

Berea, Anamaria; Tsvetovat, Maksim; Daun-Barnett, Nathan; Greenwald, Mathew;
Cox, Elena

Article

A new multi-dimensional conceptualization of individual achievement in college

Decision Analytics

Provided in Cooperation with:

Springer Nature

Suggested Citation: Berea, Anamaria; Tsvetovat, Maksim; Daun-Barnett, Nathan; Greenwald, Mathew; Cox, Elena (2015) : A new multi-dimensional conceptualization of individual achievement in college, Decision Analytics, ISSN 2193-8636, Springer, Heidelberg, Vol. 2, Iss. 1, pp. 1-15, <https://doi.org/10.1186/s40165-015-0012-8>

This Version is available at:

<https://hdl.handle.net/10419/161799>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>

RESEARCH

Open Access

A new multi-dimensional conceptualization of individual achievement in college

Anamaria Berea^{1*}, Maksim Tsvetovat², Nathan Daun-Barnett³, Mathew Greenwald⁴ and Elena Cox⁵

*Correspondence:

aberea@rhsmith.umd.edu

¹Center for Complexity in Business,
University of Maryland, College Park,
MD 20742, USA

Full list of author information is
available at the end of the article

Abstract

Individual achievement in college has been closely tied to dropout and graduation rates and these indices of completion have been considered the most important measures for gauging achievement in undergraduate education. In this study, we both define and examine a new concept of thriving in college. Both the literature and the data suggest that a multidimensional concept of thriving is better suited for predicting student achievement. We base our data analysis on the data we collected from a nationwide survey conducted in 2013-2014. The survey data show that student achievement in college does not depend solely on academic performance, and that there is no single factor that predicts thriving in college, in general, for an individual student. We show that in a heterogeneous population of students and a heterogeneous set of colleges, there is no unique pattern or factor for all students to thrive in the same college. We define and quantify the concepts of multidimensional thriving, personal traits and college ecosystems and build an algorithm that shows which is the best college ecosystem for a unique student. Furthermore, we show which traits of this individual are most responsible for the subsequent thriving in this specific college and which other traits s/he should learn or acquire in order to increase these chances of thriving.

Keywords: Education analytics, Thriving, College choice

Background

While most of the education research has been geared towards policy design for the overall student success in college, usually measured by graduation and drop-out rates, there is little research on the individual decision practices for choosing colleges and applying for college. This paper shows a integrated model - data analysis - algorithm design that aims to help prospective students not only to choose a college but also to be successful in the college of their choice. The paper is structured in 3 parts: the re-conceptualization of student thriving in college (the rest of this chapter); the data collection and analysis for thriving (chapter 2) and the algorithm for individual decision making for choosing colleges (chapter 3).

Literature review

For the better part of five decades, education researchers have set out to understand the factors that influence student's success in college. Both academic studies and current policy in practice have emphasized grades, graduation and dropout rates in order to quantify

and predict student achievement in college (Carnevale and Fry 2001; Carnevale et al. 2012).

Some early models build upon Tinto's Interactionist Framework (Tinto 1993) for understanding students decisions to leave postsecondary education. His framework hypothesized that student's decisions to leave an institution were a function of a set of pre-college attributes, a commitment to the institution and the goal of earning a degree, and student's level of social and academic integration into the institution. Tinto's model suggests that, in order for students to effectively engage in the campus community, they must first separate from their prior communities. (Tinto 1993) Researchers debate the degree to which Tinto's theory explains the complex phenomena of student departure (dropout) (Braxton 2000, but what is clear from all that has been written is that the decision to leave is influenced by some combination of individual characteristics and attributes and the institution they choose to attend. The evidence testing the degree to which the interaction between the individual and the institution has been mixed, largely because the model has always treated students as a single homogeneous group, rather than recognizing the interplay between the student and the institution may differ depending upon characteristics of both.

Astin was one of the early education scholars to point towards the importance of a student's engagement in their educational environment as predictive of the quality of the student outcomes – as measured by student learning, engagement and successful completion of a college degree (Astin 1993). His Individual – Environment – Outcome (I – E – O) model suggests that student outcomes are ultimately a function of the fit between the individual and the institution.

Another education scholar, Kuh, developed and has administered the National Survey of Student Engagement (NSSE) to thousands of students over more than a decade and his work suggests that the important predictor of student success reflects the degree to which students are engaged in their campus communities (Kuh 2004). Kuh and colleagues (Kuh 2004) identify six properties or conditions common among the most effective institutions: (1) a living mission that is clear and consistently understood by students, faculty, and staff, (2) a focus on student learning in and out of the classroom, (3) environments adapted for educational enrichment, (4) clear pathways to student success, (5) an improvement-oriented ethos across the institution, and (6) shared responsibility for educational quality and student success. The key contribution of their work is to underscore the role institutions play in the success of undergraduate students.

Towards a multi-dimensional concept of thriving

A more recent line of inquiry on college student success draws on Seligman's concept of flourishing from positive psychology, to suggest that thriving in college includes a high degree of emotional, psychological, and social well-being (Seligman 2011).

Schreiner and her colleagues have developed a thriving quotient that identifies five key constructs that approximate student thriving: (1) engaged learning, (2) academic determination, (3) positive perspective, (4) diverse citizenship, and (5) social connectedness (Schreiner 2013). Their work distinguishes between simply earning a college degree and maximizing the full potential of that experience while enrolled. The constructs Scheiner identifies are rooted in the early work of Tinto and the retention literature (Bean 2005; Dey and Astin 1993; Oseguera and Rhee 2009; Seidman 2005; Singell 2010) but they

emphasize the psychosocial predictors of student development a formal conceptualization of the role institutions play in facilitating student thriving.

Perhaps the most important development in our understanding of the degree to which students thrive in college and succeed in earning their degrees is the growing body of research that recognizes the path to success is different depending upon the student. Perna and Thomas (2008) recognize that the success of a student is contingent upon their situated context. In particular, institutional practices will have different effects on the characteristics of the student. Schreiner (2014) synthesizes recent research in student services and finds that students of color, for example, do not benefit from interacting with faculty in the same ways that white students do. Equally, Latino students are more likely to commute to college and African American students are more likely to work off campus, meaning both groups report having less time to engage in campus activities, an important predictor of student success in college. Equally, some activities are more important than others depending upon a student's class or race/ethnicity. African American students who report serving in a leadership role in student activities and organizations are more likely to thrive and institutions that are intentional about including students of color in leadership roles can positively influence student success (Prades-Collins 2012). Ultimately, this body of research recognizes that there is no single path for all students to follow in order to thrive in college; rather it is a complex interaction of the individual and their environment.

Today, educators debate the importance of institutional fit in the student enrollment decision (Bowen and Bok 1998; Roderick et al. 2008).

The debate underscores the importance placed on the fit between the individual and the institution and in this study, we extend the work of Schreiner and her colleagues (Schreiner 2013) by recognizing the important role the institution plays helping students thrive during their college experience. Figure 1 outlines our conceptual model building from these early works.

We recognize that the degree to which students thrive in college is a function of a set of individual pre-college attributes in combination with their experiences while in college

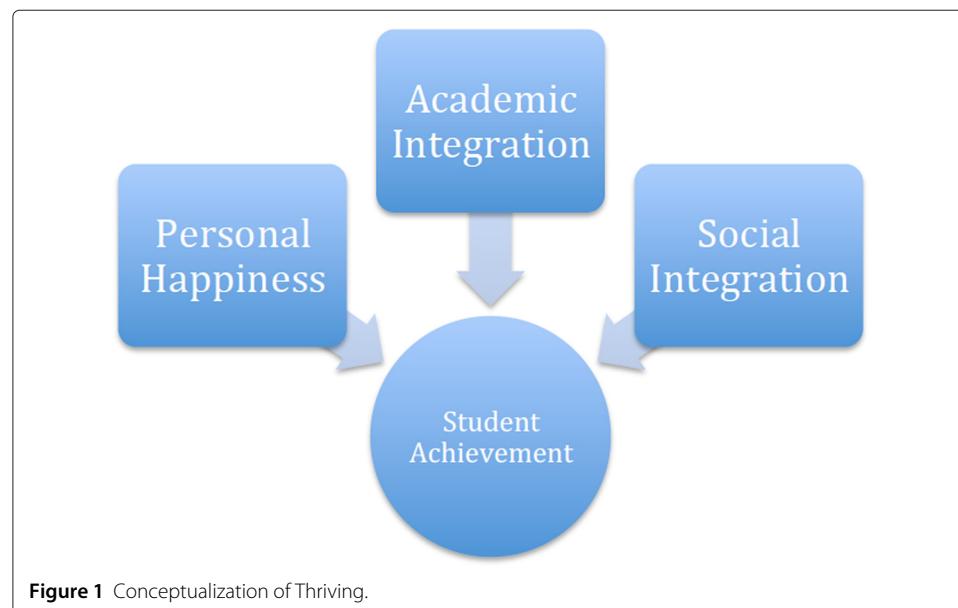


Figure 1 Conceptualization of Thriving.

and the characteristics of the institutions they attend. Our constructs at the individual level are similar to those identified by Schreiner; academic integration in our terms combines factors related to engaged learning and academic determination; social integration is similar to social connectedness; our inclusion of satisfaction/happiness is reflective of their positive perspective though it may get at more psychological constructs relative to self-efficacy and locus of control than we measure in our constructs. Schreiner discusses diverse citizenship in their conceptualization of thriving, which may be a limitation of the current investigation.

We add to our model a consideration of the role the institution plays in the degree to which students thrive while in college. This idea is well established in the early literature on student persistence and success and is critical if researchers are going to understand the complexity of fit and its relationship to the likelihood that students maximize the full potential of their college experience while in school.

Given that thriving characterizes a proportion of successful degree completers, we build from existing models of student achievement to identify what factors influence the degree to which students will thrive. We also recognize that thriving is contingent upon the environment. As such, our conceptualization of thriving must account for the interaction of the individual with their respective campus environment. Figure 1 provides the conceptual framework that informed our thinking about the thriving construct and the factors that influence whether a student is likely to thrive in a given environment. Based upon the existing literature on student persistence and achievement, we know that outcomes differ by personal characteristics (race, class, sex), pre-college attributes (student achievement, strength of curriculum, access to information, fit between the individual and the institution), which includes our three predictors of thriving – academic integration, social integration, and student satisfaction.

Methods

Our research consists of 3 phases: quantifying a multi-dimensional concept of thriving as described above; a national survey regarding thriving in college and the data analysis of the collected data; a personalized algorithm that renders the best college ecosystem for individual thriving.

Through our integrated theoretical – data collection & analysis – algorithm construction approach, we are looking mainly to answer the following questions:

1. Are there any factors that are more likely to lead to thriving in college, in general?
2. Which are the specific colleges where a student can thrive based on his or her unique characteristics?

We collected data through a nationwide panel survey in order to look for patterns of thriving in college or identify factors that are more likely to lead to thriving in college. We used these data to support the design of an algorithm that would allow any student to find the best college ecosystem where s/he would thrive in.

Data collection

To gather data for this project, an online quantitative survey was conducted using Research Now online consumer panel. To qualify for the survey, potential respondents had to be ages 18-24, living in the U.S. before entering college, and either in their

sophomore, junior or senior year at a postsecondary four-year institution or graduating within the past two years. Those obtaining their postsecondary instruction completely or mostly online were terminated, as were those who transferred or dropped out for financial reasons or external factors and those who attended two or more institutions but did not obtain their college degree. Shortly after the start of interviewing, these qualifiers were altered slightly to allow in those who had last attended college within the past four years and had not graduated, as well as transfer students who had graduated within the past four years. The purpose of these changes was to include more individuals who were not a good fit with their choice of schools. Finally, quotas were set by race/ethnicity to ensure adequate representation for analysis.

The questionnaire for the survey was designed by a reputable market research company in Washington, DC. Questions included in the instrument covered satisfaction with their college experience; college attributes including distance from home, student types, course of study and teaching methods, learning resources, preparation for the real world, student athletics and fitness, rules and structure, dorms, finances and other dimensions; respondent character traits and academic performance with a particular focus on what the student was like in high school; and pricing for the online tool. When answering questions about their college experience and college attributes, transfer students were asked to focus on the first college they attended. A series of questions was also asked to gather demographic characteristics, such as sex, age, race/ethnicity and family income. Six cognitive interviews were conducted before finalizing the questionnaire.

The interviews lasted an average of 25 minutes and were conducted between December 20, 2013 and January 6, 2014. Several methods were used to keep the respondents engaged and the majority found the survey experience extremely or very enjoyable. The large majority were able to keep their concentration on the survey questions and the median perceived elapsed time was only 20 minutes.

In total, 2,857 respondents were interviewed and included in the final data set. The data set was weighted by race/ethnicity and gender to match the distribution of these characteristics among 19-24 year-olds with at least some college in the U.S. population based on the March 2012 Current Population Survey. Data on college characteristics from the Integrated Postsecondary Education Data System (IPEDS) was merged into the data file. This resulting data set constituted a robust body of data for subsequent steps of the research process.

The institutional characteristics we included in the survey come from IPEDS (NCES 2015) and includes both basic institutional demographics (type, size, control, average net price), and unique institutional characteristics that may facilitate thriving generally (Carnegie classifications) and for specific sub-populations of students (i.e. Women's Colleges, HBCU, religious affiliation). We consulted three main sources to conceptualize institutions differently – the most recent version of the Carnegie Classification System for Colleges and Universities (NCES 2015), George Kuh's (2004) characterizations of DEEP Schools (Documenting Effective Educational Practice Project of the NSSE Institute), and the Integrated Postsecondary Education Data System (IPEDS) (NCES 2015). Some of the Carnegie classification data are included in the IPEDS system and are accessible to categorize institutions.

One of our key considerations throughout this process has been to avoid identifying matches between students and specific institutions. Rather, our emphasis has been to

identify the characteristics of colleges and universities that may facilitate or impede a student ability to thrive rather than a list of actual institutions.

The personal characteristics included in the survey included psychological traits (i.e. “ambitious”, “extroverted”), academic performance in high-school (i.e. “hard-working”, “completed projects”), economic and demographic characteristics (i.e. family income).

For each of these questions, the respondents answered on a 1-7 Likert scale. The data collected and used for the analysis consists of 605 variables, grouped as follows: demographic variables, economic variables, geo-spatial and transportation variables, high-school experience variables, behavioral variables, college campus variables and psychological traits variables.

A quantitative multi-dimensional concept of thriving

Additionally, the survey asked the students whether they considered themselves as thriving in college or not, based on 18 questions (dimensions of thriving) that would relate to the academic, personal happiness and social integration, as described in Section ‘Towards a multi-dimensional concept of thriving’ of this paper.

One of the goals of the survey was to find linkages between student characteristics and college characteristics that could then be used to predict thriving, based on self-reporting of the respondents regarding their own perception of thriving in college. In order to alleviate for the self-reporting bias regarding thriving, we conceptualized thriving in college based on the following 18 dimensions (variables):

- Q1. How satisfied are you with your overall experience attending college?
- Q2. How would you evaluate your choice of college?
- Q3a. How well does the following describe your experience at your college: you feel/felt you belong(ed) there
- Q3b. How well does the following describe your experience at your college: You can/could find support from friends, if you need(ed) it
- Q3c. How well does the following describe your experience at your college: you are/were satisfied with the number of friendships you have/had
- Q3d. How well does the following describe your experience at your college: outside the classroom itself, you have/had people you look(ed) up to
- Q3e. How well does the following describe your experience at your college: you enjoy(ed) involvement in non-academic student organizations
- Q3f. How well does the following describe your experience at your college: You have/had plenty of good times outside of class
- Q3g. How well does the following describe your experience at your college: your academic experience is/was satisfying
- Q3h. How well does the following describe your experience at your college: you have/had academic discussions with faculty outside of class
- Q3i. How well does the following describe your experience at your college: your classes are/were exciting to you
- Q3j. How well does the following describe your experience at your college: your college experience is helping/helped you develop intellectually
- Q3k. How well does the following describe your experience at your college: your college experience is helping/helped you learn to be more creative

- Q3l. How well does the following describe your experience at your college: your college experience is helping/helped you become comfortable talking about your ideas with others
- Q3m. How well does the following describe your experience at your college: college is helping/helped you learn how hard you can work to achieve a goal
- Q3n. How well does the following describe your experience at your college: college is helping/helped you acquire concrete skills that are useful in the real world
- Q3o. How well does the following describe your experience at your college: you are/were happy with life in college
- Q3p. How well does the following describe your experience at your college: your college experience is helping/helped you develop as a person beyond academics

Data analysis

Effect of demographic variables

First, we analyzed the data collected from the survey in order to find if there are any correlations of thriving with race, gender or family income, on each of the 18 dimensions of thriving. We found no significant correlations or relationship between these and thriving (see Figure 2).

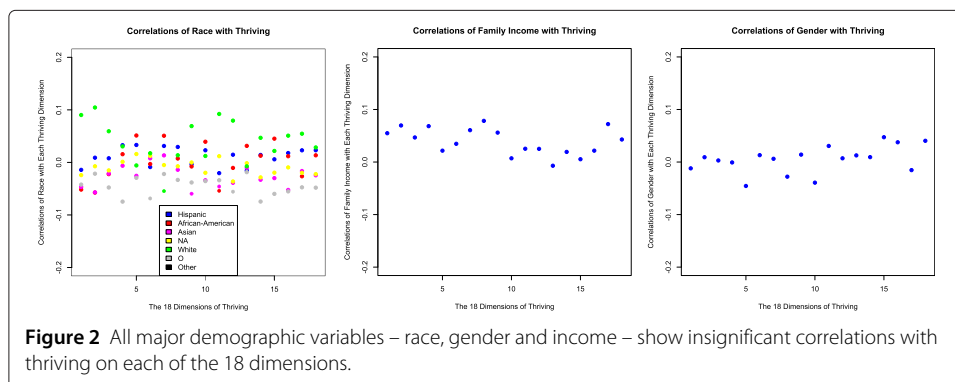
No general prediction of thriving

After we eliminated any demographic variables as potential predetermined factors for thriving, we also tested whether any of the other variables – personal traits and college traits – are correlated with thriving.

In order to do this, we analyzed not only the pairwise correlations of the each of the other variables with the 18 dimensions of thriving, but also calculated an aggregate measure of thriving based on the 3 supra-dimensions: academic, social and happiness, that we discussed in Section ‘Background’. The aggregation of the 18 thriving dimensions into 3 supra-dimensions was based on an exploratory factor analysis using the principal components method and calculates an overall raw score of thriving based on the following derived formulas:

The academic thriving score:

$$\begin{aligned}
 academic = & (.522 * Q3d) + (.539 * Q3g) + (.692 * Q3h) + (.624 * Q3i) \\
 & + (.648 * Q3j) + (.671 * Q3k) + (.611 * Q3l) + (.676 * Q3m) \\
 & + (.651 * Q3n) + (.579 * Q3p)
 \end{aligned}
 \tag{1}$$



The social thriving score:

$$social = (.519 * Q3a) + (.757 * Q3b) + (.785 * Q3c) + (.557 * Q3d) + (.636 * Q3e) + (.756 * Q3f) + (.548 * Q3o) \tag{2}$$

The happiness score:

$$happiness = (.771 * Q1) + (.881 * Q2) + (.636 * Q3a) + (.644 * Q3g) + (.639 * Q3o) \tag{3}$$

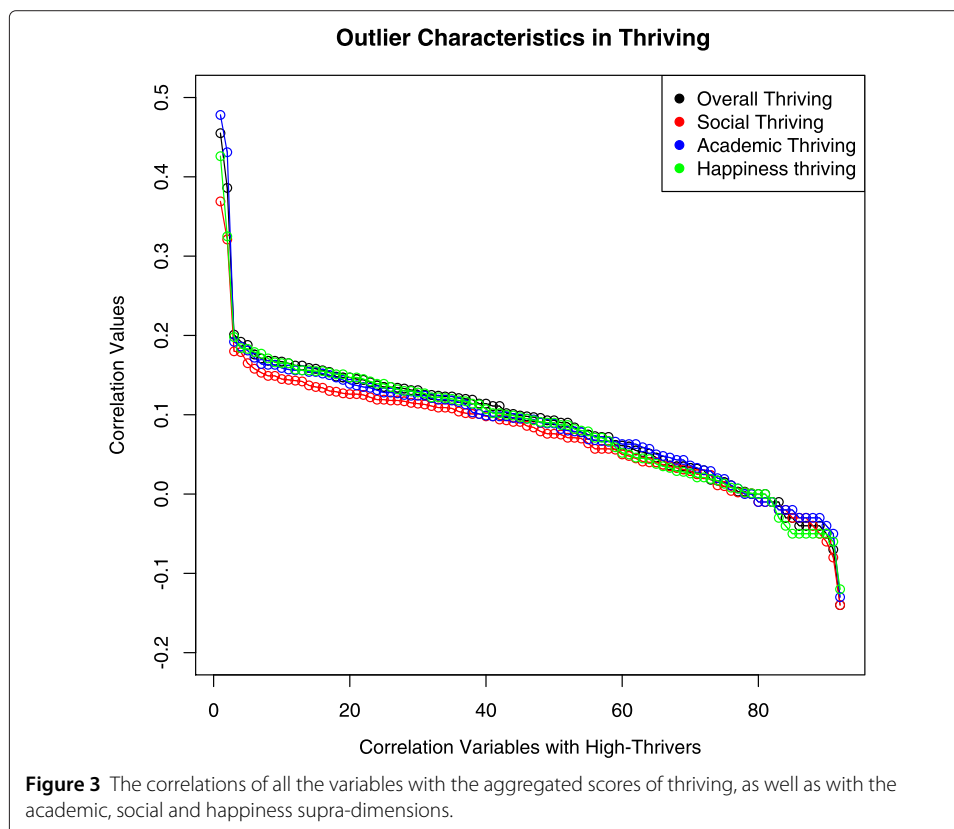
where Q1 – Q3p are each of the 18 thriving dimensions described in Section ‘A quantitative multi-dimensional concept of thriving’. Based on these aggregated 3 scores, we calculated the aggregated raw overall thriving score by normalizing the above 3 raw scores, as follows:

$$overall = (.24819 * academic) + (.21833 * happiness) + (.21601 * social) \tag{4}$$

Figure 3 shows that the correlations between an aggregated dimension of thriving (overall) and all the other variables in the data set are insignificant.

We also ran the correlations of all the other personal and college traits variables with each of the 18 dimensions of thriving, without any significant results. Additionally, we ran a k-cluster analysis in order to identify those clusters of variables, particularly college variables, that are predictive to thriving.

All these analyses proved that there is not one single variable that is significantly correlated with thriving. The analyses above were done by 3 independent teams and they show that both the student and the college universes are very diverse and heterogeneous



and that aggregating the data and looking for general patterns of thriving does not render any variable for predicting thriving in general.

Vibeffect algorithm

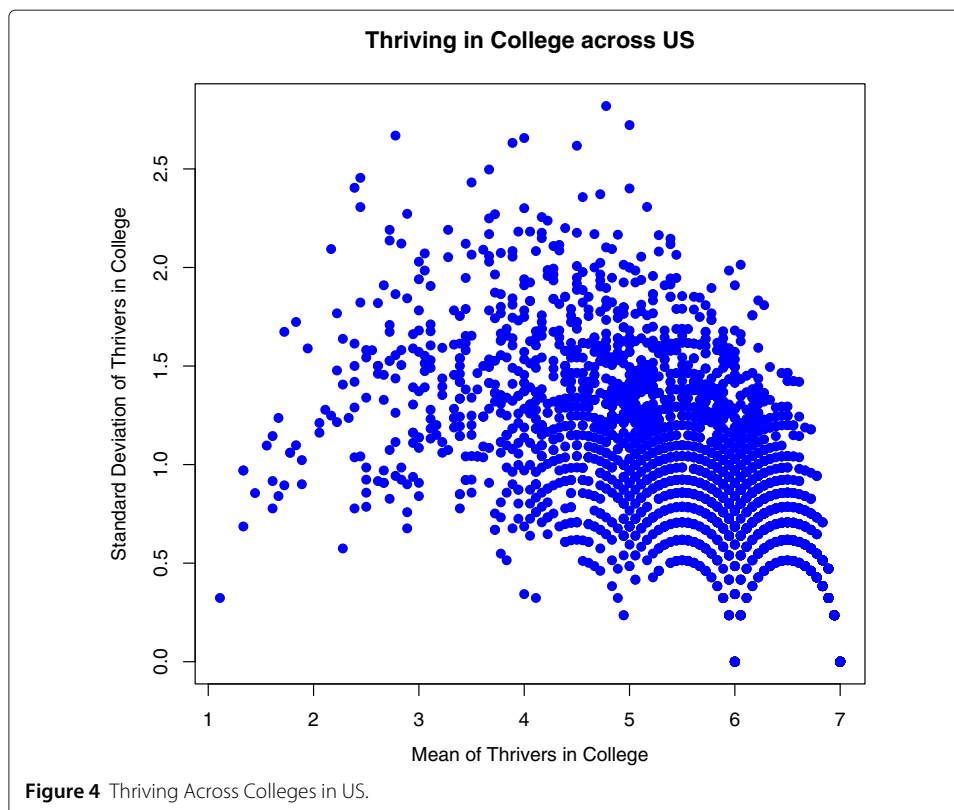
Description of algorithm

Our exploratory data analysis shows that student thriving in the US colleges is not determined by any general personal characteristic of students (such as academic scores or extroversion in high school) or by any general characteristic of college (such as technology on campus or campus size). Based on the data analysis above, we are arguing that there is no general pattern that is a good predictor for thriving in college.

This means that students and colleges should be treated individually, not aggregately, and that a recommending algorithm should be able to assign specific and unique college traits (a unique college ecosystem) to specific and unique individual traits (an unique student).

In this respect, we built a personalized algorithm that matches various combinations of personal traits with various combinations of college traits, i.e college ecosystems, and ranks them according to the best chances of thriving, thus rendering the best college ecosystem that fits any given individual. The algorithm answers the question: which combination of college traits gives the best predictability of thriving for a given combination of personal traits, on a case by case situation, for one person.

We select the data set into the “high thrivers” and “no thrivers” based on a mean above 5 and a standard deviation below 1. Figure 4 shows how the high thrivers in US are clustering and the low thrivers in US are dispersing in their thriving on all 18 dimensions.



Only two sets of information – the personal traits survey data (variables Q33 through Q42) and the college traits survey data (variables Q10 through Q28), are used to construct our algorithm. With these, a pairwise Pearson correlation matrix is constructed. This matrix is a 66 (personal trait) \times 100 (college trait) dimensions of correlation coefficients.

Taking into account only the subsetted correlation matrix for “high thrivers”, we look at any combination of personal traits and rank the correlations with the college traits for each of the personal traits variable, from the highest value to the smallest value. The strongest college trait is then selected for each student trait.

In this way, for each input combination of personal traits, the algorithm renders an output combination of college traits that has the highest ranked correlation with each input variable. This combination of college traits forms the college ecosystem where that respective individual is more likely to thrive.

Moreover, the algorithm can identify which personal traits the individual may wish to develop or emphasize, in order to enable him or her to get even more benefit out of the same collegiate institution.

Implementation of algorithm

A. Selection of the data for high-thrivers based on mean and standard deviation thresholds:

First, we calculate the mean and standard deviation of the 18 dimensions of thriving for each of 2857 students in the data:

```
mea <- mean(as.numeric(thrivingdata[1,]))
sta <- sd(as.numeric(thrivingdata[1,]))
for(i in 1:2857){
  mea[i] <- mean(as.numeric(thrivingdata[i,]))
  sta[i] <- sd(as.numeric(thrivingdata[i,]))
}
```

The plot of the mean and standard deviations of the students shows us an interesting clustering effect of the high-thrivers and the sparsity of the low-thrivers (see Figure 4). Based on this, we subset the high-thrivers – we select the high-thrivers as the students with the mean above 5 and the standard deviation below 1:

```
topthriving <- mydata[ which(rowMeans(mydata[1:18]) >
5 & rowSds(as.matrix(mydata[1:18])) < 1), ]
```

B. Computation of pairwise correlation matrix:

Second, we calculate the pairwise correlations between the college factors and the personal factors for high-thrivers:

```
Ptcollege <- cor(collegedata, personaldata)
N <- as.matrix(cor(collegedata, personaldata))
for(i in 1:length(colnames(personaldata))){
  N[,i] <- rownames(Ptcollege[ order(as.numeric(-Ptcollege[,i])), ])
}
```

Third, we either transpose the matrix N above or correlate the personal factors with the personal factors for high-thrivers and store it in a separate data frame:

```

Ctoperson < -cor(personaldata, collegedata)
P < -as.matrix(cor(personaldata, collegedata))
for(i in 1:length(colnames(collegedata))){
  P[, i] < -rownames(Ctoperson[ order(as.numeric(-Ctoperson[, i])), ]) }

```

C. The outputs of the algorithm:

The user chooses the personal traits from the list of questions from the data, that s/he feels are the best personal description. The selection for the input is made by the user from questions Q33 - Q42 in the data.

Example: *input* < -c("Q34A", "Q34C", "Q34E", "Q37F")

The algorithm searches for the highest correlation with each of the input variable among the Q10 - Q28 questions in the data, as described in C.1 below:

C.1. A unique college ecosystem for each student:

We order decreasingly and select the best college trait for each personal trait:

```
bestcollegetraits < -order(as.numeric(-N[, ]))
```

We select the top college trait for each of the personal trait and clump them together into the college ecosystem; we print the actual names of the variables (i.e. "campus that is technologically advanced, campus that is close to outdoors, etc.):

```
collegeecosystem < -as.data.frame(bestcollegetraits)[ input ] [ 1, ]
```

```
ecosystem < -as.vector(t(collegeecosystem))
```

```
mycollege < -t(variablesnames[ ecosystem ])[ , 2]
```

```
mycollege
```

mycollege output is a unique set of college characteristics - the college ecosystem - corresponding to the unique set of input. This college ecosystem is the best ranked college ecosystem out of any other possibilities that will help the student thrive.

C.2. The "ideal student" for an input of college traits:

The input here is the college ecosystem *mycollege* above. Separately from the individual choice above, we can use as input any set of college traits (Q10 - Q28) when we need to find out the "ideal student" for a different college eco-system.

We order decreasingly and select the best personal traits for each college trait:

```
bestpersonaltraits < -order(as.numeric(-P[, ]))
```

We select the top personal trait for each of the college traits and clump them together into the make up of the "ideal student"; we print the actual names of the variables (i.e. "student that is extrovert, student that participates in varsity sports", etc.):

```
optimaltraits < -as.data.frame(bestpersonaltraits)[ input ] [ 1, ]
```

```
personality < -as.vector(t(optimaltraits))
```

```
mystudent < -t(variablesnames[ personality ])[ , 2]
```

```
mystudent
```

mystudent output is a unique set of personal traits - the "ideal student" - that is most likely to be thriving in this college eco-system.

C.3. The characteristics that the student should enhance and those he should acquire in order to increase his or her chances of thriving:

The algorithm can also print the additional traits the student should get, that s/he does not currently possess, as the difference in traits between the “ideal student” and the current student:

```

optimiz <- -as.vector(t(as.data.frame(bestpersonaltraits)[ecosystem][1,]))
newtraits <- -setdif(input, optimiz)
ntraits <- -t(variablesnames[newtraits])[,2]
ntraits

```

Similarly, it can output which current traits the student should enhance, namely those common traits between the “ideal student” and the current user:

```

besttraits <- -intersect(input, optimiz)
btraits <- -t(variablesnames[besttraits])[,2]
btraits

```

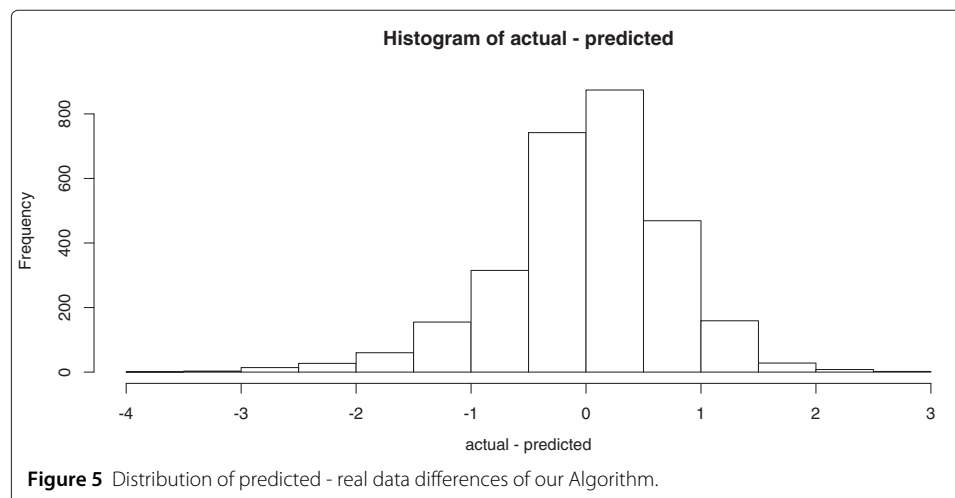
The algorithm matches the combinations of individual traits and college ecosystems bi-directionally: it can also be a recommender for colleges about the student traits that are more likely to thrive under their ecosystem. And by intersecting the input from C1 *input* with the output from C2 *mystudent*, we find those characteristics that are more likely to help a student thrive in the college environment of their choice, distinguishing between those he has and should enhance versus those he should get.

Results and discussion

Predictive power and validation

Currently our algorithm is based on the variables and correlations from the survey. In order to calculate its predictive power, we used A|B testing and randomly split the data in 2 data sets, for training and testing.

For the same input of personal traits, randomly sampled from sets of min 3 to max 66 personal traits, the college ecosystem output shows a predictive power of 53% for exact matching of output college ecosystem traits, a predictive power of 56% for 90% matching in outputs of college ecosystem traits (this means that there are 10% traits that do not match exactly between the training and testing data) and a predictive power of 88% for 80% matching in outputs of college ecosystem traits. Figure 5 shows the differences



in the values between the actual data and the predicted data for exact matching. The predictability errors follow a Bell curve distribution.

The algorithm shows that predicted values tend to be slightly optimistic, but not significantly.

The algorithm is currently implemented into a commercial digital product. After the launch of the product, we will be able to collect data from the real users and assess the commercial validity and customer satisfaction of the algorithm.

The richer the selection of variables in the input is, the more refined and unique the combinations of outputs are and the higher the predictability of the outputs.

Algorithm applications, examples and extentions

Our algorithm can combine 66 personal traits into clumps of 1, 2, ... 66; this means that we can create 6.45^{146} "persons" with different psychological traits in the lab.

Additionally, it can show which of the current traits of the student are more likely to lead to his/her thriving in the recommended college ecosystem. It can also show which of the traits the student does not have, but are also desirable to his or her specific college ecosystem.

We asked several prospective college students to pick a set of personal traits and we answered them with the thriving college ecosystem, the strengths they have and the traits they should develop. For example, John picked as his personal traits the following: need for solitude, caring and supportive, self-centered, artsy and creative, calm and emotionally stable and hard working. The best college ecosystem for him is one who is academically rigorous, encourages students to meet new people, has a student body that is self-centered, easy going and creative and where the campus is not well connected with places of interest.

On a case by case study, we have tested the algorithm on approximately a dozen students. For example, some of the unexpected thrivers thrive on campuses where there are outdoor activities, there are off campus distractions but also there is a lack of transportation to go off campus. Perhaps these are the students who are confined to campus, but have the outdoor activities as an outlet and also have all the resources they need on campus to keep them focused. And the expected thrivers would thrive on campuses that have inclusion, by perhaps being exposed to other students than just those like themselves. Another way to understand this is that unexpected thrivers would thrive better in campuses where they have different activities in one place (the actual logistical space is more important), while for the expected thrivers it is about being exposed to various people and students (the social possibilities are more important).

Conclusion

The main conclusion we are drawing from this research is that there is no staple measure that we can apply to all and any potential student (i.e. not all extrovert students would be thriving under any conditions as well as not all introvert students would be thriving under any conditions). This conclusion forms the basis for our personalized approach to an algorithm that can take any combination of personal traits to give a combination of college traits (ecosystem) where this unique person is most likely to thrive.

The data analysis, when looking at the general effects, shows that there is no variable that distinctively influences thriving, whether these variables are demographic or personal traits. If there is no general trait or demographic that someone should possess in order to thrive, this means that every person can be treated differently and assess where they are more likely to thrive based on their unique makeup. Our algorithm does that by ranking the correlations of each personal trait with each college trait and selecting only the ones ranked on top. In other words, we are providing a tool that helps organize low effects into ecosystems with the best likelihoods of thriving, by using a highly personalized approach.

Our research is currently limited to the data we have collected and to the fact that the correlation factors from this data set were very small. Our research is also limited to the fact that, being in the first stages of implementation, we do not have much feedback from the users. We are planning to track the users for the next 4 years of their college experience in order to assess the validity of our model on an individual basis. We have also collected a second data set through the same panel survey method as the data analyzed here and we are integrating the two data sets in order to increase our predictable value.

The results of this research are being implemented into an index that the parents, prospective students, education policy practitioners, and virtually anybody interested in the topic of thriving in college will be able access open source. The algorithm itself has been implemented into a commercial digital product that the prospective students can access in order to find out more information about themselves and the college ecosystem that would support their thriving.

The education policy and data science researchers can find a number of implications and may advance our research. Education is a very complex phenomenon and while there has been research in learning analytics, the use of computational tools, multiple methodologies and data sources for education policy research is still very incipient. Analytics has the power to transform not only the learning environment, but also the greater educational ecosystem. While current research has not yet moved from basic descriptive applications to real predictive and prescriptive models, we are among the first to provide a model that is ready to implement and be used. We hope that our research will move toward more advanced analytical models and share the lessons learned, findings and future directions in a way that invites debate, input and collaboration.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AB is responsible for the computational analysis and the design of the algorithm, as well as for the large part of the writing of the paper. MT is responsible for the A|B testing and the predictive analytics ideas. MG carried out the panel and survey research. NB is responsible for Background and Literature Review. EC is responsible for the definition of thriving and part of the panel survey question creation.

Acknowledgments

The authors would like to thank Howard Everson, Wayne Camara, Peter Negroni, Ruth Helman, Evan Radko, Merrilea Mayo and Linda Naiditch for very useful comments and insights into this research.

Author details

¹Center for Complexity in Business, University of Maryland, College Park, MD 20742, USA. ²Department of Computational Social Science, George Mason University, Fairfax, VA 22030, USA. ³Educational Leadership and Policy Program, University At Buffalo, Buffalo, New York 14260, USA. ⁴Greenwald & Associates, Washington, DC 20008, USA. ⁵vibeffect, Washington, DC 20036, USA.

Received: 10 December 2014 Accepted: 29 April 2015

Published online: 16 May 2015

References

- Astin, AW (1993). *What Matters in College? Four Critical Years Revisited*. San Francisco: Jossey-Bass.
- Bean, J (2005). Nine themes of college student retention. In A Seidman (Ed.), *College student retention: Formula for student success* (pp. 215–243). Westport, CT: Praeger Publishers.
- Bowen, WG, & Bok, DC (1998). *The Shape of the River: Long-term Consequences of Considering Race in College and University Admissions*. Princeton, NJ: Princeton University Press.
- Braxton, JM (2000). *Reworking the Student Departure Puzzle*, vol. 1st edition. Nashville, TN: Vanderbilt University Press.
- Carnevale, AP, & Fry, RA (2001). Economics, demography and the future of higher education policy. *Higher Expectations I*, 5–6.
- Carnevale, AP, Jayasundera, T, Cheah, B (2012). *The College Advantage: Weathering the Economic Storm*. Washington, DC: Georgetown Public Policy Institute.
- Dey, EL, & Astin, A (1993). Statistical alternatives for studying college student retention: A comparative analysis of logit, probit, and linear regression. *Research in Higher Education*, 34(5), 569–581.
- Kuh, GD (2004). The national survey of student engagement: conceptual framework and overview of psychometric properties. Technical report, Indiana University. http://nsse.indiana.edu/2004_annual_report/pdf/2004_Conceptual_Framework.pdf.
- NCES (2015). College Navigator. <http://nces.ed.gov/collegenavigator/>.
- Oseguera, L, & Rhee, BS (2009). The influence of institutional retention climates on student persistence to degree completion: A multilevel approach. *Research in Higher Education*, 50, 546–569.
- Perna, LW, & Thomas, SL (2008). Theoretical Perspectives on Student Success: Understanding the Contributions of the Disciplines. <http://eric.ed.gov/?id=EJ804747>.
- Prades-Collins, K (2012). Thriving in students of color on predominantly White campuses: A divergent path? In LA Schreiner, MC Louis, DD Nelson (Eds.), *Thriving in transitions: A research-based approach to college student success* (pp. 66–86). Columbia, SC: University of South Carolina, National Resource Center for the first-year experience and students in transition.
- Roderick, M, Nagaoka, J, Coca, V, Moeller, E (2008). *From High School to the Future: Potholes on the Road to College*. Research Report. 1313 East 60th Street, Chicago, IL 60637: Consortium on Chicago School Research.
- Seligman, MEP (2011). *Flourish: a Visionary New Understanding of Happiness and Well-being*, 1st hardcover edn. New York: Free Press.
- Schreiner, LA (2013). Positive psychology and appreciative inquiry in higher education. *New Directions for Student Services*, 143, 41–52.
- Schreiner, LA (2014). Different pathways to thriving among students of color: An untapped opportunity for success. *About Campus*, 19.5, 10–19.
- Seidman, A (2005). *College Student Retention: Formula for Student Success*. Westport, CT: Praeger Publishers.
- Singell, LD (2010). Modeling retention at a large public university: Can at-risk students be identified early enough to treat? *Research in Higher Education*, 51, 546–572.
- Tinto, V (1993). *Leaving College: Rethinking the Causes and Cures of Student Attrition*, vol. 2nd edition. Chicago, London: University of Chicago Press.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
