

Exhibit A

**IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MASSACHUSETTS**

STUDENTS FOR FAIR ADMISSIONS, INC.,

Plaintiff,

v.

PRESIDENT AND FELLOWS OF HARVARD
COLLEGE (HARVARD CORPORATION),

Defendant.

Civil Action No.
1:14-cv-14176-
ADB

DECLARATION OF PETER ARCIDIACONO

Peter Arcidiacono, pursuant to 28 U.S.C. § 1746, declares the following:

1. I am an individual over twenty-one years of age, and of sound mind, who has never been convicted of a felony, is capable of making this declaration, and am fully competent to declare as to matters stated herein.

2. I make this Declaration in connection with what I understand to be a motion to compel the production of certain documents that Plaintiff Students for Fair Admissions, Inc. (“SFFA”) has requested from Defendant President and Fellows of Harvard College (Harvard Corporation) (“Harvard”). I have reviewed the requests for production and responses thereto that are relevant here.

A. Qualifications

3. I earned a bachelor’s degree in Economics from Willamette University, and I earned a Ph.D. in Economics from the University of Wisconsin, where I was awarded a Sloan Dissertation Fellowship.

4. I am a Professor in the Department of Economics at Duke University. I joined the Duke Economics faculty as an Assistant Professor in 1999, was promoted to Associate Professor (with tenure) in 2006, and became a Full Professor in 2010.

5. I have taken multiple Ph.D.-level courses in econometrics and regularly teach a Ph.D.-level class on the estimation of dynamic models.

6. My primary fields of interest are Labor Economics, Applied Econometrics, and Applied Microeconomics. These fields all involve the quantitative analysis of economic data through the application of mathematics and statistical methods in order to draw reliable inferences that give empirical content to economic relations.

7. I have served as an editor or associate editor for several economics journals, including serving as editor for the *Journal of Labor Economics*, the top field journal in labor economics, an associate editor for the *Journal of Applied Econometrics*, and a foreign editor for *The Review of Economic Studies*, one of the top five general interest journals in economics, and one of the two top-5 economics journals that publishes pieces on econometrics. I have published dozens of works in peer-reviewed academic and economics journals, and have given presentations across the country and around the world on topics in applied economics and econometrics. I also have two forthcoming survey papers on affirmative action in higher education, including one in the *Journal of Economic Literature*, widely regarded as the top journal for works synthesizing the literature on a particular topic.

8. In connection with my work and my research in economics and econometrics, I regularly employ statistical methods and conduct statistical analyses. I have applied discrete choice analysis, where the dependent variable is binary, in much of

my work, including using it to characterize the role of race in both undergraduate and law school admissions.

9. I have been awarded numerous grants for research in these areas generally and in particular with regard to the nature, impacts, and effects of the use of race as a factor in admissions decisions in American colleges and universities.

10. I was retained in this matter by counsel for SFFA to provide economic and statistical analysis of Harvard's use of race as a factor in undergraduate admissions decisions. The rate for my services in this matter is \$450/hour.

B. Background of the Present Dispute

11. It is necessary to review and analyze application files in order to analyze the effect of Harvard's use of race as a factor in its admissions calculus and determine how much weight this factor is given in admissions decisions.

12. Electronic databases with application scores will provide evidence as to whether or not Harvard discriminates conditional on the application score. However, these databases will not be informative for understanding whether Harvard discriminates in the formation of the score itself. The raw application files are necessary for this latter test.

13. In its First Set of Requests for Production, SFFA requested from Harvard a small fraction of the approximately 148,000 application files reviewed by Harvard over the last four years. Specifically, SFFA requested from Harvard:

A preliminary sample, stratified by race, of the complete application files (including summary sheets, notes and other documents You prepared to supplement the application as received) for admission You received and evaluated [in the last four admission cycles], to be used by the parties to determine the number of complete application files needed to form a

statistically significant sample of complete application files from the [last four admissions cycles]. This preliminary sample shall include 400 complete application files from each of the four major racial categories (Asian, Black or African-American, Hispanic or Latino, and White) from each [of the last four] admissions cycle[s]. These complete application files will be randomly selected in a manner mutually agreed upon by the parties.

14. In total, SFFA sought a preliminary sample of less than 5% of the application files (6,400 out of approximately 148,000) over the last four admissions cycles.

15. Harvard objected to this request and offered as a counterproposal to produce “80 applicant files from the graduating class of 2019 chosen by SFFA ... in addition to 80 applicant files of Harvard’s choosing.”

16. In total, Harvard proposed to produce a total of less than $\frac{1}{2}$ of 1% of the application files (160 application files out of approximately 37,000) from the most recent admission cycle and no other application files from any other admissions cycle.

C. Statistical Sampling

17. To understand the nature of the dispute here, some background on statistical sampling is warranted.

18. In statistical analysis, sampling relates to the selection of a subset of individuals from statistical population to estimate characteristics of the whole population.

19. Analyzing the whole population is preferable but tends to be costly. The goal of statistical sampling thus is to generate a sample that approximates the characteristics of the population from which it is taken such that reliable conclusions (or “statistically significant” findings, in statistics parlance) may be drawn from the sample about the whole population.

20. The sampling of a population for the purpose of drawing inferences regarding that population necessitates both relatively large samples and attention to how the sampling is performed. For example, a sample of a baseball player's at-bats that consisted of the player's first forty plate appearances in a season would not provide a sound basis for inferring the player's batting average over an entire season.

21. The appropriate sample size for producing statistically significant findings depends upon the number of variables in the population, as well as the relative frequency of the outcome of interest. A higher number of variables necessitates a larger sample size in order to yield statistically significant findings. And the relative infrequency of the outcome of interest tends to necessitate a larger sample size.

22. A sampling method is biased if it favors certain outcomes over others or otherwise excludes a segment of the population from possible selection. In particular, it would be especially inappropriate to draw conclusions about a baseball player's batting average by selectively choosing which at-bats to include in the sample. Simply put, a biased sample is not representative of the entire population, and the inferences drawn from such a sample cannot reliably characterize for the entire population.

23. Random sampling eliminates these problems by ensuring that each member of the population has an equal chance of being selected for the sample; this ensures that the sample is representative of the whole population.

24. Stratified sampling is a sampling method whereby the entire population is divided into different subgroups or "strata," and the final subjects are selected proportionally from the different strata. Characteristics of the whole population can then

be reconstructed from the strata given the knowledge of the fraction of each strata that are in the whole population.

25. Stratified random sampling is especially important here given Harvard's low acceptance rates. Random sampling would produce very few acceptances making it difficult to know what factors were important in determining admissions. A sufficient number of both rejected and accepted applicants is necessary to draw any meaningful conclusions.

26. Stratified random sampling is commonly used to discern differences or relationships between subgroups within a population and thus is often employed when studying economic phenomena across characteristics such as age, gender, socioeconomic status, religion, race or ethnicity, and educational attainment.

D. SFFA's Request for a Preliminary Sample of Application Files is Reasonable.

27. In order to examine whether discrimination is present against a particular group in Harvard admissions, data is needed on either the population of students or a sample of students from which the population characteristics can be derived. As noted above, having the full population of students would be ideal, but is costly. Hence, SFFA has requested a sample of applicants from which reliable conclusions can be drawn regarding the presence or absence of discrimination among the whole population of applicants.

28. By examining the raw files, we can see, for example, whether application files are systematically scored differently on the basis of race. With the raw files in hand, we can code the various factors Harvard describes as important in determining application subscores (for example, creating an indicator variable for whether the student

was a valedictorian). We can then use regression analysis to see, for example, whether Asians received lower subscores conditional on the factors that Harvard describes as important for that subscore.

29. In general, the ability to conclude that there is a “statistically significant” association between two variables (for example, a student’s race and her admissions outcome or application score) depends on two things: the size of the effect, and the size of the random sample available for analysis.

30. Harvard indicates that it employs a multi-factored admissions calculus in which a large number of other factors influence admissions decisions. Identifying how both major and relatively minor factors influence admissions is critical to understanding what role race may play. This is particularly true because some admissions factors tend to interact with race, e.g., low socioeconomic status positively interacts with being African-American to favor admissions. These effects require larger samples to detect.

31. Another fundamental issue involved in the analysis of Harvard admissions calculus is an understanding of the “competitive pool.” Harvard has suggested there is a vast number of applicants who are “qualified” for admission, and that the admissions process picks and chooses among this very large pool to achieve the right class mix (in terms of racial diversity and other factors). If this were true, then out of a sample of, say, 1000 files, one might find 700 that are “qualified,” and all of these files would be relevant in an analysis of which factors provide the extra boost into the admitted category.

32. In order to test this claim, the sample must be divided into equal halves: half of the files to be sampled from the 5-6% of students who are admitted; the other half to be sampled from the entire pool of rejected students. Sampling in this manner will

enable one to determine whether a very large proportion of applicants are qualified and competitive for admission. If it turns out that a relatively small number of rejected applicants have even a small chance of admission, then a larger sample would be necessary for evaluation. For example, if only 10% of all applicants have even a small chance of acceptance, then a sample of 400 applicants, divided equally between admitted and rejected applicants, would be expected to produce less than twenty rejected applicants that are really relevant to our statistical comparisons.

33. This is part of the reason why (a) substantial preliminary samples are needed, and (b) a two-stage process is necessary, the first of which is simply to demonstrate how large a sample is needed for the final analysis.

34. Because Harvard's admissions calculus includes the consideration of not only objective factors such as SAT scores and class rank but also certain subjective factors, evaluation of Harvard's use of race requires reducing the subjective factors to an algorithm that will give enable one to draw conclusions about how these factors are accounted for and how much the "weigh" in admissions decisions. For example, it is my understanding that Harvard (like most colleges and universities) gives weight to an applicant's extracurricular activities such that the level and extent of his or her participation in such activities bears on whether the applicant is admitted or rejected. In order to discern the effect that extracurricular activities have in admissions decisions, SFFA will need to "score" the extracurricular activities of numerous applicants along various dimensions, and then determine which of these "scores" matter (in conjunction with all the other characteristics of the applicants) in predicting admissions. Such an analysis requires a substantial sample of applications to produce reliable results.

35. SFFA has proposed to employ stratified random sampling in part because it would require the sampling of fewer files (relative to simple random sampling) and thus would reduce the burden of production on Harvard.

36. For all these reasons, SFFA's request for a random sample of 400 applicant files for each of four major racial groups for each of four admissions cycles is methodologically well-grounded. Such a sample may or may not be sufficient itself to allow for reliable conclusions to be drawn about the whole population of applicants. But in any case, it should be sufficient to at least determine what factors appear to be driving the admissions process, and to accurately measure what overall sample would be necessary to measure the influence of those factors with statistical confidence.

E. Harvard's Counter-Proposal is Unreasonable and Could Not Possibly Yield Reliable Findings About the Use of Race in its Admissions Decisions.

37. A non-random sample, which is what Harvard would produce with "its" 80 files, would not be of any use for statistical analysis. Even a large sample of unknown selection criteria cannot provide a basis for sound conclusions about the population as a whole.

38. Size also matters. If one were to draw a random sample of 80 (or 160) files out of a population of approximately 40,000 applications, it is highly doubtful that any analysis could yield meaningful conclusions about the application process – particularly given Harvard's low acceptance rate.

39. Even in a situation where one were interested in only one association, a sample of 80 would rarely be sufficient. For example, if one were trying to predict an election and randomly sampled voters, and the 80 voters are split 48 / 32 in favor of Candidate A, one could not estimate with 95% confidence that Candidate A will win the

election. That is to say, very large effects in a very simple model cannot be statistically distinguished in a sample of 80.

40. On top of that, the Harvard model is not simple. Rather, as explained above, it involves numerous variables of potentially differing weight in the admissions calculus. A sample of 80 files thus would be woefully insufficient for statistical analysis.

41. Adding 80 files selected by SFFA (as Harvard proposes), would make no difference for the very same reasons.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on this day, July 16, 2015.

/s/ Peter Arcidiacono
Peter Arcidiacono