Determining an Appropriate Sample Size in an Outcomes Assessment Project

John Regan, Faculty, Division of Rhetoric, College of General Studies

"America is shamefully inadequate at teaching statistics. A student can travel from kindergarten to a Ph.D. without ever encountering the subject. Yet statistics are ubiquitous in life, and so should be statistical reasoning." Alan S. Binder, Princeton economist, New York Times Dec.27, 2013.

Our Project:

In 2009 we began our assessment project to document student learning in seven outcome areas through four semesters in our program. In consultation with our science faculty and BU's Office of Institutional Research, we determined that 20% (110 ePortfolios) would be an appropriate sample size given the purpose of our study. With two rounds of funding from the Davis Educational Foundation, we were able to provide stipends to a team of ten faculty to rate student work during the summers of 2012 and 2013. Beginning in 2009, each CGS student maintained a single ePortfolio, and we found ePortfolio works well for our purposes because it provides a readily accessible, comprehensive collection of student work. Our committee holds norming sessions throughout the year in which we evaluate and rate student ePortfolios.

Note abour rubrics: If you are using a rubric, be sure to provide a full description of the competencies associated with each rubric area. Those descriptions are invaluable in norming sessions and will facilitate important discussions among raters. (See page 3 example).

Our Current Question:

We are approaching the third summer of our full-scale assessment work. Given the amount of faculty work hours that our project involves, do we continue to evaluate one hundred ePortfolios annually or can we reduce our sample size and still maintain statistical validity?

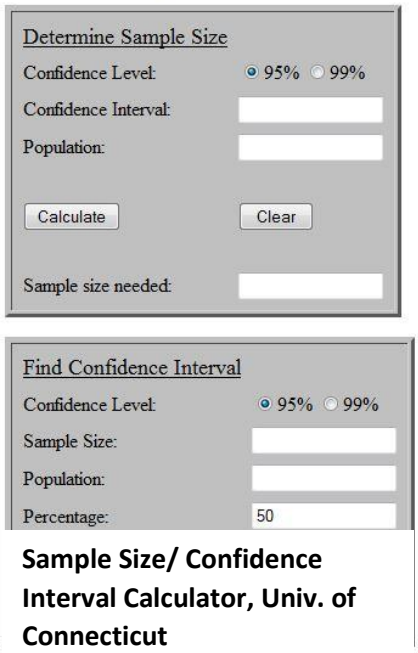
Factors to Consider:

" From a statistical perspective, sample size depends on the following factors: type of analysis to be performed, desired precision of estimates, kind and number of comparisons to be made, number of variables to be examined, and heterogeneity of the population to be sampled. Other important considerations include feasibility, such as ethical limitations on access to a population of interest and the availability of time and money." Preface to Patrick Dattalio's Determining Sample Size, Oxford University Press, 2008.

Time and Money: Assessing student ePortfolios as we have done is labor intensive and we no longer have stipends to offer our faculty raters. That said, assessment work is very highly valued at our college.

Purpose and Desired Precision: Confidence Interval and Confidence Level

Confidence Interval is a distribution of the margin of error in any measurement. **Confidence level** is the degree of probability that your result is accurate; many studies use a 95% or a 99% confidence level—a study with a 90 % confidence level, for example, means you have a 10% chance of being wrong. As you consider data collection in your assessment, you need to determine how much margin of error and possibility of being wrong can be tolerated. Certainly if you are performing clinical trials on a new drug, you want an extremely high confidence level, but in the assessment of student learning, achieving such high levels may be unnecessary and prohibitively expensive.



Sample Size/ Confidence Interval Calculator, Univ. of Connecticut

Assessment projects may have multiples purposes that could impact the desired confidence level. In our project, a summative assessment of student learning in our program, we also had an eye on presenting our work "externally" at academic conferences, publications, addition grants, etc. Perhaps befitting a general education program, some of our faculty maintain research interests in pedagogy.

Yet other assessment projects at the university may not be interested in such "external" components and may be solely focused on using their data strictly for "internal purposes" such assessing program outcomes and making curricular changes. Thus while one statistical rule of thumb is that for a population of 100 or more, the sample size must be at least 10%, perhaps the sample size and thus confidence level could be smaller in some assessments because conclusions based on analysis of the data could be measured against faculty experiences in the classroom. If the results of an assessment can produce an "a-ha" moment when a faculty member says, "You know, maybe students do a little more work in that area," perhaps there will be more faculty buy-in regarding assessment.

Possible Changes for Our Project Moving Forward

1) Reduce sample population to 15% and use a split sample collection of data. The data could be compiled in halves and then the two halves could be compared to see if there is any variance. High variance would indicate that the sample size is too small; low variance would suggest an appropriate sample size was used.
2) Use two raters for a small percentage of ePortfolios during summer assessment period. We would then double-check that our norming methods are holding, strengthen our inter-rater reliability, and perhaps allow us to hold fewer norming sessions throughout the year.
3) Focus on a specific population and trace their progress (or lack thereof) in specific rubric areas. For example, of those students scoring 3 or better in a given rubric area, do they progress at the same rate as those scoring lower? Such an example may help us determine if we are meeting the needs of our best-prepared students in achieving specific learning outcomes.

Rubric Area "Gathering, Analyzing, and Documenting Information" with Descriptions

Excellent: Synthesizes in-depth information from a range of high-quality, credible, relevant sources that are appropriate for the discipline and genre to develop ideas and documents these sources fully using MLA or Chicago style.	**Competent:** Consistently presents in-depth information from credible, relevant sources appropriate to the discipline and genre to support ideas. Documents sources with few errors or exceptions using MLA or Chicago style.	**Developing:** Demonstrates an attempt to use credible and/or relevant sources to support ideas and to document these sources properly using MLA or Chicago style.	**No Mastery:** Minimally attempts to use sources to support ideas in the writing; these sources may not be correctly documented using an acceptable style manual and/or may not be fully relevant to the task at hand.

CGS Assessment Data 2012 (100 ePortfolios)

Assessment Score Sheet: Averages 2012

Skill Levels: 1--no mastery; 2--developing; 3--competent; 4--excellent

	Term 1	Term 2	Term 3	Term 4	+/- Change
Written & oral communication	2.23	2.50	2.62	2.83	.60: **+27%**
Analyzing & documenting information	2.32	2.56	2.74	2.93	.67: **+29.5%**
Awareness of historic & cultural contexts	2.27	2.46	2.75	2.88	.62: **+27%**
Awareness of rhetorical & aesthetic conventions	2.29	2.49	2.60	2.80	.51: **+22%**
Critical thinking & perspective taking	2.34	2.48	2.66	2.92	.58: **+25%**
Integrative and applied learning	2.22	2.41	2.66	2.89	.67: **+30%**
Quantitative methods	2	1.53	2.4	2.5	.54: **+27.5%**

Required Sample Size†

Population Size	Confidence = 95% Margin of Error				Confidence = 99% Margin of Error			
	5.0%	3.5%	2.5%	1.0%	5.0%	3.5%	2.5%	1.0%
10	10	10	10	10	10	10	10	10
20	19	20	20	20	19	20	20	20
30	28	29	29	30	29	29	30	30
50	44	47	48	50	47	48	49	50
75	63	69	72	74	67	71	73	75
100	80	89	94	99	87	93	96	99
150	108	126	137	148	122	135	142	149
200	132	160	177	196	154	174	186	198
250	152	190	215	244	182	211	229	246
300	169	217	251	291	207	246	270	295
400	196	265	318	384	250	309	348	391
500	217	306	377	475	285	365	421	485
600	234	340	432	565	315	416	490	579
700	248	370	481	653	341	462	554	672
800	260	396	526	739	363	503	615	763
1,000	278	440	606	906	399	575	727	943
1,200	291	474	674	1067	427	636	827	1119
1,500	306	515	759	1297	460	712	959	1376
2,000	322	563	869	1655	498	808	1141	1785
2,500	333	597	952	1984	524	879	1288	2173
3,500	346	641	1068	2565	558	977	1510	2890
5,000	357	678	1176	3288	586	1066	1734	3842
7,500	365	710	1275	4211	610	1147	1960	5165
10,000	370	727	1332	4899	622	1193	2098	6239
25,000	378	760	1448	6939	646	1285	2399	9972
50,000	381	772	1491	8056	655	1318	2520	12455
75,000	382	776	1506	8514	658	1330	2563	13583
100,000	383	778	1513	8762	659	1336	2585	14227
250,000	384	782	1527	9248	662	1347	2626	15555
500,000	384	783	1532	9423	663	1350	2640	16055
1,000,000	384	783	1534	9512	663	1352	2647	16317
2,500,000	384	784	1536	9567	663	1353	2651	16478
10,000,000	384	784	1536	9594	663	1354	2653	16560
100,000,000	384	784	1537	9603	663	1354	2654	16584
300,000,000	384	784	1537	9603	663	1354	2654	16586