Pareto Improvements in the Contest for College Admissions

Wojciech Olszewski and Ron Siegel*

December 2016

Abstract

We model high school students' competition for college admissions as an all-pay contest with many players and prizes, and investigate how reducing the information revealed to colleges about students' performance can improve students' welfare in a Pareto sense. Less information reduces the assortativity of the resulting matching, which reduces welfare, but also mitigates competition and reduces student effort, which increases welfare. We characterize the Pareto frontier of Pareto improving policies, and also identify improvements that are robust to the distribution of college seats and students' utilities from the same. We apply our results to data from the National Longitudinal Survey of Youth to identify Pareto improving policies in the United States.

1 Introduction

Every year, millions of high school students compete for college admissions.[1] Higher-performing students, who are presumably of higher ability, are admitted to more desirable colleges, which leads to beneficial assortativity. But the effort students exert to signal their ability is costly. As competition for college admissions intensifies, a growing concern is that this cost outweighs the benefits of assortativity. In South Korea, for example, it is not uncommon for high school students to spend sixteen hours a day on their studies, and the high stakes competition for college admissions is seen as one of the main causes for the high rates of unhappiness and suicide among teenagers.[2] Similar concerns regarding the negative consequences of

*Department of Economics, Northwestern University, Evanston, IL 60208 (e-mail: wo@northwestern.edu) and The Pennsylvania State University, University Park, IL 16802 (e-mail: rus41@psu.edu). We are indebted to Bela Szabadi for outstanding research assistantship. We thank Shengwu Li, Emir Kamenica, and seminar participants at Bocconi, Cambridge, Hong Kong, Johns Hopkins, Kyoto, Leicester, Northwestern, Oxford, Penn State, and Wisconsin for very helpful comments and suggestions. Financial support from the NSF (grant SES-1325968) is gratefully acknowledged.

[1] In 2012, for example, 4-year colleges in the United States received more than 8 mln applications and enrolled approximately 1.5 mln freshmen.

[2] See, for example, Matthew Carneys' discussion "South Korean education success has its costs in unhappiness and suicide rates" from June 15, 2015 on the Australian Broadcasting Corporation. The high suicide rate among teenagers is frequently attributed to their and others' expectations for them to do well in the competition for college admissions (Ahn and Baek (2013)).

intense competition have also been raised in the United States.[3]

In an attempt to reduce the detrimental effects of competition, the South Korean Ministry of Education is changing how the standardized test that determines college admissions will be graded. Currently, each part of the College Scholastic Ability Test (CSAT) is graded on a 0-100 or 0-50 scale. Starting in 2018, scores of 90-100 in the English component of the CSAT will be reported as one grade, scores of 80-89 as another grade, etc. "Students in the same graded classification will all be considered on an equal playing field in the college admissions process, regardless of their numerical scores.[4]" The goal is to reduce competition between students,[5] while recognizing that the assortativity of the admissions process will be reduced as well.[6]

This paper investigates how different performance disclosure policies affect students' welfare in the competition for college admissions. We model college admissions as a contest with a large, but finite, number of players (students) and prizes (college seats). Students may be ex-ante asymmetric in their scholastic ability (type), and invest costly effort to signal their ability (as in Spence (1973)). Their effort determines their performance, the students with the best performance are admitted to the best college, those with the best performance among the remaining students are admitted to the best college among the remaining colleges, etc.[7]

The combination of heterogeneous players, incomplete information, and heterogeneous prizes makes a direct equilibrium analysis of this large contest intractable. We overcome this difficulty by using results from Olszewski and Siegel (2016a), which relate the equilibrium outcome of large contests to a particular mechanism in a single-agent setting. This mechanism implements the assortative allocation of prizes to agent types, where the prize distribution is the empirical distribution of prizes in the contest and the type distribution is the average of the type distributions of the players in the contest. The results show that this mechanism closely approximates the equilibrium outcome of the contest. The intuition is that in a large contest players effectively compete against players with similar types, and this leads to prizes being allocated assortatively and pins down players' effort. This approximation allows us to investigate how different performance disclosure policies affect the contest's outcome by studying the effects of these policies

[3]For example, Hsin and Xie (2014) report that the high academic effort Asian-Americans exert leads to lower subjective well being and to psychological and social difficulties.

[4]Korea JoongAng Daily, October 10, 2015. "CSAT English section to take absolute grade scale."

[5] This "... grading system aims to reduce excessive competition among test-takers." "'We are trying to alleviate unnecessary and exorbitant competition between students who are competing with one another to gain one or two points more,' said Kim Doo-yong, a ministry official." Ibid.

[6]"In the last mock exam..., 23 percent of total examinees scored in the first grade... but only 4.64 percent of total examinees received perfect scores, which by current standards means they would have been the only ones to classify in the first grade." Ibid.

[7]Notice that, unlike in Spence (1973), the ability to signal via costly effort can increase welfare through its effect on assortativity. In addition, our model is not a signaling game, since students' payoffs depend on their effort and rank order of performance, and not on colleges' beliefs about their ability.

on the approximating single-agent mechanism.

We begin by considering “top pooling,” in which the highest performing students are pooled and randomly assigned to the corresponding fraction of the best college seats. The approximating mechanism is characterized by a threshold type, such that all higher types choose the same level of effort and are randomly awarded one of the top prizes, and all lower types choose effort and obtain a prize as in the approximating mechanism without top pooling. Types slightly above the threshold exert more effort and obtain a prize lottery that is better than in the approximating mechanism without top pooling, and the reverse holds for types close the maximal type. Our key finding is that top pooling is Pareto improving, i.e., weakly improves the welfare of all types and strictly improves the welfare of some types, if and only if the average prize utility from the lottery over college seats assigned to types above the threshold type weighted according to the conditional type distribution truncated below the threshold type exceeds the one from the lottery over these college seats weighted according to the uniform distribution truncated below the threshold type. This holds, for example, if the conditional type distribution first-order stochastically dominates (FOSD) the conditional uniform distribution. One implication of this is that randomly assigning students to colleges Pareto dominates the admissions contest when the type distribution FOSD the uniform distribution. We characterize the Pareto frontier of top pooling policies, and also show that in large contests top pooling is equivalent to a policy that restricts the maximal effort that students can exert.[8] Since some top pooling policies are Pareto improving, this shows that the insight that leveling the playing field by restricting effort helps weaker players by hurting stronger players does not always hold.[9]

We then turn to more general performance disclosure policies, which we term “category rankings.” A category ranking consists of one or more ordinal ranking intervals, such that all students whose performance ranking lies within an interval are pooled and obtain the same lottery over college seats. Our key finding for top pooling generalizes to category rankings by considering each interval separately. We show that category rankings sometimes Pareto dominate top pooling policies, and provide a characterization of the Pareto frontier of category rankings.

The conditions under which top pooling and category ranking are Pareto improving depend on the particular distribution of college seats, distribution of student abilities, and students’ valuation function for college seats. Since obtaining reliable estimates for all these parameters may be difficult, it is desirable to obtain conditions that are independent of at least some of these parameters. The ranking in terms of conditional FOSD described above delivers such conditions, which apply to all distributions of college seats and student valuation functions and depend only on the distribution of student abilities. An added benefit of these robust conditions is that the Pareto frontier of the robust top poolings and category rankings are singletons and easy to characterize. For top pooling it is the largest fraction of highest performing students such that the conditional FOSD holds. For category rankings it is the union of the maximal intervals on

[8] One example of such a policy is to limit the number of AP classes that students can take or AP credits they can use.

[9] Che and Gale (1998) demonstrated this insight in the context of political lobbying.

which the conditional FOSD holds.

We apply this characterization of the Pareto optimal robust category ranking to grading policies in the United States. To do this we assume that aggregate grade distributions correspond to the population ability distributions. The practice of "curving" grades underlies this assumption, since curving can be interpreted as an attempt to make the grade distribution better reflect the ability distribution. We consider grade distributions based on the grade data in the National Longitudinal Survey of Youth 1997, which includes SAT, ACT, and PSAT scores, high school grades, and AP credits. For each performance measure we identify the Pareto optimal category ranking. We find that for every performance measure coarsening the grading policy would lead to Pareto improvements. One similarity across all performance measures is that the optimal robust category ranking for each prescribes a large "bottom pooling" interval. One difference is that the upper endpoint of this interval for English, foreign language, and social sciences is so high that nearly all students are pooled, whereas for math and overall GPA it is optimal to include additional grading categories that include a substantial fraction of the students.

Finally, we use the data on AP credits to identify the Pareto improving top poolings and the corresponding Pareto frontier. We first estimate a linear relationship between AP credits (proxy for ability) and average post-graduation income. We then use this estimate to identify all the Pareto improving top pooling thresholds, using the condition for top poolings described earlier. The Pareto frontier consists of four top pooling thresholds, such that pooling students with a higher number of AP credits is Pareto optimal.

The rest of the paper is organizes as follows. Section 1.1 reviews the related literature. Section 2 introduces the model. Section 3 presents the approximation technique. Section 4 investigates top pooling. Section 5 investigates category rankings. Section 6 examines the effect of positive externalities, which are abstracted from in the majority of the analysis. Section 7 derives the robust conditions for Pareto improvements. Section 8 considers the Pareto improving grading policies for the US grade data. Appendix A contains proofs. Appendix B extends our results for top pooling to more general student utility functions. The Data Appendix describes the grade distributions and Pareto improving category rankings.

1.1 Relation to the literature

College admissions feature prominently in the matching literature, beginning with Gale and Shapley's (1962) seminal contribution. The focus of much of this work is on stability and efficiency in the presence of heterogeneous student preferences, while abstracting from the effort students exert. Approximating continuum models have been recently used to investigate large matching markets (for example, Azevedo and Leshno (2016)). Since we are interested in Pareto improvements, endogenous effort choice is an important feature of our framework. To obtain a tractable model, we impose the same ordinal (but not necessarily cardinal) preferences for all students. In reality there are differences in students' preferences over colleges, but homogeneity may be a reasonable simplifying assumption, at least for students in the same geographical area.

A more closely related paper is Ostrovsky and Schwarz's (2010) investigation of information disclosure policies in matching markets. In their setting, each school knows the ability of each of its students and chooses how much information to reveal about these abilities by pooling together students with different abilities. Students are passive and exert no effort, and the results focus on the amount of information schools reveal in equilibrium. In our analysis, in contrast, performance disclosure policies affect students' efforts, which in turn determines which policies are Pareto preferred.

Frankel and Kartik (2016) consider the amount of information about students' abilities revealed by standardized testing, and study a signaling game in which a player can manipulate the market's beliefs about her ability by sending a costly signal. Players are heterogenous in their ability and their cost of sending a signal, and this two-dimensional heterogeneity leads to pooling in equilibrium. Our model is not a signaling game, and we characterize when performance disclosure policies that induce partial or complete pooling are Pareto optimal.

Dubey and Geanakoplos (2010) consider a game of status between students. A student's status is equal to the difference between the number of students with a lower grade and the number of students with a higher grade. In particular, the aggregate allocation value of status is always 0. A student's performance is a noisy measure of his costly effort, and, similarly to our model, a grading policy pools intervals of performance. The focus is on characterizing grading policies that maximize effort. Such policies involve some pooling, and with heterogeneous students necessary conditions for such policies are derived.

Our paper also contributes to the theory of all-pay contests. Most papers in this literature focus on settings with two players, ex-ante symmetric players, or identical prizes. Olszewski and Siegel (2016a) introduced the approximation approach to large contests, which makes it possible to study contests with many ex-ante asymmetric players and heterogeneous prizes, as we do here. Olszewski and Siegel (2016c) use this approach to study effort-maximizing contests, and Bodoh-Creed and Hickman (2015) use a similar (and independently developed) approach to study affirmative action in college admissions. While there are several technical differences between their model and ours,[10] the main differences are in the design instrument they consider (affirmative action) and their focus on aggregate welfare as opposed to our focus on Pareto improvements. One of their findings is that using a lottery to assign students to colleges would generate higher welfare than a contest for college admissions. Our investigation of optimal category rankings shows that a pure lottery can be improved upon in a Pareto sense by partitioning the set of students into several categories based on their effort and using a separate assignment lottery for each category.

[10] They consider a more general utility function but assume that the limit distribution of college seats is atomless, and restrict attention to two groups of students, the minority and the majority, such that students within each group are ex-ante symmetric.

2 Contest economy

A large number of high-school students compete for university admissions by investing effort. We model this environment as an economy that consists of n players (students) and n prizes (university seats), where n is large. A player is characterized by a type $x \in X = [0,1]$, and a prize is characterized by a number $y \in Y = [0,1]$. Prize 0 can be interpreted as "no prize." The prizes are known, and commonly ranked by all players: $y_1 \leq y_2 \leq \cdots \leq y_n$ (some of them may be 0, i.e., no prize). As specified below, prizes' cardinal values can vary across players. Since each player can obtain at most one prize (a student enrolls at no more than one university) and some prizes can be "no prize," it is without loss of generality to have the number n of players equal to the number of prizes. Player i's privately-known type x_i is distributed according to a cdf F_i, and these distributions are commonly known and independent across players.[11] Notice that the distributions need not be identical, which accommodates ex-ante asymmetry across players. A player's type affects her effort cost or prize valuation.

After privately observing her type, each player chooses an effort t (which can be interpreted as performance), and the player with the highest effort obtains the highest prize, the player with the second-highest effort obtains the second-highest prize, and so on.[12] Ties are resolved by a fair lottery. While effort may be productive, we assume that its cost outweighs the benefit it generates, so it is wasteful on net, as in Spence's (1973) signalling model.

The utility of a type x player from providing effort $t \geq 0$ and obtaining prize y may have one of the two forms:

$$xh(y) - c(t), \tag{1}$$

in which the player's type affects her prize valuation, or

$$h(y) - \frac{c(t)}{x}, \tag{2}$$

in which the player's type affects her effort cost, where $h(0) = c(0) = 0$ and $\lim_{t\to\infty} c(t) = \infty$, the prize valuation function h and cost function c have second-order continuous derivatives, c is strictly increasing in t, and h is strictly increasing in y. These forms are strategically equivalent, because the former utility of each type x is obtained from the latter utility by multiplying it by x. In particular, the set of equilibria is the same for both forms. Some of the welfare analysis will, however, differ between the two forms. Every contest has at least one (mixed-strategy) Bayesian Nash equilibrium.[13]

While we focus on (1) and (2) (as does most research on contests), notice that the strategic equivalence extends to any form

$$f(x)\,h(y) - \frac{c(t)}{g(x)},$$

[11] All probability measures are defined on the σ-algebra of Borel sets.

[12] We assume that effort fully determines a player's performance, and abstract from any other factors that may realistically affect this performance.

[13] This follows from Corollary 5.2 in Reny (1999), because the mixed extension is better-reply secure.

with functions $f \geq 0$ and $g \geq 0$ that satisfy $f(x)\,g(x) = x$. This can accommodate types that affect both prize valuations and effort cost.

In Appendix B, we show that our results on top pooling generalize to separable utility functions of the form $h(x, y) - c(t)$ and $h(y) - c(x, t)$ that satisfy some conditions.[14]

3 Mechanism-design approach to studying contests

As pointed out in the introduction, a direct analysis of heterogeneous contests of the kind described in Section 2 is difficult or impossible. Since the contests we consider have many players and prizes, we can make use of the tractable approach to studying the equilibria of large contests, which was developed in Olszewski and Siegel (2016).

The idea underlying this approach is as follows. In equilibrium, the prize a player obtains by choosing effort t is determined by the ranking of t relative to the other players' equilibrium effort choices. Since these choices are random from the player's point of view, his effort leads to a probability distribution over prizes. When the number of players and prizes in the contest is small, these prize distributions can differ significantly even across players who choose the same level of effort. In addition, the variance of each prize distribution can be substantial. For an illustration, consider two players with different valuations who compete for one prize in an all-pay auction with complete information. This corresponds to a contest with $y_1 = 0$, $y_2 = 1$, prize valuation function h that satisfies $h(1) = 1$, and cost function $c(t) = t$. Players' publicly observed, deterministic types satisfy $0 < x_1 < x_2 < 1$. It is well known (Hillman and Riley (1989)) that this contest has a unique equilibrium. Player 2 chooses an effort by mixing uniformly on the interval $[0, x_1]$, and player 1 chooses effort 0 with probability $1 - x_1/x_2$ and with the remaining probability mixes uniformly on the interval $[0, x_1]$. By choosing an effort $t \in (0, x_1)$, player 1 wins the valuable prize with probability t/x_1 and wins "no prize" with probability $1 - t/x_1$. Both probabilities are non-negligible for t that is not close to 0 or x_1. The corresponding probabilities for player 2 are $1 - x_1/x_2 + t/x_2$ and $x_1/x_2 - t/x_2$, which are, respectively, higher and lower than those of player 1. This asymmetry reflects the asymmetry in players' valuations.

In a large contest, however, all players face essentially the same competition, even if they are asymmetric. More precisely, the ranking of two players who choose the same effort level differs by at most 1, which is very small in percentile terms when the number of players is large. That is, the mappings between effort levels and the distributions of players' percentile rankings (given the other players' equilibrium strategies), and therefore the distributions of prizes they obtain, are similar across players.[15] Moreover, because players'

[14] Similar results can be obtained for category rankings, and even for some more general utility functions of the form $h(x, y) - c(x, t)$. But the form of such general results is more involved.

[15] For some effort levels the distributions of prizes may not be similar for all players even when the percentile rankings are, but these levels are negligible in equilibrium (see Olszewski and Siegel (2016) for details).

types' are independent and their equilibrium strategies are independent, by the law of large numbers these mappings are almost deterministic, so any effort choice leads to an almost deterministic prize. Thus, in an equilibrium of a large contest players can be thought of as facing a common, deterministic "inverse tariff," which maps every effort level to a prize. Now, notice that players' utilities (1) and (2) satisfy strict single crossing, that is, if a player prefers to obtain a higher prize at a higher effort level, then any player with a higher type strictly prefers to obtain the higher prize at the higher effort level.[16] This implies that a player with a higher type chooses a weakly higher prize from the inverse tariff. Therefore, the equilibrium allocation of prizes to player types is close to the unique assortative allocation, in which higher types obtain higher prizes. Finally, well-known results from the mechanism design literature show that the allocation pins down players' effort levels, provided the lowest type obtains utility 0 (which each player can guarantee by choosing effort 0). To summarize, any equilibrium of a large contest is approximated by the unique mechanism that implements the assortative allocation of prizes to player types and gives the lowest type utility 0.

To formalize this, let $F = \left(\sum_{i=1}^{n} F_i^n\right)/n$ be the average distribution of players' types, so $F(x)$ is the expected percentile ranking (up to $1/n$) of type x given the random vector of players' types. Let G be the empirical distribution of prizes, which assigns a mass of $1/n$ to each y_j (recall that each prize y_j is known). We assume that F is strictly increasing and continuous, but G can be any distribution. This assumption on F is made primarily for expositional simplicity.[17] It is satisfied, for example, when the type distribution F_i of each player i is continuous and strictly increasing. The assortative allocation of prizes to types assigns to each type x prize $y^A(x) = G^{-1}(F(x))$, where $G^{-1}(r) = \inf\{z : G(z) \geq r\}$. It is well-known (see, for example, Myerson (1981)) that for the utilities (1) and (2) the unique mechanism that implements the assortative allocation and gives type $x = 0$ utility 0 specifies efforts

$$t^A(x) = c^{-1}\left(xh(y^A(x)) - \int_0^x h\left(y^A(z)\right) dz\right). \tag{3}$$

This implies that for utility (1) type x obtains utility

$$U(x) = xh(y^A(x)) - c(t^A(x)) = \int_0^x h\left(y^A(z)\right) dz \tag{4}$$

in this mechanism, and the same expression divided by x for utility (2).

Theorem 1 below, which restates Corollary 2 in Olszewski and Siegel (2016a) and their discussion in Section 2.3, shows that the equilibria of a contest with many players and prizes can be approximated by the unique mechanism that implements the assortative allocation. The larger the contest, the better the approximation.

Theorem 1 *For any $\varepsilon > 0$, if the number n of players and prizes is sufficiently large, then in any equilibrium*

[16] Formallyis, for any $x_1 < x_2$, $t_1 < t_2$, and $y_1 < y_2$ we have that $x_1h(y_2) - c(t_2) \geq x_1h(y_1) - c(t_1)$ implies $x_2h(y_2) - c(t_2) > x_2h(y_1) - c(t_1)$.

[17] The assumption can be generalized to requiring that, for larger and larger contests, F assign a vanishing probability to each type. This holds, for example, for complete-information contests in which player i's type is i/n.

each of a fraction of at least $1-\varepsilon$ of the players i obtains with probability at least $1-\varepsilon$ a prize that differs by at most ε from $y^A(x_i)$, and with probability at least $1-\varepsilon$ chooses an effort level that is within ε of $t^A(x_i)$.

Theorem 1 and related results below imply that to study equilibrium behavior in large contests it suffices to study the corresponding mechanism that implements the assortative allocation. This is what we will do in our analysis of Pareto improving measures.

4 Top pooling

We now consider how the information provided to colleges about students' performance affects the equilibrium outcome and students' welfare. One possibility, which we dub "top pooling," is to pool together a fraction of the highest performing students by giving them all the same grade. The top ranking students still obtain the highest prizes, but the allocation of these prizes to the students is random. That is, if the top performing fraction $q = L/n$ of the students are pooled together, then each of them obtains one of the prizes $y_{n-L+1}, \ldots, y_n$, where $y_1 \leq y_2 \leq \cdots \leq y_n$, each with probability $1/L$. Relative to the original contest, this leads to better prizes for some students and worse prizes for other students. In addition, top pooling reduces the incentive of high-ability students to exert wasteful effort, because it limits the amount of separation they can achieve from other students by putting in high effort. We will see that together the two effects lead to a Pareto improvement in some large contests. We will also see that in large contests the effect of top pooling is similar to that of imposing an effort cap, which limits the effort students can put in (by, say, limiting the number of AP classes a student is allowed to take).

4.1 The corresponding approximation result for contests with top pooling

To analyze the effect of top pooling in large contests, we need an approximation result that corresponds to Theorem 1. To this end, notice that a contest that pools together the $q = L/n$ fraction of the top performing students ("top q pooling") is strategically equivalent to a contest in which the fraction q of the highest prizes are replaced with identical prizes $y(q)$, where $y(q)$ satisfies

$$h(y(q)) = \frac{1}{L}h(y_{n-L+1}) + \cdots + \frac{1}{L}h(y_n).$$

Since function h is continuous and strictly increasing, $y(q)$ is well defined and unique. Denote by G^q the empirical distribution of the modified set of prizes $y_1, \ldots, y_{n-L}, \underbrace{y(q), \ldots, y(q)}_{L}$.

The assortative allocation (for top pooling) assigns to each type x prize

$$y^{A,q}(x) = (G^q)^{-1}(F(x)). \tag{5}$$

The unique incentive-compatible mechanism that implements this allocation and gives type $x = 0$ a utility of 0 specifies efforts

$$t^{A,q}(x) = c^{-1}\left(xh(y^{A,q}(x)) - \int_0^x h\left(y^{A,q}(z)\right)dz\right). \tag{6}$$

Consider how this mechanism compares with the one in Section 3, which is given by y^A and t^A. Let $x^* = F^{-1}(1-q)$ be the type whose percentile location in the type distribution is where pooling starts. By definition of G, we have that

$$h(y(q)) = \frac{\int_{x^*}^1 h(y^A(z))dF(z)}{1-F(x^*)}.$$

By definition of G^q, we have that $(G^q)^{-1}(F(x)) = G^{-1}(F(x))$ for $x \le x^*$ and $(G^q)^{-1}(F(x)) = y(q)$ for $x > x^*$. Therefore, $y^{A,q}(x) = y^A(x)$ and $t^{A,q}(x) = t^A(x)$ for $x \le x^*$, and $y^{A,q}(x) = y(q)$ and $t^{A,q}(x) = M$ for $x > x^*$, where

$$M = c^{-1}\left(x^* h(y(q)) - \int_0^{x^*} h\left(y^A(z)\right)dz\right) = c^{-1}\left(x^* \frac{\int_{x^*}^1 h(y^A(z))dF(z)}{1-F(x^*)} - \int_0^{x^*} h\left(y^A(z)\right)dz\right). \tag{7}$$

To gain some intuition for the value of M, notice that (4) and (7) imply that

$$x^* h(y^A(x^*)) - c(t^A(x^*)) = x^* \frac{\int_{x^*}^1 h(y^A(x))dF(x)}{1-F(x^*)} - c(M), \tag{8}$$

that is, M is such that type x^* is indifferent between choosing effort $t^A(x^*)$ and obtaining prize $y^A(x^*)$ and choosing effort M and obtaining a prize randomly from the mass $1-F(x^*)$ of the highest prizes. (note that $t^A(x^*) < M$.) The following result is an immediate corollary of Theorem 1.

Corollary 1 *For any $\varepsilon > 0$, if the number n of players and prizes is sufficiently large, then in any equilibrium of the contest with top q pooling:*

(a) each of a fraction of at least $1-\varepsilon$ of the players i with $x_i < x^*$ with probability at least $1-\varepsilon$ chooses effort within ε of $t^A(x_i)$, and obtains with probability at least $1-\varepsilon$ a prize that differs by at most ε from $y^A(x_i)$;

(b) each of a fraction of at least $1-\varepsilon$ of the players i with $x_i > x^*$ with probability at least $1-\varepsilon$ chooses effort within ε of M and obtains with probability at least $1-\varepsilon$ the prize $y^A(x)$ for a randomly chosen $x > x^*$ distributed according to F contingent on $x > x^*$.

Corollary (1) shows that type x^* is a threshold type, above which pooling occurs in large contests: all higher types choose the same effort M and obtain the same lottery over prizes. Since there is a one-to-one correspondence between q and x^*, in what follows we also refer to top q pooling as "top pooling with threshold x^*."

Before turning to the welfare effects of top pooling, let us see the connection between top pooling and imposing an effort cap. Consider a contest in which no player is allowed to choose an effort higher than some $M > 0$. If the contest is large, the intuition given in Section 3 indicates that a player's chosen effort weakly increases in her type. Thus, in the approximating mechanism types up to some type x^* choose effort according to (3), just like they do without a cap. Types higher than x^* choose effort M. These types are awarded a prize randomly from the mass $1-F(x^*)$ of the highest prizes. Type x^* is indifferent between obtaining prize $y^A(x^*)$ by choosing effort $t^A(x^*)$ and obtaining the aforementioned random prize, which is higher than $y^A(x^*)$, by choosing effort M. That is, x^* and M satisfy (8). Olszewski and Siegel (2016b) show that this mechanism indeed approximates the equilibrium outcome in large contests when the bid cap M is imposed. Thus, as contest grows large, the effects of imposing a cap M and top pooling with threshold x^* become similar and coincide in the limit.

4.2 Welfare comparisons

Theorem 1 Corollary 1 show that in large contests the effect of top pooling with threshold x^* on the efforts of players with types $x < x^*$ and the prizes they obtain is negligible. Thus, to understand the welfare effect of top pooling in large contests, we compare the effort and prizes of types $x > x^*$ in the approximating mechanisms with and without top pooling. These are, respectively, given by $t^A(\cdot)$ and $y^A(\cdot)$ and by M and $y^A(z)$ for a randomly chosen $z > x^*$. The following results and examples concern this comparison. We use the term "Pareto-improving" in reference to the utility of the types in the approximating mechanisms. Such an improvement implies that in sufficiently large contests some players are strictly better off and no player is worse off by more than an arbitrarily small amount; moreover, the sum of these small amounts across all players who are worse off is arbitrarily small compared to the gains of the players who are strictly better off.

Proposition 1 *Consider top pooling with threshold x^*.*

(a) The utility of types $x < x^*$ is not affected.

(b) The utility of type $x > x^*$ increases if and only if

$$\frac{\int_{x^*}^{1} h(y^A(z))dF(z)}{1-F(x^*)} \geq \frac{\int_{x^*}^{x} h\left(y^A(z)\right)dz}{x-x^*}. \tag{9}$$

(c) The gain in utility for types $x > x^*$ first increases and then decreases in type. Thus, there is a type $x^{**} \geq x^*$ such that the utility of types x in (x^*, x^{**}) strictly increases, and the utility of types $x > x^{**}$ strictly decreases.

(d) Top pooling is Pareto improving if and only if it increases the utility of type 1, that is,

$$\frac{\int_{x^*}^{1} h(y^A(z))dF(z)}{1-F(x^*)} \geq \frac{\int_{x^*}^{1} h\left(y^A(z)\right)dz}{1-x^*}. \tag{10}$$

Proposition 1 shows that the effect of top pooling on players' welfare depends on their types. Players with low types (less than x^*) are unaffected, since their effort does not change and is lower than the effort of the other players. Consequently, the prize they obtain also does not change. Players with intermediate types (in (x^*, x^{**})) benefit, but the reason for this varies across the players. Players with intermediate types close to x^* exert more effort with top pooling (since $t^A(x^*) < M$), and obtain a prize lottery that is better than the prize they obtain without top pooling (because prize $y^A(x^*)$, which is in percentile q of the prize distribution, is the lower bound of the support of the prize lottery). If type x^{**} is close to 1, then intermediate types close to x^{**} benefit even though they obtain a prize lottery that is worse than the prize they obtain without top pooling (because prize $y^A(1)$ is the best prize); they benefit because they exert less effort with top pooling than without it (since $M < t^A(1)$). This tradeoff goes in the other direction for players with high types (above x^{**}), so they are made worse off. The tradeoff is captured by (9), which compares the gain in utility of type $x > x^*$ relative to type x^* with and without top pooling. Note that the left-hand side of (9) is the average of $h(y^A(z))$ across types z that choose effort M. This term is independent of x. The right-hand side of (9) is the average of $h(y^A(z))$ across all types lower than x that choose effort M. In addition, the average in the first term is taken with respect to the actual (truncated) distribution of types, while the average in the second term is taken with respect to the (truncated) uniform distribution.

Proposition 1 shows that types slightly higher than x^* benefit from top pooling, but high types may or may not benefit. This depends on whether type 1 benefits, in which case all types higher than x^* do. The two possibilities are depicted in Figure 1, which illustrates the utility gain resulting from top pooling as a function of type. The left-hand side corresponds to top pooling with $x^{**} < 1$, so it is not Pareto improving, and the right-hand side corresponds to top pooling that is Pareto-improving, so $x^{**} = 1$.

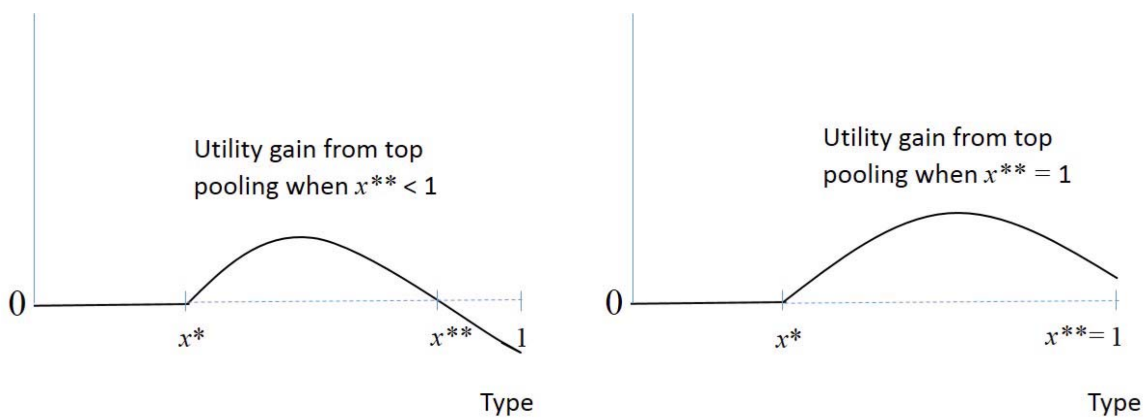


Figure 1: Utility gain from top pooling

Remark 1 *Some previously existing literature compares contests and lotteries from the perspective of contestants' welfare (see Chakravarty and Kaplan (2013), Koh et al. (2006), and Taylor et al. (2003)). Since a fair lottery can be thought of as a contest that pools all players ($x^* = 0$), part (d) of Proposition 1 shows that as n grows large, a lottery Pareto dominates a contest when*

$$\int_0^1 h(y^A(z))dF(z) \geq \int_0^1 h(y^A(z))dz.$$

This happens, for example, when the distribution F first-order stochastically dominates the uniform distribution.

The following example illustrates the results from Proposition 1.

Example 1 *Suppose that F and G are uniform, and $h(y) = y$ and $c(t) = t$. Then $y^A(x) = x$ and $t^A(x) = \frac{1}{2}x^2$, and the payoff of type x, for utility (1), is $\frac{1}{2}x^2$.*

With top pooling with threshold x^*, every type $x < x^*$ exerts effort $\frac{1}{2}x^2$ and obtains prize x, and every type $x > x^*$ exerts effort M and obtains a prize from $[x^*, 1]$, all with equal probability. For utility (1), the effort M is given by

$$\frac{1}{2}(x^*)^2 = x^* \frac{1 + x^*}{2} - M,$$

so $M = x^*/2$. Thus, the payoffs are $\frac{1}{2}x^2$ for $x < x^*$ and

$$x \frac{1 + x^*}{2} - \frac{x^*}{2} \geq \frac{1}{2}x^2$$

for $x > x^*$. All top pooling thresholds are Pareto-improving.[18]

Example 1 shows that there may exist multiple Pareto-improving pooling thresholds. It is therefore reasonable to ask whether these thresholds can be Pareto ranked. In Example 1, the derivative of the payoff of every type $x > x^*$ with respect to the threshold type x^* is $(x - 1)/2$, so all types prefer $x^* = 0$, i.e., the Pareto preferred top pooling leads to a lottery. In general, however, Pareto-improving pooling thresholds are not Pareto ranked, as the following example shows.

Example 2 *Let $F = G$ have density $f = g = 7/4$ on intervals $[0, 1/4]$ and $[3/4, 1]$. On interval $(1/4, 3/4)$, let $f = g = 1/4$. Consider utility (1) with $h(y) = y$ and $c(t) = t$.*[19] *Then $y^A(x) = x$ and $t^A(x) = \frac{1}{2}x^2$, and the payoff of type x is $\frac{1}{2}x^2$.*

Top pooling with threshold $x^* = 0$, which is a pure lottery, is Pareto improving, since the expected prize of $1/2$ at no effort exceeds utility $\frac{1}{2}x^2$ for all x. Now consider top pooling with threshold $x^* = 1/2$. For this threshold we have $M = 19/64$, since

$$\frac{1}{2}\left(\frac{1}{2}\right)^2 = \frac{1}{2}\left(\frac{1}{8}\frac{5}{8} + \frac{7}{8}\frac{7}{8}\right) - \frac{19}{64}.$$

Type $x = 1$ benefits from this top pooling, since $\left(\frac{1}{8}\frac{5}{8} + \frac{7}{8}\frac{7}{8}\right) - \frac{19}{64} > \frac{1}{2}$. It is therefore Pareto improving, by part (d) of Proposition 1. Type $x = 1$ type (as well as types close to it) also prefer this top pooling to a pure lottery. However, types in the interval $(0, 1/2)$ have the opposite preference, because a pure lottery gives each of them an expected utility of $\frac{1}{2}x$, and top pooling with threshold $x^* = 1/2$ gives each of them an expected utility of $\frac{1}{2}x^2$.

The following result clarifies when Pareto-improving pooling thresholds are Pareto-ranked.

Proposition 2 *Suppose that $x_1^* < x_2^*$ are Pareto-improving pooling thresholds.*

(a) If type $x = 1$ weakly prefers x_1^* to x_2^*, then types x in $(x_1^*, 1)$ strictly prefer x_1^* to x_2^*, so x_1^* is Pareto preferred to x_2^*.

---

[18] The same results hold for utility (2).

[19] The same results hold for utility (2).

(b) If type $x = 1$ strictly prefers x_2^* to x_1^*, then x_1^* and x_2^* are not Pareto ranked. There is an $x^{**} \in (x_2^*, 1)$ such that types x in (x_1^*, x^{**}) strictly prefer x_1^*, and types $x > x^{**}$ strictly prefer x_2^*.

Figure 2 illustrates the two parts of Proposition 2. The left-hand side corresponds to part (a), and the right-hand side corresponds to part (b).

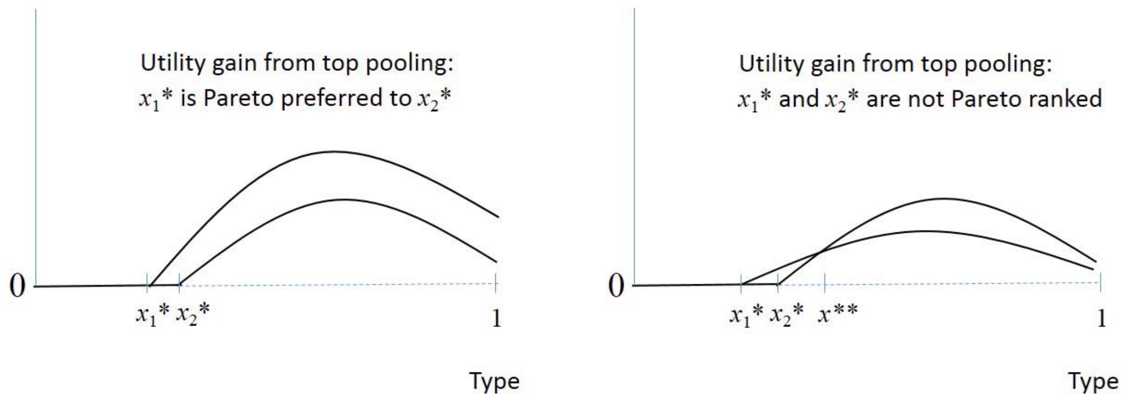


Figure 2: Comparing the utility gain from pooling thresholds $x_1^* < x_2^*$

Proposition 2 leads to a simple description of the Pareto frontier of pooling thresholds. To see this, consider the function φ that assigns to any threshold x^* the utility of type $x = 1$ in the approximating mechanism with this threshold. Denote by ϕ the lowest monotone function that is pointwise weakly higher than φ. Then, the Pareto frontier consists of all the thresholds M at which $\varphi(x^*) = \phi(x^*)$ (see Figure 3).

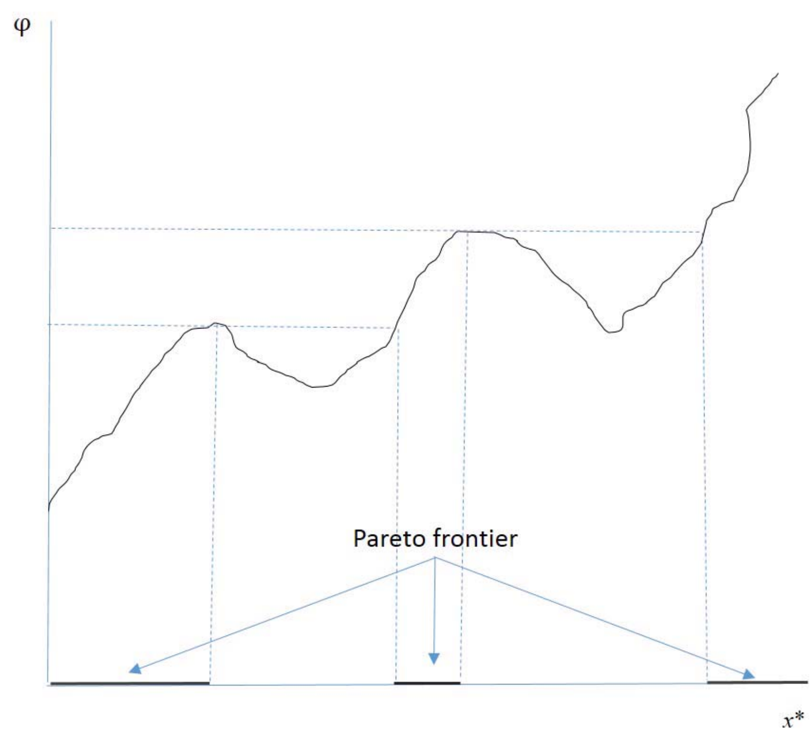


Figure 3: Pareto frontier of pooling thresholds

When we consider only a finite set of pooling thresholds X^*, Proposition 2 implies that the frontier of X^* consists of the threshold x^* that is most preferred by type 1 among all the thresholds X^*, the threshold that is most preferred by type 1 among the thresholds in X^* that are lower than x^*, and so on.

5 Category Rankings

We now consider more general performance disclosure policies, which pool students whose performance lies within some interval of performance ranking. Different policies provide colleges with different information about students' performance. For example, coarser grading categories (e.g., A-F) provide less information than finer ones (e.g., 0-100). We will see how different policies affect the equilibrium outcome and students' welfare, and identify Pareto improving ones. We will use the term "category rankings" to describe such policies.

A category ranking is a monotone partition of the players according to the ranking of their effort. One example is partitioning them above and below the median effort. Another example is partitioning them according to whether their effort is below the 10-th percentile, between the 10-th percentile and the 20-th percentile, etc. Partitioning the players in this way conveys coarser or finer information about their effort and resulting rank order. A category ranking induces a partition of the set of prizes, and the prizes within each element of the partition are randomly assigned to the players in the corresponding element of the category ranking.

Formally, a category ranking is a monotone partition J of the set $[0,1]$ into singletons and left-open intervals. The intervals are $J_k = \left(q_k^l, q_k^h\right]$ for $1 \leq k \leq K \leq n$, where $0 \leq q_1^l < q_1^h < \cdots < q_K^l < q_K^h \leq 1$ are fractions with denominator n. The interpretation is that the $\left(q_k^h - q_k^l\right) n$ players whose effort's ordinal percentile rank lies in J_k are grouped together (any rule can be used to break ties in the ranking of two or more players who choose the same effort). Only the element of the partition to which a player belongs is revealed; in particular, players' precise efforts and their ranking within an interval partition element are not revealed. Prizes are assigned in decreasing value to the partition elements, and distributed according to a fair lottery among the players in each partition element. That is, each of the players in partition element J_k obtains one of the prizes $y_{nq_k^l+1}, \ldots, y_{nq_k^h}$ with equal probability.

5.1 The approximation result for contests with category rankings

To analyze the effect of category rankings, we need an approximation result that corresponds to Theorem 1. To this end, denote by $y(J_k)$ the certainty equivalent prize of the prize lottery for interval J_k, that is,

$$h(y(J_k)) = \frac{1}{\left(q_k^h - q_k^l\right) n} h(y_{nq_k^l+1}) + \cdots + \frac{1}{\left(q_k^h - q_k^l\right) n} h(y_{nq_k^h}). \tag{11}$$

Since function h is continuous and strictly increasing, $y(J_k)$ is well defined and unique. Denote by $G^{\mathcal{J}}$ the empirical distribution of the certainty equivalents induced by the partition J.

The assortative allocation (for category rankings) assigns to each type x prize

$$y^{A,\mathcal{J}}(x) = \left(G^{\mathcal{J}}\right)^{-1}(F(x)). \tag{12}$$

The unique incentive-compatible mechanism that implements this allocation and gives type $x = 0$ a utility of 0 specifies efforts

$$t^{A,\mathcal{J}}(x) = c^{-1}\left(xh(y^{A,\mathcal{J}}(x)) - \int_0^x h\left(y^{A,\mathcal{J}}(z)\right)dz\right). \tag{13}$$

We therefore obtain the following result as an immediate corollary of Theorem 1.

Corollary 2 *For any $\varepsilon > 0$, if the number n of players and prizes is sufficiently large, then in any equilibrium of the contest with category ranking J, each of a fraction of at least $1-\varepsilon$ of the players i with probability at least $1-\varepsilon$ exerts effort within ε of $t^{A,\mathcal{J}}(x_i)$, and obtains with probability at least $1-\varepsilon$ a lottery over prizes with a certainty equivalent that differs by at most ε from $y^{A,\mathcal{J}}(x_i)$.*

Note that each type in the approximating mechanism chooses a deterministic effort, and higher types choose weakly higher efforts. Therefore, a category ranking induces a partition I of the set of types $X = [0,1]$ into singletons and K intervals $I_k = \left(F^{-1}\left(q_k^l\right), F^{-1}\left(q_k^h\right)\right]$, such that all types in interval I_k choose the same effort and obtain the same fair lottery over the prizes $y_{nq_k^l+1}, \ldots, y_{nq_k^h}$ in the approximating mechanism, and singleton types obtain the prize they did in the original approximating mechanism. The certainty equivalent $y(J_k)$, given by (11), can thus also be described by the equation

$$h(y(J_k)) = \frac{\int_a^b h(y^A(x))dF(x)}{F(b) - F(a)} \tag{14}$$

for $a = F^{-1}\left(q_k^l\right)$ and $b = F^{-1}\left(q_k^h\right)$. Thus, the assortative allocation and approximating mechanism can be equivalently defined from the partition I of types (instead of the partition J) by letting $G^{\mathcal{I}}$ (which coincides with $G^{\mathcal{J}}$) be the empirical distribution of the set of certainty equivalents, and defining $y^{A,\mathcal{I}}(x)$ and $t^{A,\mathcal{I}}(x)$ as in (12) and (13) with I instead of J.

Thus, from the perspective of the approximating mechanism, a category ranking J corresponds to a partition I of the set of types into singletons and a finite number of intervals. In what follows, it will be convenient to consider such partitions of the set of types and the corresponding approximating mechanisms.[20] We will abuse terminology slightly by also referring to such partitions I of the type interval $[0,1]$ as category rankings.

5.2 The added value of category rankings

Top pooling is a particular kind of category ranking: top pooling with threshold x^* is the category ranking $I = \{(x^*, 1]\} \cup \{\{x\} : x \leq x^*\}$. Thus, the existence of Pareto-improving top poolings implies the existence of

[20] Any such partition can clearly be approximated arbitrarily closely by category rankings when n grows large.

Pareto-improving category rankings. In fact, the richer set of outcomes that can be generated by category rankings may contain outcomes that are Pareto preferred to all outcomes that can be generated by top poolings. This is what the following example demonstrates.

Example 3 *Let $F = G$ have density $f = g = 5/4$ on the interval $[0, 3/4]$, and density $f = g = 1/4$ on the interval $[3/4, 1]$. Consider utility (1) with $h(y) = y$ and $c(t) = t$.*[21] *Then, $y^A(x) = x$ and $t^A(x) = \frac{1}{2}x^2$, and the payoff of type x is $\frac{1}{2}x^2$.*

Top pooling with threshold $x^* = 3/4$ is Pareto improving. Indeed, the corresponding M is given by

$$\frac{1}{2}\left(\frac{3}{4}\right)^2 = x^*\left(\frac{1+\frac{3}{4}}{2}\right) - M,$$

which gives $M = 3/8$. Types in $(3/4, 1]$ choose effort M, and each of them obtains a prize randomly chosen from the interval $(3/4, 1]$. The utility of type $x = 1$ is equal to $1/2$ both with and without top pooling. So, by part (d) of Proposition 1, top pooling with threshold $x^* = 3/4$ is Pareto improving. One can readily check that top pooling with any threshold $x^* > 3/4$ is also Pareto improving and gives type $x = 1$ utility $1/2$. We will show that top pooling with any threshold $x^* < 3/4$ is not Pareto improving. So, by part (a) of Proposition 2, the threshold $x^* = 3/4$ is the Pareto preferred one.

To see that no threshold $x^* < 3/4$ is Pareto improving, recall that the effort $M < 3/8$ is implicitly determined by

$$\frac{1}{2}(x^*)^2 = x^* E[y \mid t = M] - M,$$

where $E[y \mid t = M]$ is the expected prize contingent on providing effort M. The utility of type $x = 1$ is thus

$$E[y \mid t = M] + \underbrace{\frac{1}{2}(x^*)^2 - x^* E[y \mid t = M]}_{-M} < \frac{1}{2},$$

because $E[y \mid t = M] < (1 + x^*)/2$ for $x^* < 3/4$.

Top pooling with threshold $x^* = 3/4$ is the category ranking that pools together the top $1/16$ of the types and leaves the other types as singletons. However, this category ranking is Pareto inferior to the category ranking that pools together the top $1/16$ of the types, and pools together the bottom $15/16$ of the types. Indeed, under this category ranking, the bottom $15/16$ of the types exert no effort and obtain an expected prize of $3/8$, while the top $1/16$ of the types exert effort $3/8$ and obtain an expected prize of $7/8$. Under the former category ranking, the top $1/16$ of the types also exert effort $3/8$ and obtain an expected prize of $7/8$, but the bottom $15/16$ of the types x obtain a lower utility of $x^2/2$.

5.3 Welfare comparisons

Consider first single-interval category rankings, that is, category rankings of the form $I = \{(x^*, x^{**}]\} \cup \{\{x\} : x \leq x^* \text{ or } x > x^{**}\}$ for some types $0 \leq x^* < x^{**} \leq 1$; top pooling is a special case in which $x^{**} = 1$. In Appendix A, we prove the following generalization of Proposition 1.

[21] The same results hold for utility (2).

Proposition 3 *(a) The utility of type $x \in (x^*, x^{**}]$ increases as a result of the category ranking $\mathcal{I}$ if and only if*

$$\frac{\int_{x^*}^{x^{**}} h(y^A(z))dF(z)}{F(x^{**}) - F(x^*)} \geq \frac{\int_{x^*}^{x} h\left(y^A\left(z\right)\right)dz}{x - x^*}.$$

(b) The category ranking $\mathcal{I}$ is Pareto improving if and only if

$$\frac{\int_{x^*}^{x^{**}} h(y^A(z))dF(z)}{F(x^{**}) - F(x^*)} \geq \frac{\int_{x^*}^{x^{**}} h\left(y^A\left(z\right)\right)dz}{x^{**} - x^*}. \tag{15}$$

The intuition for Proposition 3 is similar to the one underlying Proposition 1, applied to types in the interval $(x^*, x^{**}]$. In particular, if a type $x \in (x^*, x^{**}]$ benefits from the category ranking, then all types in the interval $(x^*, x]$ benefit as well. Types $x \leq x^*$ are clearly not affected by the category ranking, and the derivative of the utility of types $x > x^{**}$ is equal to $h\left(y^A\left(x\right)\right)$ both in the original contest and under the category ranking. Thus, if type x^{**} is better off under the category ranking, then so are all types higher than x^{**}, which gives part (b).

Using Proposition 3, we can provide a method for checking whether a category ranking belongs to the Pareto frontier of category rankings. For this, we will need another concept. Let $\mathcal{I}$ be a category ranking, and let $x^* < x^{**}$ be an arbitrary pair of types that belong to two different elements $I \neq I'$ (intervals or singletons) of $\mathcal{I}$, so $x^* \in I \in \mathcal{I}$ and $x^{**} \in I' \in \mathcal{I}$. We define a new category ranking $\mathcal{I}(x^*, x^{**})$ that groups all types between x^* and x^{**} into one category as follows: (i) if $I = (a, b]$ and $I' = (a', b']$, replace I, I', and all elements of $\mathcal{I}$ between I and I' with $(a, x^*]$, $(x^*, x^{**}]$, and $(x^{**}, b']$; (ii) if $I = \{x^*\}$ and $I' = (a', b']$, replace I' and all elements of $\mathcal{I}$ between I and I' with $(x^*, x^{**}]$, and $(x^{**}, b']$; (iii) if $I = (a, b]$ and $I' = \{x^{**}\}$, replace I, I', and all elements of $\mathcal{I}$ between I and I' with $(a, x^*]$ and $(x^*, x^{**}]$; (iv) if $I = \{x^*\}$ and $I' = \{x^{**}\}$, replace I' and all elements of $\mathcal{I}$ between I and I' with $(x^*, x^{**}]$.

Proposition 4 *A category ranking $\mathcal{I}$ belongs to the Pareto frontier of category rankings if and only if there is no pair of types $x^* < x^{**}$ such that*

$$x^* = a \text{ for some } I = (a, b] \in \mathcal{I} \text{ or } x^* = d \text{ for some } I = \{d\} \in \mathcal{I} \text{ and } x^{**} \in I' \neq I \in \mathcal{I},$$
$$\text{and type } x^{**} \text{ weakly prefers ranking } \mathcal{I}(x^*, x^{**}) \text{ to ranking } \mathcal{I}.$$

Proposition 4 is proven in Appendix A. The proposition helps in characterizing the Pareto frontier by substantially reducing the set of category rankings to which any given ranking must be compared, as the following example demonstrates.

Example 4 *Revisit Example 3. It is easy to verify that any interval that satisfies condition (15) must be contained in* $(0, 3/4]$ *or* $(3/4, 1]$*. Thus, any candidate for any Pareto-improving interval category ranking consists of an interval partition of* $(0, 3/4]$ *and an interval partition of* $(3/4, 1]$*. But by the general payoff* *formula (17) in the proof of Proposition 4, if a partition of* $(0,3/4]$ *(or a partition of* $(3/4,1]$*) comprises more than one element, then* $(x^{*},x^{**}]=(0,3/4]$ *(*$(x^{*},x^{**}]=(3/4,1]$*, respectively) violates the condition from Proposition 4. Indeed the payoffs for* x^{**} *are equal under* I *and under* $I(x^{*},x^{**})$*. Thus, the Pareto frontier has only one element, namely, the category ranking* $I=\{(0,3/4],(3/4,1]\}$.

6 Externalities

In order to focus on the competitive aspect of college admissions, we modeled effort as a costly choice variable that is used to rank students. Students' effort may have additional, external effects, both because it affects the learning environment, which affects all students, and because the level of education may affect society more generally.[22] Our analysis showed that top pooling decreases the effort of players with high types and increases the effort of some players with lower types. The aggregate effect on effort is therefore ambiguous.

When effort costs are convex, one sufficient condition for top pooling to increase aggregate effort in a large contest is that the marginal effort cost at the effort M is sufficiently high. The intuition is as follows. In the original contest, players with different types can separate by choosing different levels of effort. Around M (in the original contest), it suffices to choose slightly higher effort levels in order to separate from players with lower types, because the marginal cost of effort is high. Top pooling prevents this separation, but the slight reduction in effort from high types who now choose M is outweighed by lower types who face a lower marginal cost of increasing their effort to reach M. This and other sufficient conditions[23] can be combined with the conditions under which top pooling is Pareto improving to obtain conditions under which top pooling is socially beneficial even in the presence of effort externality.

Another form of externality, which we dub "match externality," concerns the social cost (beyond students' individual utility) of the departure of the allocation from assortative matching. Our results above show that when all information about students' performance is released, assortative matching is obtained. The associated effort is, however, costly. It may be socially beneficial to impose a category ranking that reduces the match quality along with the associated effort. To highlight this possibility, consider utility (2) with linear functions c and h and suppose that types and prizes are distribution uniformly. Suppose also that society suffers a loss $-c\,|x-y|^{q}$ if type x obtains prize y. Then, if $q>2$, then any category ranking that maximizes welfare comprises a finite number of intervals, none of them degenerate, and the length of these intervals decreases when one moves from the left to the right. The intuition is that intervals reduce competition and therefore reduce effort, which increases welfare, but decrease the match quality, which decreases welfare. The

[22] This is also true for college students, who often learn from one another, and also affects the classes they take, because classes are often adjusted to the students taking them. Similarly, the effort invested in preparing grant applications makes it easier to disseminate ideas across fields, and even within fields.

[23] Additional sufficient conditions are given by Olszewski and Siegel (2016b). They also show that with linear or concave costs effort caps (which are equivalent to top pooling in large contests) do not increase aggregate effort.

second effect is more important when players' types are higher, so the intervals shrink as types increase.[24]

7 Robust Pareto improvements

We would like to use the results of the previous sections to find Pareto improvements in the contest for college admissions. One difficulty is that the results and implied improvements involve expressions that include the type distribution F, the prize valuation h, and the prize distribution G (via the assortative allocation y^A). Since obtaining good estimates for all of these functions may be difficult or impossible, we would like to derive similar results that are independent of the values of at least some of them. We now present such results, which rely only on properties of the type distribution F and correspond to Propositions 1, 2, 3, and 4. We use the term "robust" as shorthand for "for any functions c, h, and G."

The first result characterizes robust Pareto improving top pooling thresholds. It follows immediately from part (4) of Proposition 1 and the definition of first-order stochastic dominance (FOSD).

Corollary 3 *Type x^* is a robust Pareto improving top pooling threshold if and only if the distribution F truncated below type x^* FOSD the uniform distribution truncated below this x^*.*

Consider two robust Pareto improving top pooling thresholds $x_1^* < x_2^*$. By definition of top pooling, the effect of top pooling with threshold x_2^* on the approximating mechanism is identical to the effect of, instead of top pooling, changing the prize distribution on $\left[y^A\left(x_2^*\right), y^A\left(1\right)\right]$ from what is specified by G to a mass $1 - F\left(x_2^*\right)$ of the certainty equivalent of awarding the prizes in $\left[y^A\left(x_2^*\right), y^A\left(1\right)\right]$ randomly. And because Corollary 3 holds for any prize distribution, top pooling with threshold x_1^* leads to a further robust Pareto improvement. This proves the following result.

Corollary 4 *If $x_1^* < x_2^*$ are robust Pareto improving top pooling thresholds, then top pooling with threshold x_1^* is robust Pareto preferred to top pooling with threshold x_2^*. Thus, the Pareto frontier of robust Pareto improving top pooling thresholds is a singleton, which is the lowest robust Pareto improving top pooling threshold.*

Corollary 4 explains why in Example 1 lower top pooling thresholds are Pareto preferred to higher ones, and why a lottery is Pareto preferred to any positive top pooling threshold.[25]

The next result characterizes robust Pareto improving single-interval category rankings. It follows immediately from part (b) of Proposition 1 and the definition of FOSD.

Corollary 5 *The category ranking $I = \{(x^*, x^{**}]\} \cup \{\{x\} : x \leq x^* \text{ or } x > x^{**}\}$ is robust Pareto improving if and only if the distribution F truncated below x^* and above x^{**} first-order stochastically dominates the uniform distribution truncated below x^* and above x^{**}.*

[24] For $q = 2$ the optimal category ranking consists of an interval with lower endpoint 0 and singletons above the upper bound of the interval.

[25] Notice that Example 2 fails this condition, because F does not FOSD the uniform distribution on $[0, 1]$.

Consider two robust Pareto improving single-interval category rankings I_1 and I_2, with corresponding pooled type intervals $I_1 = [x_1^*, x_1^{**}]$ and $I_1 = [x_2^*, x_2^{**}]$. Suppose $I_1 \subseteq I_2$. The effect of I_2 on the approximating mechanism is identical to the effect of, instead of applying the category ranking, changing the prize distribution on $\left[y^A(x_2^*), y^A(x_2^{**})\right]$ from what is specified by G to a mass $F(x_2^{**}) - F(x_2^*)$ of the certainty equivalent of awarding the prizes in $\left[y^A(x_2^*), y^A(x_2^{**})\right]$ randomly. And because Corollary 5 holds for any prize distribution, applying the category ranking I_2 leads to a further robust Pareto improvement. Thus, I_2 is Pareto preferred to I_1.

Similarly, if I_1 and I_2 are disjoint, then the two-interval category ranking with pooled type intervals I_1 and I_2 is Pareto preferred to I_1 and I_2. Finally, suppose that I_1 and I_2 intersect. In this case, the following lemma and our result for the case $I_1 \subseteq I_2$ imply that the single-interval category ranking with pooled type interval $I_1 \cup I_2$ is Pareto preferred to I_1 and I_2.

Lemma 1 *Consider two intervals $I_1 = (x_1, x_3]$ and $I_2 = (x_2, x_4]$ for $0 \leq x_1 < x_2 < x_3 < x_4 \leq 1$. If for each interval F restricted to the interval FOSD the uniform distribution restricted to the interval, then F restricted to the union of the intervals $I_1 \cup I_2$ FOSD the uniform distribution restricted to $I_1 \cup I_2$.*

These observations lead to the following result.

Corollary 6 *The Pareto frontier of robust Pareto improving category rankings is a singleton I^{PF}, which consists of the maximal intervals on which F restricted to each interval FOSD the uniform distribution restricted to the interval, along with singletons for all other types.*

The robust improvements described by the results in this section rely only on properties of the type distribution F, a feature that we will use in Section 8. The results also show that the comparison between robust improvements and the structure of the Pareto frontier of robust improvements are much simpler that those of improvements that are not robust. This also facilitates the analysis in Section 8.

8 Application to college admissions

Our theoretical results can help us understand how to modify the contest for college admissions in a way that benefits all students. The following observations and assumption will allow us to use the robust Pareto improvements of Section 7. First, note that students compete for admissions regionally, and sometimes nationally or internationally, and colleges receive a large number of applications.[26] This allows us to view the competition for college admissions as a large contest. College seats are the prizes in this contest, which implicitly assumes that students have the same ordinal ranking over colleges. While in reality these rankings

[26] For example, in 2015-2016 Berkeley received close to 80,000 applications and admitted more than 13,000 students, and Northwestern received more than 33,000 applications and admitted 13% of them. Overall, there were approximately 8.2 mln freshman applications in 2011.

may differ across students, especially for lower ranked colleges, abstracting from these differences may be a reasonable simplification, at least for students in the same geographical area. Colleges rank students according to their performance, so students with a better performance are admitted to better colleges. Second, since the contest is large, our results above show that students with higher ability choose higher effort or, equivalently, higher performance.[27] Thus, students with higher ability are admitted to better colleges. This still leaves a great deal of flexibility in the performance distribution. We assume that the performance distribution reflects the student population ability distribution. This is consistent with "curving," which is a common practice. Thus, the location of each student's performance in the performance distribution corresponds to the location of his ability in the population ability distribution.

This last observation implies that we can apply Corollaries 3 and 5, and especially Corollary 6 to students' grade distributions. When the conditions in the corollaries hold, coarsening the performance disclosure policy, that is, pooling together students with different performance levels (and therefore different abilities), is Pareto improving. Such coarsenings reduces the amount of information revealed to colleges. This reduces the assortativity of the matching between students and colleges, but also leads to a corresponding change in students' efforts, which results in an overall Pareto improvement.

We obtain grade distributions from the grade data in the National Longitudinal Survey of Youth 1997 (NLSY97), which in turn is part of the National Longitudinal Surveys (NLS) program. This data set is based on annual interviews conducted over a sixteen-year period, beginning in 1997, with a cohort of individuals born between 1980 and 1984. The cohort whose data we consider consists of a cross-sectional sample of 6748 respondents designed to be representative of people living in the United States during the initial survey round. We use the survey data regarding education, training, and achievement scores, which includes SAT, ACT, and PSAT scores, high school grades, and AP credits. We exclude incomplete transcripts, which leaves 800-1200 observations for each type of test and approximately 4500 GPA and academic credit observations from high-school years.

The Data Appendix depicts three figures for each performance measure. The first is a histogram of the scores/grades. The second is a plot of the upper and lower endpoints of every interval on which the score/grade distribution restricted to the interval FOSD the uniform distribution restricted to the interval. The third depicts the maximal such intervals as colored regions. The figures show that for each performance measure the maximal intervals cover most or all of the range of scores/grades. This shows that there is a large scope for improvement by pooling together different grades, as by Corollary 6 the union of the maximal intervals forms the optimal category ranking among the robust Pareto improving ones.

Comparing these optimal category rankings across the different performance measures reveals similarities and differences. One feature common to all the performance measures is that pooling a substantial score/grade interval with lower endpoint 0 is Pareto improving. That is, students with abilities in the lower part of the ability distribution would be better off not investing effort and obtaining entry to a random col-

[27] Ability can affect the marginal cost of effort, marginal valuations, or a combination of the two, as discussed in Section 2.

lege among those to which they are (in the aggregate) currently admitted than investing effort and obtaining college admissions as they currently do. A key difference between the performance measures is the upper endpoint of this pooling interval. For SAT, ACT, and PSAT the upper endpoint is substantially lower than the highest possible score, and for math, overall GPA, and life sciences the upper endpoint is lower than 4.0. Additional pooling intervals, corresponding to top pooling and intermediate intervals, are part of the optimal category ranking. In contrast, for English, foreign language, and social sciences, the upper endpoint is at least 4.0. That is, it is optimal to pool together almost all students, with some minor pooling of the top performing students. Note, however, that this analysis ignores positive externalities, as discussed in Section 6. In practice, we would advocate a pass/fail policy for the latter high school topics, with the possibility of commending exceptional performance (which would correspond to grades higher than 4.0 today). This would incentivize students who are not at the very top of the ability distribution to invest some effort and acquire some language skills and learn some social sciences, which may have a positive societal influence.

8.1 AP Credits

The robust Pareto improvements we identified apply to any valuation function h and distribution G of prize. When estimates for these variables exist, the results of Sections 4 and 5 can be used to identify Pareto improvements and their Pareto frontier. We now do this for top poolings by considering the data regarding students' AP credits and the salary they earned after graduation. We consider AP credits, because those students who take AP classes tend to be higher ability students, so the omitted variable bias that plagues any "return to education" estimate is likely to be less severe than when estimating the affect of other performance measures on salary. The distribution of AP credits and salaries, a scatter plot relating the two, and the correlation between them are depicted in the "Correlation" figure in the Data Appendix. We consider the per-person average salary (across the available years post college graduation), assume that students' utility h is linear in the salary they earn, and estimate a linear relationship between students' ability (taken as the number of AP credits they earned) and the salary they earned, i.e., between x and $h\left(y^A\left(x\right)\right)$. This relationship is depicted in the "Tobit estimation" figure in the Data Appendix.[28] The estimation shows that an extra credit is associated with an increase of almost a thousand dollars in income. We then use (10) to determine the Pareto improving top pooling thresholds, and use the procedure described at the end of Section 4 to determine the Pareto frontier.[29] These are depicted in the final figure in the Data Appendix. The Pareto frontier consists of four elements, which correspond to pooling students with at least 26, 25, 20, or 19 AP credits.

[28] This required a Tobit regression, since the top 2 percent of salaries are top coded and replaced in the data by their average.

[29] Identify the top pooling threshold that type 1 prefers most, the threshold type 1 prefers most among the remaining thresholds, etc.

A Proofs

Proof of Proposition Part (a) follows because with top pooling types $x < x^*$ choose effort $t^A(x)$ and obtain prize $y^A(x)$. For part (b), note that the utility of type x^* is the same in the approximating mechanisms of the original contest and in the one with top pooling. Consider first the utility of a type $x > x^*$ in the approximating mechanism of the original contest. By (4), this utility exceeds that of type x^* by

$$\int_{x^*}^{x} h\left(y^A(z)\right) dz.$$

In the approximating mechanism with top pooling, the utility of type x exceeds that of type x^* by

$$(x - x^*)\frac{\int_{x^*}^{1} h(y^A(z))dF(z)}{1 - F(x^*)},$$

since both types bid M and obtain a prize randomly from the mass $1 - F(x^*)$ of the highest prizes. Thus, top pooling increases the utility of type x if and only if (9) holds.

For part (c), note that the derivative with respect to x of the utility gain of type x is

$$\frac{\int_{x^*}^{1} h(y^A(z))dF(z)}{1 - F(x^*)} - h\left(y^A(x)\right). \tag{16}$$

The fraction in (16) is a weighted average of $h\left(y^A(z)\right)$ over types in $[x^*, 1]$, so (16) is positive for types x close to x^*, monotonically decreases as x increases, and becomes negative for types x close to 1. Thus, the utility gain resulting from top pooling for types $x > x^*$ first increases and then decreases in the type. In particular, the utility of all types $x < 1$ strictly increases if the utility of type 1 weakly increases, which gives part (d).

Proof of Proposition 2. Recall that given a threshold x^*, the utility with top pooling of type $x > x^*$ is

$$U^{x^*}(x) = (x - x^*)\frac{\int_{x^*}^{1} h(y^A(z))dF(z)}{1 - F(x^*)} + \int_{0}^{x^*} h\left(y^A(z)\right) dz.$$

This means that $U^{x^*}(x)$ can be represented as $x\phi(x^*) + \psi(x^*)$, where

$$\phi(x^*) = \frac{\int_{x^*}^{1} h(y^A(z))dF(z)}{1 - F(x^*)}$$

and

$$\psi(x^*) = \int_{0}^{x^*} h\left(y^A(z)\right) dz - x^*\frac{\int_{x^*}^{1} h(y^A(z))dF(z)}{1 - F(x^*)}.$$

Notice that $\phi(x^*)$ increases in x^*, because as x^* increases $\phi(x^*)$ becomes an average over higher values of $h(y^A(z))$.

Thus, $(x-1)\phi(x_1^*) > (x-1)\phi(x_2^*)$ for all $x < 1$, and if $\phi(x_1^*) + \psi(x_1^*) \geq \phi(x_2^*) + \psi(x_2^*)$, then

$$x\phi(x_1^*) + \psi(x_1^*) = (x-1)\phi(x_1^*) + \phi(x_1^*) + \psi(x_1^*) >$$

$$> (x-1)\phi(x_2^*) + \phi(x_2^*) + \psi(x_2^*) = x\phi(x_2^*) + \psi(x_2^*),$$

which yields part (a) for types $x > x_2^*$. Lower types higher than x_1^* also strictly prefer pooling with threshold x_1^* to x_2^*, since they strictly prefer pooling with threshold x_1^* to no pooling (part (c) of Proposition 1 with

$x^{**}=1$), and their utility with threshold x_2^* is equal to their utility with no pooling (part (a) of Proposition 1). Types lower than the threshold x_1^* obtain the same prize and provide the same effort with both thresholds.

Suppose now that type 1 strictly prefers x_2^* to x_1^*. Observe first that there exists an $x^{**} > x_2^*$ such that types x in (x_2^*, x^{**}) strictly prefer x_1^* and types $x > x^{**}$ strictly prefer x_2^*. Indeed, since $\phi(x_1^*) < \phi(x_2^*)$, this follows from the fact that

$$[x\phi(x_2^*) + \psi(x_2^*)] - [x\phi(x_1^*) + \psi(x_1^*)] = x[\phi(x_2^*) - \phi(x_1^*)] + [\psi(x_2^*) - \psi(x_1^*)]$$

strictly increases in x, and type x_2^* strictly prefers x_1^* to x_2^*. This observation about type x_2^*'s preferences follows from the fact that x_1^* is Pareto improving, part (c) of Proposition 1 applied to x_1^*, and the fact that type x_2^* is indifferent between pooling with threshold x_2^* and no pooling. Since all types close to 1 strictly prefer x_2^* to x_1^*, we have that $x^{**} < 1$, and because type x_2^* strictly prefers x_1^* to x_2^*, we have that $x^{**} > x_2^*$. Types in (x_1^*, x_2^*) strictly prefer x_1^* to x_2^* for the same reason that type x_2^* does, and types lower than x_1^* are indifferent between x_1^*, x_2^*, and no pooling.

Proof of Proposition 3. Let $r^* = F(x^*)$ and $r^{**} = F(x^{**})$. Then, any type $x < x^*$ provides effort $t(x) = t^A(x)$ and obtains prize $y(x) = y^A(x)$. Types in $(x^*, x^{**}]$ provide a certain effort t and obtain a fair lottery over prizes $y \in (G^{-1}(r^*), G^{-1}(r^{**})]$. Since type x^* is indifferent between the two options, we have that

$$x^* h(y^A(x^*)) - c(t^A(x^*)) = \int_0^{x^*} h\left(y^A(z)\right) dz = x^* \frac{\int_{x^*}^{x^{**}} h(y^A(z)) dF(z)}{F(x^{**}) - F(x^*)} - c(t).$$

By (4) we have that

$$U^{\mathcal{I}}(x) - U(x) = (x - x^*) \frac{\int_{x^*}^{x^{**}} h(y^A(z)) dF(z)}{F(x^{**}) - F(x^*)} - \int_{x^*}^{x} h\left(y^A(z)\right) dz = (x - x^*) \left[\frac{\int_{x^*}^{x^{**}} h(y^A(z)) dF(z)}{F(x^{**}) - F(x^*)} - \frac{\int_{x^*}^{x} h\left(y^A(z)\right) dz}{x - x^*} \right].$$

Thus, (15) is a necessary and sufficient condition for types x in $(x^*, x^{**}]$ to be better off. To show that (15) is a necessary and sufficient condition for Pareto improvement, observe that types $x \le x^*$ are indifferent. Any type $x > x^{**}$ obtains prize $y^A(x)$ and has payoff

$$U^{\mathcal{I}}(x) = \int_0^{x^*} h\left(y^A(z)\right) dz + (x^{**} - x^*) \frac{\int_{x^*}^{x^{**}} h(y^A(z)) dF(z)}{F(x^{**}) - F(x^*)} + \int_{x^{**}}^{x} h\left(y^A(z)\right) dz,$$

which is no lower than $U(x)$ if and only if condition (15) is satisfied.

The last equality is obtained directly from (4) by noticing that the contest under our category ranking I can be viewed as a contest in which prizes $y^A(z)$, for z in $(x^*, x^{**}]$, are replaced with the certainty equivalents of the lottery faced by types z in $(x^*, x^{**}]$ under our category ranking I, with no additional constraints.

Proof of Proposition 4. It will be helpful to provide first a general formula for the utility of type $x \in [0,1]$ under category ranking I. This utility exceeds $U(x)$ given by (4) by the expression

$$\sum_{(a,b] \in \mathcal{I}, a < b < x} \left[(b-a) \frac{\int_a^b h(y^A(z)) dF(z)}{F(b) - F(a)} - \int_a^b h\left(y^A(z)\right) dz \right] + \tag{17}$$

$$(x-a)\frac{\int_a^b h(y^A(z))dF(z)}{F(b)-F(a)} - \int_a^x h\left(y^A(z)\right)dz \text{ for } x \in (a,b] \in \mathcal{I}.$$

This formula follows directly from the fact that types $z \in (a,b] \in I$ obtain a fair lottery over prizes $y^A(z')$ for $z' \in (a,b]$.

We will first show that when a pair $x^* < x^{**}$ satisfies the condition in Proposition 4, the category ranking $J = I(x^*, x^{**})$ Pareto improves over I. Since types $x \in [0, x^*]$ are unaffected by these category rankings, they are indifferent between them. By assumption, the utility of type x^{**} is no lower under J than under I. We will now show that the utility of types $x \in (x^*, x^{**})$ is strictly higher under J than under I. Indeed, the derivative on $(x^*, x^{**}]$ of type x's utility under J, $U^{\mathcal{J}}(x)$, is constant and equal to

$$\frac{\int_{x^*}^{x^{**}} h(y^A(z))dF(z)}{F(x^{**})-F(x^*)}.$$

In turn, the derivative on $(x^*, x^{**}]$ of type x's utility under I, $U^{\mathcal{I}}(x)$, is equal to $h\left(y^A(x)\right)$ if x does not belong to any non-degenerate interval $(a,b] \in I$, and is equal to

$$\frac{\int_a^b h(y^A(z))dF(z)}{F(b)-F(a)}$$

if $x \in (a,b] \in I$. This means that the derivative increases in x, and increases strictly except on intervals $(a,b] \in I$. So, $U^{\mathcal{I}}(x)$ is a convex non-linear function. Since $U^{\mathcal{J}}(x)$ is linear on $(x^*, x^{**}]$, $U^{\mathcal{I}}(x^*) = U^{\mathcal{J}}(x^*)$, and $U^{\mathcal{I}}(x^{**}) \leq U^{\mathcal{J}}(x^{**})$, we obtain that $U^{\mathcal{I}}(x) \leq U^{\mathcal{J}}(x)$ for all $x \in (x^*, x^{**}]$, and the inequality is strict for all types $x \in (x^*, x^{**}]$. Similarly, the derivative of $U^{\mathcal{J}}(x)$ on $(x^{**}, b']$ exceeds that of $U^{\mathcal{I}}(x)$ if $a' < x^{**} < b'$ for some $(a', b'] \in I$, and the two derivatives are equal for $x > b'$, which completes the proof that J Pareto improves over I.

Suppose now that another category ranking I' Pareto improves over I. Recall that I consists of singletons and a finite number of intervals $[x_1, x_1'], (x_2, x_2'], ..., (x_k, x_k']$, with $x_i' < x_{i+1}$. Denote by x' the highest type such that I and I' coincide up to x', and suppose that x' is the lower endpoint of an interval $(x_l, x_l']$ in I. (A similar argument to the one that follows applies if x' is a singleton.)

Now, x' must be the endpoint of a non-trivial interval in I', which we denote by $(x^*, x^{**}]$, where $x' = x^* < x^{**}$. Otherwise, for types x slightly higher than x_l the utility of these types under I would exceed their utility under I' by (17). It also cannot be that $x^{**} < x_l'$, since it would then follow from (17) that x^{**} strictly prefers I to I'.

Thus $x_l' < x^{**}$, and since I' Pareto improves over I, type x^{**} weakly prefers I' to I. And since (by (17)) the payoff of type x^{**} under any ranking depends only on the intervals up to the one that contains x^{**}, type x^{**} is indifferent between ranking I' and ranking $J = I(x^*, x^{**})$, and therefore prefers ranking J to ranking I.

Proof of Lemma 1. We have to show for every $x \in I_1 \cup I_2$ that

$$\frac{F(x)-F(x_1)}{F(x_4)-F(x_1)} \leq \frac{x-x_1}{x_4-x_1}. \tag{18}$$

Consider first $x \in I_1$. By definition of FOSD on I_1 we have

$$\frac{F(x)-F(x_1)}{F(x_3)-F(x_1)} \leq \frac{x-x_1}{x_3-x_1},$$

so for (18) it suffices to show that

$$\frac{F(x_3) - F(x_1)}{F(x_4) - F(x_1)} \leq \frac{x_3 - x_1}{x_4 - x_1}.$$

Suppose instead that

$$\frac{F(x_3) - F(x_1)}{F(x_4) - F(x_1)} > \frac{x_3 - x_1}{x_4 - x_1}. \tag{19}$$

This implies that

$$\frac{F(x_4) - F(x_3)}{F(x_4) - F(x_1)} < \frac{x_4 - x_3}{x_4 - x_1} \tag{20}$$

and also that

$$\frac{F(x_3) - F(x_2)}{F(x_4) - F(x_1)} > \frac{x_3 - x_2}{x_4 - x_1}, \tag{21}$$

where the last inequality is the product of (19) and the implication of FOSD on I_1 that[30]

$$\frac{F(x_3) - F(x_2)}{F(x_3) - F(x_1)} \geq \frac{x_3 - x_2}{x_3 - x_1}. \tag{22}$$

Dividing (20) by (21) we obtain

$$\frac{F(x_4) - F(x_3)}{F(x_3) - F(x_2)} < \frac{x_4 - x_3}{x_3 - x_2} \Rightarrow \frac{F(x_3) - F(x_2)}{F(x_4) - F(x_3)} + 1 > \frac{x_3 - x_2}{x_4 - x_3} + 1 \Rightarrow$$

$$\frac{F(x_4) - F(x_2)}{F(x_4) - F(x_3)} > \frac{x_4 - x_2}{x_4 - x_3} \Rightarrow \frac{F(x_4) - F(x_3)}{F(x_4) - F(x_2)} < \frac{x_4 - x_3}{x_4 - x_2}.$$

This last inequality is a contradiction, since FOSD on I_2 implies the opposite weak inequality, similarly to (22). Therefore, (18) holds for $x \in I_1$.

Now consider $x \in [x_3, x_4]$. Instead of (18) we will show the equivalent inequality

$$\frac{F(x_4) - F(x)}{F(x_4) - F(x_1)} \geq \frac{x_4 - x}{x_4 - x_1}.$$

From FOSD on I_2 we have

$$\frac{F(x_4) - F(x)}{F(x_4) - F(x_2)} \geq \frac{x_4 - x}{x_4 - x_2}.$$

Thus, it suffices to show that

$$\frac{F(x_4) - F(x_2)}{F(x_4) - F(x_1)} \geq \frac{x_4 - x_2}{x_4 - x_1}.$$

This inequality holds, because otherwise we would have

$$\frac{F(x_2) - F(x_1)}{F(x_4) - F(x_1)} > \frac{x_2 - x_1}{x_4 - x_1},$$

which would violate (18) for $x = x_2 \in I_1$.

[30] This inequality holds because

$$\frac{F(x_3) - F(x_2)}{F(x_3) - F(x_1)} = 1 - \frac{F(x_2) - F(x_1)}{F(x_3) - F(x_1)} \geq 1 - \frac{x_2 - x_1}{x_3 - x_1} = \frac{x_3 - x_2}{x_3 - x_1}.$$

B Extensions

Consider the following, separable utility functions: $h(x,y)-c(t)$ and $h(y)-c(x,t)$, where $h(x,0)=c(x,0)=0$ for all x, c is strictly increasing in t when $x>0$ and decreasing in x when $t>0$, and h is strictly increasing in y when $x>0$ and strictly increasing in x when $y>0$. These utilities generalize utilities (1) and (2), respectively. We will now describe how the results of Section 4 can be extended to these utility functions.[31] Part (a) is immediate from Theorem 1 and Corollary 1, as these results hold for the more general utility functions. (Of course, the formula defining M will be different for the more general utility functions.)

Consider first the utility $h(x,y)-c(t)$. The efforts $t^A(x)$ satisfy the following equations

$$t^A(x)=c^{-1}\left(h(x,y^A(x))-\int_0^x h_1\left(z,y^A(z)\right)dz\right), \tag{23}$$

and

$$h(y^A(x))-c(x,t^A(x))=-\int_x^1 c_1(z,t^A(z))dz.$$

respectively. The former equation determines $t^A(x)$. By differentiating the latter equation, assuming that F and G are differentiable, we obtain the following differential equation

$$h'(y^A(x))(y^A)'(x)=c_2(x,t^A(x))(t^A)'(x).$$

This equation, together with the initial condition $t^A(0)=0$, uniquely determines $t^A(x)$, assuming that the involved functions satisfy the Lipschitz condition.

For the utility $h(x,y)-c(t)$ we have that

$$U(x)=h(x,y^A(x))-c(t^A(x))=\int_0^x h_1\left(z,y^A(z)\right)dz$$

and

$$U^{x^*}(x)=\frac{\int_{x^*}^1 h(x,y^A(z))dF(z)}{1-F(x^*)}-c(M)=\frac{\int_{x^*}^1 h(x,y^A(z))dF(z)}{1-F(x^*)}-\frac{\int_{x^*}^1 h(x^*,y^A(z))dF(z)}{1-F(x^*)}+\int_0^{x^*} h_1\left(z,y^A(z)\right)dz.$$

This yields

$$U^{x^*}(x)-U(x)=\frac{\int_{x^*}^1 h(x,y^A(z))dF(z)}{1-F(x^*)}-\frac{\int_{x^*}^1 h(x^*,y^A(z))dF(z)}{1-F(x^*)}-\int_{x^*}^x h_1\left(z,y^A(z)\right)dz.$$

Comparing this expression to 0, we obtain an analogue of condition (9) from part (b) of Proposition 1.

The derivative of $U^{x^*}(x)-U(x)$ is equal to

$$\frac{\int_{x^*}^1 h_1(x,y^A(z))dF(z)}{1-F(x^*)}-h_1\left(x,y^A(x)\right).$$

If we assume that $h_1(x,y)$ increases in y, then this derivative is positive at $x=x^*$ and negative at $x=1$. Assuming that the derivative changes its sign only once, we obtain part (c) of Proposition 1; in addition, this yields the following analogue of the condition from part (d):

$$\frac{\int_{x^*}^1 h(1,y^A(z))dF(z)}{1-F(x^*)}-\frac{\int_{x^*}^1 h(x^*,y^A(z))dF(z)}{1-F(x^*)}-\int_{x^*}^1 h_1\left(z,y^A(z)\right)dz\geq 0.$$

[31] The results of Section 4 can be extended to a more general separable utility function $h(x,y)-c(x,t)$. However, the results for this more general function would be a combination the results for the two more specific functions, and the analysis would be less transparent.

For the utility $h(y) - c(x,t)$ we have that

$$U(x) = h(y^A(x)) - c(x, t^A(x)) = -\int_0^x c_1(z, t^A(z))dz$$

and

$$U^{x^*}(x) = \frac{\int_{x^*}^1 h(y^A(z))dF(z)}{1 - F(x^*)} - c(x, M) = c(x^*, M) - c(x, M) - \int_0^{x^*} c_1(z, t^A(z))dz,$$

which yields

$$U^{x^*}(x) - U(x) = c(x^*, M) - c(x, M) + \int_{x^*}^x c_1(z, t^A(z))dz.$$

Comparing this expression to 0, we obtain an analogue of condition (9) from part (b) of Proposition 1.

The derivative of this expression is $c_1(x, t^A(x)) - c_1(x, M)$. Assume that $c_1(x,t)$ decreases with t. Then, $c_1(x, t^A(x)) - c_1(x, M) \geq 0$ when $t^A(x) \leq M$, and $c_1(x, t^A(x)) - c_1(x, M) \leq 0$ when $t^A(x) \geq M$. Since $U^M(x) - U(x) = 0$ when $x = x^*$, part (c) of Proposition 1 must hold; in addition, this yields the following analogue of the condition from part (d):

$$c(x^*, M) - c(1, M) + \int_{x^*}^1 c_1(z, t^A(z))dz \geq 0.$$

Proposition 2 also generalizes to utilities $h(x,y) - c(t)$ and $h(y) - c(x,t)$, and its proof requires only minor changes. Consider first the utility $h(x,y) - c(t)$. we have

$$U^{x^*}(x) = \frac{\int_{x^*}^1 h(x, y^A(z))dF(z)}{1 - F(x^*)} - \frac{\int_{x^*}^1 h(x^*, y^A(z))dF(z)}{1 - F(x^*)} + \int_0^{x^*} h_1\left(z, y^A(z)\right)dz,$$

so, $U^{x^*}(x)$ can be represented as $\phi(x, x^*) + \psi(x^*)$, where

$$\phi(x, x^*) = \frac{\int_{x^*}^1 h(x, y^A(z))dF(z)}{1 - F(x^*)}$$

and

$$\psi(x^*) = \int_0^{x^*} h_1\left(z, y^A(z)\right)dz - \frac{\int_{x^*}^1 h(x^*, y^A(z))dF(z)}{1 - F(x^*)}.$$

Assume that for all $x'' > x'$, the difference $h(x'', y) - h(x', y)$ strictly increases in y. This implies that $\phi(x, x_1^*) - \phi(1, x_1^*) > \phi(x, x_2^*) - \phi(1, x_2^*)$. Together with $\phi(1, x_1^*) + \psi(x_1^*) \geq \phi(1, x_2^*) + \psi(x_2^*)$, this yields part (a).

To show part (b), suppose that type 1 strictly prefers x_2^* to x_1^*. By the assumption that $h(x'', y) - h(x', y)$ strictly increases in y, it follows that $U^{x_1^*}(x) - U^{x_2^*}(x)$ strictly decreases in x. Thus, there exists an x^{**} such that types $x < x^{**}$ strictly prefer x_1^* and types $x > x^{**}$ strictly prefer x_2^*. And since all players with types close to 1 strictly prefer x_2^* to x_1^*, we have that $x^{**} < 1$. Finally, it can be readily checked that the derivative of ψ with respect to x^* is negative. Hence, $\psi(x_1^*) > \psi(x_2^*)$, and so $x^{**} > 0$.

For the utility $h(y) - c(x,t)$, we have that

$$U^{x_1^*}(x) = c(x^*, M) - c(x, M) - \int_0^{x^*} c_1(z, t^A(z))dz.$$

So, $U^{x_1^*}(x)$ can be represented as $\phi(x, x_1^*) + \psi(x_1^*)$, where

$$\phi(x, x_1^*) = -c(x, M_1)$$

and

$$\psi(x_1^*) = c(x^*, M) - \int_0^{x^*} c_1(z, t^A(z))dz.$$

Assume that for all $x'' > x'$, the difference $c(x', M) - c(x'', M)$ strictly increases in M. This implies that $\phi(x, x_1^*) - \phi(1, x_1^*) > \phi(x, x_2^*) - \phi(1, x_2^*)$. Together with $\phi(1, x_1^*) + \psi(x_1^*) \geq \phi(1, x_2^*) + \psi(x_2^*)$, this yields part (a).

To show part (b), suppose that a player of type 1 strictly prefers x_2^* to x_1^*. By the assumption that $c(x', M) - c(x'', M)$ strictly increases in M, it follows that $U^{x_1^*}(x) - U^{x_2^*}(x)$ strictly decreases in x. Thus, there exists an x^{**} such that types $x < x^{**}$ strictly prefer x_1^* and types $x > x^{**}$ strictly prefer x_2^*. And since all players with types close to 1 strictly prefer x_2^* to x_1^*, we have that $x^{**} < 1$. Finally, it can be readily checked that the derivative of ψ with respect to x^* is negative. Hence, $\psi(x_1^*) > \psi(x_2^*)$, so $x^{**} > 0$.

References

[1] Ahn S-Y and Baek H-J (2013): "Academic achievement-oriented society and its relationship to the psychological well-being of Korean adolescents." In: Yi CC (eds) The psychological well-being of East Asian youth, quality of life in Asia, 265–279.

[2] Azevedo, E. and J. Leshno (2016): "A Supply and Demand Framework for Two-Sided Matching Markets," Journal of Political Economy, 124(5), 1235-1268.

[3] Bodoh-Creed, A. and B. Hickman (2015): "College Assignment as a Large Contest," Mimeo.

[4] Budish, E. (2011) "The Combinatorial Assignment Problem: Approximate Competitive Equilibrium from Equal Incomes," Journal of Political Economy, 119 (6), 1061 -1103.

[5] Chakravarty, S. and T. R. Kaplan (2013): "Optimal allocation without transfer payments," Games and Economic Behavior, 77, 1-20.

[6] Che, Y.-K. and I. Gale (1998): "Caps on Political Lobbying," American Economic Review, 88, 643-651.

[7] Dubey, P. and J. Geanakoplos (2010): "Grading exams: 100, 99, 98, . . . or A, B, C?," Games and Economic Behavior, 69, 72-94.

[8] Frankel, A. and N. Kartik (2016): "Muddled Information," Mimeo.

[9] Gale, D. and L. Shapley (1962): "College Admissions and the Stability of Marriage," American Mathematical Monthly, 69 (1), 9-15.

[10] Hsin, A. and Y. Xie (2014): "Explaining Asian Americans' academic advantage over whites," Proceedings of the National Academy of Sciences, 111(23), 8416–8421.

[11] Koh, W. T., Z. Yang and L. Zhu (2006): "Lottery rather than waiting-line auction," Social Choice and Welfare, 27, 289-310.

[12] Olszewski, W. and R. Siegel (2016a): "Large Contests." Econometrica 84, 835-854.

[13] Olszewski, W. and R. Siegel (2016b): "Bid Caps in Contests." Mimeo.

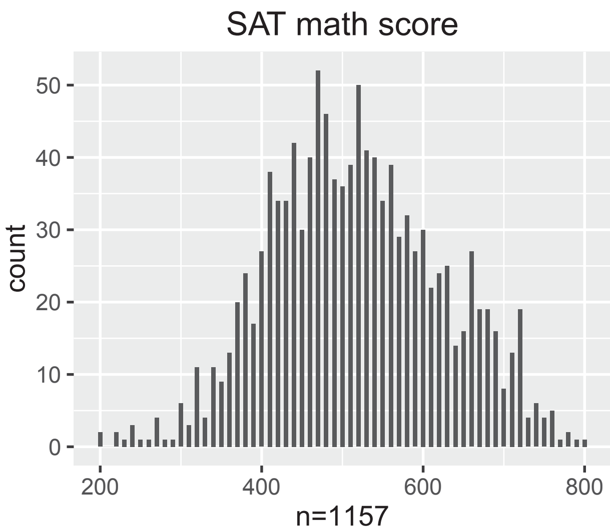
[14] Olszewski, W. and R. Siegel (2016c): "Effort-Maximizing Contests." Mimeo.

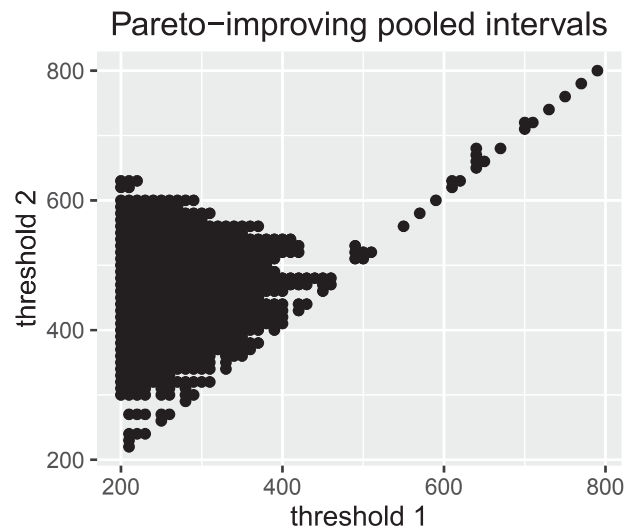
[15] Spence, M. (1973). "Job Market Signaling." Quarterly Journal of Economics, 87, 355-374.

[16] Taylor, G. A., K. K. Tsui, and L. Zhu (2003): "Lottery or waiting-line auction?," Journal of Public Economics, 87, 1313-1334.

Data Appendix

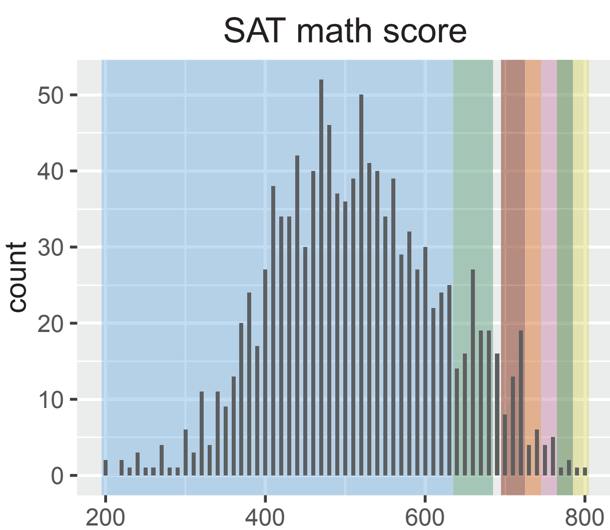
SAT math score



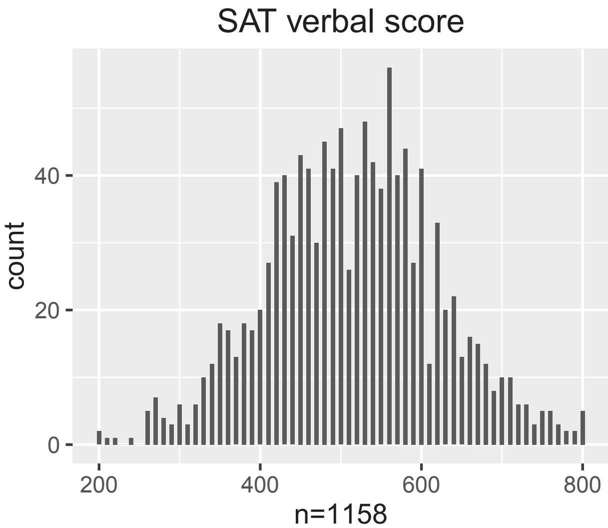


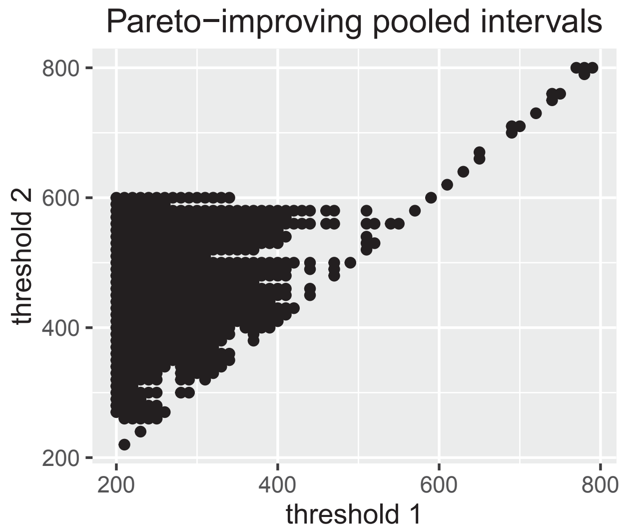
Maximal Pareto-improving pooled intervals

```
##                 [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## lower_endpoint  200  640  700  730  750  770  790
## upper_endpoint  630  680  720  740  760  780  800
```



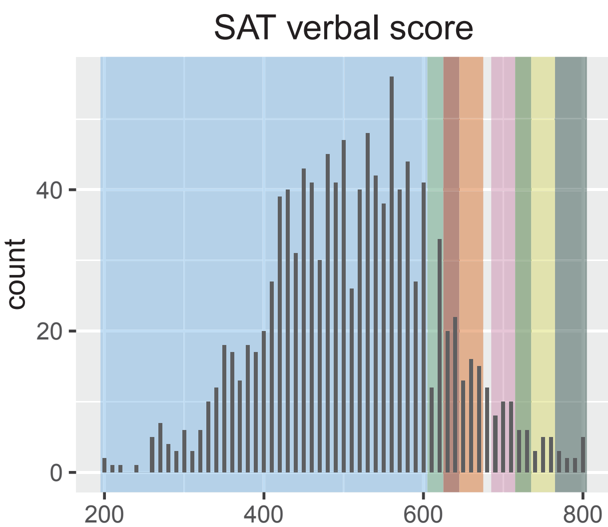
SAT verbal score



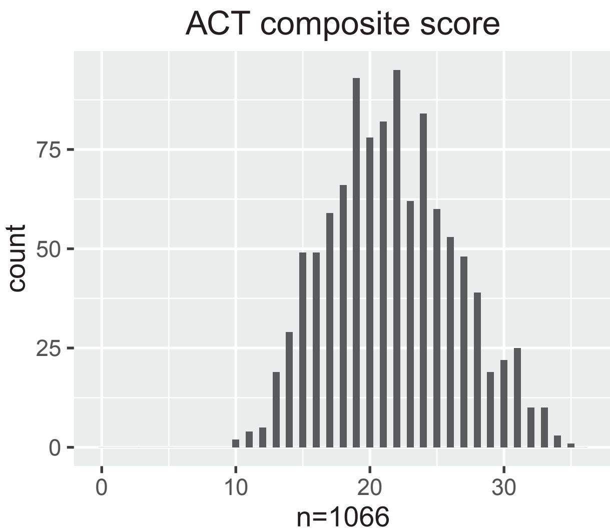


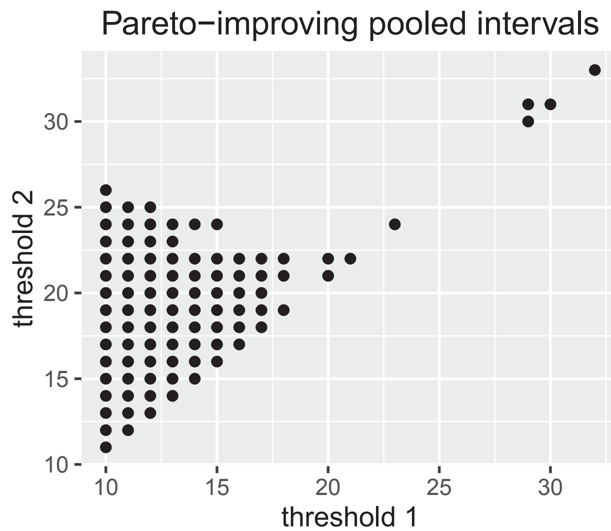
Maximal Pareto-improving pooled intervals

```
##                 [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
## lower_endpoint  200  610  630  650  690  720  740  770
## upper_endpoint  600  620  640  670  710  730  760  800
```



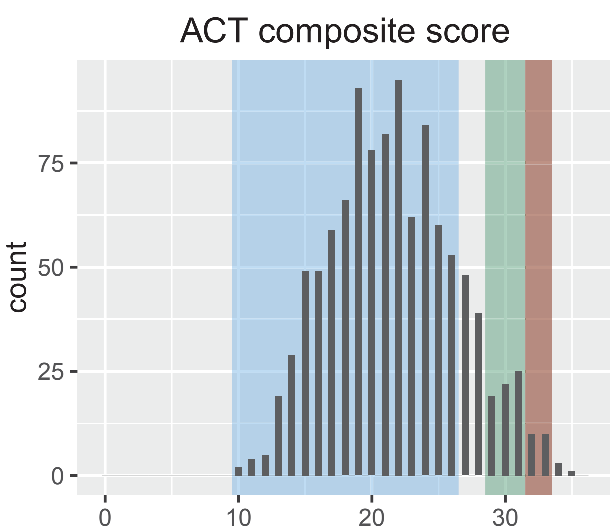
ACT composite score



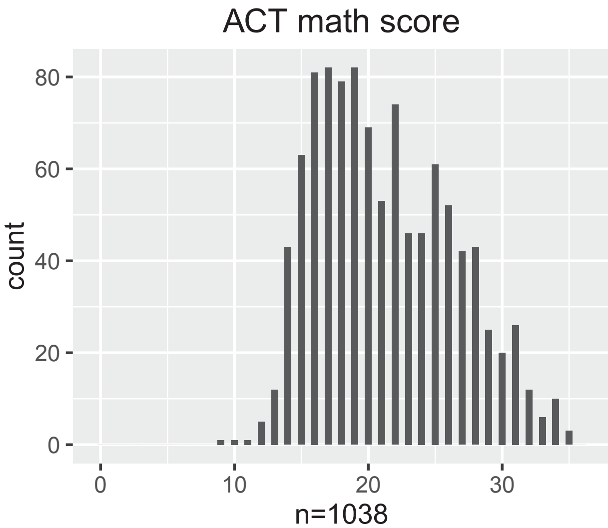


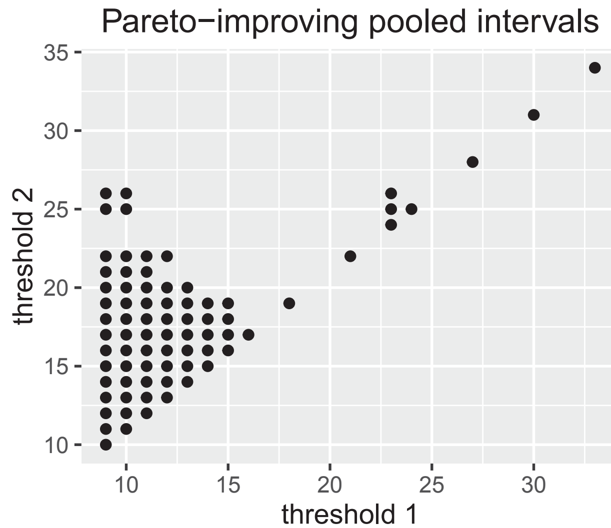
Maximal Pareto-improving pooled intervals

```
##                 [,1] [,2] [,3]
## lower_endpoint    10   29   32
## upper_endpoint    26   31   33
```



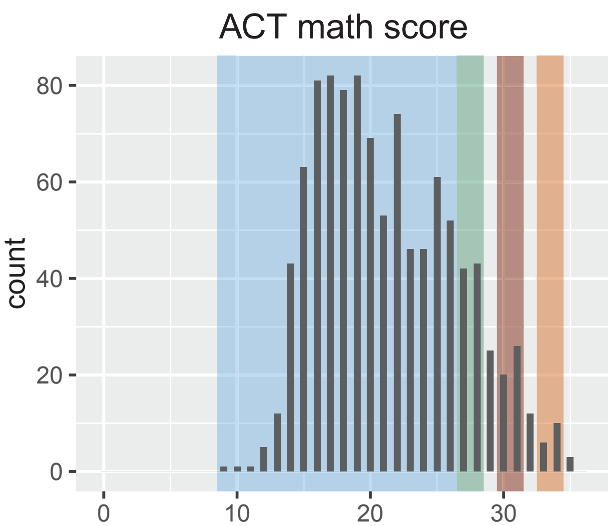
ACT math score



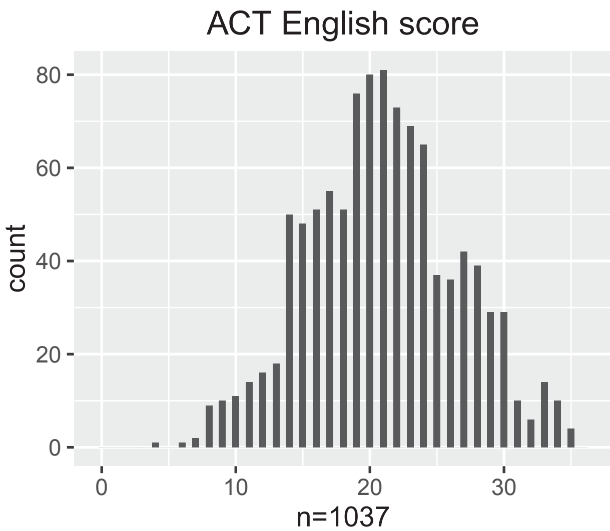


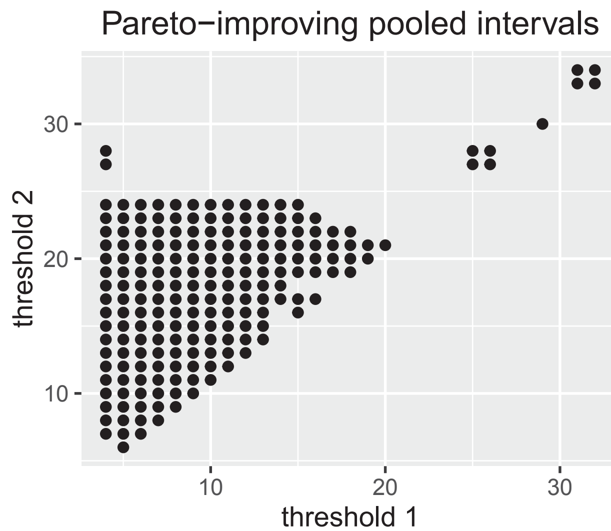
Maximal Pareto-improving pooled intervals

```
##                 [,1] [,2] [,3] [,4]
## lower_endpoint     9   27   30   33
## upper_endpoint    26   28   31   34
```



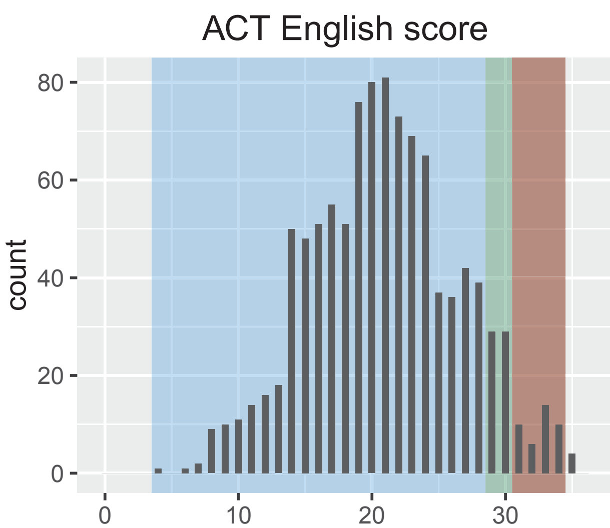
ACT English score



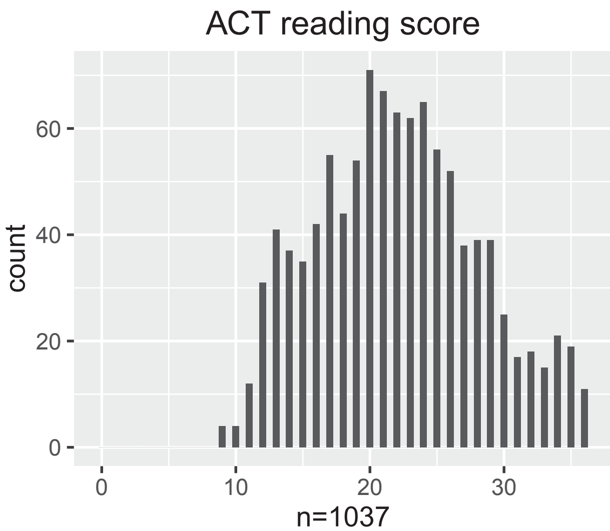


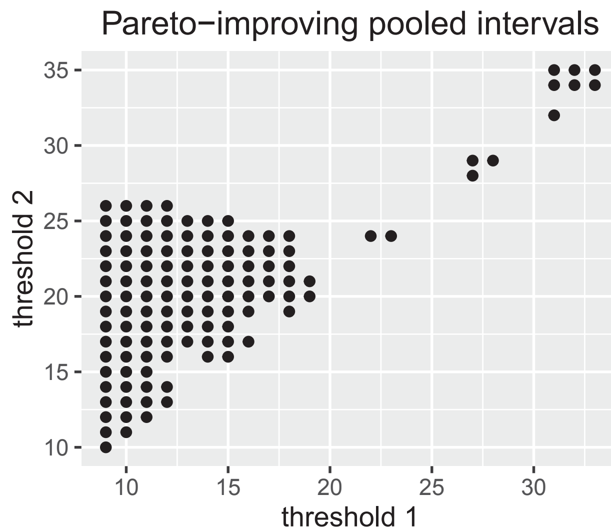
Maximal Pareto-improving pooled intervals

```
##                 [,1] [,2] [,3]
## lower_endpoint    4   29   31
## upper_endpoint   28   30   34
```



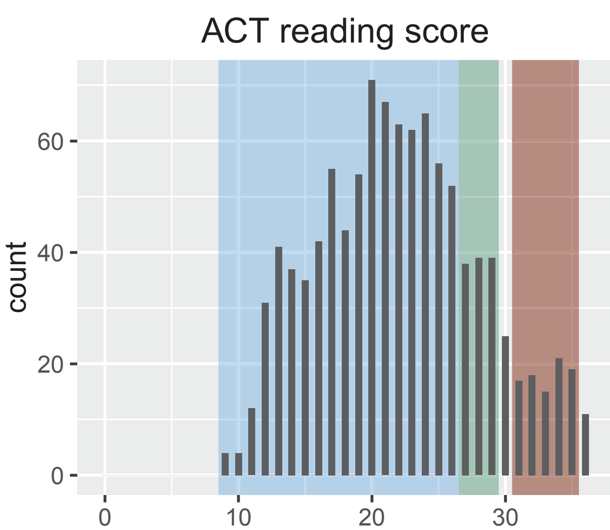
ACT reading score



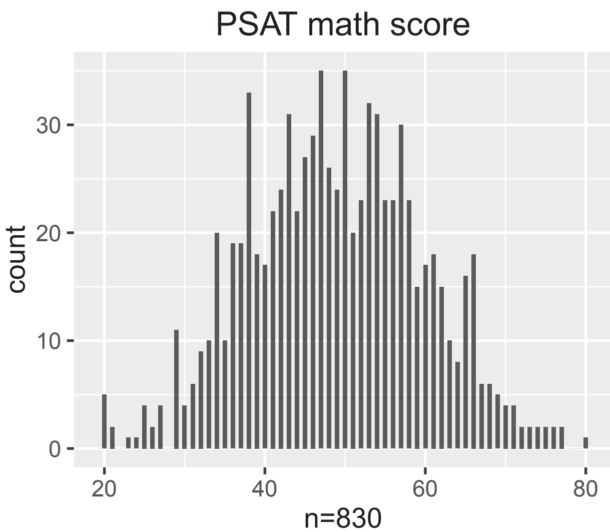


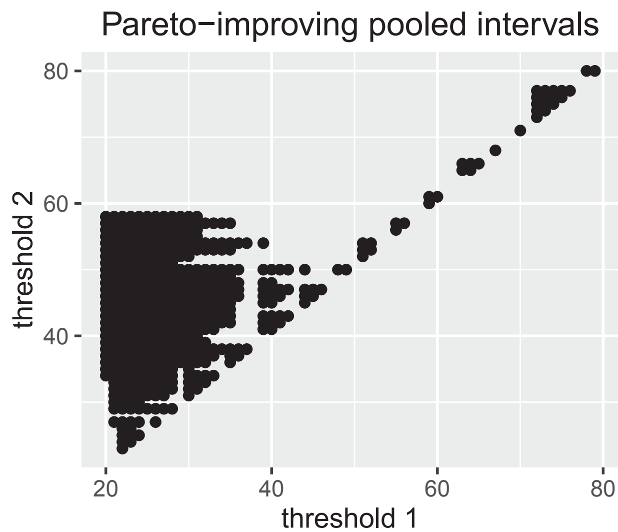
Maximal Pareto-improving pooled intervals

```
##                 [,1] [,2] [,3]
## lower_endpoint     9   27   31
## upper_endpoint    26   29   35
```



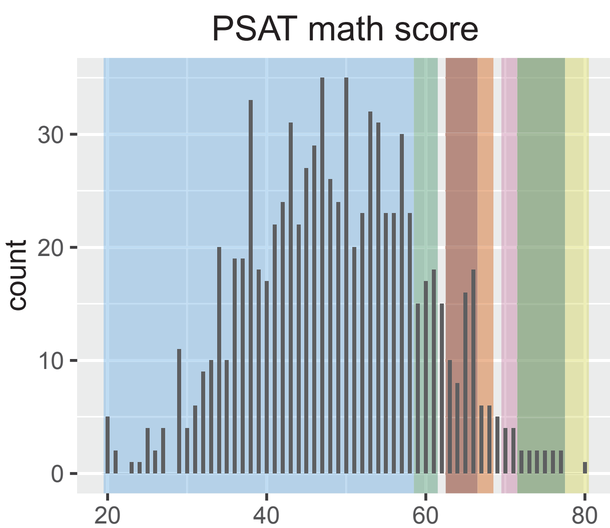
PSAT math score



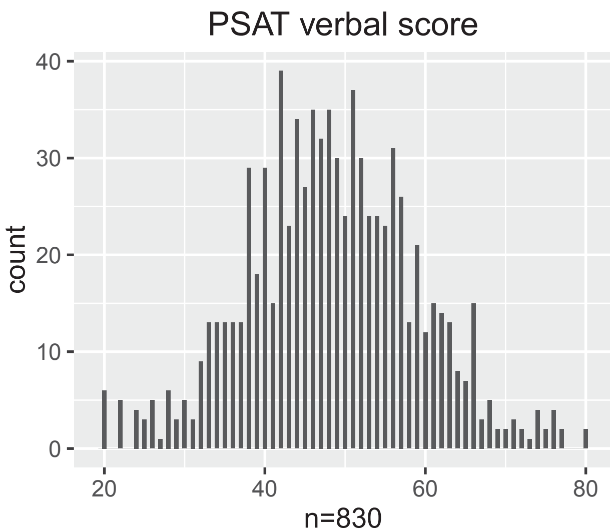


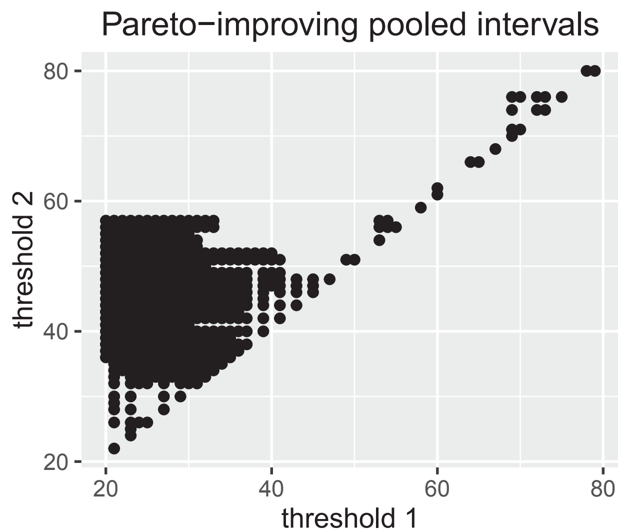
Maximal Pareto-improving pooled intervals

```
##                 [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## lower_endpoint    20   59   63   67   70   72   78
## upper_endpoint    58   61   66   68   71   77   80
```



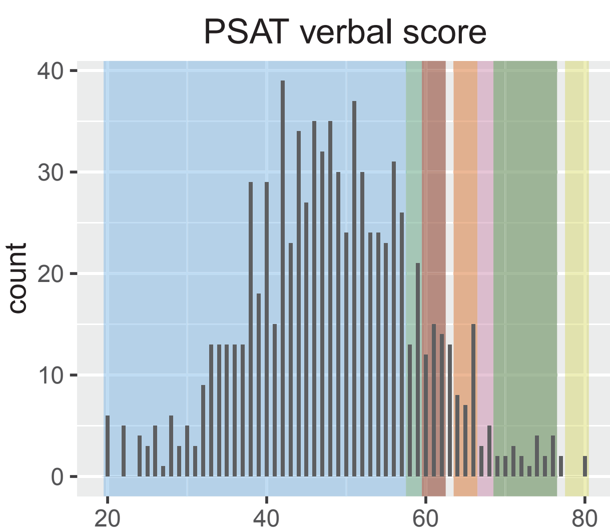
PSAT verbal score



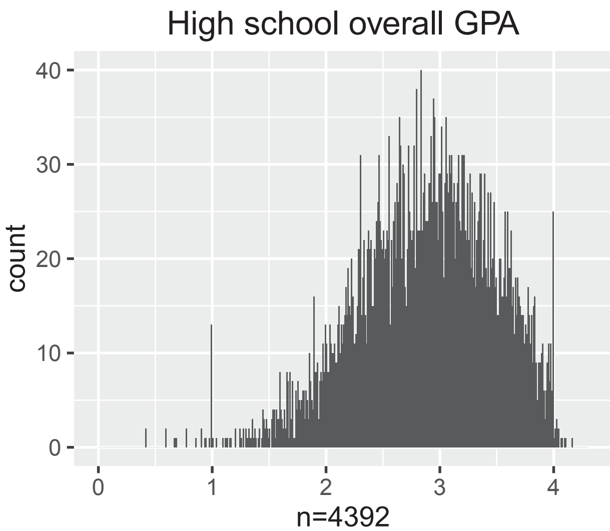


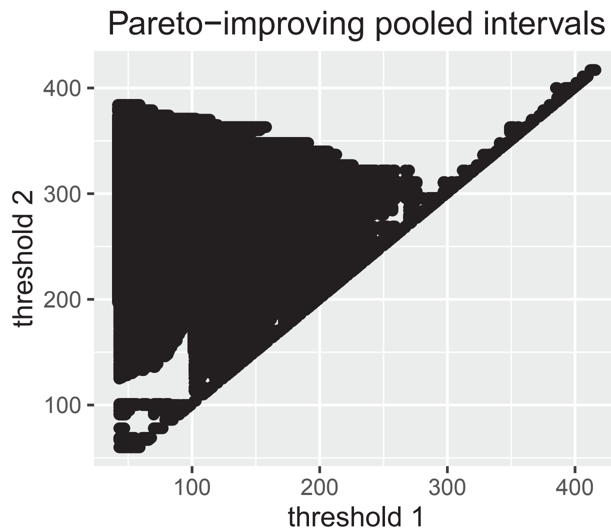
Maximal Pareto-improving pooled intervals

```
##                 [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## lower_endpoint    20   58   60   64   67   69   78
## upper_endpoint    57   59   62   66   68   76   80
```



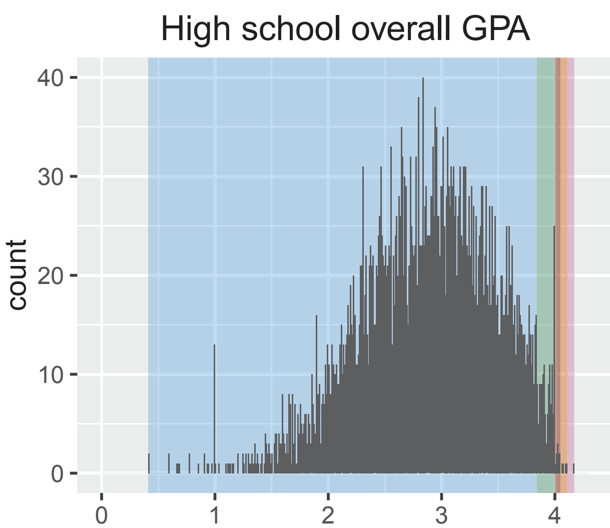
High school overall GPA



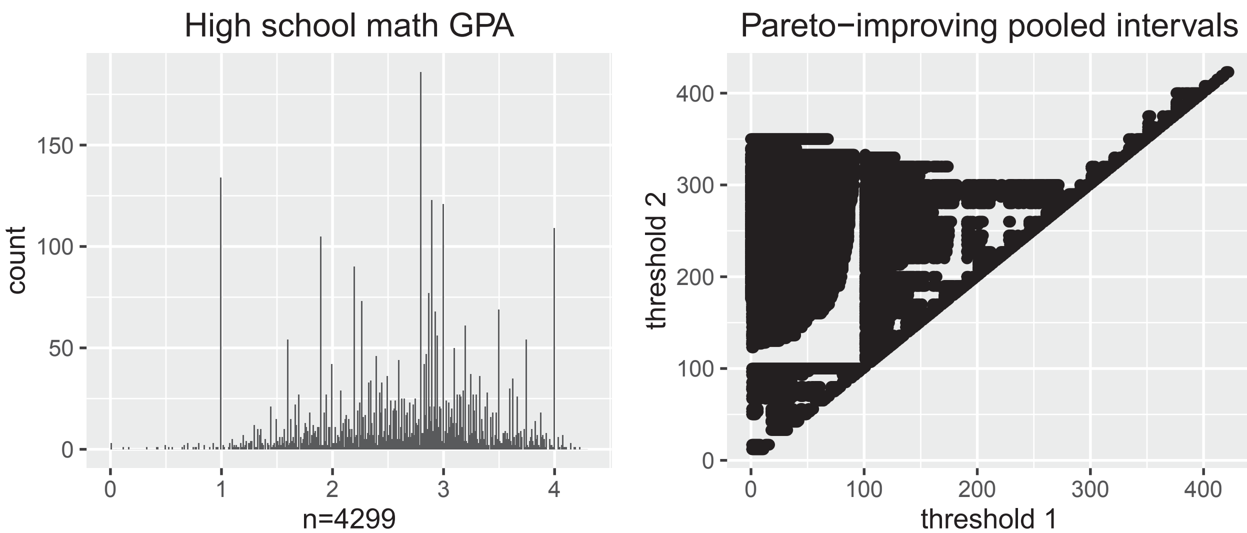


Maximal Pareto-improving pooled intervals

```
##                  [,1] [,2] [,3] [,4] [,5]
## lower_endpoint 0.42 3.85 4.01 4.06 4.12
## upper_endpoint 3.84 4.00 4.05 4.11 4.17
```

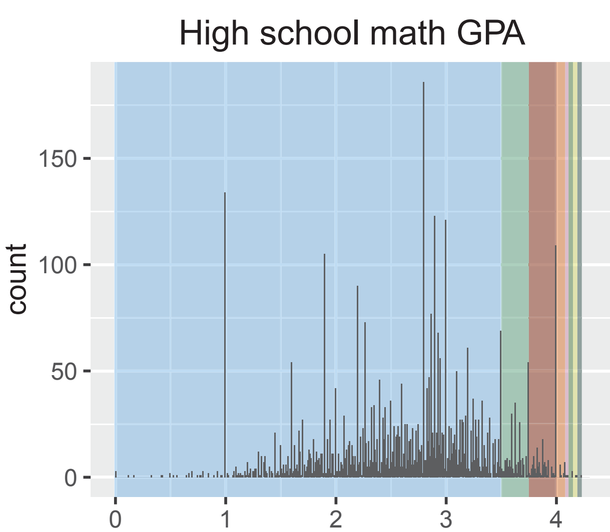


High school math GPA

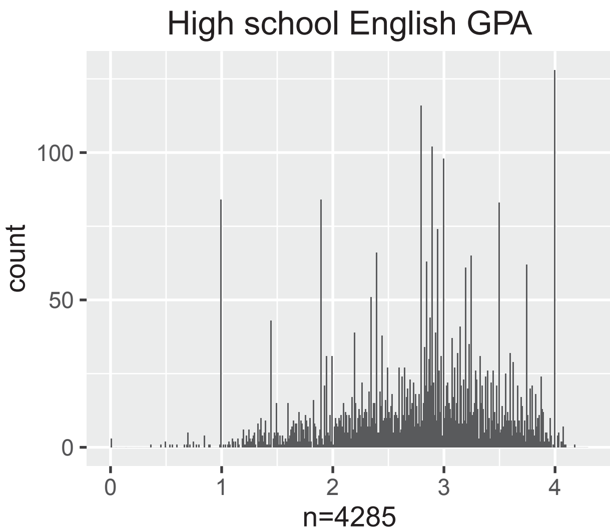


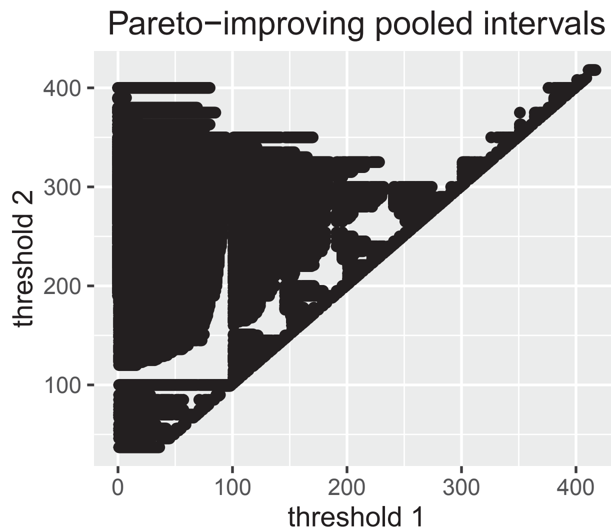
Maximal Pareto-improving pooled intervals

```
##                  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
## lower_endpoint  0.0 3.51 3.76 4.01 4.09 4.12 4.16 4.20
## upper_endpoint  3.5 3.75 4.00 4.08 4.11 4.15 4.19 4.23
```



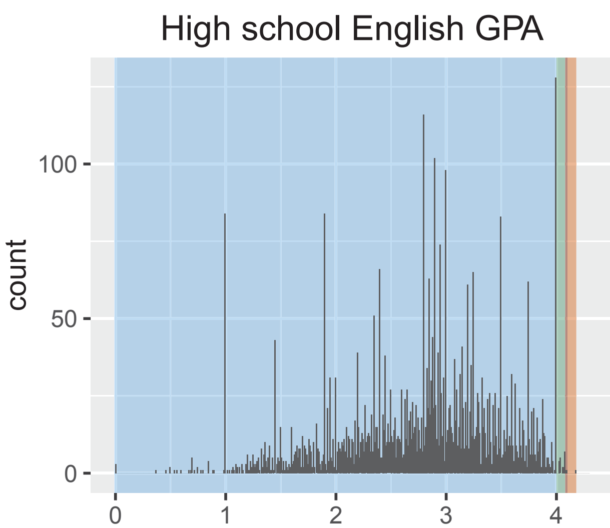
High school English GPA



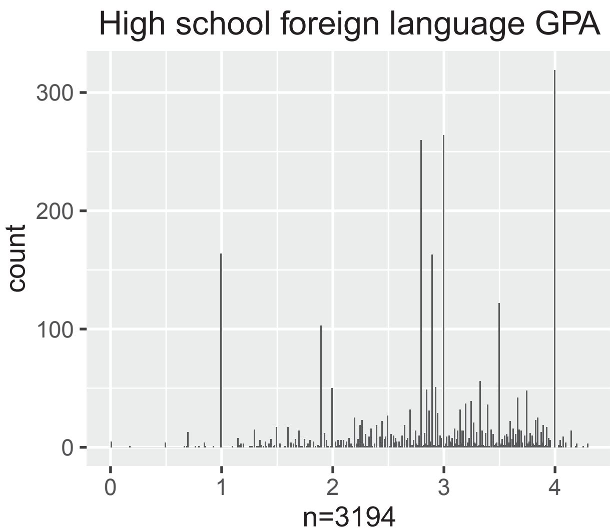


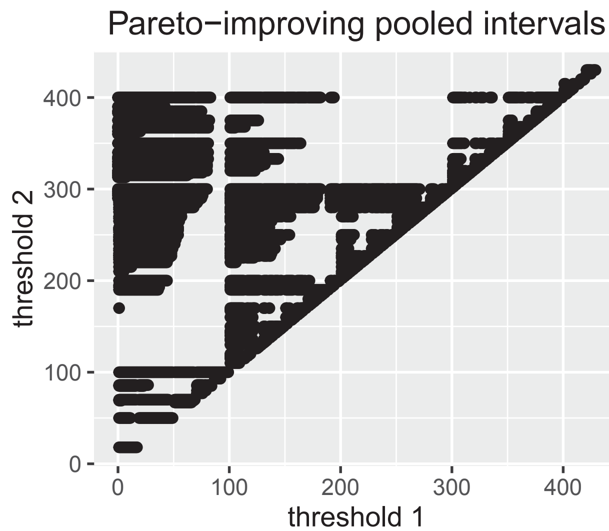
Maximal Pareto-improving pooled intervals

```
##                 [,1] [,2] [,3] [,4]
## lower_endpoint     0 4.01 4.09 4.11
## upper_endpoint     4 4.08 4.10 4.18
```



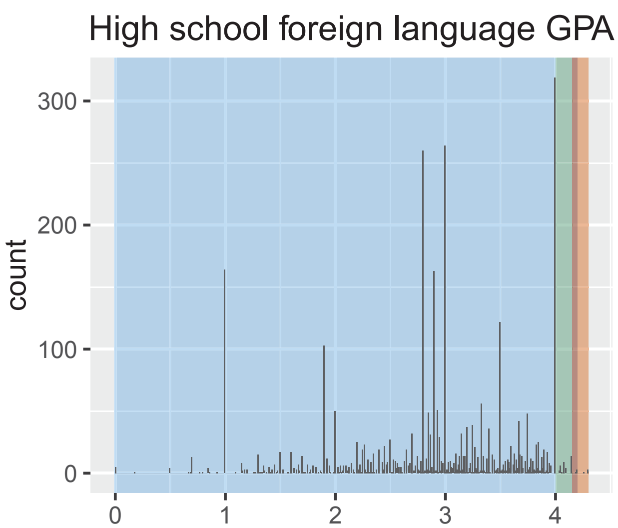
High school foreign language GPA



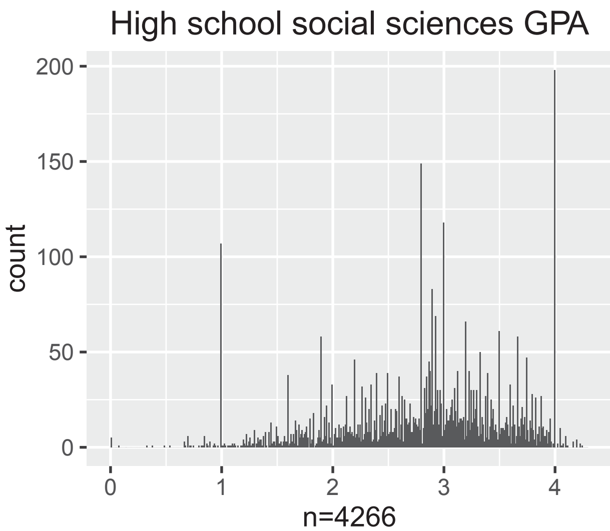


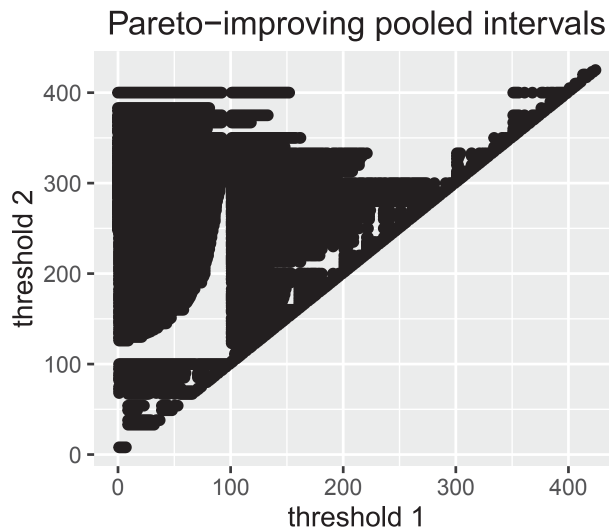
Maximal Pareto-improving pooled intervals

```
##                 [,1] [,2] [,3] [,4]
## lower_endpoint     0 4.01 4.16 4.21
## upper_endpoint     4 4.15 4.20 4.30
```



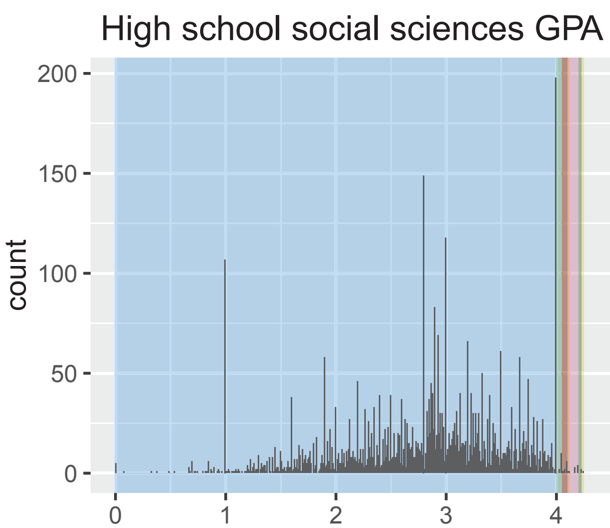
High school social sciences GPA



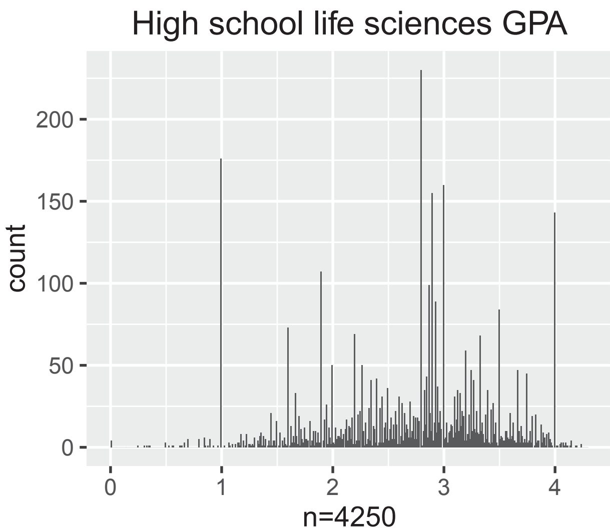


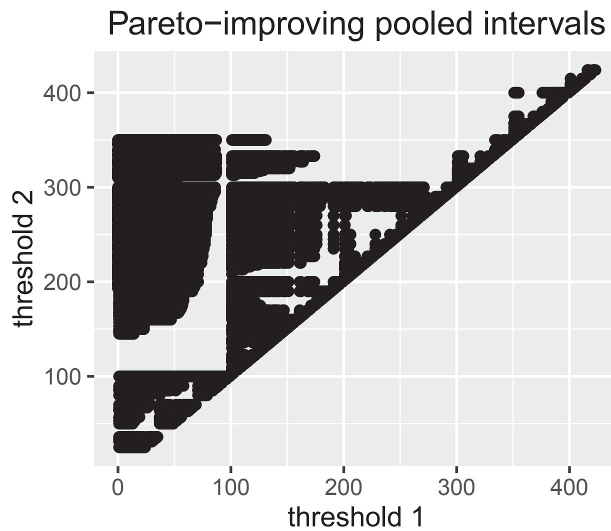
Maximal Pareto-improving pooled intervals

```
##                  [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## lower_endpoint     0 4.01 4.06 4.11 4.13 4.21 4.24
## upper_endpoint     4 4.05 4.10 4.12 4.20 4.23 4.25
```



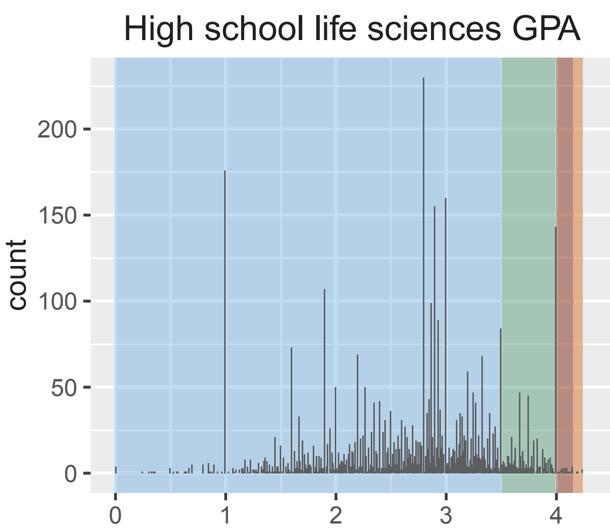
High school life sciences GPA





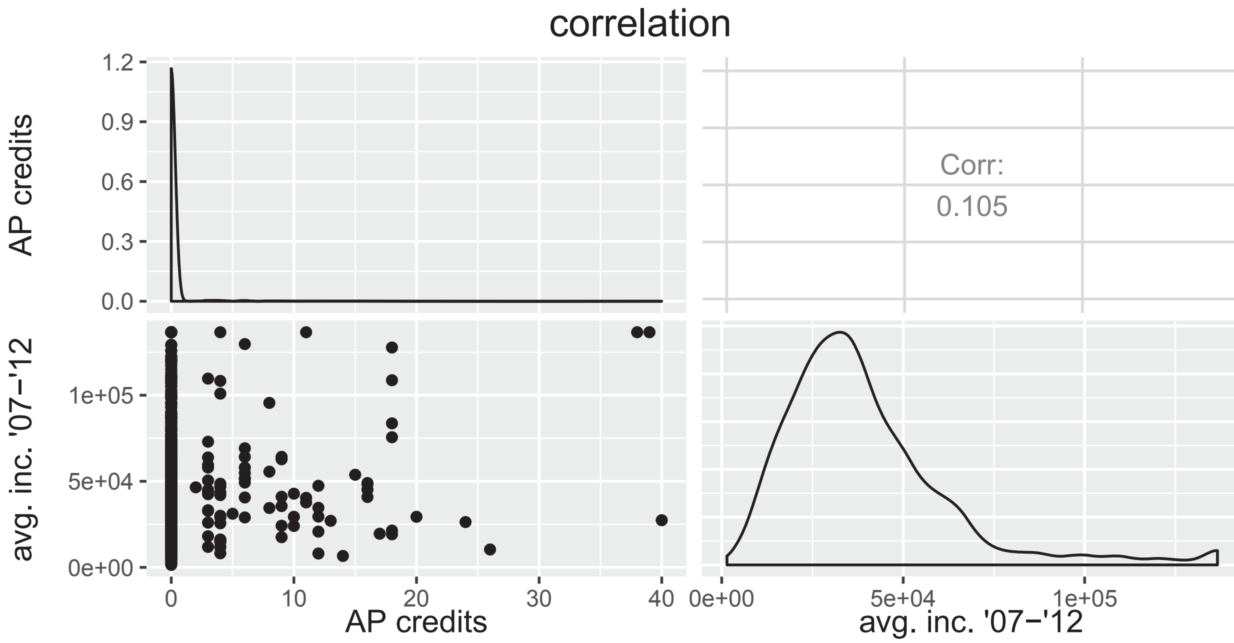
Maximal Pareto-improving pooled intervals

```
##                 [,1] [,2] [,3] [,4]
## lower_endpoint  0.0 3.51 4.01 4.16
## upper_endpoint  3.5 4.00 4.15 4.24
```



AP credits earned

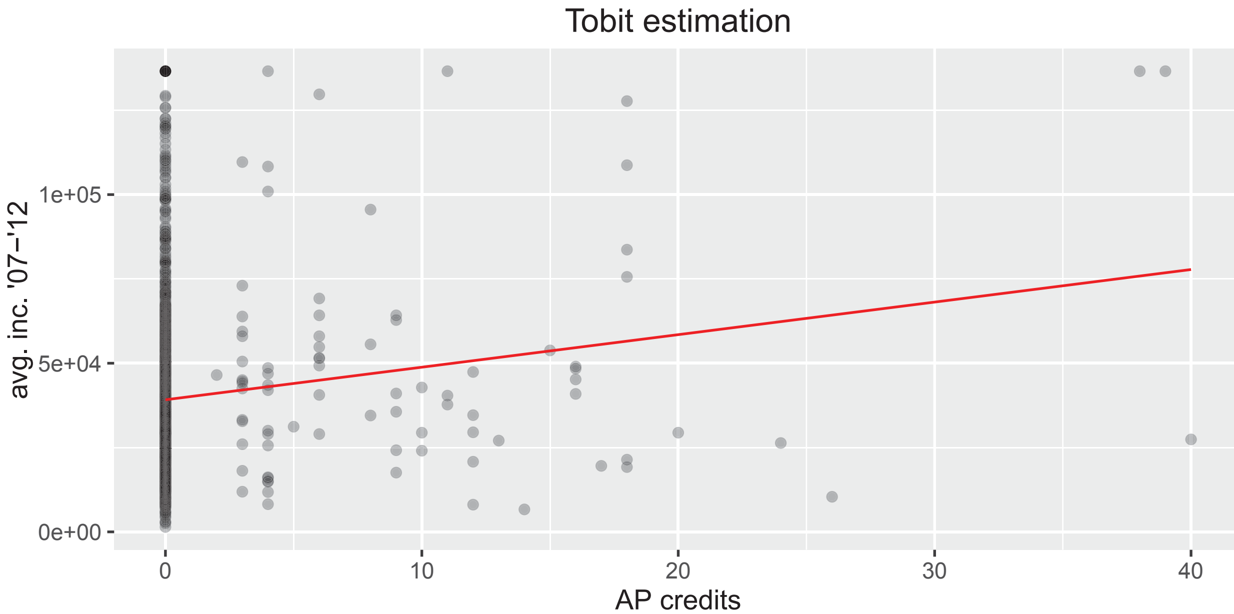
- correlation with average income:



- Tobit regression:

```
##
## Call:
## vglm(formula = average_income ~ indepvar, family = tobit(Upper = max(average_income,
##     na.rm = TRUE)), data = df)
##
## Pearson residuals:
##                Min      1Q  Median      3Q    Max
## mu       -2.2653 -0.6243 -0.1757  0.3894  4.049
## loge(sd) -0.7352 -0.6905 -0.5214 -0.1057 13.125
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept):1 3.916e+04  5.778e+02  67.780  < 2e-16 ***
## (Intercept):2 1.007e+01  1.762e-02 571.618  < 2e-16 ***
## indepvar      9.654e+02  2.115e+02   4.565 4.99e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of linear predictors:  2
##
## Names of linear predictors: mu, loge(sd)
##
## Dispersion Parameter for tobit family:   1
##
## Log-likelihood: -19761.81 on 3475 degrees of freedom
##
## Number of iterations: 4
```

- actual and fitted variables:



- Pareto-improving caps: 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40
- Pareto frontier: 19, 20, 25, 26
- top agent's utility for different caps: