Synthesizing Measures of Institutional Success

Eric Hsu, Vilma Mesa, and the Calculus Case Collective[1]

The purpose of this document is to describe in as much detail as possible the steps that the Collective took to select the final 18 institutions that were part of the case studies for the Characteristics of Successful Programs in College Calculus [CSPCC] project, funded by NSF (Award #0910240.) administered by the Mathematical Association of America [MAA]. We provide some background on the study, followed by the quantitative analysis performed on the survey data by Drs. Philip Sadler and Gerhard Sonnert from the Harvard-Smithsonian Center for Astrophysics, and then discuss additional data collected that led to the final selection of case study institutions. The final sample included four associate's, five bachelor's, four master's, and five doctoral-granting institutions.

The CSPCC Study

This study was funded by the National Science Foundation and led by the MAA in collaboration with a team of researchers. The goals of the project are:

- To improve our understanding of the demographics of students who enroll in calculus.
- To measure the impact of the various characteristics of calculus classes that are believed to influence student success.

The first phase of the project involved online surveys of Calculus I students and instructors from across the country in the fall of 2010. This was followed by a second phase involving case studies of selected institutions. This technical report focuses on how the analyses conducted with the Phase 1 survey data and with other derived measures were used to select institutions for participation in the case study research.

Five major online surveys were constructed: one for the calculus coordinator, two for the calculus instructors of which one was administered immediately before the start of the course and the other immediately after it ended, and two for the students in the course (one at the start of the term and the other at the end of the term). In addition, instructors reported on the distribution of final grades and were asked to submit a copy of the final exam. One year after the surveys were administered, a short follow-up survey was sent to those students who had volunteered their email addresses. No incentives were given for completing the surveys. Links to the surveys can be found at www.maa.org/cspcc (Bressoud, Carlson, Mesa, & Rasmussen, 2013, p. 2).

The surveys addressed issues such as plans for continued study of calculus, attitudes about mathematical ability, the role of the instructor, instructional practices, and student effort.

For the purposes of surveying post-secondary mathematics programs in the United States, the Conference Board of the Mathematical Sciences (CBMS) characterizes colleges and universities into four types determined by the highest mathematics degree that is offered: associate's degree, bachelor's degree, master's degree and doctoral degree. Because enrolments vary so greatly

---

[1]Suggested APA Citation: Hsu, E., Mesa, V., & The Calculus Case Collective. (2014). *Synthesizing measures of institutional success CSPCC-Technical Report #1*. Washington DC: Mathematical Association of America.

within each type of institution, CBMS further stratifies these institutions according to the number of full-time equivalent (FTE) undergraduate students. We sampled most heavily at the institutions with the largest enrolments. No for-profit colleges or universities were included in the study. In all, we selected 521 colleges and universities. Table 1 has the total number of institutions in the United States, the number of institutions that were sampled, and the number of institutions for which we have data for analysis.

Table 1: U.S. Institutions, sampling, response rates for institutions in the study.

Institution type	Number of US institutions[a]	Sample size (sample rate)	Participant Institutions[b] (response rate)	Number of substrata by FTE	Range of response rates[c]
Associate's granting	1,121	207 (18%)	40 (19%)	8	17%–42%
Bachelor's granting	1,015	134 (13%)	41 (31%)	5	29%–52%
Master's granting	181	60 (33%)	21 (35%)	4	33%–54%
PhD granting	197	120 (61%)	66 (55%)	6	46%–88%
Total	2,514	521 (21%)	168 (32%)		

Notes: a. As counted by the CBMS. b. Number of colleges or universities that provided data and percentage of the sample that provided data. c. Range of percentages, by substrata, of the sampled institutions that provided data. See Bressoud et al., (2013, p. 3).

Initial Analysis

Philip Sadler and Gerhart Sonnert conducted the analysis of the quantitative data generated from the surveys with the goal of suggesting variables that could be used as indicators for selecting institutions as potential cases for being studied in more depth.

A main issue we faced was low response rates by institutions. We have four main analytical data sets that we have been using in our analyses. The MAA_ALL_data includes all the records for which there are responses from students, instructors, or coordinators. The CSPCC_Data file includes all the records for which there are responses from students, instructors, or coordinators, and includes the pilot institutions that were later added to the study. The MAA_Long_data excludes those records for which there is coordinator information but no student data. It also excludes records from students that could not be matched to an instructor (perhaps the coordinator sent the link to the survey, the faculty asked the students to complete the survey, but the instructor did not get the survey completed within the time frame allocated). The third file, MAA_short_data, includes only subjects with pre- and-post surveys with instructors who also completed both pre- and post-surveys, and with the course coordinator survey. The number of students, instructors, and coordinators is given in Table 2.

Table 2: Four different files available for various analyses.

File Name	# of students	# of instructors	# of coordinators	# of institutions
ALL_data	14,247	1,168	260	504
CSPCC_Data	14,279	503	144	168
MAA_long_data	13,965	496	169	168
MAA_Short	3,103	308	123	123

With the current design of the surveys, it is not possible to determine whether the observed attrition is normal (what should be expected in these large studies) or whether better students are responding to the post-test. In either case, the sample itself is important. If the sample is biased towards higher achieving students, this will constitute important baseline data that can then be used to contrast with future studies with lower attrition rates. For more details on the differences in response rates, see (Sonnert, Sadler, & Bressoud, 2014).

As a next step, Sadler and Sonnert sought to identify statistically significant changes in student responses from pre- to post-surveys the following outcome variables that were included in the proposal, regarding attitudes and intention to take calculus II. They also included one additional variable: increased interest in mathematics, which was only included on the post-survey. The outcome variables for the study are given in Table 3.

Table 3: Outcome Variables for the Study.

Variable	**How measured**	**Start of term**	**End of term**
Confidence	6-option Likert scale, strongly disagree to strongly agree (e.g., "I am confident in my mathematics abilities)	X	X
Choice to take more math	4-option Likert scale (e.g., "If I had a choice: I would never take another mathematics course / I would continue to take mathematics")	X	X
Enjoyment of Math	6-option Likert scale, strongly disagree to strongly agree (e.g., "I enjoy doing mathematics)	X	X
Final grade of C or higher	Instructor reported grades		X
Increased interest	6-option Likert scale, strongly disagree to strongly agree "This course has increased my interest in taking more mathematics")		X
Intention to take Calculus II	Options: no, not sure, yes ("Do you intend to take Calculus II?")	X	X

Table 4 shows the means, standard deviations, standard errors, for the pre- and post-survey variables, with the change and their effect size. These were calculated with the MAA_Short_data file.

Table 4: Mean, Standard Deviation, and Standard Error of Pre and Post-survey measures of the Outcome variables, with Difference and Effect Size.

Outcome Variables	**PRE**		**POST**		**Difference**	**Effect Size**
	Mean (SD)	**SE**	**Mean (SD)**	**SE**		
Confidence in Math	3.89 (1.01)	0.02	3.42 (1.18)	0.02	-0.47	-0.46
Choice to take more Math	1.93 (1.02)	0.02	1.84 (1.08)	0.02	-0.09	-0.09
Enjoyment of Math	3.63 (1.27)	0.02	3.28 (1.37)	0.02	-0.35	-0.27
Choice to take more math			2.66 (1.43)	0.03		0.11
Intention to take Calculus II	0.81 (0.33)	0.01	0.74 (0.44)	0.01	-0.07	-0.20

Sources: (Sadler & Sonnert, 2011; Sonnert et al., 2014).

The negative differences in Table 4 indicate that taking the calculus I course in general had a negative impact on these outcome variables. In addition, the effect size in four of the variables is negative: the distribution of the variables is “moving” to the left at least one fifth of a standard deviation (except for the “If I had a choice” variable”, which is about one tenth). Confidence in ability to do mathematics has the most significant drop, with almost a half standard deviation. The effect sizes are represented in Figure 1.

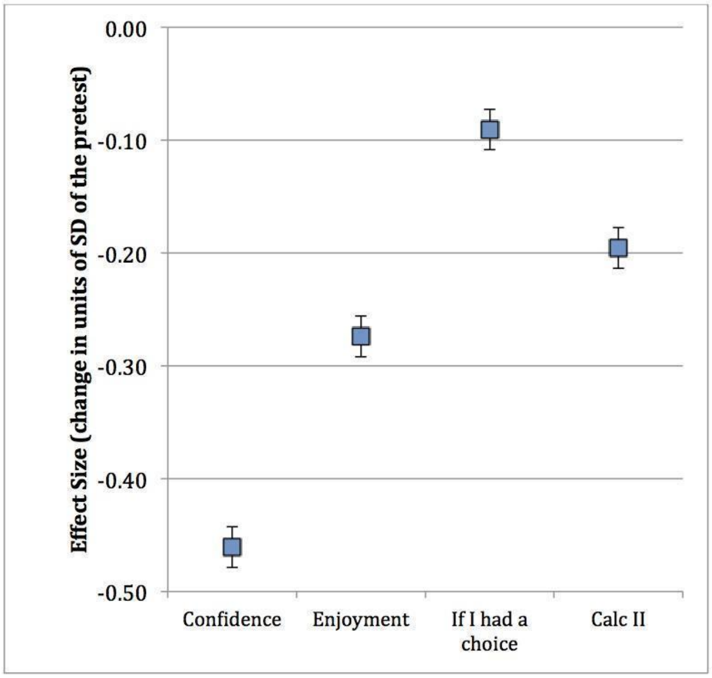

Figure 1: Effect sizes for outcome variables of interest.

As a next step, Sadler and Sonnert used this information to estimate the extent to which student level factors, instructor level factors, and institution level factors in the dataset could explain these outcomes. In order to do this, the analysts performed a hierarchical linear model analysis (to account for the nested nature of the data: students within in classrooms and classrooms within institutions) after imputing missing data. Multiple imputation was preferred to case deletion because there were 28% of subjects had had one or more missing variables (Sonnert et al., 2014).

The data have a three-level hierarchical structure: students nested within instructors, and these nested within institutions. As a first step, the analysis estimated the variance in outcome variables that could be explained by different levels:

> This is done by running "unrestricted" models without any predictor variables. These models thus simply indicate how the overall variance is divided across the levels—indicating the proportion of variance available to predictor variables at the respective levels. If there is a lot of variance at a certain level, there is great potential for predictor variables of that level to reduce or "explain" variance. Conversely, if there is very little variance at a certain level, this already indicates that no predictor variable at that level will make much of a difference (in terms of a main effect). (Sadler & Sonnert, 2011, p. 1)

Their analysis of the unrestricted models for Desire to continue studying mathematics and Increased interest suggest that most of the variance in outcome variables are due to student factors, followed by institutional factors; instructor factors played a negligible role in explaining the variance in the outcome variables (see Table 5).

Table 5: Variance explained by various levels, unrestricted model, for two outcome variables.

	Student Level	Instructor Level	Institution Level	Key Predictor
Desire to continue studying mathematics	92%	2%	6%	Intention at the beginning of term
Intention to take Calculus II	81%	3%	16%	Intention at the beginning of term

When using Intention to take calculus II at the beginning of the semester to predict desire to continue studying mathematics at the end of the semester, the analysts found:

> This is a very powerful predictor, explaining 42% of the overall variance. The instructor level variance gets reduced the most (by 95%). Student level variance is reduced by 38.5%, and institution level variance by 80%. This results in a model in which variance is almost exclusively concentrated at the student level (98%). The variances at the instructor level (0.2%) and at the institutional level (2%) are negligible by comparison (p. 3)

Using Intention to take calculus II at the beginning of the semester to predict intention to take Calculus II at the end of the semester, the analysts found:

> initial intention explains almost a third (32%) of the variance in final intention. Looking at individual levels, the student level variance has been reduced by 23%, the instructor level by 82%, and the institution level by 67%. As a result, the remaining variance is even more concentrated at the student level (91%) and the instructor level variance has become virtually eliminated (1%). (p. 3)

The analysts conclude:

> This analysis of baseline models does not warrant optimism in regard to identifying instructor or institutional level variables that make any difference as main effects. (p. 5)

Thus, the recommendation was to seek out institutions that had a positive shift on student outcome questions, controlling for factors such as:

- Demographics (Gender, SES, Race/Ethnicity);
- HS Math (Calculus type, Grad in Highest Course, SAT/ACT scores), and
- College Variables (Prior college math, year, career intention and Pre-survey value)

As an example, using the "Change in student confidence" variable they identified the following seven institutions which had the largest effect sizes (in absolute value) in this variable:

Table 6: Possible set of successful institutions, by "Change in student confidence."

Institution	N	Effect Size	Standard Error	Probability
Most Negative Effect Size				
Institution 1	13	-0.41	0.18	0.024
PhD3	17	-0.39	0.17	0.023
Institution 2	27	-0.33	0.15	0.027
Institution 3	43	-0.33	0.13	0.011
Institution 4	37	-0.29	0.14	0.034

Institution	N	Effect Size	Standard Error	Probability
Most Positive Effect Size				
Institution 5	22	0.39	0.16	0.015
PhD0	13	0.40	0.18	0.029

A notable point in Table 6 is the small sample size. These are estimates run with all the students who had a full set of data and for which it was possible to include the most controls. Small sample size was a very problematic aspect of the data that affected all decisions in the project. This analysis resulted in a list of all the institutions with all the estimated effect sizes for each of the outcome variables. As we explain in the next section, this analysis was augmented with a number of analyses that included variables at the student, departmental, and institutional level, that were necessary to make a final decision. In that process we tried to honor the quantitative work (by looking at these effect sizes) and also the local conditions of the available data.

Identification of Relevant Features

Our charge, after receiving the effect sizes for the outcome variables, was to choose the most promising institutions to study given a lack of access to student learning outcomes and in many cases, the small sample size that was used to estimate those effects. Initially, we only considered institutions that participated in the survey above a minimum level of 30 responses in order to get a feel for the data. Because of differential response rates for bachelor's and associate's degree granting institutions, we lowered the threshold to 20 and 10 responses respectively.

In selecting institutions for further case studies, we sought to balance three notions of "success:" affect outcomes, persistence outcomes, and achievement outcomes. Each of these are discussed in turn.

1. Affect Outcomes Student Outcomes in Survey Data. This is a cluster of variables that correspond to the statistical analysis of the survey data conducted by Sadler and Sonnert: Confidence, Desire to continue studying mathematics, Enjoyment, Increased interest, and Intention to take Calculus II (see Table 5).

2. Persistence Outcomes. In addition to the information on Intention to take Calculus II provided by the Sadler and Sonnert analysis, we wanted to classify student intention more carefully by studying four different groups of students, depending on the values of two variables: Intention to take calculus II at the beginning of the study (Yes/No) and Intention to take calculus II at the end of the study (Yes/No, see Table 7). By doing this, we were able to enlarge the pool of data to run the analyses and we were able to use multiple questions in the survey to correctly classify students with less clear intentions. These analyses were conducted with the CSPCC_Data set.

We were particularly interested in students in the "Converter" and "Switcher" groups, as we felt that could capture the difference an institution was making on students' choices, because Intention to take Calc2 might be influenced by the career goals of the student population of each institution (some majors do not require a calculus II course). Because in all, the number of students in the Converter group was small (1%), we concentrated on the Switcher group (9%) (Rasmussen & Ellis, 2013).

Table 7: Four groups of students depending on their change of Intention of taking Calculus II

		Intention of taking Calculus II, Beginning of Term		Total
		YES	NO	
Intention of taking Calculus II, End of Term	YES	Persister 4,699	Converter 90	4,789
	NO	Switcher 648	Culminater 1,783	2,341
Total		5,347	1,873	7,220

Source: (Rasmussen & Ellis, 2013)

3. Achievement Outcomes. We considered a number of other variables that could allow us to include measures of student learning to include institutions. One such measure was student self-reported final grades, derived from responses to post-surveys administered during the final week of the course. Because students tend to overestimate their grades, the student reported grades are probably higher compared to the actual grades they obtain. However, because this bias was present across all students, we believed that this measure of anticipated, self-reported grade could help us in comparing institutions: the bias would be equally distributed across all respondents.

Other measures considered were: pass rates and the average final grades reported by institutions and instructors. Initially we discarded these possible measures because it was not clear that they could be compared reliably across the different institutions. It was difficult for example, to say how a pass rate of 60% at institution A compared to a pass rate of 80% at institution B, as we lacked historical perspective for these measures within the institutions. Likewise an average grade of B might be better in the absolute than an average of C, yet, when comparing across institutions it is not clear that such comparisons are valid. One way to handle this was to determine the extent to which specific institutional features could help predict pass rates.

To do this, we obtained several variables from the Integrated Post-secondary Education Data System (IPEDS) to gather information about the student body. Specifically we obtained:

- 75th and 25th percentile SAT scores (not available for Two-Year institutions)
- Admission rate
- Four-year graduation rate (most useful for two-year institutions)
- Six-year graduation rate (most useful for four-year institutions)
- Tuition

We also computed a smoothed version of the 75th percentile SAT score by using the 3-year average. Using Excel and R we used linear regressions to identify relationships between these variables and the pass-rate described in the student end-of-term survey.We found that the relationship was weak between most variables and calculus passing rate except for three:

- 75th percentile SAT
- 4-year
- 6-year graduation rate

We created a rough linear model using the SAT and six-year grade variables, modeling the expected calculus pass rate as 0.0003589*SAT75 + 0.0042560*SIXYEAR + 0.1959819. We then looked for institutions that over-performed and under-performed the model. We did not have relevant data for all institutions, but we used this regression model whenever possible for case study selection.

Case Study Selection

In Table 8 we list the selected sites along with their performance on the three measures relative to other sites of the same institution type (the performance is described qualitatively on a scale from ++ high, + above average, 0 average, - below average, -- low). Specific details are provided in the following section.

It would have been a clean selection process if a number of institutions "rose to the top" on each of the three measures identified above. Interestingly, institutions that excelled in one measure did not necessarily excel in the other measures. It was impossible to cleanly draw a line between the top two cases and the rest of the choices.

Because of this, we decided to increase the number of case study sites in order to capture the diversity of success profiles, and to use further judgment to encourage a diverse mix of institution types, sizes and urban/rural settings. The original plan for two cases studies at each of the four institution types was increased to four. Each research team developed an internal process to nominate institutions for site visits that took into account response rate, affect, persistence, and achievement outcomes, and the information from the regression, among other criteria (e.g., balancing geographical distribution). Each team also chose a nearby institution where all the instruments and procedures for data collection were tested. Two of these institutions were later included as selected institutions.

Table 8: Institutions selected for case study and "success" measures.

	Affect Outcomes	Persistence Outcomes	Achievement Outcomes	Additional Criteria
BA 0				Pilot[a]
BA 1	++	+	++	
BA 2	+	++	n/a	
BA 3	+	0	++	
BA 4	-	+	++	
MA 0				Pilot
MA 1	+++	++	n/a	
MA 2	++	+	- -	
MA 3	++	- -	+	
MA 4	+	+	+	
PhD 0	++	0	+[c]	Pilot[a]
PhD 1	++	-	0	In best practices literature
PhD 2	+++	++	+	In best practices literature
PhD 3	- - -	- -	0	In best practices literature
PhD 4	++	++	--	
TY 0				
TY 1	+	- -		Responsiveness[b]
TY 2	+	+		Responsiveness
TY 3	0	+		Responsiveness
TY 4	+	-		Responsiveness

Notes: ++ high, + above average, 0 average, - below average, -- low relative to the possible institutions that fit the outcome criteria. a. The pilot institution was included as part of the case study. b. The response rate for two-year institutions was generally low, which has serious implications for the reliability of the Student Outcomes in Survey Data variables. The regression model for passing rate was only constructed for four-year colleges as the data was not systematically available for two-year institutions. Therefore, the research team leader for the two-year college team replaced that measure with the practical measure of responsiveness to a second round of data requests for passing rates disaggregated by ethnicity. c. The PhD team attended to literal passing rates and graduation rates and not over/underperformance of the simple linear model.

Specific details from each team regarding their selection of each of these institutions follows.

Individual Case Selection

Baccalaureate Institutions, Sean Larsen and Estrella Johnson

We used multiple criteria to select schools for our case studies:

- The difference between the school's predicted Calc I pass rate (based on Eric an Vilma's regression analysis using 6 year graduation rates and SAT scores) and the school's reported Calc I pass rate from the survey data
- The Harvard team's analysis of data collected during the survey phase – we considered the overall performance on the dependent variables (in one case we looked more closely at the responses that lead to a marginal score and found that the "low" score was the result of a lack in extreme responses)
- We looked at the switcher analysis from Chris and Jess, identifying schools that had 1) a high percentage none switchers (students who originally intended to take Calc II and then did) and 2) a low percentage of switchers (students who originally intended to take Calc II and the did not)
- Other interesting results from survey data, especially diversity of reported percentages of lecture experienced by students

BA1 – This school out performed it's predicted pass rate for Calc I and performed well in regards to the Harvard group's survey analysis data. This school had a high percentage of non-switchers and a low percentage of switchers. There was also a high diversity in the amount of lecture time reported by students.

- + 12% from predicted passing rate (with Calc II)
- +1.19592 for +/- of Phil's data
- NS1 80.6% S1 8.3%
- Diverse amount of lecture time

BA2 - This school performed well in regards to the Harvard group's survey analysis data. This school had a very high percentage of non-switchers and a very low percentage of switchers.

- + .44872 for +/- of Phil's data (with Calc II)
- NS1 90% S1 6%
- (They were not included in the REESE Calc Institute Data, so we don't know how they performed based on what was predicted)

BA3 - This school out performed it's predicted pass rate for Calc I and performed well in regards to the Harvard group's survey analysis data. This school had an above average percentage of non-switchers and an average percentage of switchers. This school is also a large urban commuter school and, as such, likely faces challenges faced by other schools we considered.

- + 14% from predicted passing rate
- +.43694 for +/- of Phil's data (with Calc II)
- NS1 78% S1 15%

BA4 - This school out performed its predicted pass rate for Calc I. This school performed about average in 3 out of the 4 variables in regards to the Harvard group's survey analysis data, with a low reported *interest*- this is mostly because of there were a large number of weak agrees bringing down the average. This school had a high percentage of non-switchers and a low percentage of switchers. There was also a high diversity in the amount of lecture time reported by students.

- + 20% from predicted passing rate
- -.4217 for +/- of Phil's data (with Calc II)
- This is mostly because of the interest number but there are not very many disagrees, just weak agrees bringing down the numbers
- NS1 85% S1 11%
- Diverse amount of lecture time

Master's Institutions, Eric Hsu and Addie Evans

We first selected institutions that had at least two of the following positive measures:

- Positive sum of Sadler indexes
- Below median numbers of S1 (stem intending switchers), minimum 10 respondents
- Over-performance of expected calculus passing rate as modeled by school SAT 75th Percentile and 6 Year Graduation Rate. This linear regression was calculated from available data for our set of 49 schools with high response rates.

Five institution's met these requirements. One institution was deleted because it had mostly negative Sadler indexes, and in fact its Sadler index sum was more than twice as negative as the next most negative.

PhD Institutions, Chris Rasmussen and Jess Ellis

In general, we wanted to choose institutions that showed success using the Harvard data and low numbers of STEM intending switchers. Also, we want to have a variety in size and STEM trajectory students in order to provide case study data that is applicable to a wider range of institutions.

PhD0: This institution rose to the top of many of the statistical analyses from Phil and Gerhard. Specifically, they have:

- high increased confidence,
- choice of taking more math,
- intention to take Calc II, and high expected grades.
- They have average S1 and NS1 numbers,
- low STEM students (38%),
- and high pass rate in calculus.
- We have a contact person, and it is clearly a cost effective place for a pilot.

- We have high response rates and they responded to our request for more data.

PhD1: This institution rose to the top of many of the statistical analyses from Phil and Gerhard. Specifically, they have:

- high increased in confidence,
- increased enjoyment,
- choice of taking more math,
- and high interest in mathematics.
- Although they have slightly higher switchers and less stem intending non switchers,
- we have very high response numbers and are interested in what they are doing.
- Also, they responded to our request for more data.
- Additionally, PhD1 has been previously recognized as having best practices in the literature (personal communication, Peter Ewell).

PhD2: This institution had the highest numbers from the statistical analysis form Phil and Gerhardt.

- They gave very high stem intending non switchers (91.5%) and very low stem intending switchers (1.1%).
- They have fairly high response numbers,
- and high pass rates and graduation rates.
- Additionally, they have been previously recognized for having best practices through Teagle grants for innovation (http://www.teagle.org/grantmaking/grantees/systematiclac.aspx; personal communication, Peter Ewell).

PhD3: Although this institution was at the bottom of the statistical analysis from Phil and Gerhardt and

- had high STEM intending switchers and non STEM intending non switchers, we know that at this institution, many non STEM intending students are required to take Calculus I. Thus it makes sense that these numbers (in confidence, interest in math, etc.) are lower that other institutions.
- We know that they implement innovative practices in their Calculus program and have GTAs teaching a large proportion of calculus classes.
- Additionally, they have been previously recognized for having best practices through George Kuh's High Impact practices (http://generaled.unlv.edu/cctl/terry_rhodes_unlv.pdf; personal communication, Peter Ewell).

PhD4: This institution rose to the top of many of the statistical analyses from Phil and Gerhard, and had positive gains in every variable reported.

Similar to PhD2, they had very high STEM intending non-switchers and low STEM intending switchers. In addition, they have high response rates, but very low graduation rates (4 years is 26%; 6 years is 61%). Additionally, they have a self-reported (by the course coordinator) effective TA training program.

Two-Year Institutions, Vilma Mesa

Using the Sadler Data Organized spreadsheet, which had 123 records, I selected all the records corresponding to the "associates" LEVEL category. This yielded 24 colleges. The file also had a column for the number of respondents. From this list, I selected all colleges that had more than 6 responses. This resulted in 12 colleges.

The Sadler file included six additional columns, from self-reported scores to various items on (1) Confidence in own ability to do mathematics, (2) enjoyment of mathematics, (3) would choose to take more math, (4) course increased interest in math, (5) intends to take Calc 2, and (6) expected grade. In general, we wanted these coefficients to be positive, as they suggest that there was a gain from the beginning of term measure to the end of term measure. From the list of 12 colleges, I identified five for which the coefficients were all positive, except for one college for which intention to take Calculus II as negative.

In parallel to this, and using the SwitcherInfo_CC.doc file, I considered students' changes in their responses to their intentions to take Calculus II, measured through a question asked at the beginning of the term and at the end, and taking into account whether the students had initially intended to take Calculus II and took it (Non-Switcher in), Did not intend to take Calculus II, and did not take it (Non-Switcher out), intended to take Calculus II but did not (Switcher out), and did not intend to take Calculus II but enrolled in it (Switcher in). We were interested in a high percentage in the Non-Switcher In, and Switcher In columns, and small percentage in the Switcher Out column. Note: shaded cells correspond to the highest percentage of non-switcher students. This yielded a second list of five colleges, in which the retention is greater than 90% for Non-Switchers. The two combined lists yielded seven possible institutions to select from.

Another consideration regarded whether the colleges submitted pass rates information as requested from the original 12 institutions in late December, 2011. This information is valuable, because it establishes a connection between the project team and the colleges. Thus we sought to work with colleges that provided the requested information.

A final consideration had to do with the size of the calculus class (estimated through the number of sections offered in the Fall 2011) and three characteristics from the IPEDS data set, their level of urbanization, the Carnegie classification for 2-yr colleges, and Enrollment figures for 2009. Highlighted rows correspond to colleges under 5,000 students, considered small. We sought to have a variety in terms of size, location, and diversity of student body.

The final selection was made through the combination of the various criteria described above (see Table 9) The index column is simply an addition of all the positive cells in the columns.

Table 9: Indicators for the Two-Year Colleges for final selection.

	Ten or more student responses Sadler's estimates	Mostly positive coefficients in Sadler's estimates	High percentage of Students staying in calculus	Responded to request for more data	College size	Index
TY1	1	1	1	1	Small	**4**
TY2	1		1	1	Large	**3**
TY3	1	1		1	Medium	**3**

	Ten or more student responses Sadler's estimates	Mostly positive coefficients in Sadler's estimates	High percentage of Students staying in calculus	Responded to request for more data	College size	Index
TY4	1	1			Large	2
TY5		1	1		Large	2
TY6	1			1	Large	2
TY7				1	Large	1
TY8	1				Large	1
TY9			1		Medium	1
TY10	1				Large	1
TY11		1	1	1	Small	**3**
TY12	1			1	Small	2

The colleges with the higher indexes were good candidates for belonging to the case studies. TY11 was a possible candidate, but TY1 was chosen instead to provide geographic variety. We included TY4 in spite of the lower score relative to the other institutions, because the state in which this college is located has a unified system at the city level, which is very different from the other colleges. TY4 is urban campus in a large town, which also provided variety to the sample. This list provided also geographical distribution.

References

Bressoud, D. M., Carlson, M., Mesa, V., & Rasmussen, C. L. (2013). The calculus student: Insights from the MAA national study. *International Journal of Mathematical Education in Science and Technology*. doi: 10.1080/0020739X.2013.798874

Rasmussen, C. L., & Ellis, J. (2013). *Who is switching out of calculus and why.* Paper presented at the Annual Meeting of the Psychology of Mathematics Education-North American Chapter, Chicago.

Sadler, P., & Sonnert, G. (2011). *Progress on Characteristics of Successful Programs in College Calculus (CoSPiCC) data analysis*. Report to David Bressoud. Harvard Smithsonian Center for Astrophysics, Cambridge, MA.

Sonnert, G., Sadler, P., & Bressoud, D. M. (2014). The impact of instructor pedagogy on college calculus students' attitude toward mathematics (under review). Boston, MA: Harvard University.