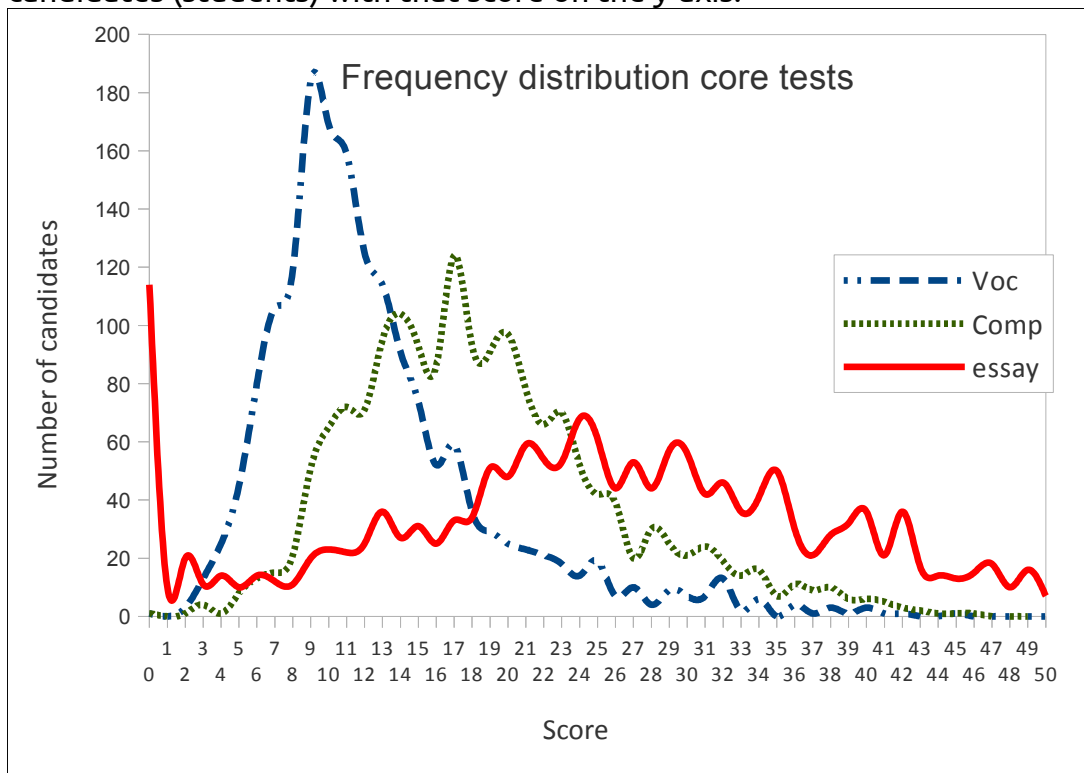


COMET 2012 Focused Exploration of Performance on English Subsections

In the spring of 2003 a principal at a high school asked questions that led to a careful statistical review of the spring 2003 COMET results. As a result of the review process a faulty conversion table was discovered and students were admitted that should not have been admitted under the system in place at that time. The lesson learned in 2003 was that admissions letters should not go out until a statistical exploration of the COMET data has been done to look for statistically unusual changes from past values and other anomalies in the data.

Spring term 2012 outside actors again asked questions about the COMET results, in this case the questions arose from parents in regards the results at Nanpei Memorial High School (NMHS). This data exploration was a result of these questions and hence focuses on NMHS. As admissions depends more heavily on the three English subsections, vocabulary, comprehension, and the essay, these were the sections that were explored.

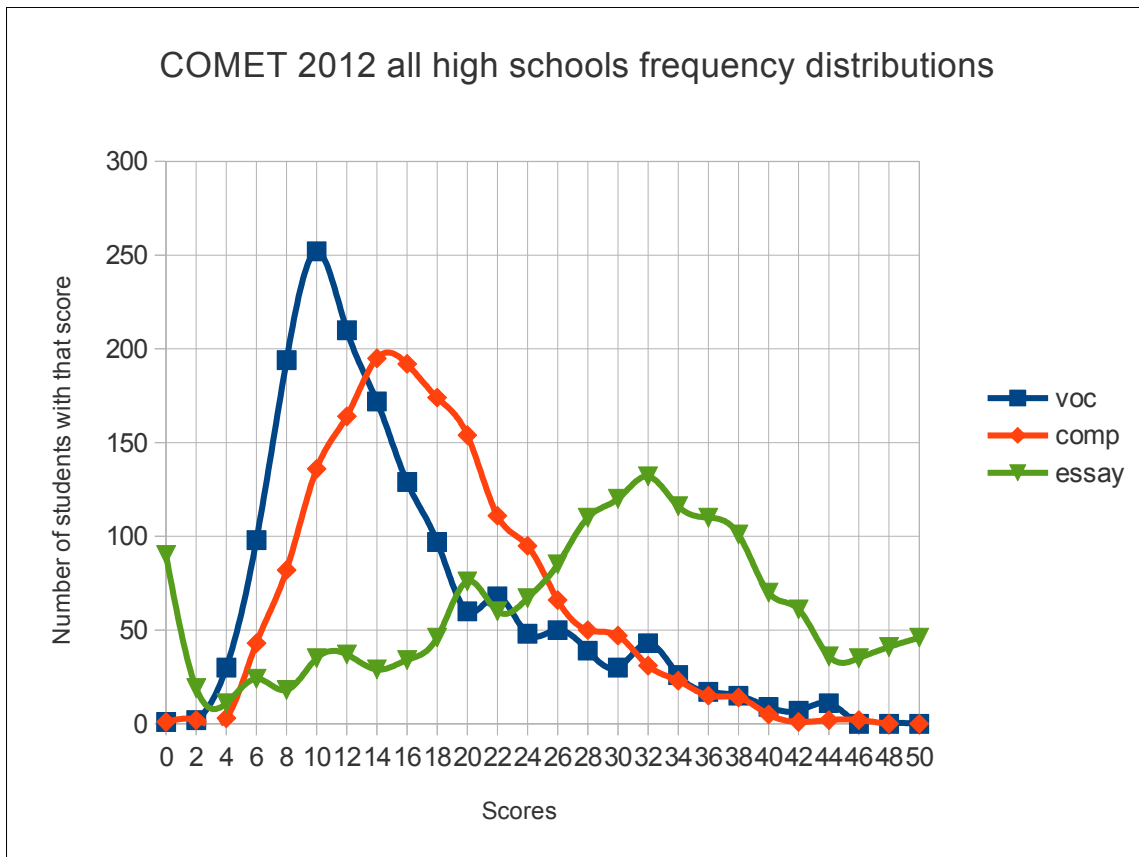
As noted in the first paragraph, understanding how a test has behaved in the past is important to determining what changes are unusual in the present. The distribution of the test scores on the English subsections has been very stable. In statistics the word "distribution" means the shape of the data plot when plotted as frequencies of occurrence. The test score distribution spring 2010 shows the score on the x-axis and the number of candidates (students) with that score on the y-axis.



Note that the peak for the vocabulary section is lower than the peak for the comprehension

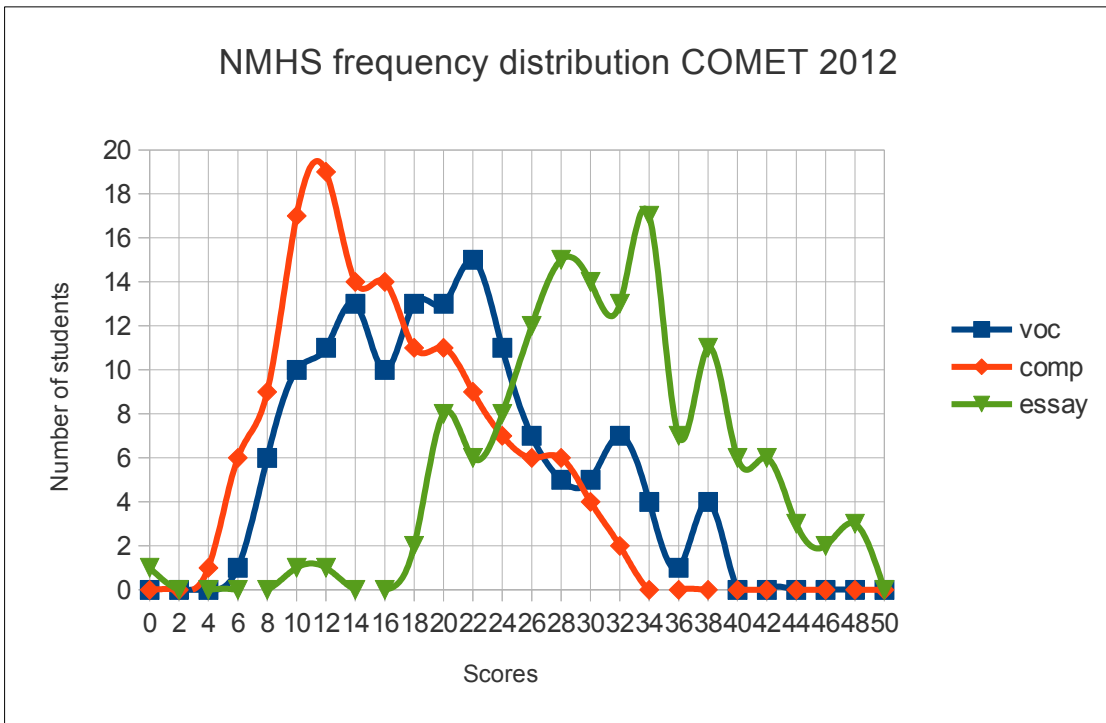
section which in turn is lower than the peak for the essay. This order, vocabulary-comprehension-essay has been stable since these three subsections have been administered in their present form.

The score distribution spring 2012 is no exception, the vocabulary peak is the lowest, comprehension is in the middle, and the peak of the essay is the highest.

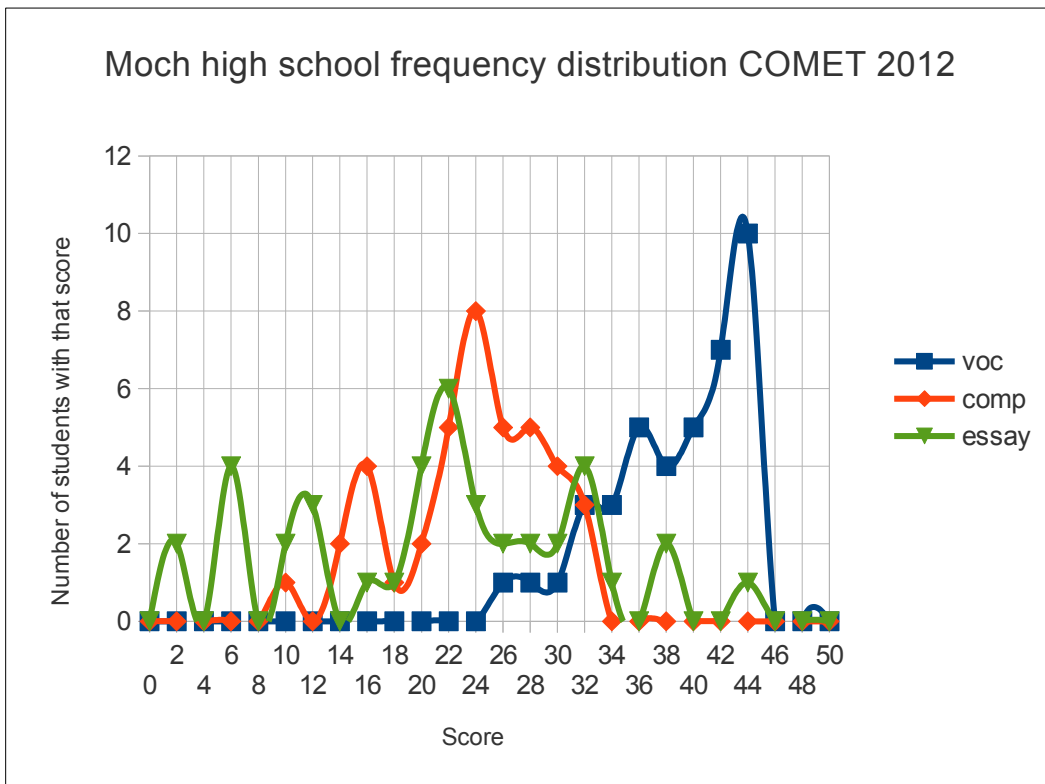


In part, differences in the total possible points and the scoring rubrics generate this very stable pattern. Another factor is that students working in a second language (L2) tend to do better at comprehension than vocabulary. This author is familiar with this effect. I often know what a sentence means in the Kosraean language without being able to decode and define every single word.

The distribution for Nanpei Memorial High School spring 2012, however, is anomalous. The vocabulary distribution peaks higher than the comprehension distribution peak. The students vocabulary capabilities appear to exceed their comprehension.



A check for other schools with a vocabulary distribution peak higher than the comprehension turned up other schools of which Moch was the most extreme example.



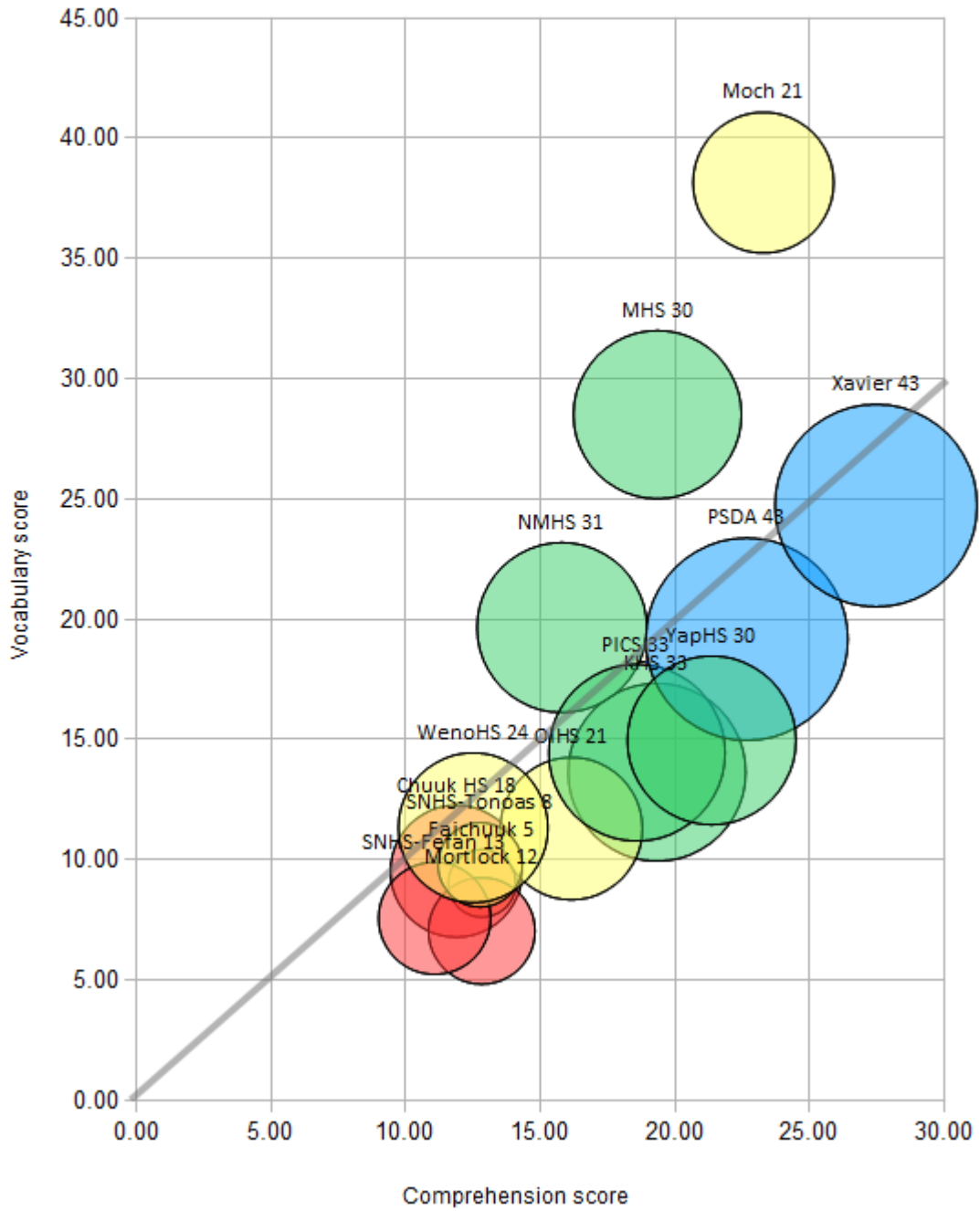
Moch has attained a near perfect score as the peak of their highly skewed vocabulary distribution. Comprehension is significantly lower and the essay spread across a wide range of values with an average of 21 out of 50. The students (candidates) are nearly perfect in vocabulary but can neither comprehend English nor put together an organized, coherent, syntactically correct essay.

This report is only a data exploration and provides no answers as to why there was an anomaly. There are many hypotheses that one could form. The test could have been compromised in some way. The students could have drilled on vocabulary using rote memorization to the exclusion of any other skills in English.

The following chart is a single chart that plots the average comprehension score versus the average vocabulary score. The average score is at the peak for a symmetric distribution, and moves away from the peak and towards the tail for asymmetric distributions. Thus a plot of comprehension on the x-axis versus vocabulary on the y-axis should result in schools falling in the region where comprehension is greater than vocabulary, or where $x > y$, below a 45° line on the chart. To provide additional information, the following chart plots the data as "bubbles" where the bubble radius is proportional to the average essay score for the school. This average is reported above the bubble after the high school abbreviation.

Comprehension versus vocabulary

Score above bubble is essay: radius is proportional



In the above bubble chart Moch, Madolehnihmw HS (MHS), and NMHS are above a 45° line which is shown in gray. In the on line version of the chart, schools with an essay average of 40 or more are in blue, essay averages of 30 to 39 are in green, essays averages of 20 to 29 are yellow, and below 20 is red. Note that Moch (essay average 21), in yellow, is far from the cluster of other schools which scored 20 to 29 on the essay. That the other schools clustered together reflects the underlying statistical fact that there is relationship between vocabulary, comprehension, and the essay score.

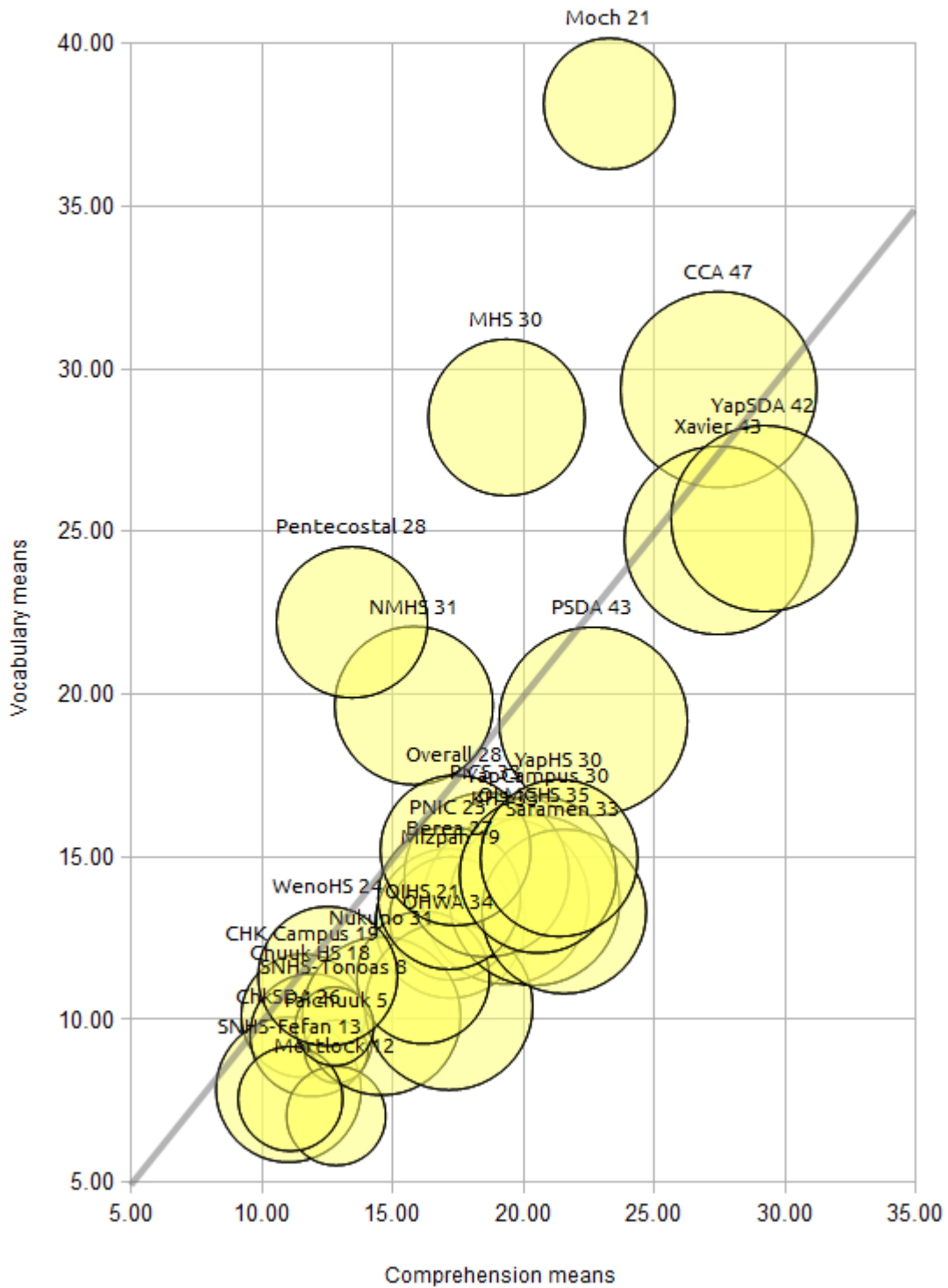
Madolenihmw HS is also above the 45° line and while not as far from the other schools with an average of 30 to 39 as Moch was from schools in its essay cohort, MHS is still anomalous. MHS performed better on the vocabulary subsection than schools that were far stronger on the essay section and comprehension sections (Xavier and Pohnpei SDA).

Nanpei Memorial HS is still above the 45° line, but the performance on the vocabulary subsection is not as unusual for schools with an essay average between 30 and 39. NMHS is near the cluster of schools with similar essay averages. While the vocabulary score for NMHS is anomalous, the result is not as extreme as those for MHS and Moch.

In the preceding chart only a few schools were selected for display to reduce clutter. In the following chart all schools are depicted. Although cluttered, this chart again shows that for most schools the comprehension average exceeds the vocabulary average. The chart also indicates that Pentecostal Lighthouse Academy with an essay average of 28 performed anomalously in regards to vocabulary and comprehension.

Calvary Christian Academy (CCA) is not necessarily anomalous. The sample size is very small, only eleven students. The essay score average is all but perfect at 47. Comprehension is also very high. These students could be expected to have strong command of vocabulary.

Comprehension versus vocabulary, essay average above bubble



If the anomalies were to be found to be a result of the vocabulary test being compromised, past reaction has been only to increase security and increase the rate at which the test contents are refreshed. A security response alone is not a complete solution. The reality is that the Pell grant refund provides a real cash incentive to gain entrance to the college. This turns the COMET into a high stakes test with real financial rewards for success. The complete solution must include moving beyond the COMET as the one and only criteria for college admission (a high school diploma is also required, but that is a minimal barrier at best).

The following are a series of data exploratory charts that were run originally to look for other anomalies. These charts also provide insight into performance on the COMET test.

The college should also look at transcripts, grade point averages, courses taken, the high school attended, and the high school class the student was enrolled in. In the following portion of this report I will use the term "class" for high school section to avoid confusion with the use of sections to refer to parts of the COMET test.

High school class information is important. The essay test, being non-multiple choice and requiring command of complex grammar, vocabulary, organization, and cohesion skills in English, appears to be a fairly valid measure of written English skills. Of the three subsections in English, the essay has the most value.

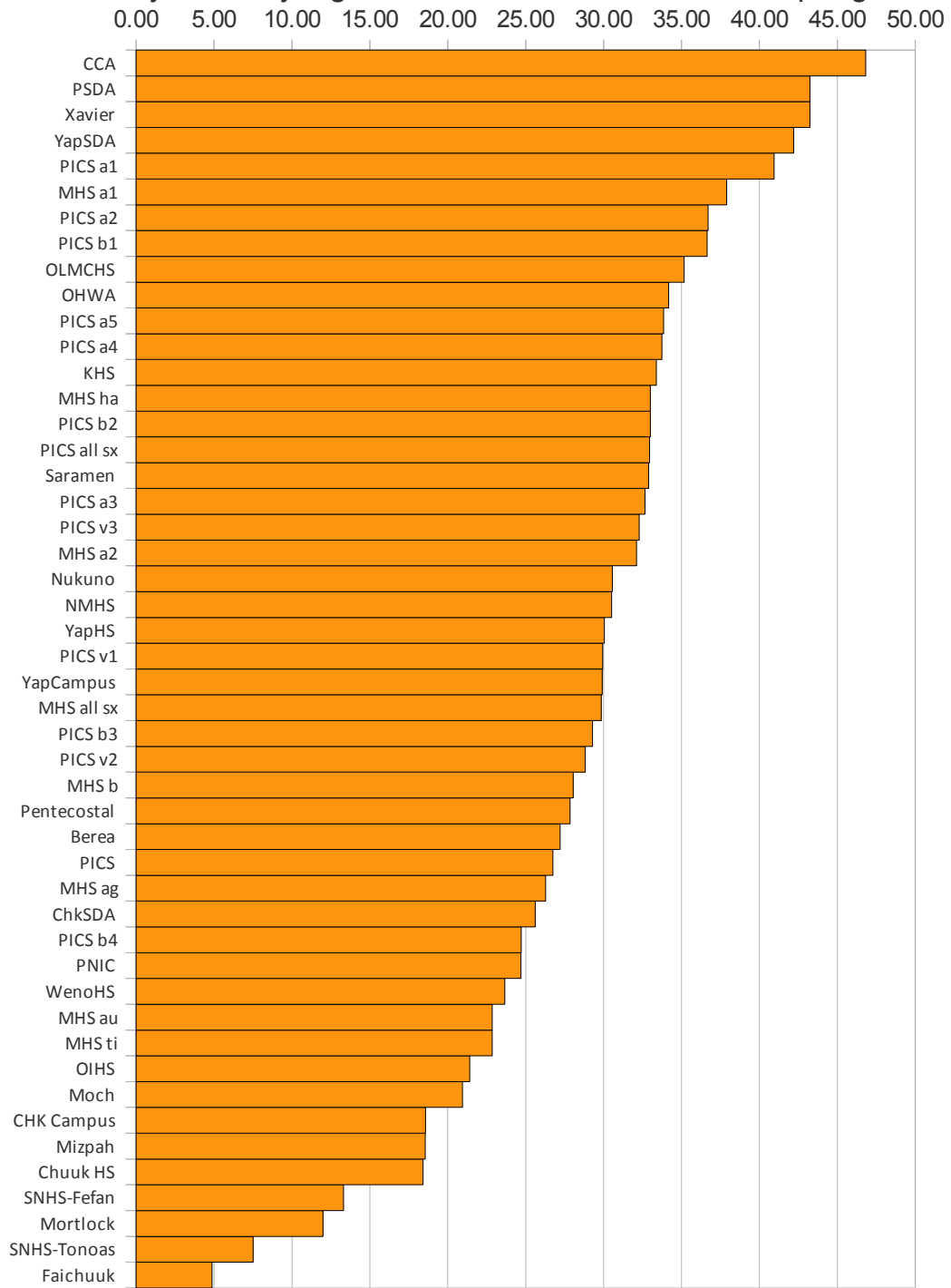
While the top four high schools in rank order for the essay subsection are CCA, Pohnpei SDA, Xavier, and Yap SDA, at number five in rank order is Pohnpei Islands Central School (PICS) class a1 with an average of 40.94. At number six is Madolehnihmw high school a1 with an average of 37.89.

The next chart shows the average for each high school on the essay section. For PICS and MHS, class averages are also provided. I was unable to obtain class lists for Nanpei Memorial HS. At PICS the "a" sections are academic, the "b" sections are business, and the "v" sections are vocational. At MHS the A and B sections (labeled a1 and a2 in the chart) are academic. MHS C (labeled b in the chart) is their business section. MHS also has sections for trades and industry (ti), auto mechanics (au), agriculture (ag), and home/art [sic] (ha).

Following the chart is a table showing the average essay score for the high schools from 2007 to 2012. Data for 2011 is missing, a year during which this author was not involved in the COMET analysis. For schools such as PICS, MHS, NMHS, KHS, and others, the averages are gradually improving. Year-to-year the averages are not far apart, over the five year span essays have improved.

There has been a tendency by some over the past couple of years to engage in disparaging remarks about the school systems here in the FSM. The COMET test is not a standardized test and the only purpose for the test is to assist the college in making admissions decisions. That said, the improvements seen in the essay scores reflect improvement in the capabilities of the candidates sitting the COMET test.

Mean essay score by high school and section COMET Spring 2012



Mean essay score

Spring 2007		Spring 2008		Spring 2009		Spring 2010		Spring 2012	
HS essay	mean	HS essay	mean	HS essay	mean	HS essay	Mean	HS essay	Mean
Berea	15.7	BEREA	26.73	Berea	23.33	BEREA	34	Berea	27.21
CCA	42	CCA PNI	39.25	CCA	45.3	CCA	40.31	CCA	46.82
CHS	9.97	CHS	17.04	CHS	15.32	CHS	13.61	Chuuk HS	18.41
CSDA	20.8	CSDA	28.38	CSDA	32.19	CSDA	30.6	ChkSDA	25.62
Faichuk	4.95	FHS	6.18	Faichuk	5.57	Faichuuk	2.35	Faichuuk	4.87
KHS	26.91	KHS	25.99			KHS	28.72	KHS	33.39
				KHS non-a	25.24				
KHS adv	37.27	KHS a1	40.71	KHS advanced	32.71	KHS A	40.05		
KHS a2	38.31	KHS a2	29.23	KHS a2	26.76	KHS B	34.45		
Mizpah	21.05	MCHS	20.1	Mizpah	22.91	MCHS	6.5	Mizpah	18.56
Mado HS	26.36	MHS	24.59	MHS	20.62	MHS	26.4	MHS	29.86
								MHS a1 (A)	37.89
								MHS a2 (B)	32.11
Mortlocks			9.77			Mortlock	9.38	Mortlock	12.00
								Moch	20.95
NICHS	19.58	NICHS	13.98						
NMS	27.75	NMHS	22.58	NMHS	25.07	NMHS	25.15	NMHS	30.51
NMS a	36.74	NMHS a1	30.95						
		NMHS a2	22.43						
NMS b	23.74	NMHS b	20.2			NMHS B	26.8		
NMS v1	19	NMHS h	18.85						
NMS v2	20.91	NMHS v	18.81						
Nukuno	12.91					Nukuno	11.89	Nukuno	30.56
Ohwa	23.33	OCHS	16.17	Ohwa	26	OHWA	30.54	OHWA	34.17
OIHS	21.3	OIHS	18.87	OIHS	18.15	OIHS	20.09	OIHS	21.41
		OLMVTs	33.56	OLMS	27.33	OLMCHS	38.43	OLMCHS	35.17
PICS	25.16	PICS	28.73	PICS	27.44	PICS	28.02	PICS	32.95
PICS a1	34.48					PICS a1	36.72	PICS a1	40.94
						PICS a2	34.69	PICS a2	36.71
PLHA	14.69	PLHA	18.67	PLHA	17.42	PLHA	24.17	Pentecostal	27.86
PSDA	37.22	SDA PNI	41	PSDA	38.63	PSDA	35.66	PSDA	43.24
Saramen	28.69	SARAM	37	Saramen	22.7	SCA	36	Saramen	32.89
SNHS	14.05	SNHS	8.02						
				SNHST	10.18			SNHS-Tonoas	7.52
		SNHS-F	9.18	SNHSF	8.61	SNHSF	9.83	SNHS-Fefan	13.32
Weipat			5.59						
Weno	14.81	WHS	17.65	Weno	14.57	WHS	20.87	WenoHS	23.67
Xavier	40.27	XHS	43.63	Xavier	44.65	XAVIER	44.66	Xavier	43.24
YSDA	40.44	SDA (YAP)	30	YSDA	28.2	YSDA	24.2	YapHS	30.06
YHS	23.86	YHS	28.99	YHS	29.33	YHS	26.86	YapSDA	42.20
Overall	22.03	Overall	24.35	Overall	23.21	Overall	24.16	Overall	27.54

Averages such as the essay averages above are sensitive to extreme values. A box plot was used to explore the actual distribution of scores for the top 12 classes on the preceding chart. As a box plot may be new to some, an explanation of a box plot follows. This section can be skipped if one already knows how to read a box plot.

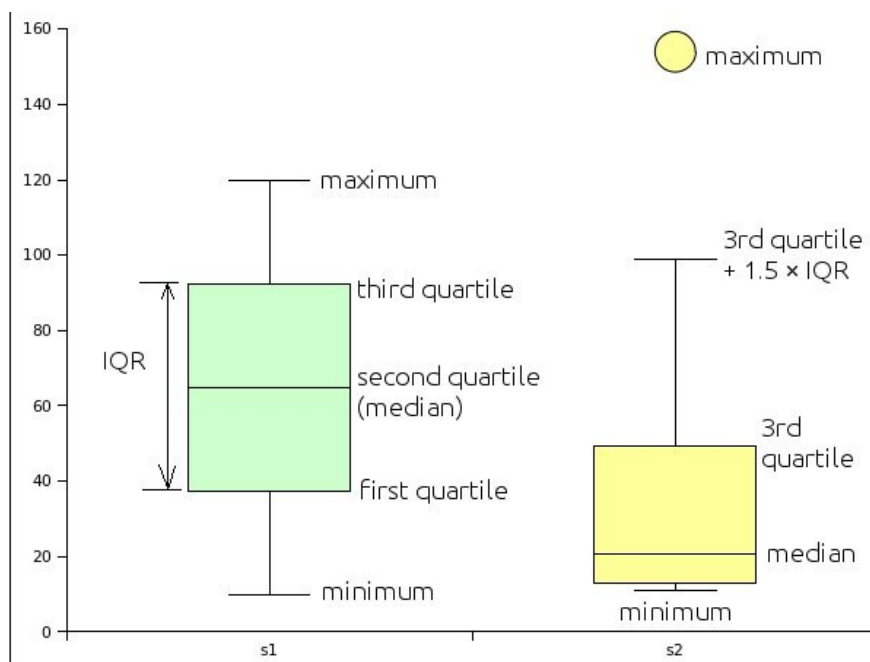
A box plot is built around a box that runs from the value at the 25th percentile (first quartile) to the value at the 75th percentile (third quartile). The length of the box spans the distance from the value at the first quartile to the third quartile, this is called the Inter-Quartile Range (IQR). A line is drawn inside the box at the location of the 50th percentile. The 50th percentile is also known as the second quartile and is the median for the data. Half the scores are above the median, half are below the median. Note that the 50th percentile is the median, not the mean.

The basic box plot described above has lines that extend from the first quartile down to the minimum value and from the third quartile to the maximum value. These lines are called "whiskers" and end with a cross-line called a "fence". If, however, the minimum is more than $1.5 \times \text{IQR}$ below the first quartile, then the lower fence is put at $1.5 \times \text{IQR}$ and the values below the fence are marked with a round circle. These values are referred to as potential outliers – data is unusually far from the median in relation to the other data in the set.

Likewise, if the maximum is more than $1.5 \times \text{IQR}$ beyond the third quartile, then the upper fence is located at $1.5 \times \text{IQR}$ above the 3rd quartile.

An example with hypothetical data sets is given to illustrate box plots. The data consists of two samples. Sample one (s1) is a uniform distribution and sample two (s2) is a highly skewed distribution.

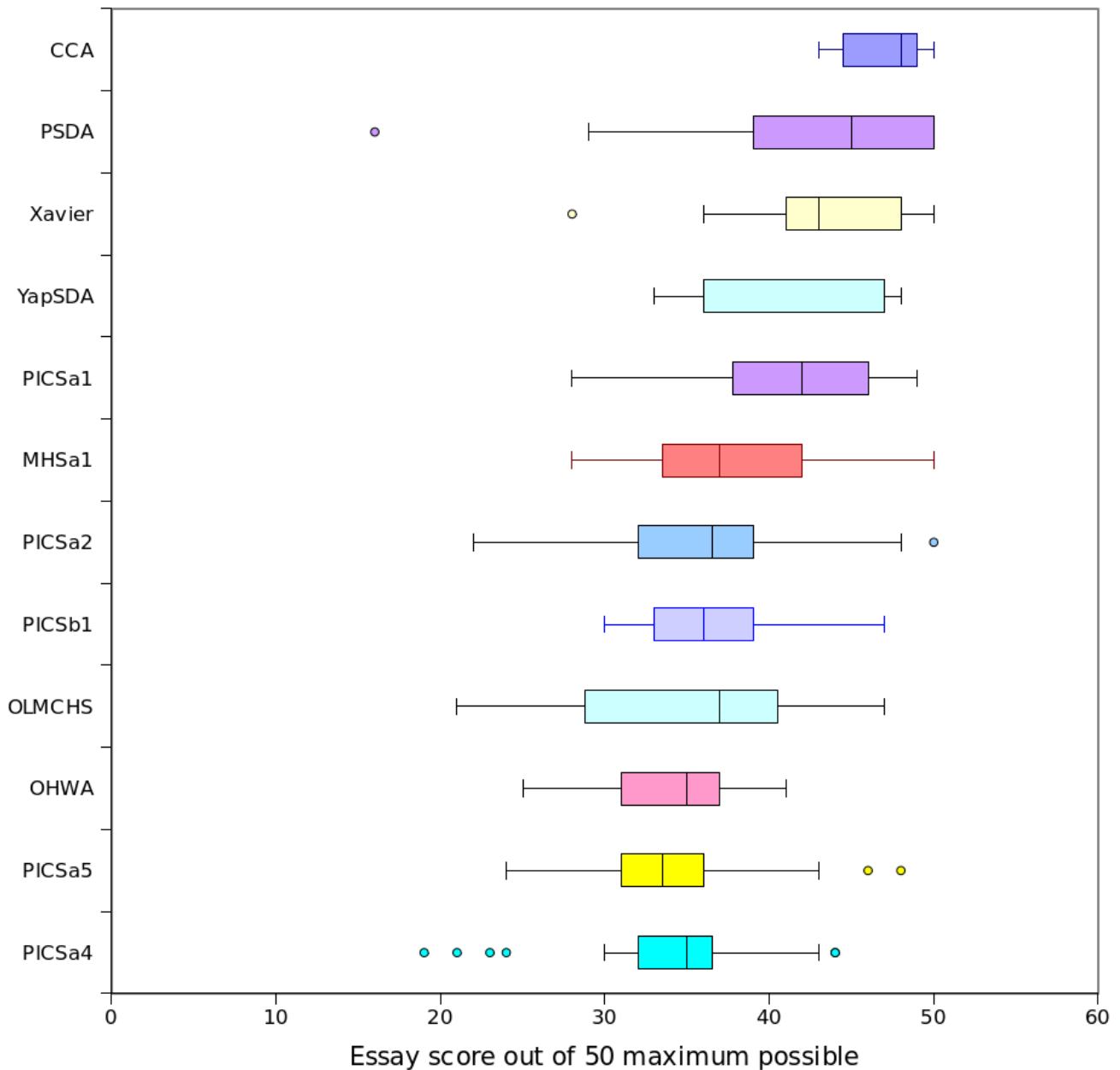
s1	s2
10	11
20	11
30	12
40	13
50	15
60	18
70	23
80	31
90	44
100	65
110	99
120	154



The use of 1.5 times the Inter-Quartile Range beyond the first or third quartiles to determine outliers goes back to the inventor of the box plot, John Tukey, who chose 1.5 times the IQR in 1977. For distributions that are not necessarily normal distributions and involve discrete data, 1.5 times the IQR has proven to work well for identifying outliers.

The box plot includes only the distribution of the essay scores for the top 12 classes on the COMET essay subsection. Note that the order is in descending order for the mean (average), but the line in the middle of the box plot is the median. Thus the chart is not in median descending order.

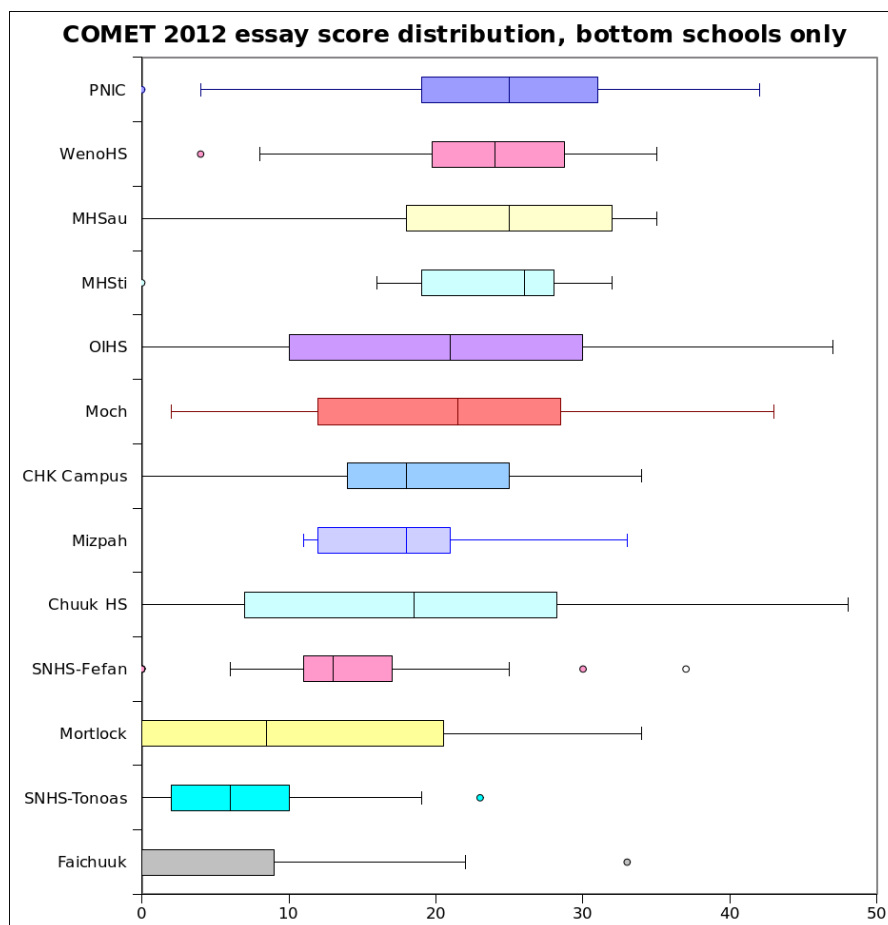
COMET 2012 essay score distributions, top classes only



The chart suggests that 50% of the students in a class score within two to five points of the median. Half of the students in a class are fairly closely clustered around the median. The whiskers extend another three to five points out to the fences. Classes are somewhat coherent in their performance on the essay. Pohnpei SDA and Xavier both had a single low outlier, while PICS a4 had four low outliers and one high suggesting higher variability in this section. Sixty-three percent of the PICS a1 students scored above Xavier high school's 25th percentile. Given that Xavier pre-selects the top eighth grade students using a test, this performance is very strong.

The fairly symmetric distribution of the essay scores on either side of the median and the lack of large numbers of outliers suggest the students are an academic cohort with similar capabilities. The box plot does not indicate any obvious anomalies in the data.

Although the box plot analysis was run for all schools and sections, the chart has too many box plots packed together. The following chart is a box plot for the bottom schools and includes tests taken at the two state campuses spring 2012.



The Pohnpei campus data and the MHS ti have outliers at a score of zero on the essay. The

schools with whiskers that extend to zero indicate schools where students scored a zero on the essay and that score is not more than 1.5 IQR below the 25th percentile.

Mortlock high schools 25th percentile is zero. More than 25% of the students scored a zero on essay. This could mean the essay was off-topic, which is believed by some essay graders to reflect an inability to read the essay question. Zeros also result from completely incoherent essays.

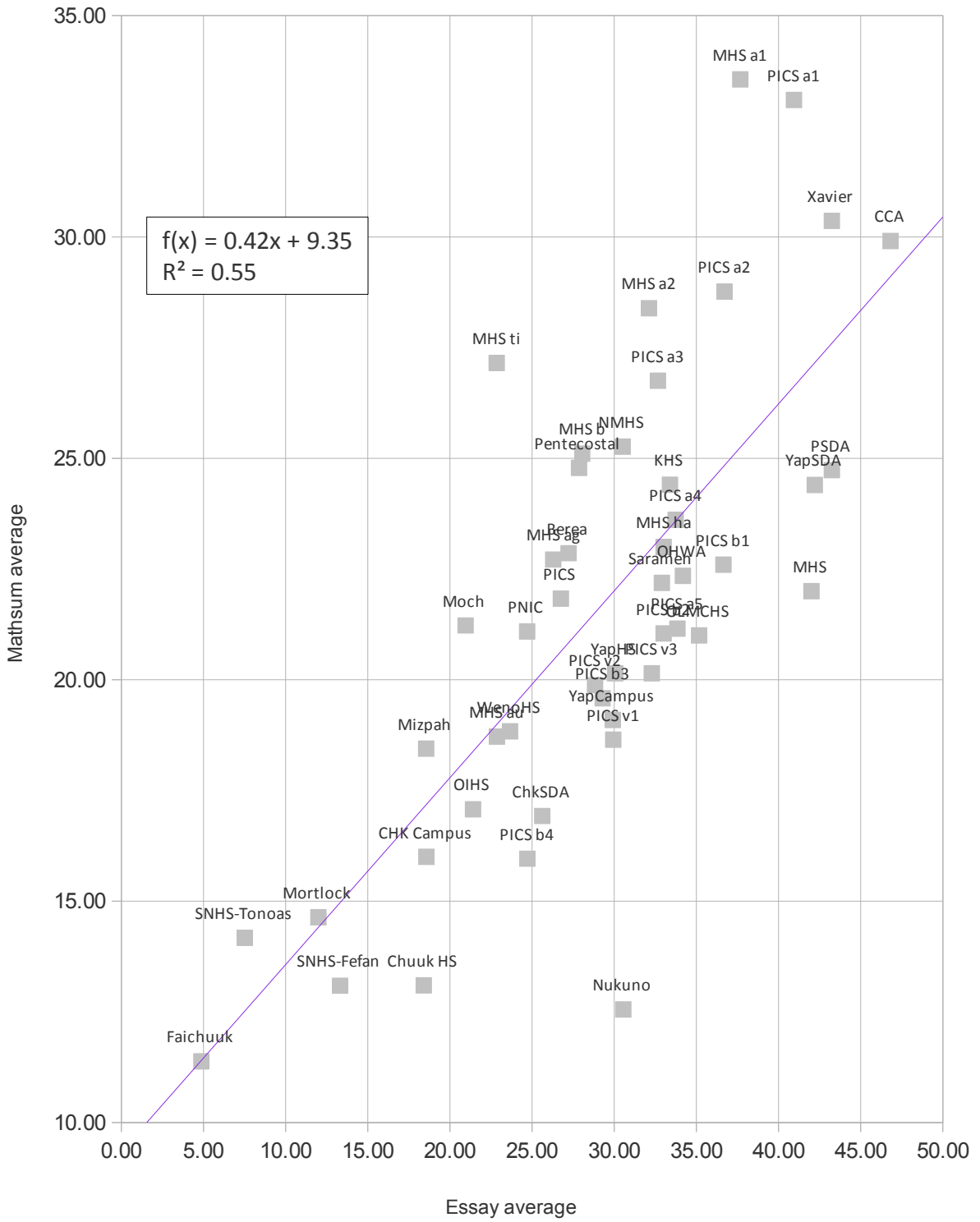
Faichuuk high school has a median of zero – over fifty percent of the students scored a zero. This author recalls Jean-Jacques Rousseau writing in his book *Émile, or on Education* that one should consider the value of an education to a child not by the final end result, but by whether – if the child passed away tomorrow – the academic journey thus far had been of value and pleasure to the child.

Faichuuk HS has long been the weakest high school among the also weak other Chuuk lagoon high schools. At the end of 12 years of "school" the student has no academic skills to show for their effort (note their math average in a chart further below). School has been a waste of 12 years of their precious lives. The high school "education" is meaningless. Set the children free. Terminate the teachers, close Faichuuk high school

In the past there has been a moderate correlation between scores on the English subsections and the mathematics average sum. Not a strong enough one to use the English score to place students in mathematics courses, but strong enough to detect anomalous behavior, if any exists.

There is a moderate to strong correlation between the essay average and the average of the sum of the math subsection scores with a correlation coefficient r of 0.74 for the spring 2012 data. Again, the strength of the performance of the a1 classes can be seen in the upper right hand corner of the following chart.

Correlation between essay and math score COMET spring 2012



Where data is not correlated, the data produces more information about a student than correlated data. Perfect correlation means that one could have wholly inferred the one variable from the other, and thus the second measurement is not an effective use of time and resources. The vocabulary, comprehension, and essay sections all tend to produce correlated data. As seen above, even the mathematics section correlates well with the essay section. There might not be a need to run three English sections. The essay section places students in writing classes, the comprehension could place students into reading classes.

Another statistical idea to grasp is that symmetric, heap-like distributions produce more information about a student than highly skewed distributions. Think of it this way. If every student passes a test with 100%, then one knows nothing about differences in the capabilities of the students. Slightly more subtle, if every student scores 0% one again knows nothing about the differential capabilities of the students. The most information about differential student abilities is gained from distributions that are not highly skewed, from information that is relatively independent of each other.

Internal to the COMET adding sections that test natural and social sciences while possibly deleting the highly skewed vocabulary sections (see the first two charts) might add information to the COMET results and should provide a more complete academic picture of a student.

External to the COMET, the college should also evaluate transcripts courses successfully completed, obtain GPAs, and class information for use in admissions decisions.

Recommendations

1. Admissions letters should not be sent out until statistical analyses of the COMET have been run to look for anomalies.
2. Other admissions criteria should be utilized in conjunction with the COMET to make admissions decisions.

Author and contact information:

All errors are solely those of the author. Please contact Dana Lee Ling at dleeling@comfsm.fm or 691-320-2480 extension 228 if you have questions, corrections, or unmet data needs in regards the COMET test. If there is break-out aggregate data you require such as class level data not broken out above, please send me a list of the names of the students/candidates and I can generate the aggregate statistics for those students/candidates.