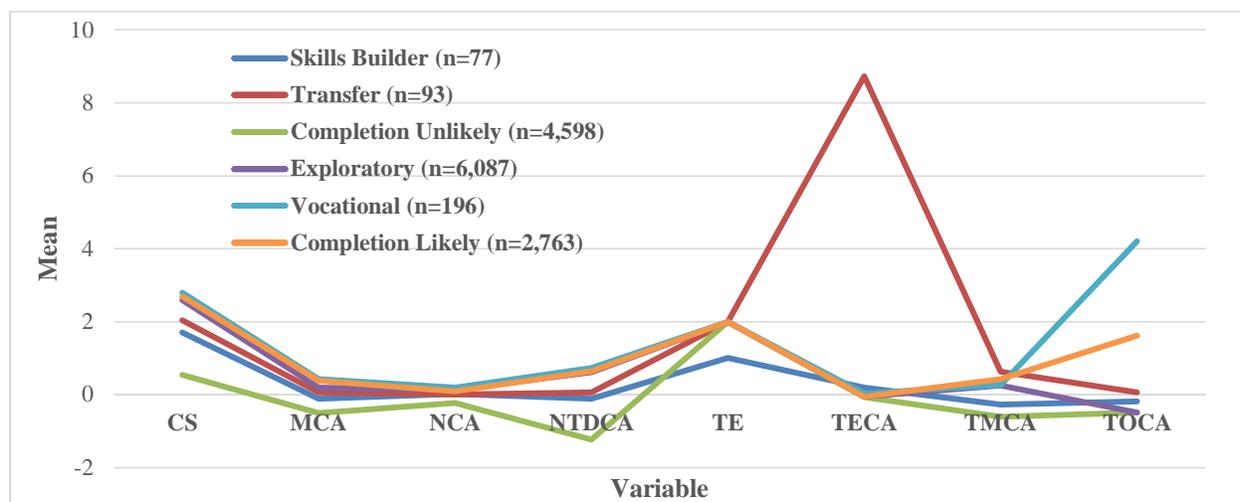


Figure 4.2. Latent class means for a 6-class cross-validation model.

Table 14. Standardized Latent Class Means and 95% Confidence Intervals for a 6-Class Cross-Validation Model

Variable (SD)	Class 1 Skills Builder n=77 (.5%)	Class 2 Transfer n=93 (.5%)	Class 3 Completion Unlikely n=4,598 (33%)	Class 4 Exploratory n=6,087 (44%)	Class 5 Vocational n=196 (2%)	Class 6 Completion Likely n=2,763 (20%)
Terms Enrolled (TE; 0.000)	1.002 (1.001/1.003)	2.000 (2.000/2.000)	2.000 (2.000/2.000)	2.000 (2.000/2.000)	2.000 (2.000/2.000)	2.000 (2.000/2.000)
Course Success (CS; 1.071)	1.801 (1.409/1.993)	2.039 (1.754/2.324)	0.537 (0.059/0.565)	2.590 (2.558/2.622)	2.789 (2.639/2.940)	2.700 (2.661/2.739)
Mean Credits Attempted (MCA; 0.750)	-0.106 (-.317/0.104)	0.057 (-0.145/0.260)	-0.506 (-0.530/-0.481)	0.199 (0.174/0.224)	0.424 (0.289/0.559)	0.381 (0.344/0.417)
Non-transferrable Degree Credits Attempted (NTDCA; 0.055)	-0.110 (-0.333/0.112)	0.065 (-0.133/0.263)	-1.232 (-1.251/-1.214)	0.618 (0.605/0.631)	0.736 (0.670/0.803)	0.644 (0.624/0.664)
Noncredit Courses Attempted (NCA; 0.949)	0.015 (-0.237/0.267)	0.007 (-0.158/0.173)	-0.232 (-0.274/-0.191)	0.131 (0.113/0.150)	0.192 (0.103/0.280)	0.085 (0.061/0.109)
Transferrable Math Credits Attempted (TMCA; 0.650)	-0.274 (-0.434/-0.115)	0.629 (0.336/0.923)	-0.612 (-0.624/-0.600)	0.254 (0.226/0.281)	0.264 (0.124/0.404)	0.429 (0.389/0.469)
Transferrable English Credits Attempted (TECA; 0.237)	0.189 (0.035/0.343)	8.720 (5.187/12.253)	-0.069 (-0.083/-0.055)	-0.059 (-0.071/-0.047)	0.029 (-0.065/0.123)	-0.056 (-0.102/-0.010)
Transferrable Other Credits Attempted (TOCA; 0.001)	-0.180 (-0.192/-0.168)	0.067 (-0.066/0.199)	-0.495 (-0.497/-0.492)	-0.495 (-0.496/-0.493)	4.199 (3.871/4.527)	1.611 (1.597/1.626)

Frequencies in gender and race were fairly consistent between the two analyses as well (Table 15). The only marked difference was the number of minority students in Latent Class 2 (Transfer) of the validation LPA ( $n=46$ ), which was higher than that of Latent Class 3 (Transfer) in the initial LPA ( $n=36$ ), resulting in a more even split between groups.

Table 15. Frequencies in Gender and Race for a 6-Class Cross-Validation Model

	Class 1 Skills Builder ( $n=77$ )	Class 2 Transfer ( $n=93$ )	Class 3 Completion Unlikely ( $n=4,598$ )	Class 4 Exploratory ( $n=6,087$ )	Class 5 Vocational ( $n=196$ )	Class 6 Completion Likely ( $n=2,763$ )
Male	27	36	2,061	2,243	111	945
Female	50	57	2,537	3,844	85	1,818
Minority	41	46	1,947	3,109	83	1,306
Nonminority	36	47	2,651	2,978	113	1,403

**Logistic Regression.** Using the Skills Builder class (Latent Class 1) as the reference category as in the initial LPA, logistic regression results for the cross-reference LPA (Table 16) also demonstrated a statistically significant relationship between latent class membership and the student's gender and race. Similar to the initial LPA, in the Completion Unlikely class (Latent Class 3; OR 0.778; 95% CI: 0.706-0.856;  $p = 0.048$ ), the odds of being male ( $n = 2,061$ ) were 0.778 times that of being female ( $n = 2,537$ ). The logistic regression results for the impact of race mirrored those from the initial analysis as well, although in the Vocational class, a student's gender was not found to impact the student's classification, only race. In the cross-reference LPA, membership in the Vocational class (Latent Class 5; OR 2.481, 95% CI: 1.846-3.334;  $p = 0.002$ ) showed a significant impact on class membership, with the odds of being nonminority ( $n = 113$ ) nearly 2.5 times that of being minority ( $n = 83$ ).

Table 16. Logistic Regression and Odds-Ratio Model Results for a 6-Class Cross-Validation Model

	Est.	SE	Est./SE	<i>p</i> -Value	OR	95% CI
Class 2 (Transfer)						
Gender	-0.191	0.318	-0.600	0.548	1.011	0.649/1.576
Race	0.157	0.322	0.487	0.627	1.217	0.791/1.872
Class 3 (Completion Unlikely)						
Gender	-0.454	0.230	-1.974	0.048	0.778	0.706/0.856
Race	0.395	0.241	1.642	0.101	1.545	1.399/1.706
Class 4 (Exploratory)						
Gender	-0.128	0.229	-0.556	0.578	1.078	0.984/1.180
Race	0.077	0.240	0.321	0.748	1.124	1.022/1.236
Class 5 (Vocational)						
Gender	-0.441	0.271	-1.627	0.104	0.788	0.585/1.061
Race	0.869	0.279	3.112	0.002	2.481	1.846/3.334
Class 6 (Completion Likely)						
Gender	-0.202	0.231	-0.876	0.381	0.823	-0.635/0.245
Race	-0.040	0.242	-0.163	0.870	0.752	-0.730/0.160

*Note:* Reference Category = Latent Class 1 (Skills Builder)

**Summary.** Based on the similarity of results for the two 6-class analyses, it would seem that a cross-validation of the initial LPA using a different sample of student enrollment data does support the initial results. However, some differences in the impact of gender and race on class membership did exist between the two solutions. The logistic regression results for the Completion Unlikely class suggested that males may be more likely to complete. Whereas, nonminority males tend to dominate the Vocational group. These outcomes suggest that while a student's enrollment behavior can discern the student's latent class assignment, the student's gender and race may impact the student's latent class assignment as well.

## CHAPTER V

### DISCUSSION

Student classification, or the creation of student typologies, allows educational institutions to better target those groups of students most in need of intervention. By focusing on at-risk populations, schools are able to maximize results with the use of often-limited resources. Likewise, having an improved understanding of their students can enable higher educators to better determine student outcomes. Establishing the best method for creating student typologies, however, has been the subject of considerable research. The current study contributes to the literature by examining whether typological models can be used across institutions and if student enrollment behavior can serve as an appropriate grouping mechanism. The use of latent profile analysis as a means for student classification is also addressed.

To summarize the findings of the four research questions, the results found that a) in validating Bahr's 2011 cluster analysis, similar typological categories could not be replicated using latent profile analysis (LPA) with a Midwestern community college student sample, b) a 6-class LPA model comprised of Bahr's eight enrollment-oriented variables was determined to be the best fitting, c) the 6-class LPA model was confirmed via LPA with a cross-validation sample of students, and d) gender and race were found to be significantly related to class membership in each of the samples.

#### **LPA Cross-Validation of Bahr's Cluster Analysis**

A cross-validation of Bahr's 2011 cluster methodology was conducted using latent profile analysis (LPA) with a Midwestern community college sample, but was not supported. As with Bahr's analysis, the LPA model was comprised of eight enrollment-behavior related variables and utilized a sample of first-time attendees only. And because Bahr's process

promoted a 5-class solution, the LPA was similarly modeled. However, a subsequent comparison of variable means, group sizes, and demographic composition suggested that the 5-latent-class LPA solution did not support Bahr's findings.

The most notable difference was that Bahr's model included a Noncredit cluster, which was not evident in the LPA solution. Variations in institutional processes may be key here. In Bahr's California Community College (2011) sample, the eight enrollment-behavior variables apply to both credit and noncredit students alike, which was not the case in the Midwestern sample. In the latter case, credit and noncredit students are distinct student types. Although a credit student might take noncredit courses, this individual would not be identified as a noncredit student. This difference in student definition serves to highlight the need for each educational institution to develop its own classification process.

In 2011, Bahr published his process for developing a student typology, suggesting that his methodology could be utilized by other institutions. However, the numerous but highly varied typological research available in the literature (Adelman, 2005; Ammon, Bowman, & Mourad, 2008; Astin, 1993; Attinasi, Stahl, & Okun, 1982; Bahr, 2010, 2011; Boughan, 2000; Clagett, 1995; Cox, 2011; Dugan, 2011; Hagedorn & Prather, 2005; Horn, 2009; Houle, 1961; Hu & McCormick, 2011; Knight, 2013; Kuh, 1995; Kuh, Hu, & Vesper, 2000; Luan, 2006; Luan, Zhao, & Hayek, 2009; Marti, 2008; Pascarella & Terenzini, 1980; Robinson, 2004; Taber & Hackman, 1976; VanDerLinden, 2002; Zhao, Gonyea, & Kuh, 2003) suggests that while many institutions agree that a need does exist to classify students, the consensus is that the process should be locally driven.

In addition, the LPA model should be theorized for the population from which the data sample is obtained (Lazarsfeld, & Henry, 1968), which, again, supports the use of a local

solution. Bahr's sample was comparable to that of the overall California Community College system student population. Comparing Bahr's sample with that of the Indiana community college, which also mirrored its overall student population, the number of minority students included was much higher, at 55% versus 32% for the Indiana sample. With respect to gender and age, however, the two samples were fairly similar. For both groups of students, the majority were age 25 or less (Bahr's = 64%; Indiana = 60%). Regarding gender, Bahr's sample was 50% female versus the Indiana sample's 55% female.

The use of enrollment-behavior variables does seem to be a less subjective means of categorizing students (Astin 1993; Kuh, 1995), though. Drawbacks to the use of demographic and self-report variables in student classification can include categories that are neither exhaustive nor mutually exclusive, are descriptive rather than explanatory or predictive, and are based on subjective criteria (Bailey, 1994; Fenske et al, 1999). Typologies based on students' performance and engagement patterns, however, have been shown to operate more from a data-driven, empirically derived orientation (Adelman, 2005; Bahr, 2010; Clark & Trow, 1966; Horowitz, 1987; Katchadourian & Boli, 1985; Luo & Jamieson-Drake, 2005; Zhao, Gonyea, & Kuh, 2003). More recent typological studies based on student activities and behaviors were found to be consistent with earlier research showing that students learn and develop from their engagement in educationally oriented activities (Bahr, 2010, 2011; Hu & McCormick, 2011; Kuh, Hu, & Vesper, 2000; Pascarella & Terenzini, 1991, 2005).

### **Cross-Validation of Initial LPA Results**

Because Bahr's 5-class cluster solution was not supported by latent profile analysis (LPA), a subsequent series of LPAs were conducted to establish the number of latent classes existing in the data. As before, the LPA models were comprised of Bahr's eight enrollment-

behavior related variables, but instead utilized a sample of all student types, rather than just first-time only, which was more reflective of the local, total population. Bahr's analysis focused on first-time attendees only, which is not representative of the heavily non-traditional Midwestern population sampled. For this group, not including all student types (e.g. transfer, continuing, and return) greatly affects the results. Consequently, those students who would seem most at risk would fail to benefit from any possible class-based interventions.

Two series of models were examined: those including the covariates gender and race, and those without. In each group, the 6-latent-class solution was shown to be the best-fitting as determined by aBIC, BLRT, LMR, and posterior classification probabilities. An ensuing aBIC comparison of the non-nested 6-latent-class covariate and non-covariate models found the covariate solution to fit best. Based on the pattern of variable means and class sizes, the six latent classes were characterized as Skills Builder (n = 77), Transfer (n = 93), Completion Unlikely (n = 4,598), Exploratory (n = 6,087), Vocational (n = 196), and Completion Likely (n = 2,763).

Historically, cluster analysis has been widely used in the classification of students (Adelman, 2005; Ammon, Bowman, & Mourad, 2008; Bahr, 2010, 2011; Bahr, Bielby, & House, 2011; Boughan, 2000; Hagedorn & Prather, 2005; Horn, 2009; Hu & McCormick, 2011; Knight, 2013; Ku, Hu, & Vesper, 2000; Luan, 2006; Luan, Zhao, & Hayek, 2009; Marti, 2008; VanDerLinden, 2002), but it is limited in that it is not a statistical test and it requires the a priori specification of the number of clusters (Likas, Vlassis, & Verbeek, 2003). The LPA approach includes tests of statistical significance and probability (Magidson & Vermunt, 2002),

Dugan (2011) noted that LPA is an increasingly popular alternative to cluster analysis in determining student typologies. To describe differences between k-means clusters on one or

more possible covariates, post-cluster analysis is needed. With LPA, however, the model can be broadened easily to include demographic and other exogenous variables, allowing for the classification and cluster description to be performed at the same time (Magidson & Vermunt, 2002).

### **Cross-Validation of the 6-Class LPA Model**

To confirm the 6-latent-class covariate LPA model as best fitting, a new LPA was conducted with a different sample of Midwestern community college students' enrollment behavior as defined by Bahr's eight variables. As before, this sample included all student types, and the cross-validation analysis was shown by BLRT and posterior classification probabilities to be a good fit to the data. Using variable means and class sizes, the results of this analysis were then compared to those of the initial 6-latent-class LPA and, with the exception of demographic composition (e.g. gender and race), were found to support its structure.

A number of student classification studies have used cluster analysis to categorize community college students (Adelman, 2005; Ammon, Bowman, & Mourad, 2008; Bahr, 2010, 2011; Hagedorn & Prather, 2005; Horn, 2009; Luan, 2006; Marti, 2008; VanDerLinden, 2002). Some of this research, however, was rather limited in scope. For example, several studies were based on first-time students (Bahr, 2010, 2011; Boughan, 2000; Horn, 2009; Kuh, Hu, & Vesper, 2000; Marti, 2008; Robinson, 2004), or those who had not previously enrolled in college. While the intent is to prevent duplication of students, this method can cause certain student populations (e.g. returning or career and technical/noncredit) to appear underrepresented in the results.

### **Gender and Race as Predictors of LPA Class Membership**

For the initial and cross-validation 6-latent-class LPA models, gender and race were shown to impact a student's class assignment. However, the results were somewhat inconsistent

between analyses. For both samples, logistic regression results for only one latent class (Vocational) showed a significant relationship between the class to which the student belongs and the student's race. With respect to gender, however, two latent classes (Completion Unlikely and Vocational) in the initial LPA found a significant relationship to class membership, as opposed to only one latent class (Vocational) in the cross-validation analysis. Regardless, these outcomes differ from previously published research, which determined no statistically significant impact of gender or race (Horn, 2009; Kuh, 1995; Kuh, Hu, & Vesper, 2000; Luan, 2006) on a student's inclusion in a particular typology.

For student typographical research outcomes related to the inclusion of demographic variables, the results are mixed. Several studies have shown no statistically significant impact of gender or race on a student's inclusion in a particular typology (Horn, 2009; Kuh, 1995; Kuh, Hu, & Vesper, 2000; Luan, 2006). Similar investigations have determined academic behaviors (Ammon, Bowman, & Mourad, 2008; Hagedorn & Prather, 2005; Luan, 2006), course intensity, and transfer intentions (Hagedorn & Prather, 2005) to be better student classifiers than gender and ethnic demographics. However, some researchers have seen an impact. Mauss (1967) and Pascarella and Terenzini, (1980) determined only gender to have an effect, while others (Dugan, 2011, 2013; Hagedorn & Prather, 2005; Hu & McCormick, 2012; Knight, 2013; Luan, Zhao, & Hayek, 2009; VanDerLinden, 2002) found both gender and race to be significant.

Despite these conflicting outcomes, it would seem advisable to include gender and race as covariates in an LPA of student types due to the impact of these variables on class assignment. When categorizing students by enrollment behavior, knowing as much about the students in each latent class is integral to the development of effective intervention approaches.

### **Limitations and Future Directions**

Several limitations of the current study should be addressed in future work. First, the use of CLPA to validate Bahr's cluster results would have been the ideal; however, all attempts to run the analysis resulted in a lack of model convergence. It could be that too many constraints were needed in the CLPA model to allow it to function properly. Another possibility is that Bahr's pattern of group differences did not exist in the Midwestern community college student population, such that the CLPA model was not able to assign students to specific classes. Thus, a locally developed LPA model for classification of student enrollment behavior would seem more prudent, and might allow for CLPA validation.

Second, although convention requires that, in k-means, variables be standardized to ensure equal variance to avoid clusters dominated by variables with the most variation (Magidson & Vermunt, 2002), Bahr failed to do so. This not only caused difficulty with respect to comparison of the LPA and cluster means, it might have influenced his results by inflating some of his variable means as well.

Third, a validation of students' latent class assignments should be included as part of the classification process. In addition to the confirmation of model structure with new samples, a validation of class assignment using student outcomes such as retention, completion, and transfer seems advisable. Earlier studies have used direct assessment, grade point averages, and retention (Hu & McCormick, 2011) as well as self-reported student gains (Hu & McCormick, 2011; Kuh et al., 2000).

Because the Indiana community college population sampled for the current study clearly delineates between credit and noncredit students, any future research looking at these students should not include the Noncredit Courses Attempted (NCA) variable in the analysis.

Additionally, validation of typological outcomes is key, so classification results should be authenticated prior to the development or application of student interventions. Examples of data to utilize for this purpose can include persistence, retention, graduation, and transfer records.

Clearly, from an intervention standpoint, the earlier a student is classified the better. Thus, the next step with respect to future research would be to investigate whether determining student typologies based on incoming data, such as prior performance, demographics, etc., can produce groupings similar to several terms of enrollment behavior. The sooner educators can identify struggling students and apply strategies to improve their success, the better the possible outcome. Categorizing students well into their post-secondary education, although it can serve to inform the development of interventions for future students, will greatly limit potential benefit of those interventions for any current students.

### **Conclusion**

Classification of students using LPA has been shown to be an effective and reliable method, as does categorizing students by their enrollment behavior, rather than by demographic variables alone. It is important to note, however, that any model used for this purpose should be developed locally and driven by that institution's policies and student types. Additionally, much thought and planning should go into model creation and validation to ensure that all students are included and classified appropriately. Ultimately, in order to be useful, a student typology must be logical and reliable in how it classifies students.

Research has shown that engagement-based typologies have the diagnostic potential to identify at-risk students (Hu & McCormick, 2012). By determining student types and their possible associated outcomes, an institution can design targeted interventions to promote increased learning and college success. For example, by identifying poorly performing students,

educators can look more closely at these students to better understand their demographics, backgrounds, and other potential factors that might be impacting their level of achievement. These findings can then inform targeted interventions such as creating flags in the system to indicate when a student is faltering, requiring more advising appointments, or an increase in faculty contact. It is important to note, however, that a typology must be able to highlight the distinct differences between subgroups' limitations and strengths, which are unlikely to be met by a single intervention (Janosz et al, 2000). Likewise, initiatives designed to increase areas such as student achievement, persistence, retention, and motivation should vary with each category of students and be tailored to those students' specific needs.

Looking specifically at the Indiana outcomes, most of the students in the sample were classified as either Exploratory or Completion Unlikely. Of these two groups, females dominated males, and females were more shown to be more likely to be non-completers. It is interesting to note that the number of females in the Completion Likely class was greater than the number of males; however, there was no significant difference seen regarding the impact of gender on student assignment to this latent class. It does support, though, that although gender and race may have an impact on a student's classification, the driver for that student's class assignment is the student's course enrollment and success. The results do beg the question, however, as to why females in the Indiana sample were less successful. For future research, delving deeper into these two latent classes may help to determine what additional factors might be at play, such as SES or being a single parent. Realities such as not being able to afford reliable transportation or childcare can greatly impact whether a female student is successful.

## REFERENCES

- Adelman, C. (1999). *Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment*. Washington, DC: U.S. Department of Education.
- 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.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (Eds B. N. Petrov and F. Cásiki), pp. 267–281, Budapest. Akademiai Kaidó. Reprinted in *Breakthroughs in Statistics*, eds Kotz, S. & Johnson, N. L. (1992), Volume I, pp. 599–624. New York: Springer.
- Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis*. Thousand Oaks, CA: Sage.
- Alfred, R. L., Peterson, R. O., & White, T. H. (1992). *Making community colleges more effective: Leading through student success*. Ann Arbor, MI: Community College Consortium.
- 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.
- Andrews, H. A. (2003). *Enrollment trends in community colleges*. ERIC Clearinghouse for Community Colleges.
- Astin, A. W. (1993). An empirical typology of college students. *Journal of College Student Development*, 34, 36-46.

- Attinasi, L. C., Stahl, V. V., & Okun, M. A. (1982). A preliminary typology of motivational orientations of community college students. *Community/Junior College Quarterly*, 6, 371-390.
- Bahr, P. R. (2010). The bird's eye view of community colleges: A behavioral typology of first-time students based on cluster analytic classification. *Res High Educ*, 51, 724-749.
- Bahr, P. R. (2011). A typology of students' use of the community college. *New Directions for Institutional Research*, 51, 33-48.
- Bahr, P. R., Bielby, R., & House, E. (2011). The use of cluster analysis in typological research on community college students. *New Directions for Institutional Research*, 51, 67-81.
- Bahr, P. R., Booth, K., & Willett, T. (2012). *Segmentation model for assessing course-taking patterns: Research methodology and discussion guide*. Berkeley, California: Research and Planning Group for California Community Colleges.
- Bahr, P. R., Gross, J. L., Slay, K. E., & Christensen, R. D. (2012). *First in line: Student registration priority in community colleges*. Center for the Study of Higher and Postsecondary Education, University of Michigan, Ann Arbor, Michigan.
- Bahr, P. R., Hom, W., & Perry, P. (2004). Student readiness for postsecondary course work: Developing a college-level measure of student average academic preparation. *Journal of Applied Research in the Community College*, 12, 7-16.
- Bahr, P. R., Hom, W., & Perry, P. (2005). College transfer performance: A methodology for equitable measurement and comparison. *Journal of Applied Research in the Community College*, 13, 73-87.
- Bailey, K. D. (1994). *Typologies and taxonomies: An introduction to classification techniques*. Thousand Oaks, CA: Sage.

- Bailey, T. R., & Alfonso, M. (2005). Paths to persistence: An analysis of research on program effectiveness at community colleges. *Lumina Foundation for Education: New Agenda Series, 6(1)*.
- Bailey, T., Calcagno, J. C., Jenkins, D., Leinbach, T., & Kienzl, G. (2006). Is student-right-to-know all you should know? An analysis of community college graduation rates. *Research in Higher Education, 47*, 491-519.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bauer, D. J., & Shanahan, M. J. (2007). Modeling complex interactions: Person-centered and variable-centered approaches. In T. D. Little, J. A. Bovaird, and N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 255–283). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Borden, V. M. (1995). Segmenting student markets with a student satisfaction and priorities survey. *Research in Higher Education, 36*, 73-88.
- Borden, V. M. (2005). Identifying and analyzing group differences. In M.A. Coughlin (Ed.), *Intermediate/advanced statistics in institutional research* (pp. 132-168). Tallahassee, FL: Association for Institutional Research.
- Boughan, K. (2000). The role of academic process in student achievement: An application of structural equation modeling and cluster analysis to community college longitudinal data. *AIR Professional File, 74*.
- Bradburn, E. M., & Hurst, D. G. (2001). *Community college transfer rates to 4-year institutions using alternative definitions of transfer (NCES 2001-197)*. Washington, DC: National Center for Education Statistics.

- Clagett, C. A. (1995). Community college strategies: An outcomes typology for community colleges. *Assessment Update*, 7(4), 10-11.
- Clark, B., & Trow, M. (1966). The organizational context. In T. M. Newcomb & E. K. Wilson (Eds.), *College peer groups: Problems and prospects for research* (pp. 17-70). Chicago: Aldine.
- Cormack, R. M. (1971). A review of classification. *Journal of the Royal Statistical Society, Series A*, 134, 321-367.
- Cox, B. E. (2011). A developmental typology of faculty-student interaction outside the classroom. *New Directions for Institutional Research*, 2011(S1), 49-66.
- Dellow, D. A., & Romano, R. M. (2002). Measuring outcomes: Is the first-time, full-time cohort appropriate for the community college? *Community College Review*, 30(2), 42-54.
- Dowd, A. C. (2003). From access to outcome equity: Revitalizing the democratic mission of the community college. *Annals of the American Academy of Political and Social Science*, 586, Community Colleges: New Environments, New Directions, 92-119.
- Dowd, A. C. (2007). Community colleges as gateways and gatekeepers: Moving beyond the access "Saga" toward outcome equity. *Harvard Educational Review*, 77, 407-419.
- Dowd, A. C., & Tong, V. P. (2007). Accountability, assessment, and the scholarship of "best practice." In J.C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. XXII, pp. 57-119). New York: Springer.
- Dugan, J. P. (2011). Students' involvement in group experiences and connections to leadership development. *New Directions for Institutional Research*, 2011(S1), 17-32.
- Dugan, J. P. (2013). Patterns in group involvement experiences during college: Identifying a taxonomy. *Journal of College Student Development*, 54(3), 229-246.

- Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Hoboken, NJ: Wiley.
- Fenske, R. H., Keller, J. E., & Irwin, G. F. (1999). Toward a typology of early intervention programs. *Advances in Education Research, 4*, 117-133.
- Finch, W. H., & Bronk, K. C. (2011). Conducting confirmatory latent class analysis using Mplus. *Structural Equation Modeling, 18*, 132-151.
- Firebaugh, G. (2008). *Seven rules of social research*. Princeton, NJ: Princeton University Press.
- Flacks, R., & Thomas, S. L. (2007). "Outsides," student subcultures, and the massification of higher education. In J.C. Smart (Ed.), *Higher education: Handbook of theory and research* (pp. 181-218). New York: Springer.
- Fowlkes, E. B., Gnanadesikan, R., & Kettenring, J. R. (1988). Variable selection in clustering. *Journal of Classification, 5*, 205-228.
- Gibson, W. A. (1959). Three multivariate models: Factor analysis, latent structure analysis, and latent profile analysis. *Psychometrika, 24*(3), 229-252.
- Gillmore, G. M., & Hoffman, P. H. (1997). The graduate efficiency index: Validity and use as an accountability and research measure. *Research in Higher Education, 38*, 677-697.
- Goldrick-Rab, S. (2010). Challenges and opportunities for improving community college student success. *Review of Educational Research, 80*(3), 437-469.
- 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, 25*, 181-218.
- Hagedorn, L. S., & Kress, A. M. (2008). Using transcripts in analyses: Directions and opportunities. *New Directions for Community Colleges, 143*, 7-17.

- Hagedorn, L. S., & Prather, G. (2005). *The community college solar system: If university students are from Venus, community college students must be from Mars*. Paper presented at the 2005 annual forum of the Association for Institutional Research. San Diego.
- Hagerty, M. R. (1985). Improving the predictive power of conjoint analysis: The use of factor analysis and cluster analysis. *Journal of Marketing Research*, 22(2), 168-184.
- Hoachlander, G., Sikora, A. C., & Horn, L. (2003). *Community college students: Goals, academic preparation, and outcomes* (NCES 2003-164). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Hojtink, H. (2001). Confirmatory latent class analysis: Model selection using Bayes factors and (pseudo) likelihood ratio statistics. *Multivariate Behavioral Statistics*, 36, 563-588.
- Hom, W. C. (2009). The denominator as the “target.” *Community College Review*, 37(2), 136-152.
- Horn, L. (2009). *On track to complete? A taxonomy of beginning community college students and their outcomes 3 years after enrolling: 2003-04 through 2006* (NCES 2009-152). Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Horowitz, H. L. (1987). *Campus life: Undergraduate cultures from the end of the eighteenth century to the present*. New York: Knopf.
- Houle, C. O. (1961). *The inquiring mind*. Madison: University of Wisconsin Press.
- Hu, S., Katherine, L., & Kuh, G. D. (2011). Student typologies in higher education. *New Directions for Institutional Research*, 51, 5-15.
- Hu, S., & Li, S. (2011). Typological research on college students for better outcomes. *New Directions for Institutional Research*, 2011(S1), 83-87.

- Hu, S., & McCormick, A. (2012). An engagement-based student typology and its relationship to college outcomes. *Res High Educ, 53*, 738-754.
- Jain, A. K., & Dubes, R.C. (1988). *Algorithms for clustering data*. Upper Saddle River, NJ: Prentice Hall.
- Janosz, M., Le Blanc, M., Boulerice, B., & Tremblay, R. E. (2000). Predicting different types of school dropouts: A typological approach with two longitudinal samples. *Journal of Educational Psychology, 92(1)*, 171-190.
- Katchadourian, H. A., & Bolin, J. (1985). *Careerism and intellectualism among college students: Patterns of academic and career choice in the undergraduate years*. San Francisco: Jossey-Bass.
- Kaymak, U., & Setnes, M. (2000). *Extended fuzzy clustering algorithms*. Discussion paper of Erasmus Research Institute of Management (ERIM, Erasmus University Rotterdam).
- Keniston, K. (1973). The faces in the lecture room. In R. Morrison (Ed.), *The contemporary university: U.S.A.* (pp. 315-349). Boston: Houghton Mifflin.
- Kim, K. A. (2002). ERIC Review: Exploring the meaning of “nontraditional” at the community college. *Community College Review, 30(1)*, 74-89.
- Kim, K. A., Sax, L. J., Lee, J. J., & Hagedorn, L. S. (2010). Redefining nontraditional students: Exploring the self-perceptions of community college students. *Community College Journal of Research and Practice, 34(5)*, 402-422.
- Knight, D. B. (2013). Reversing the logic: An outcomes-based student typology for determining “what works” in promoting an array of engineering-related student learning outcomes. *Educational Evaluation and Policy Analysis, 0162373713504987*.

- Krathwohl, D. R. (1998). *Methods of educational and social science research: An integrated approach* (2nd ed.). Longman.
- Kuh, G. D. (1990). Assessing student culture. In W. G. Tierney (Ed.), *Assessing academic climates and cultures. New Directions for Institutional Research*, No. 68. San Francisco: Jossey-Bass.
- Kuh, G. D. (1995). The other curriculum: Out-of-class experiences associated with student learning and personal development. *The Journal of Higher Education*, 123-155.
- Kuh, G. D., & Whitt, E. J. (1988). *The invisible tapestry: Culture in American colleges and universities. ASHE-ERIC Higher Education Report No. 1*. Washington, DC: Association for the Study of Higher Education.
- Kuh, G. D., Hu, S., & Vesper, N. (2000). They shall be known by what they do: An activities-based typology of college students. *Journal of College Student Development*, 41(2), 228-244.
- Kuh, G. D., Kinzie, J., Schuh, J. H., & Whitt, E. J. (2005). *Student success in college: Creating conditions that matter*. San Francisco, CA: Jossey-Bass.
- Laanan, F. S. (2000). Community college students' career and educational goals. *New Directions for Community Colleges*, 112, 19-33.
- Lane, J. E. (2003). Studying community colleges and their students: Context and research issues. *New Directions for Institutional Research*, 118, 51-67.
- Laudy, O., Boom, J., & Hoijsink, H. (2005). Bayesian computational methods for inequality constrained latent class analysis. In L.A. van der Ark, M.A. Croon, & K. Sijtsma (Eds.), *New developments in categorical data analysis for the social and behavioral sciences* (pp. 63–82). Mahwah, NJ: Lawrence Erlbaum Associates.

- Layzell, D. T. (1999). Linking performance to funding outcomes at the state level for public institutions of higher education: Past, present, and future. *Research in Higher Education, 40*, 233-246.
- Lazarsfeld, P. F. & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- Lewis-Beck, M. S. (1994). Series editor's introduction. In K.D. Bailey, *Typologies and taxonomies: An introduction to classification techniques*. Thousand Oaks, CA: Sage
- Likas, A., Vlassis, N., & Verbeek, J. (2003). The global *k*-means clustering algorithm. *Pattern Recognit, 36*(2), 451-461.
- Lo, Y., Mendell, N., & Rubin, D. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*, 767-778.
- Lorr, M. (1983). *Cluster analysis for social scientists*. San Francisco: Jossey-Bass.
- Luan, J. (2006). Using academic behavior index (AB-Index) to develop a learner typology for managing enrollment and course offerings: A data mining approach. *IR Applications*, July 2006. <http://www.airweb.org/?page=566>. Accessed October 30, 2013.
- Luan, J., Zhao, C. M., & Hayek, J. C. (2004). *Exploring a new frontier in higher education research: A case study analysis of using data mining techniques to create NSSE institutional typology*. Paper presented at the annual meeting of the California Association for Institutional Research, Anaheim, California.
- Luan, J., Zhao, C. M., & Hayek, J. C. (2009). Using a data mining approach to develop a student engagement-based institutional typology. *IR Applications, Volume 18*.
- Luo, J., & Jamieson-Drake, D. (2005). Linking student precollege characteristics to college development outcomes: The search for a meaningful way to inform institutional practice and policy. *IR Applications, Volume 7*.

- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. Lecam & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Vol. I* (pp. 281-297). Berkeley: University of California Press.
- Magidson, J., & Vermunt, J. K. (2002). Latent class models for clustering: A comparison with k-means. *Canadian Journal of Marketing Research, 20*, 37-44.
- Marti, C. N. (2008). Latent postsecondary persistence pathways: Educational pathways in American two-year colleges. *Res High Educ, 49*, 317-336.
- Masyn, K. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.) *The Oxford handbook of quantitative methods in psychology* (Vol. 2, pp. 551-611). New York, NY: Oxford University Press.
- Mauss, A. L. (1967). *Toward an empirical typology of junior college student subcultures*. Paper presented at the annual meeting of the Pacific Sociological Association, Long Beach, CA. ERIC Document Reproduction Service No. ED013076.
- McCutcheon, A. L. (1987). *Latent class analysis. Quantitative applications in the social sciences series, 64*. Thousand Oaks, California: Sage Publications.
- McCutcheon, A. L. (2002). Basic concepts and procedures in single-and multiple-group latent class analysis. In J. A. Hagenars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 57-88). Cambridge, UK: Cambridge University Press.
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
- Mezzich, J. E., & Solomon, H. (1980). *Taxonomy and behavioral science: Comparative performance of grouping methods*. London: Academic Press.

- Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, *45*, 325-342.
- Muthén, B. O. (2001). Latent variable mixture modeling. In G.A. Marcoulides & R.E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1-34). Mahwah, NJ: Lawrence Erlbaum Associates.
- Muthén, L. K., & Muthén, B. O. (2007). *Mplus. Statistical analysis with latent variables, Version 3*.
- Muthén, L. K., & Muthén, B. O. (2008). *Mplus user's guide* (5th ed.). Los Angeles: Muthén & Muthén.
- National Survey of Student Engagement. (2004). *Student engagement: Pathways to collegiate success*. Bloomington, IN: Indiana University Center for Postsecondary Research.
- National Survey of Student Engagement. (2005). *Student engagement: Exploring different dimensions of student engagement*. Bloomington, IN: Indiana University Center for Postsecondary Research.
- Newcomb, T. M., Koenig, K. E., Flacks, R., & Warwick, D. P. (1967). *Persistence and change: Bennington College and its students after twenty-five years*. Hoboken, NJ: Wiley.
- Noftinger, J. B., & Newbold, K. F. (2007). Historical underpinnings of access to American higher education. *New Directions for Higher Education*, *138*, 3-18.
- Nylund, K.L., Asparouhov, T., & Muthén, B.O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14*, 535-569.
- Pakhira, M. K., Bandyopadhyay, S., & Maulik, U. (2004). Validity index for crisp and fuzzy clusters. *Pattern Recognition*, *37*, 487-501.

- Pascarella, E. T., & Terenzini, P. T. (1980). Predicting freshman persistence and voluntary dropout decisions from a theoretical model. *The Journal of Higher Education*, 60-75.
- Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students*. San Francisco: Jossey-Bass.
- Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students* (2nd ed.). San Francisco: Jossey-Bass.
- Perry, P. (2005). *AB1417 District Performance Framework Project*. Presentation at Ohlone College. Fremont, California.
- Perry, W. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart, & Winston.
- Punj, G., & Stewart, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, 20, 134-148.
- Rapkin, B. D., & Luke, D. A. (1993). Cluster analysis in community research: Epistemology and practice. *American Journal of Community Psychology*, 21(2), 247-277.
- Robinson, R. (2004). Pathways to completion: Patterns of progression through a university degree. *Higher Education*, 47, 1-20.
- Roksa, J. (2006). Does the vocational focus of community colleges hinder students' educational attainment? *The Review of Higher Education*, 29(4), 499-526.
- Saenz, V. B., Hatch, D., Bukoski, B. E., Kim, S., Lee, K., & Valdez, P. (2011). Community college student engagement patterns: A typology revealed through exploratory cluster analysis. *Community College Review*, 39(3), 235-267.
- Samuelsen, K., & Raczynski, K. (2013). Latent class/profile analysis. *Applied Quantitative Analysis in the Social Sciences*, 304-328.

- Santibáñez, L., Gonzalez, G. Morrison, P. A., & Carroll, S. J. (2007). Methods for gauging the target populations that community colleges serve. *Population Research and Policy Review, 26*, 51-67.
- Schuck, P., & Zeckhauser, J. (2008). Targeting in social programs: Avoiding bad bets, removing bad apples. *Journal of Regional Science, 48*, 465-467.
- Scott, M. A., & Kennedy, B. B. (2005). Pitfalls in pathways: Some perspectives on competing risks event history analysis in education research. *Journal of Educational and Behavioral Statistics, 30(4)*, 413-442.
- Scott-Clayton, J. (2011). *The shapeless river: Does a lack of structure inhibit students' progress at community colleges?* (CCRC Working Paper No. 25, Assessment of Evidence Series). New York, NY: Columbia University, Teachers College, Community College Research Center.
- Seybert, J. A. (2002). Assessing student learning outcomes. *New Directions for Community Colleges, 117*, 55-65.
- Shannon, H. D., & Smith, R. C. (2006). A case for the community college's open access mission. In B.K. Townsend & K.J. Dougherty (Eds.), *Community college missions in the 21st Century. New Directions for Community Colleges*, No. 136. San Francisco: Jossey-Bass.
- Shaw, K. M., & Jacobs, J. A. (2003). Community colleges: New environments, new directions. *Annals of the American Academy of Political and Social Science, 586*, 6-15.
- Shulock, N., & Moore, C. (2007). *Rules of the game: How state policy creates barriers to degree completion and impedes student success in the California community colleges*. Sacramento, CA: Institute for Higher Education Leadership and Policy.

- Skinner, B. F. (1945). The operational analysis of psychological terms. *Psychological review*, 52(5), 270.
- Skinner, B. F. (1953). *Science and Human Behavior*. New York: Macmillan.
- Sokal, R. R., & Sneath, P. H. A. (1963). *Principles of numerical taxonomy*. San Francisco: W. H. Freeman.
- Sousa, J. M., & Kaymak, U. (2002). Fuzzy decision making in modeling and control. *Robotics and intelligent systems (vol. 27)*. River Edge, NJ: World Scientific.
- Taber, T. D., & Hackman, J. D. (1976). Dimensions of undergraduate college performance. *Journal of Applied Psychology*, 61, 546–558.
- Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45, 89-125.
- Townsend, B. K. (2002). Transfer rates: A problematic criterion for measuring the community college. *New Directions for Community Colleges*, 117, 13–23.
- Tryon, R. C. (1939). *Cluster analysis: Correlation profile and orthometric (factor) analysis for the isolation of unities in mind and personality*. Edwards brothers, Inc., lithoprinters and publishers.
- Tryon, R. C. (1958). Cumulative communality cluster analysis. *Educational and Psychological Measurement*, 18(1), 3-35.
- VanDerLinden, K. (2002). *Credit student analysis: 1999 and 2000*. Annapolis Junction, MD: Community College Press, American Association of Community Colleges.
- Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J.A. Hagenaars & A.L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89-106). Cambridge, MA: Cambridge University Press.

- Vichi, M., & Opitz, O. (Eds.). (2012). *Classification and Data Analysis: Theory and Application Proceedings of the Biannual Meeting of the Classification Group of Società Italiana Di Statistica (SIS) Pescara, July 3–4, 1997*. Springer Science & Business Media.
- Voorhees, R. A., & Zhou, D. (2000). Intentions and goals at the community college: Associating student perceptions and demographics. *Community College Journal of Research and Practice, 24*, 219-232.
- Wang, W. (2004). UCLA community college review: Community education in the community college. *Community College Review, 32*, 43-56.
- Wassmer, R., Moore, C., & Shulock, N. (2004). Effect of racial/ethnic composition on transfer rates in community colleges: Implications for policy and practice. *Research in Higher Education, 45*(6), 651-672.
- Willett, T., & Hom, W. (2007). Student flow analysis for a community college. *Journal of Applied Research in the Community College, 15*(1), 17-27.
- Watson, J. B. (1925). *Behaviorism*. Transaction Books.
- Zhao, C., Gonyea, R .M., & Kuh, G. D. (2003). *The psychographic typology: Toward higher resolution research on college students*. Paper presented at the annual forum of the Association for Institutional Research, Tampa, FL.
- Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of personality*. New York, NY: Cambridge University Press.

## APPENDIX A

## BAHR VALIDATION LPA SYNTAX

```
TITLE: Bahr LPA Validation Model
DATA: FILE IS clpabstd.csv;
VARIABLE: NAMES ARE
Gender
Race
TE
CS
MCA
NTDCA
NCA
TMCA
TECA
TOCA
;
USEVARIABLES ARE
Gender
Race
TE
CS
MCA
NTDCA
NCA
TMCA
TECA
TOCA
;
MISSING ARE .;
          CLASSES = c (5);
MODEL:
  %overall%
  c on Gender Race;
ANALYSIS: TYPE = MIXTURE;
          STARTS = 500 50;
          lrtbootstrap = 500;
          lrtstarts = 50 20 50 20;
          estimator = mlr;

SAVEDATA: File = output.sav;
SAVE = cprob;
OUTPUT: TECH7 TECH11 TECH14 cinterval;
PLOT:
TYPE = Plot2 Plot3;
Series = Gender(1) Race(2);
```

## APPENDIX B

## BAHR VALIDATION CLPA SYNTAX

TITLE: Bahr CLPA Validation Model  
DATA: FILE IS clpabstd.csv;  
VARIABLE: NAMES ARE y1-y8 x1-x2;  
USEVARIABLES ARE y1-y8;  
MISSING ARE .;  
CLASSES = c (5);  
ANALYSIS: TYPE = MIXTURE;  
STARTS=1000 100;  
LRTSTARTS=500 200 500 200;  
LRTBOOTSTRAP=1000;  
MITERATIONS=2000;

## MODEL:

%OVERALL%

%c#1%

[y1\*1] (a1);

[y2\*1] (b1);

[y3\*1] (c1);

[y4\*1] (d1);

[y5\*1] (e1);

[y6\*1] (f1);

[y7\*1] (g1);

[y8\*1] (h1);

%c#2%

[y1\*1] (a2);

[y2\*1] (b2);

[y3\*1] (c2);

[y4\*1] (d2);

[y5\*1] (e2);

[y6\*1] (f2);

[y7\*1] (g2);

[y8\*1] (h2);

%c#3%

[y1\*1] (a3);

[y2\*1] (b3);

[y3\*1] (c3);

[y4\*1] (d3);

[y5\*1] (e3);

[y6\*1] (f3);

[y7\*1] (g3);

[y8\*1] (h3);

%c#4%

[y1\*1] (a4);

[y2\*1] (b4);

[y3\*1] (c4);

[y4\*1] (d4);

```
[y5*1] (e4);  
[y6*1] (f4);  
[y7*1] (g4);  
[y8*1] (h4);  
%c#5%  
[y1*1] (a5);  
[y2*1] (b5);  
[y3*1] (c5);  
[y4*1] (d5);  
[y5*1] (e5);  
[y6*1] (f5);  
[y7*1] (g5);  
[y8*1] (h5);
```

**MODEL CONSTRAINT:**

```
a1>a5;  
a5>a2;  
a2>a3;  
a3=a4;  
b1>b5;  
b5>b4;  
b4>b2;  
b2>b3;  
c1>c2;  
c2>c3;  
c3>c4;  
c4>c5;  
d2>d1;  
d1>d5;  
d5>d3;  
d3=d4;  
e5>e1;  
e1>e2;  
e2>e4;  
e4>e3;  
f1>f2;  
f2>f5;  
f5>f3;  
f3>f4;  
g1>g2;  
g2>g5;  
g5>g3;  
g3>g4;  
h1>h2;  
h2>h3;  
h3=h5;  
h5>h4;  
OUTPUT:      TECH14;  
SAVEDATA: File = output.sav;  
SAVE = cprob;
```

## APPENDIX C

## LPA VALIDATION SYNTAX

TITLE: LPA Syntax for All Student Types Analysis

DATA: FILE IS lpaallstd.csv;

VARIABLE: NAMES ARE

Gender

Race

TE

CS

MCA

NTDCA

NCA

TMCA

TECA

TOCA

;

USEVARIABLES ARE

Gender

Race

TE

CS

MCA

NTDCA

NCA

TMCA

TECA

TOCA

;

MISSING ARE .;

CLASSES = c (6);

MODEL:

%overall%

c on Gender Race;

ANALYSIS: TYPE = MIXTURE;

STARTS = 500 50;

lrtbootstrap = 1000;

lrtstarts = 50 20 50 20;

estimator = mlr;

SAVEDATA: File = output.sav;

RESULTS ARE result.sav;

ESTIMATES ARE est.sav;

SAVE = cprob;

OUTPUT: TECH7 TECH11 TECH14 cinterval;

PLOT:

TYPE = Plot2 Plot3;

Series = Gender(1) Race(2);