

Dale Pietrzak, Ed.D., LPCMH, CCMHC
Senior Director Institutional Effectiveness and Research (IER)
dpietrzak@vinu.edu --- 812-888-4141



Cluster Analysis and Retention

September 2021

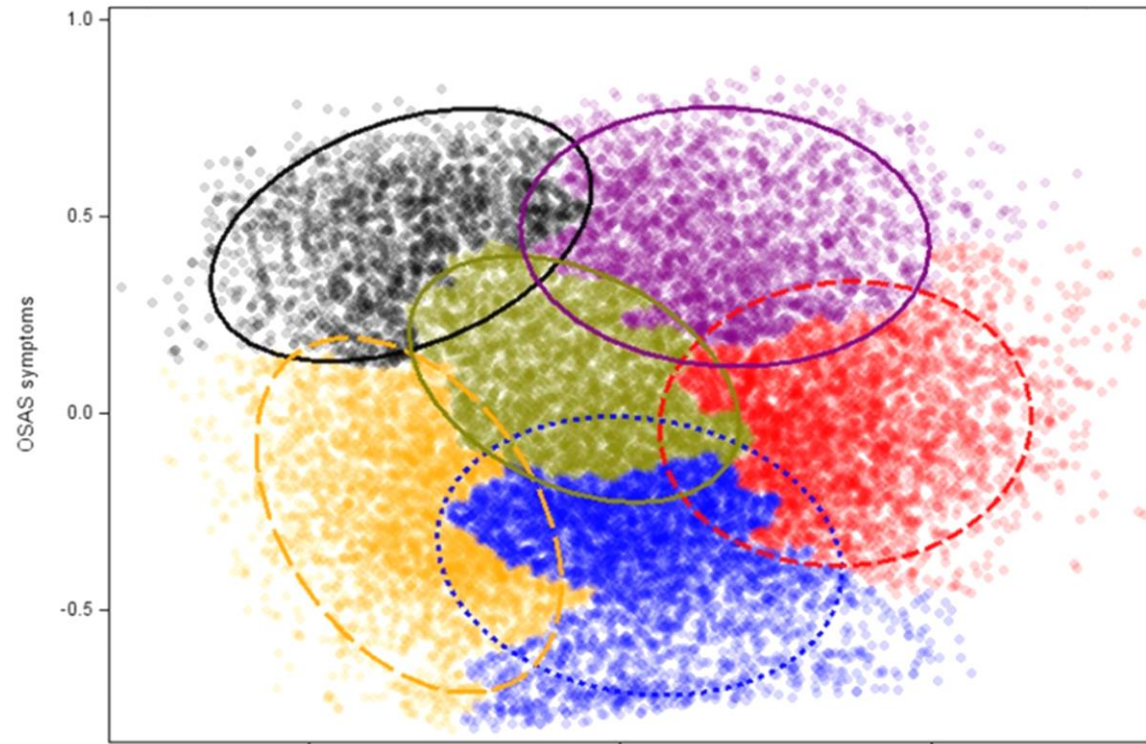
Research Team: Dale Pietrzak, Jennifer Holscher, Chad Bebee, Carmen Ridgway, Jonathan Williams, Kimela Meeks & Jonathan Williams

Overview

- What is Cluster Analysis?
- Method
- Results
 - Development of Narratives
 - Role of Self Efficacy and Skills and Abilities
- Example Narratives
- Applications
 - Retention
 - Enrollment and Admissions
 - Withdrawals

Cluster Analysis: What is it?

Axes correspond to individual coordinates for the two main dimensions of the multiple correspondence analysis. Cluster 1: the young symptomatic. Cluster 2: the old obese. Cluster 3: the multi-disease (MD) old obese. Cluster 4: the young snorers. Cluster 5: the drowsy obese. Cluster 6: the MD obese symptomatic.



Cluster Analysis is a procedure used in medicine, psychology and marketing to identify groups of people who are similar from within a larger population.

In medicine or psychology these groups generally form patient types for treatment.

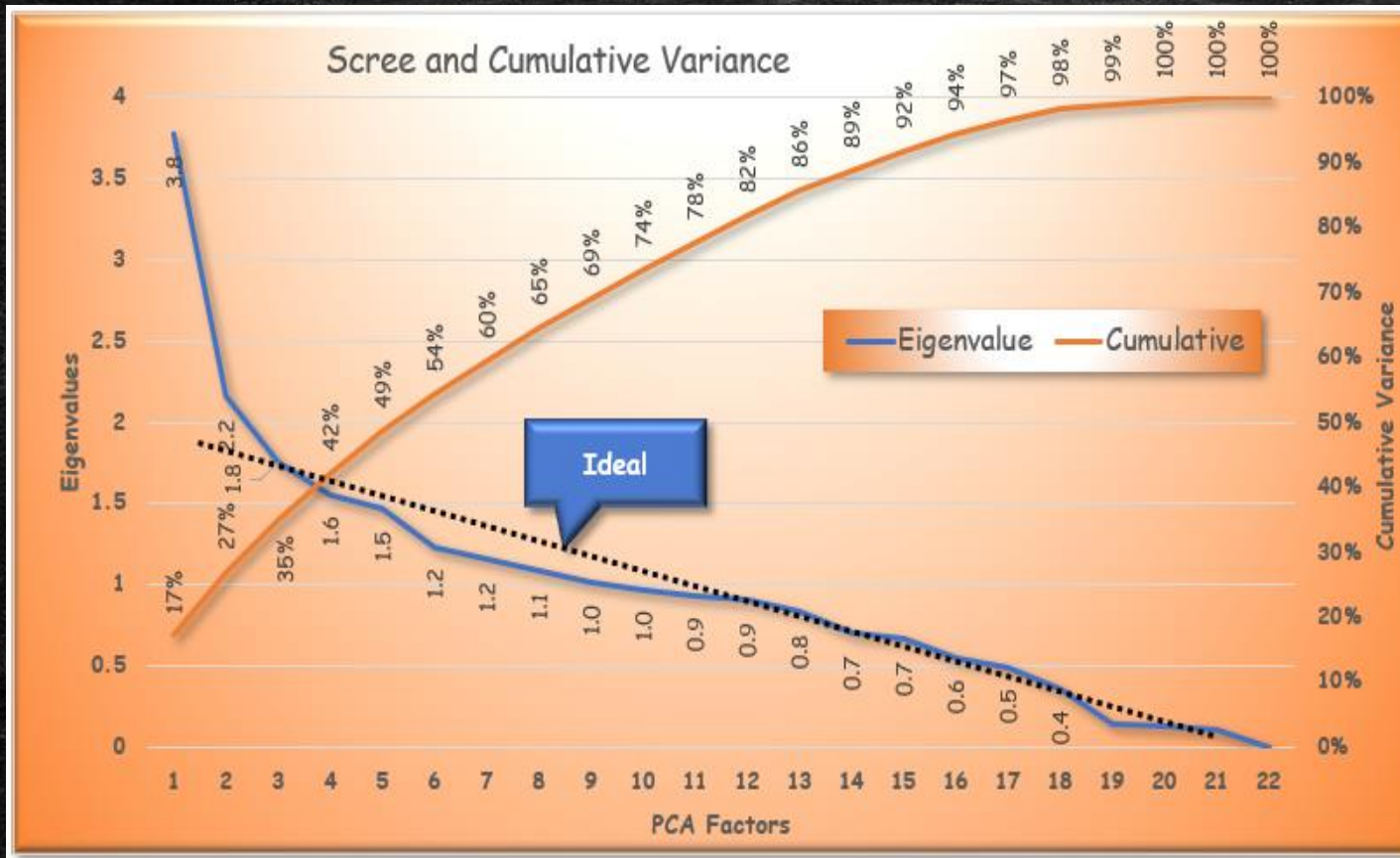
In marketing these groups become profiles to target marketing campaigns.

This scatter plot is from a study of Obstructive Sleep Apnea.

Method: Background Development

- Identify Key elements in retention and student success from Research Literature.
 - The plight of terminal uniqueness.
 - Theoretical Framework Vs the Spaghetti Noodle Toss (Correlation goes both ways)
- Examine data for Local Characteristics of import.
 - Distance
 - Financial
 - Key Points
 - Student and University Characteristics
 - Non-Academic Data
- Split Samples and Testing Proxy Elements
 - Standardize Data for use
 - 3 Years of Fulltime Non-Retained

Method: Check for Redundancy



There are many ways to check for redundancy. I used Principal Components Analysis

Ideally one would like to see a plot of eigenvalues and cumulative variance which forms a straight line "X". Here Factor 1 (Student Success - Skills and Ability) have more variance than desired. However there is planned redundancy.

Method: Proxy Variables Standardized (24)

Standardized z-score

- Age
- HS GPA
- Term 1 GPA (Year 1 Only)
- Real Time Indicator
- Student Risk Profile

Binary Coded - 1: Present 0: Not

- Vincennes Campus (Main Campus) or Other
- Continuing Student or Other
- First-Time Freshman or Other
- Citizen or Other
- Gender Male or Female
- First Generation or Other
- Pell Eligible
- White or Other
- Black or Other
- Hispanic or Other
- Core 40 HS or Other
- Indiana Urban or Other
- Indiana Rural or Other
- Indiana Mixed Urban-Rural or other
- Campus Housing or Other
- Athlete
- Completed 15 > Credits or Other
- Completed 12-15 Credits or Other
- Completed 0-12 Credits or Other

Method: K-Means Cluster Model

- A sample of 3 years of non-retained students (term 1 and year 1) was selected.
- This was randomly split into a test and validate sample.
- The CCC and Cumulative Variance Chart suggested there were 3 to 5 clusters.
- Cluster models from 2 to 10 were developed using K-means Clustering.
- Team reviewed the models and determined that the 6 Cluster model would be of more utility. For them it had the best balance of cluster sizes and common characteristics.

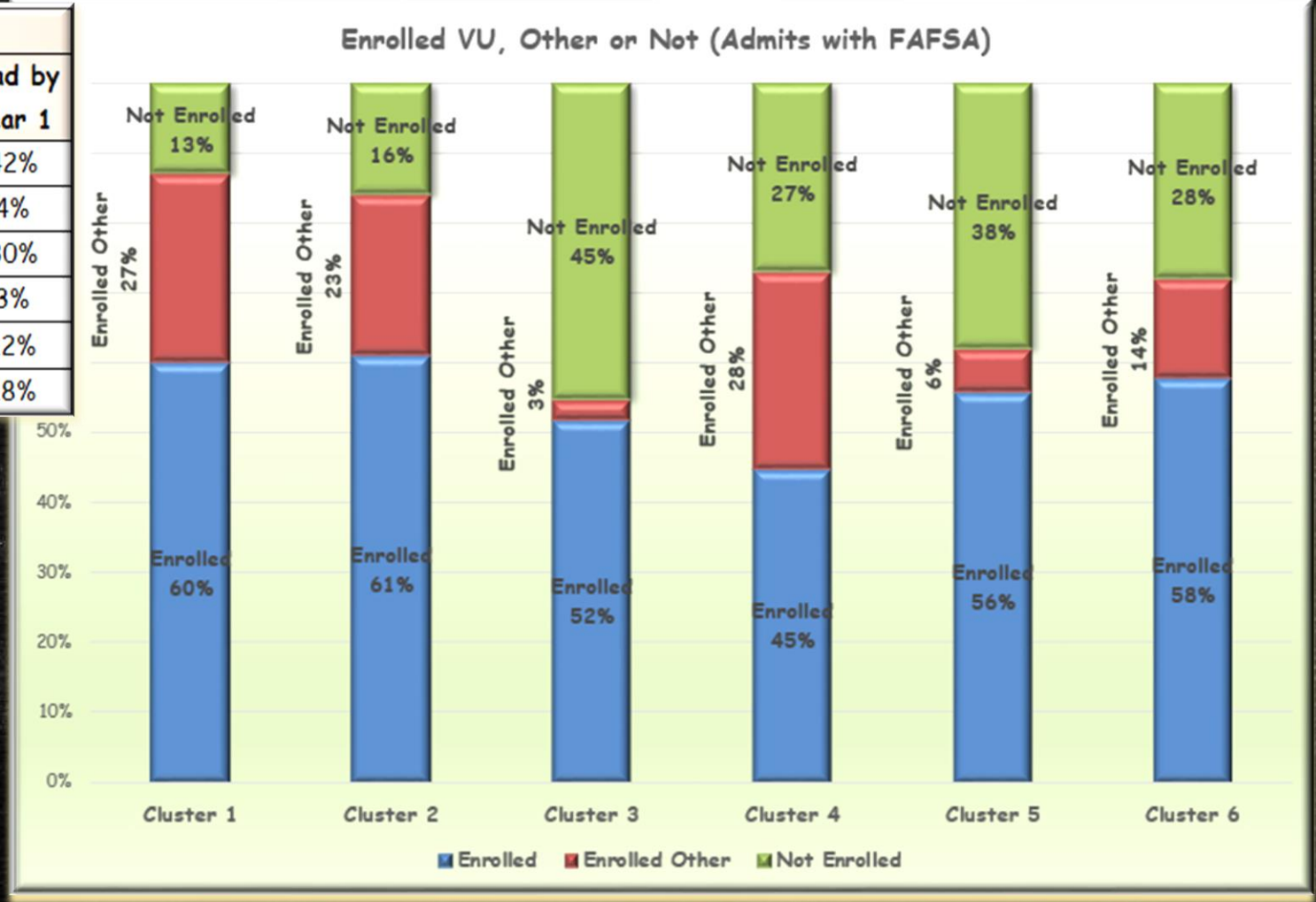
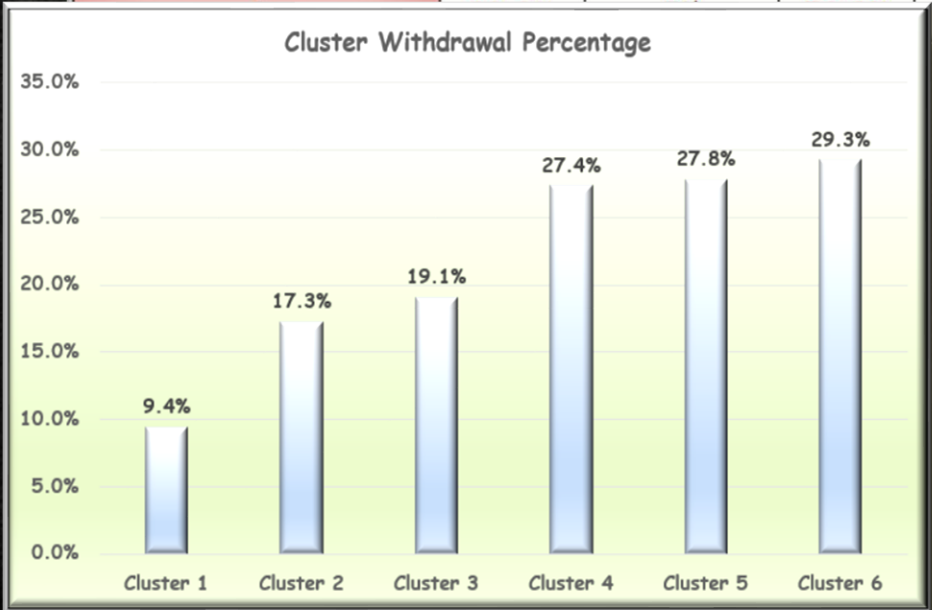
Results: Development of Base Profiles

Cluster	VM Campus	Continuing	New Freshman	White	Black	Hispanic	Urban	Rural	Mixed Rural-Urban	Campus Housing	Athlete	CORE HS	GED	15+ Credits	12-15 Credits	Credit 0-12	Citizen	1st Gen	Pell Eligible	Gender	Age	HS GPA	T1 CUM GPA
Cluster 1	H	HH	L		L							H		H						H	L	L	H
Cluster 2		LLL	HHH	H	L		L		H							H					L	L	LLL
Cluster 3	LLL		L							LL	L	LLL	HH				LL				HHH	LL	HH
Cluster 4	H	LL	HH	LLL	HHH		HHH	LL	LL	HHH			L						HH		L	LLL	LLL
Cluster 5	LLL		L							LL	L	LLL	HH	L	H				HH	L	HHH	L	HH
Cluster 6		HH	L	H	L		L			L			L			L			L		L	HHH	HHH

1. The base profiles for each cluster were developed using the overall cluster means and standard deviations as the reference (Essentially like referencing a Base Rate).
2. If the mean cluster value was 1 standard deviation higher or lower is was labeled "H" or "L" respectively. If it was 2 standard deviations from the mean it was labeled "HH" or "LL" depending on if it were high or low. Three standard deviations were labeled "HHH" or "LLL" in the same manner.
3. Note this is a Proportional Profile and not an absolute value profile. One could do such if desired.
4. Our Research Team reviewed various models and selected 6 Cluster Model.

Results: Developing Profiles - Campus Data

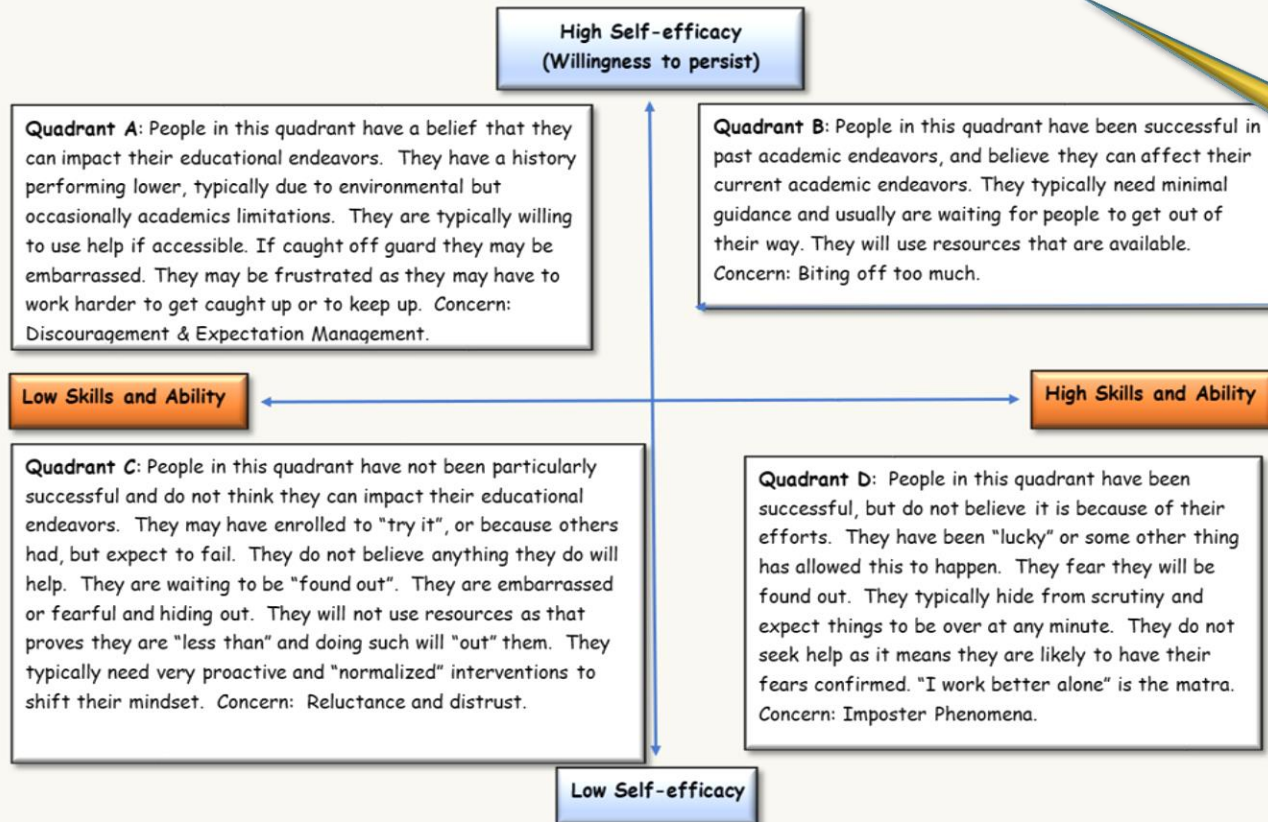
Fulltime Students: Retention				
Cluster	Term 2	Grad by Term 2	Year 1	Grad by Year 1
1	92.4%	12%	72.3%	42%
2	60.9%	0%	36.1%	4%
3	82.2%	8%	48.6%	30%
4	58.9%	1%	25.1%	3%
5	47.6%	3%	41.7%	12%
				18%



Results: Developing Profiles - Research Literature

Application of Academic Abilities and Academic Self-Efficacy.

Both Academic Self-Efficacy and Prior Skill and abilities can be thought of as continuums. One is considered the strongest "cognitive" predictor while the other is one of the strongest "psychological" predictors of university continuation and success.



2. Using Campus Performance data such as retention rates, transfer rates, graduation rates, etc. (Left for Example)

Example of using the literature.
Correlation Goes both ways.

Hayat, Shateri, Amini & Schokrpour, 2020;
Richardson, Abraham & Bond 2012; Rogers &
Summers, 2008

Example Narrative 1 (rural under performing)

Cluster 2 Description: This cluster makes up 19% of the population. This student is more commonly a white, traditional, first-time, full-time freshman student from a rural, small-town Indiana. They are similar to the general population in terms of percent Pell eligible. They more commonly have higher high school GPAs and graduate with a CORE high school diploma. This student appears academically prepared but when they arrive they do not show strong academic behaviors. They likely lack academic maturity, and while able to get by in high school at college display poor academic behaviors. They drop classes, miss classes and have difficulty adjusting to college. On average these students have an end of term GPA of 1.1. They may be engaging in inappropriate socialization (or isolation) which may include drinking, late nights, and social activities. This student doesn't come to office hours unless they are required, may respond to email but likely won't initiate contact, and this often student resents being required to utilize student support.

Cluster 2 Intervention: Only 0.4% of these students graduated between fall and spring term. About 7.0% transferred out within next year; of those, 5.1% transferred to another 2-year college, indicating these students might be looking for a fresh start in college. They likely have lower academic maturity and lower academic self-efficacy. While they did well in high school they experience the college environment as challenging. This triggers self-doubt and embarrassment. They are unlikely to seek out the needed support services. Further, they often feel that support services are not able to help. This belief is often coupled with a sense that using such services is a sign of their being "less than." This group generally needs a very proactive and intensive set of interventions. An early, intrusive and high contact advising model is suggested for these students. This may require residence hall advisor's observation and reporting.

Example Narrative 2 (urban under performing)

Cluster 4 Description: This cluster makes up about 20% of the population. These students are more commonly traditional students who are freshman on the Vincennes Campus. They are more commonly black and Hispanic and come from urban Indiana areas. This cluster is disproportionately Pell eligible and live in campus housing. They are not as well connected to the wider university or academic community. These students more often bring weaker college preparation and struggled in high school. They are typically not as academically mature. They likely intend to perform well in college by relying on the academic strategies they used in high school. This student often encounters trouble early into the semester. They are likely impacted by living in an area that has less ethnically diverse community support structures. These students most commonly feel less connected to their instructors.

Cluster 4 Intervention: About 3% of these students graduate between fall and spring terms. The end of term GPA is on average is 1.1. An additional 9.7% will have transferred out within the next year; of these, 6.5% transferred to another 2-year college, suggesting difficulty adjusting to VU or the community environment. These are students that often require intrusive advising models and additional social support. They are not only under represented, but from a very different cultural environment (urban setting). Proactive advisors, who can help with developing appropriate academic maturity and study skills are important, but the advisors will also need to help them find ways to connect to the university/community. If the community does not have indirect support services (i.e., hair care, fashion, activities, etc.) the university may wish to consider developing other ways to develop a sense of community. As they are often students with lower academic self-efficacy, they will shy from assistive services. A university "academic coach/cheerleader" who can focus on developing connections and academic self-efficacy. Advisors may find difficulty in doing this without triggering these students' self-message that they are not able to be successful.



Questions?

Selected Resources

- Hayat, Shateri, Amini & Schokrpour (2020). Relationships between academic self-efficacy, learning related emotions, metacognitive learning strategies with academic performance in medical students: a structural equation model. *BMC Medical Education*, 20:76
- Lens & Levräu (2020) Can Pre-Entry Characteristics Account for the Ethnic Attainment Gap? An Analysis of a Flemish University. *Research in Higher Education*. 61, 26-50.
- Richardson, Abraham & Bond (2012). Psychological Correlates of University Students' Academic Performance. 138, 353-387.
- Rogers & Summers (2008). African American Students at Predominantly White Institutions: A motivational and Self-Systems Approach to Understanding Retention. *Educ Psychol Rev* 20, 171-190.
- VanSickle, Schuler, Quinn, et al. (2020). Closing the Achievement Gap for Underrepresented Minority Students in Stem: A Deep Look at a Comprehensive Intervention. *Journal of STEM Education*, 21, 5-18.