An Investigation into Plagiarism in Theses of the Russ College of Engineering and Technology 1980-2006

FINAL REPORT

David A. Koonce, Ph.D.
Associate Professor, Industrial and Systems Engineering

May 7, 2007

Executive Summary

An analysis of historical theses and dissertations in the Russ College of Engineering and Technology (henceforth RCENT), dating back to 1980, shows no indication that plagiarism is above expected values. The discovery of plagiarized documents, primarily in one department, led to the question of pervasiveness of plagiarism in the college. To study the extent of plagiarism in RCENT, two fundamental problems existed: first, how to effectively sample such a large population without creating an undue burden on faculty and staff; second, the fact that plagiarism may exist between documents in the population; confounding random sampling. These two problems were solved by scanning all theses and dissertations into an electronic database and the use of a sampling protocol, acceptance sampling, to test the quality of a large population using a relatively small sample.

The expected level of plagiarism was determined by interviewing a preeminent expert in academic integrity, Dr. Don McCabe of Rutgers University[5]. He has surveyed hundreds of thousands of students across the United States about academic integrity. The level of academic misconduct he has seen is quite high. For example, of the over 700 engineering graduate students he surveyed, 23.5% admitted "...copying a few sentences of material from an electronic source ... without footnoting it in a paper." Based on these responses, the expected level of plagiarism was set at 10%.

From this level, the test was designed and conducted during the 2006-07 academic year. Thirty-two documents were selected at random from theses and dissertations not currently under investigation. Those documents were compared to all other RCENT theses and the Turnitin® database. Turnitin is a recognized leader in academic plagiarism detection and has been scanning documents since 1996 [3]. The Turnitin reports were then analyzed by the Research Integrity Committee. The number of documents in the sample found to potentially contain plagiarized material was three. This number may decrease as the Academic Honesty Hearing Committee adjudacates these cases. Because this number well below the rejection threshold, the conclusion is that the extent of plagiarism in the historical theses and dissertations of the RCENT does not appear to be above the expected value.

Contents

Chapter 1

Introduction

1.1 Problem Statement

Some theses and dissertations in the Russ College of Engineering and Technology contain text and figures that are plagiarized or appear plagiarized. And while no cases of falsified or plagiarized results have been found, the appearance of uncited material in a thesis detracts from the credibility of the document, and ultimately, the college.

The methodology employed to identify the initial documents has not been documented and the extent of the problem is therefore, not fully known. For example, if 20 documents were discovered in a search of 40, that would indicate a significant problem. However, if 500 documents were evaluated to find 20 cases, then the problem may not be as pervasive. To determine the extent of the problem, plagiarism must be further classified by identifying the source of the copied material. Depending on the source, detecting plagiarism may be more complicated. The following section will detail how plagiarism is classified and how that classification affects the ability to detect the plagiarized text.

1.2 Types of Plagiarism

Because the cases brought to the college's attention contain examples of material copied from sources internal to RCENT (other RCENT theses) and external (articles, books) sources, any limitations the source of the material places on the ability to detect plagiarism must be evaluated. For this study, any document which contains material from an externally available source will be classified as having external plagiarism. It will be assumed that any document which contains external plagiarism will have that duplicated text detected by the evaluation software employed in the study (The Turnitin® plagiarism detection tool). However, many of the currently identified cases involve material copied from other theses in the college. Thus, any document that contains material from another thesis will be classified as having internal plagiarism. It will be assumed that the software scanning tool will detect internal plagiarism if, and only if, the associated thesis has been scanned. It should be noted that these two classifications are not mutually exclusive. A document may therefore have no plagiarism, have external only, have internal only or, have both classes of plagiarism.

The complication in evaluating the pervasiveness of plagiarism is the inability of the evaluation method to detect internal plagiarism. Because the current software employed utilizes a library of existing, mostly external, documents, a case of internal plagiarism will only be detected if the base document is in its library. The means, that if a document scans as having no plagiarism, it can either have no plagiarism or have internal plagiarism from a document unavailable to the evaluation software.

Chapter 2

Designing the Test

2.1 Developing the Sampling Methodology

Testing the historical pervasiveness of plagiarism in RCENT theses proved to be more complex than initially anticipated. In cases where it is important to know the proportion of a population possessing some characteristic, random sampling is the typical approach. Random sampling relies on a uniform distribution to determine if a member of the population should be sampled. If each member has an equal chance of being selected, a selected sample is said to be representative of the population. If a sample is representative of the population then, statistics calculated from the sample (sample mean, proportion, variation...) can be used to estimate the associated population parameter. If however, the sample is not representative, remedial methods like proportional sampling may be able to weight the statistics from portions of the sample to match expected population membership. This adjustment is at the heart of political polling.

2.1.1 Statistical Techniques

The typical method for assessing the proportion of a population possessing a property is to randomly select members of the population and test for that property. Using this sample, a statistic based on the percentage of sample units possessing the property can be used to estimate the population proportion, usually termed

ρ. The quality of this estimate can be assessed using a binomial distribution. The binomial (literally two values) distribution relies on each sample unit being classified into one of two categories (termed success or failure). Knowing the proportion of the population that fall into each category, the probability a sample contains a specific number of successes (or failures) can be determined by the formula in equation 2.1.

$$p(x) = \begin{pmatrix} n \\ x \end{pmatrix} p^x q^{n-x} \tag{2.1}$$

Where x = 0, 1, 2n (the number of successes)

n = the sample size

p = the probability of success

q = (1 - p), or the probability of failure, and

$\begin{pmatrix} n \\ x \end{pmatrix} = \frac{n!}{x!(n-x)!}$

When the sample size becomes large, say over 30, the discrete binomial distribution becomes difficult to calculate. Fortunately, the shape of a large-sample binomial approximates a normal distribution. And, using the derivable mean and variance from the binomial, a standard normal (Z) distribution can be used as a very close approximation. The normal approximation is defined in equation 2.2.

$$z = \frac{x - np}{\sqrt{npq}} \tag{2.2}$$

A primary problem with using the normal approximation to the binomial to estimate the proportion of RCENT theses which contain plagiarism is the size of sample needed to produce an estimate with a high confidence. When estimating a population proportion with a sample, like political polling, it is desirable to state an interval for the estimate. This interval is called a confidence interval and is a range one would expect the true value of the population proportion (ρ) to lie. This range, also has an associated error, like 5%. That is the likelihood that the parameter lies in that range.[1]

Solving equation 2.2, for sample size (n) requires that we know the population proportion. It is common

[1]The only way to be 100% sure (no error) the range contains ρ is to extend the range from 0% to 100%

practice to assume a worst case scenario, which happens when p and q are both 0.50. Replacing both p and q with 0.5 yields equation 2.3

$$n = \frac{z^2_{\alpha/2}}{4e^2} \tag{2.3}$$

Where $z^2_{\alpha/2}$ = The standard normal Z value for an α chance for error
e = acceptable error

The problem becomes the number of units to sample to make a narrow enough interval to provide information on pervasiveness and still have a minimal acceptable amount of error. Table 2.1 shows the sample size problem using the normal approximation to the binomial distribution. For example, to create an estimate accurate to ±5%, and be 90% confident that range held the true proportion (ρ), requires a sample size of 271.

		Error α			
		1%	5%	10%	20%
Interval ±	1%	16588	9604	6764	4106
	5%	664	385	271	165
	10%	166	97	68	42

Table 2.1: Binomial Sample Size Calculations

2.1.2 Acceptance Sampling

Acceptance sampling is a method for screening a large population to see if the proportion of the population with some property is significant. It is not a good tool for estimating the true proportion of the property in the population. It is an approach to quality control and can be employed to identify if a population contains too many defective units. It is commonly used as an inspection tool to verify the quality of a lot of new materials.

Because this type of sampling problem is frequently encountered in manufacturing systems, the US military developed (in World War II) a military standard (MIL-STD) for acceptance sampling. Formally, MIL-STD 105E is the latest and has become the basis for the ANSI standard and the ISO standard[2]. These standards are available in most quality control books and online from several sources. The following sampling standards

[2]MIL-STD 105E has been abandoned by the Department of Defense, in lieu of the ANSI standard

are equivalent [7]:

- **MIL-STD 105E** - US DOD Acceptance Sampling Protocol
- **ANSI/ASQ Z1.4** - ANSI Sampling Procedures and Tables for Inspection by Attributes
- **ISO 2859** - ISO Sampling procedures for inspection by attributes – Part 4: Procedures for assessment of declared quality levels
- **BS6001** - British Standard Sampling procedures for inspection by attributes. Procedures for assessment of declared quality levels
- **DIN40.080** - Chinese Standard
- **NFX06-022** - French Standard
- **KS A 3109** - Korean Standard

The approach is to sample a small subset of units chosen at random from a large population and the number of defects in the sample is counted. If the level (count) is at or below an acceptable lower limit, the population proportion is likely below the hypothesized value. If the count is above the limit, the hypothesized proportion is highly unlikely. If the lot is rejected, worst case, the entire lot must either be returned or exhaustively sampled (100% sampling).

The steps for the process are:

1. Determine the "Acceptable Quality Level" (AQL) - This is a value that reflects the proportion of the population you will accept with defects.[3]
2. Set the inspection level - The inspection level is a determination of how thorough the inspection should be. The levels range from very thorough for populations with which you have little knowledge (or expect significant defects), to general inspection levels (recommended).
3. Determine the lot (population) size.

[3]AQL is a bit of a misnomer. No defective units are acceptable. But, realizing that defects occur, at what proportion of defects, is the population acceptable.

4. Find the sample size code from Table 1 in MIL STD 105E - Using the lot size and inspection level, a letter is assigned A-R

5. Determine if the sampling should be single, double or multiple - In this step, the plan should be made to make one single sample or take a small sample to determine if larger samples should be taken to more fully investigate the lot.

6. Choose the appropriate MIL STD105E table - Based on the AQL, inspection level and lot size, a sample size and acceptable number of defects can be determined.

7. Extract the sample and analyze. If the number of defects found is larger than the critical value, actions taken depend on the the system.

Consider, for example, that there are 1100 undergraduate students in RCENT. It has been decided that the quality of the counseling received during admissions is acceptable if no more than 4% of students feel that they chosen the wrong major. To use acceptance sampling, the AQL (say 4%) must be chosen, inspection level (General II - the recommended initial level) chosen and whether single, double or multiple sampling is appropriate. Following the MIL STD, and using a reduced sampling set due to the cost of sampling, the sample size would be 32 and the following limits are 3 or less is acceptable; 6 or more is rejection and 4 or 5 indicates that we need to test more.

If double sampling was decided upon, an initial sample of 20 would be taken. If at most 1 student chose the wrong major, then we would "accept the lot" (no further inspection needed). If more than 4 are found then "reject the lot" (consider 100% inspection). Between those two requires an additional 20 students be questioned and the lot be "rejected" if 7 or more in both samples chose the wrong major.

Because acceptance sampling provides the ability to check that a population proportion does not exceed a maximum value, while using small sample sizes, it was chosen to study the extent of plagiarism in RCENT theses.

2.2 Detecting External Plagiarism

Detecting external plagiarism was accomplished by comparing documents against the Turnitin database. RCENT has purchased an account and currently scans all submitted theses against the Turnitin database. This database is our best tool for determining external plagiarism.

2.3 Detecting Internal Plagiarism

Detecting internal plagiarism is more difficult, as comparing a selected thesis to the Turnitin database will not find the plagiarism if the base document has not been previously scanned by Turnitin. To address this problem, the decision was made to de-bind and scan all theses and dissertations since 1980.

The initial list of theses and dissertations, submitted since 1980, was shown to have 1474 documents. However, it was later determined that several of these documents were non-thesis reports. Non-thesis reports are not archived in the library and were unavailable. Also, several of the documents were published as ETD (Electronic Theses and Dissertations). These were already available in electronic format. The remaining 1218 documents were removed from the Ohio University Library and scanned by an external company into PDF files.

These documents were submitted to Turnitin and a localized database was created to scan RCENT theses against. This database will now detect internal plagiarism.

Chapter 3

Conducting the Test

The first step in conducting the test was to determine the "acceptable quality level" or, the proportion of theses we expect (will accept) to be plagiarized.

3.1 Determine the Acceptable Quality Level

The "acceptable quality level" (AQL) is a term used in acceptance sampling used to denote what proportion of the population is allowed to posses a characteristic. This is a difficult term to reconcile with plagiarism, as no plagiarized document is acceptable. However in manufacturing, natural variation in materials and processes yields defective items. When scanning a "lot of items", it should be expected that some number will be defective. If an excessive number are defective, the lot may not be acceptable for purchase or further processing.

For the RCENT thesis scanning project, it was decided to determine from external sources what a typical level of plagiarism would be and to set the AQL to that level. It is clear that in the "lot" of theses and dissertations, plagiarism exists. So, it was decided that the AQL should be the "environmental" plagiarism level.

To determine the external level, I met with Dr. Don McCabe in December, 2006. Dr. McCabe has been

studying academic dishonesty for some time and has developed a survey instrument to assess various forms of cheating by students. He has used this survey for some years and has a database of several hundred thousand responses [4]. From this database of completed surveys, he extracted those completed by graduate students in engineering. The data is from 718 engineering graduate students from a total of 23 Universities. None of these students were from Ohio University.

Regarding the validity of the statistics, we discussed accuracy. The following points should be understood before examining the statistics[5].

- The response rate is generally 10-20%, based on campus and survey method. Depending on whether the survey was on paper or the web, the response rate is good, but not great. Students tend to elect not to respond for many reasons. However, he believes that students who have committed academic misconduct are less likely to respond.
- Most of the surveys in the sample are on-line. About 90% of the surveys in the database were on-line. He has shown with parallel paper and on-line surveys that on-line respondents are less likely to admit misconduct. The students presumabily fear the traceability of on-line forms to IP addresses.
- The survey allows questions to be left blank. For a variety of reasons, students will leave a question unanswered. In some cases it is because the do not understand the question or do not think it relates to them. However, some students leave a question blank if they feel it is incriminating.

These three points all lead to the conclusion that the numbers are not accurate, as they should be biased low. What they are is precise. So, given identical sampling methodology, two schools could be compared. Or, over time, a school can be tested to see if changes (e.g. an honor code) affect student behavior. So, for our purpose, the values would be the lowest we would expect to see.

I asked about demographics and would issues, like international students, affect rates. He has purposely not collected much demographic data. Since he has not collected that data, I do not know if the sample of engineering graduate students he provided had the same proportion of international students as our college.

Here are the specific questions related to our study and the responses from the engineering graduate students in the survey.

Question 1 Have you engaged in copying material, almost word for word, from any written source and turning it in as your own work.

Response	Raw Response	Base %	Effective %
Yes	29	4.0%	4.4%
Didn't Respond	59	8.2%	
No	630	87.7%	95.6%
Total	718	100%	100%

Table 3.1: McCabe Survey Question 1 [5]

This seems to be the most closely related to our study and 12.2% admitted or refused to respond. Based on this, Dr. McCabe felt that using 10% as a threshold (AQL) would not be unreasonable. In fact, it might be quite low, depending on our threshold for plagiarism.

Question 2 Have you engaged in paraphrasing or copying a few sentences of material from a written source without footnoting or referencing it in a paper.

Response	Raw Response	Base %	Effective %
Yes	150	20.9%	23.2%
Didn't Respond	71	9.9%	
No	497	69.2%	76.8%
Total	718	100%	100%

Table 3.2: McCabe Survey Question 2 [5]

This question is more related to what the Research Integrity Committee has been seeing as improperly cited work, or borderline plagiarism. You can see that almost 21% of engineering graduate students admitted to this act and adding the non-responders raises the total to over 30%.

Question 3 Have you engaged in paraphrasing or copying a few sentences of material from an electronic source - e.g., the Internet - without footnoting it in a paper.

Response	Raw Response	Base %	Effective %
Yes	169	23.5%	26.2%
Didn't Respond	72	10.0%	
No	477	66.4%	73.8%
Total	718	100%	100%

Table 3.3: McCabe Survey Question 3 [5]

When the source was easily plagiarized (e.g. electronic) the admission rate rose to 23.5%, with 10% declining

to respond. Given the factors pointing to an estimate biased low, the true rate could approach 40% or even 50%.

Based on these numbers, it was decided to set the AQL at 10%. This is a conservative value, given the prevelance seen in the surveys and their tendency to produce estimates which are biased low.

3.2 Determine Sample Size

The determination of the sample size is based on several factors.

1. The first is the size of the **population**. In this case, our population (lot) fell in the category 1201 to 3200.
2. The next factor, the **AQL**, is set at 10%.
3. The third factor is the **inspection level**. The standard inspection level is level II. This, however, would yield an unacceptably high sample size of 125. There are special levels for systems with high inspection costs, like destructive sampling. We chose the S-4 level, which, while providing a reduced sample size, yields the largest sample size of the special inspection plans.
4. The final factor is the **type of inspection**. This factor adjusts the acceptable number of defects in the sample based on recent quality issues with the supplier. In this study, we chose the normal level.

Figure 3.1 shows the first step in determining the sample size and acceptance quantity. Based on the population size and inspection level, the Sample Size Code Letter is G. The red boxes, added to the base table in Figure 3.1 show the determination.

Figure 3.2 shows the lookup for sample size and acceptance level. The Sample Size Code Letter G is followed to the 10% AQL, showing that a sample of 32 is required and 7 is the highest number of allowable defects.

From these factors, MIL STD 105E directs that 32 units be sampled. If 7, or less are found to be defective then the lot should be accepted. If more than that are found to be defective, the lot must be rejected.

TABLE I—Sample size code letters

(see 4.9.1 and 4.9.2)

Lot or batch size			Special inspection levels				General inspection levels		
			S-1	S-2	S-3	S-4	I	II	III
2	to	8	A	A	A	A	A	A	B
9	to	15	A	A	A	A	A	B	C
16	to	25	A	A	B	B	B	C	D
26	to	50	A	B	B	C	C	D	E
51	to	90	B	B	C	C	C	E	F
91	to	150	B	B	C	D	D	F	G
151	to	280	B	C	D	E	E	G	H
281	to	500	B	C	D	E	F	H	J
501	to	1200	C	C	E	F	G	J	K
1201	to	3200	C	D	E	G	H	K	L
3201	to	10000	C	D	F	G	J	L	M
10001	to	35000	C	D	F	H	K	M	N
35001	to	150000	D	E	G	J	L	N	P
150001	to	500000	D	E	G	J	M	P	Q
500001	and	over	D	E	H	K	N	Q	R

MIL-STD-105E

Figure 3.1: Determination of 105E Sample Size Code Letter [6]

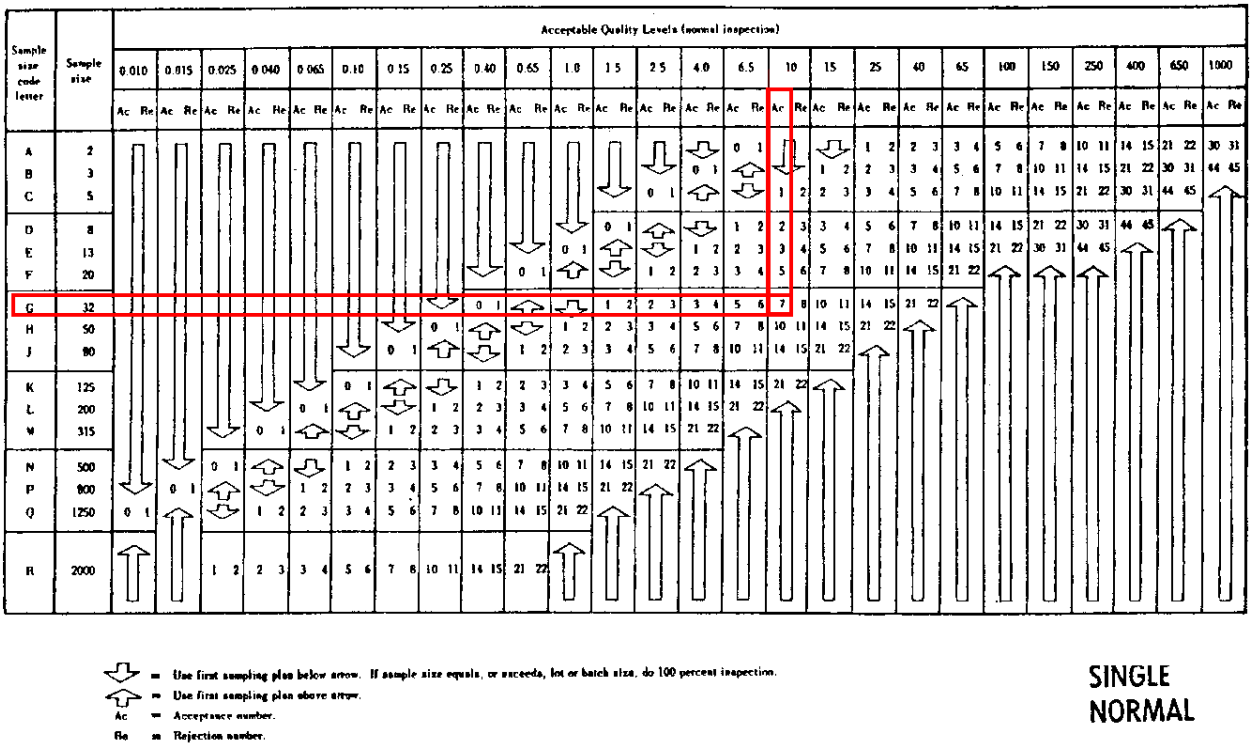
3.3 Select Documents to Examine

The documents to be examined were selected from the list of 1474 documents provided by the RCENT Dean's office. A sequence of random numbers was generated using a web based number generator [2]. The documents were then sorted using this sequence. The first 32 documents that met the following criteria were selected.

1. The document must exist in the library catalog system. This excluded non-thesis reports.
2. The document must not already be under investigation for plagiarism. This decision was based on the assumption that the previously identified problems were assumed to be isolated to specific faculty. Their documents were already being studied.

TABLE II-A—Single sampling plans for normal inspection (Master table)

(see 4.9.3 and 4.9.4)



⇩ = Use first sampling plan below arrow. If sample size equals, or exceeds, lot or batch size, do 100 percent inspection.
⇧ = Use first sampling plan above arrow.
Ac = Acceptance number.
Re = Rejection number.

Figure 3.2: Determination of Sample Size and Acceptance Level [6]

Through random sequencing, the sample of 32 documents were distributed as follows: Civil Engineering, 4; Mechanical Engineering, 9; Chemical Engineering, 4; Industrial and Systems Engineering, 5; and Electrical Engineering and Computer Science, 10. The random sample was roughly proportional to the departmental representation in the documents, with slight over representation in Mechanical Engineering. By decade, 10 documents were from the 1980s, 9 from the 1990s and 13 from the 2000s.

3.4 Examine Documents

Electronic copies of the 32 documents were submitted to the Turnitin.com software by an RCENT office administrator. The reports generated by Turnitin were then evaluated by the Research Integrity Committee (henceforth, the RIC). The RIC is chaired by Dr. Bukley, two RCENT faculty members, Drs. Sargand and

Judd, one outside faculty member (Dr. Bloomfield of the College of Arts and Sciences) and three student members.

At the time, the RIC was in the process of evaluating, or reevaluating, all of the cases submitted to RCENT, as well as former students from several individual faculty members. The 32 documents from the sample were included into the stream of documents being examined by the committee, with expectation that the documents would be reviewed at the same level of scrutiny. The RIC was briefed on the process for testing the pervasiveness of plagiarism but not the threshold for the acceptance sample.

Chapter 4

Results and Conclusions

While the process for adjudicating cases is not completed, the results of the RIC committee enable this investigation to conclude. First, the RIC has determined that, at most, three (3) of the documents from the sample of 32 have apparent plagiarism. These documents will now proceed to the case preparation phase, before finally being heard by the Academic Honesty Hearing Committee.

I discussed, in general, these documents with the chair of RIC and I understand that the three cases were deemed minor and advanced with recommendations for rewrite by the students [1]. Regardless of the severity of the cases, the total found to possibly contain plagiarism is three. Thus, this (3) is a maximum number of "defective" documents in the sample, as the Academic Honesty Hearing Committee may choose to drop any of the cases, if the document merits dismissal.

The three "defective" documents (using acceptance sampling) determined by the RIC places the number of documents well below the threshold level of seven. Therefore, the conclusion is that the proportion of documents in the historical theses of RCENT which contain plagiarized material would not appear to be above our expected level. I recommend that no further speculative investigation be done. That recommendation does not preclude investigation of prior documents that may come to light from internal or external sources.

Bibliography

[1] Angie Bukley. Personal conversation. March 29, Athens, OH, 2007.

[2] Mads Haahr. Random sequence generator. Online, 2006. http://www.random.org/sequences/.

[3] LLC iParadigms. Turnitin.com. Online, 2007. http://turnitin.com/static/company.html.

[4] Donald McCabe Linda Trevino and Kenneth Butterfield. Cheating in academic institutions: A decade of research. *Ethics & Behavior*, 11:219–232, 2001.

[5] Don McCabe. Personal conversation. December 5, Newark, NJ, 2006.

[6] Department of Defense. *MIL-STD-105E Sampling Procedures and Tables for Inspection by Attributes*. Department of Defense, Washington, DC, USA, 1989.

[7] SQC Online. Sampling by attributes - military standard 105e tables. Online, 2007. http://www.sqconline.com/military-standard-105e-ansi-z1.4.html.