Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 25-1294

The Design of Promises: The Structure of Local College Affordability Programs in the United States

David B. Monaghan
Shippensburg University

Alice Armstrong
Shippensburg University

Elizabeth Hawke
Shippensburg University

We analyze 314 local college affordability programs (i.e., "Promise" or "free college" programs) using a novel dataset detailing, for each program, rules stipulating what programs provide (provision), where they may be used (applicability), and who may use them (eligibility). We perform three sets of analyses. First, we ask whether programs can be cogently described as involving greater or lesser provision (i.e., "generosity"), broader/narrower applicability, and more/less universal eligibility by examining the internal consistency of these sets of program rules. That is, are provision, applicability, and eligibility better thought of as coherent dimensions according to which programs vary, or just buckets of conceptually related rules? Second, we inquire into whether program design is strongly imitative by examining the relationship between program similarity and both temporal and geographic proximity. Finally, we perform a data-driven inquiry through cluster analysis. We discover that program structure is most strongly influenced by the level(s) of colleges at which programs may be used: only community colleges, only public four-year colleges, or both community colleges and public four-year colleges. We provide an interpretation of why this may be the case. We thereby contribute robustly to the slight but growing literature on how college affordability programs are designed.

VERSION: September 2025

Suggested citation: Monaghan, David B., Alice Armstrong, and Elizabeth Hawke. (2025). The Design of Promises: The Structure of Local College Affordability Programs in the United States. (EdWorkingPaper: 25-1294). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/892a-cy43

The Design of Promises:

The Structure of Local College Affordability Programs in the United States

Abstract

We analyze 314 local college affordability programs (i.e., “Promise” or “free college” programs) using a novel dataset detailing, for each program, rules stipulating what programs provide (provision), where they may be used (applicability), and who may use them (eligibility). We perform three sets of analyses. First, we ask whether programs can be cogently described as involving greater or lesser provision (i.e., “generosity”), broader/narrower applicability, and more/less universal eligibility by examining the internal consistency of these sets of program rules. That is, are provision, applicability, and eligibility better thought of as coherent *dimensions* according to which programs vary, or as buckets of conceptually related rules? Second, we inquire into whether program design is strongly imitative by examining the relationship between program similarity and both temporal and geographic proximity, as well as similarity to two well-publicized programs, the Kalamazoo Promise and the Tennessee Promise. Finally, we perform a data-driven inquiry through cluster analysis. We discover that program structure is most strongly influenced by the level(s) of colleges at which programs may be used: only community colleges, only public four-year colleges, or both community colleges and public four- year colleges. We provide an interpretation of why this may be the case. We thereby contribute robustly to the slight but growing literature on how college affordability programs are designed.

The Design of Promises:

The Structure of Local College Affordability Programs in the United States

Local and state college affordability initiatives, sometimes called “Promise” or “free college” programs, have diffused throughout U.S. higher education over the past 20 years. Exact numbers vary depending on who is counting, with reports of between eighteen (Mishory 2018) and twenty-three (Jones et al., 2020) programs operated by states, 219 operating locally (Miller-Adams et al., 2023), and 425 (College Promise, 2023; Perna & Leigh, n.d.) operating either statewide or locally. Taken together, these programs represent one of the most significant developments in the U.S. financial aid landscape this century. Despite considerable scholarly investigation (for databases of research, see College Promise, 2020; Upjohn Institute, 2021; for reviews, see Swanson et al., 2020; Anderson, 2021; Monaghan, 2025), elementary questions regarding these programs remain unanswered and even largely unasked.

For instance, researchers agree that these programs vary considerably in terms of design (Perna & Leigh, 2018; Willard et al., 2019). And most would likely agree that differences in design could affect programs’ impacts. However, there has been little empirical examination of program design variation, let alone research connecting this to variation in outcomes. Additionally, there is little investigation of why or how programs come to adopt their specific designs.

Our work here addresses both these matters—the first directly and the second by extension. We assembled a novel dataset of 314 local college affordability programs operating in the 2022-23 academic year. Ours surpasses earlier databases of such programs (Perna & Leigh, n.d.; Miller-Adams et al., 2023) through far more extensive detail on rules relating to what programs provide, where they may be used, and who may use them. We perform three sets of analyses. First, we ask whether, in terms of program rules, we can describe programs as

varying continuously in how much they provide, in how broadly applicable they are, and in how restrictive they are in terms of eligibility. Can one use rules to place programs along spectra or continua in these terms? Second, we examine whether programs appear to be designed largely by copying recently developed, geographically proximate, or very well-publicized programs. Third, we ask how to sort programs into maximally similar groups according to design rules (that is, through clustering), and how best to account for this revealed similarity. What do similar programs have most in common? Our results considerably advance empirical inquiry into the design of college affordability programs.

Theory and prior research

Despite the prominence of "free college programs" and "Promise programs" in politics, media and scholarship, there is no consensus regarding how to define or distinguish them from neighboring categories of programs (for conflicting definitions, see Miller-Adams, 2015:11; Perna & Leigh, 2018: 155-6; Millet et al., 2020:18-20; for history, see Miller-Adams 2015, 2021; for the emergence and changing meaning of the terms, see Monaghan et al., 2024). Nor even is there consensus as to whether these two terms cover the same sets of "things". To avoid both controversy and confusion, we employ the more inclusive "college affordability programs", for which we provide operationalization below (see Methods). This term highlights these programs' continuity with the broader universe of student financial aid.

Long ago, March & Simon (1958) defined organizational *programs* as sets of rules for responding to contingencies. Understood thus, college affordability programs certainly involve staff, funding streams, and publicity, but are at core *sets of rules for the distribution of resources*. They vary essentially in their constituent rules. These can be sorted into three categories, establishing 1) what resources will be provided, 2) where these resources may be used, and 3) to whom the resources will be granted. We refer to these as rules governing *provision*, *applicability*, and *eligibility*, respectively.

Making sense of college affordability programs therefore requires examining their rules thoroughly and in detail. This is not standard practice. Most commonly, researchers discuss financial aid programs using folk categories based on one or two rules, such as “need-based” and “merit-based” grants. Even prior attempts to generate data-driven categorizations (Perna et al., 2008; Custer & Akaeze, 2021; Perna & Leigh, 2018) leverage small numbers of rules selected without clear rationales. Indeed, attempts to think systematically about how financial aid programs vary began only in response to the recent proliferation of local and state affordability programs in the current century. The tripartite division of rules we make use of here (i.e., provision, applicability, and eligibility) was developed by Miller-Adams (2015) and Iriti et al. (2016), with additional input by Perna and Leigh (2018), Swanson et al. (2020), and Miller-Adams et al. (2023) (see Monaghan & Hawke, 2025 for details).

We built on this in previous work (Monaghan & Hawke, 2025) in two ways. First, we suggested that this rule categorization is applicable not just to “Promise” programs but to any postsecondary grant[1]. Second, we enumerated rules in far more exhaustive detail. Here, these advances enable us to examine two sets of assumptions common in the literature on college affordability programs.

The first has to do with the relationship of rules to underlying dimensions of variation. It is common for researchers to refer to programs as varying in terms of “generosity” (e.g., Iriti et al., 2016; Miller-Adams & Iriti, 2022; Bell, 2021). By this, they seem to mean either average per-participant spending (*spending intensity*) or total spending. As the latter is the former times the number of participants, we focus on the former. Programs certainly vary in spending intensity, but *program rules* are insufficient to measure this. Rules determine spending in interaction with the distributions of spending-relevant traits among participants. This in turn is a function of the program’s targeted population and of differential selection into participation.

[1] We believe that with modifications, this framework can be extended to loan programs. In fact, we suspect this three-category framework may be applicable to programs far beyond college aid.

Similarly, researchers have referenced how programs vary in terms of *universality* (Perna & Leigh, 2018) or *saturation* (Miller-Adams et al., 2023), by which they seem to mean the share of the targeted population that is, via rules, eligible for the scholarship. For instance, Miller-Adams (2015) classes programs as "universal" or "limited" based on the presence or absence of income-based and merit-based eligibility criteria. There are two difficulties here. First, the exclusionary impact of any criterion (e.g., a grade point average or income threshold) depends on the distribution of the relevant quantity within the targeted population. The same criterion could exclude a large share of one population but virtually no one in a second population. Secondly and more fundamentally, the "target population" (e.g., the denominator when calculating eligible share) is itself a function of eligibility rules—rules the analyst has taken for granted. Most often taken for granted are rules restricting geographically (by legal residence) or temporally (e.g., limiting use to a certain period after high school completion). More accurately, programs can be said to vary in terms of the *absolute size* of the potentially eligible population. But even then, program rules could only be used to calculate this in combination with trait distributions within the rule-specified population. Doing even this requires one to disregard programs' procedural eligibility requirements, which condition eligibility on actions by intending participants.

This is to say that rules, by themselves, are invalid measures of programs' most interesting and important characteristics: how much they are reducing college costs, how many people are potentially eligible to use them, and how many take them up.

Regardless, since rules constitute programs, they remain of interest. One set of researchers attempted to measure spending intensity continuously (Miller-Adams et al., 2023) through a subset of the rules we group under "provision". This is consistent with a general tendency among researchers to think about programs as varying along three continua in terms of what they provide, where they can be used, and who can use them (e.g., Bell, 2021). That rules can establish variation along such continua is predicated on a testable empirical

assumption: that there is considerable covariance among conceptually related rules. This in turn arguably implies a further theoretical assumption: that programs are intentionally designed, through their rules, to be variously generous, broadly applicable, and universally available. We investigate these assumptions below, along with an alternative thesis: that provision, applicability, and eligibility are better thought of as *buckets of conceptually similar rules* than as dimensions of variation.

Secondly, we can examine assumptions about why and how programs are designed as they are. Programs are designed by specific people in specific contexts, and "designing" is a set of explicit or tacit decisions generating the program's rules. One theory common in literature advocating these programs is what we term the *altruistic thesis*: programs are designed to optimally meet the specific needs of the targeted population (e.g., Miller-Adams & Iriti, 2022: 2; U.S. Department of Education, 2016). This could be operationalized by stating that program design is a function of the material conditions in the targeted community.

We find this to be less than convincing. First, *community needs* have no objective existence, nor is there ever consensus regarding what they are. Instead, "needs" are the product of both conditions and observers' interpretations of these conditions, as guided by implicit or explicit causal theories and moral judgments. At minimum, program design will be impacted also by the beliefs and assumptions (e.g., political ideology) of designers. Additionally, depending on the sponsoring organization's situation, designers may respond to the (inferred) preferences of actors on whom they are dependent, such as potential donors or political authorities (Monaghan & Michaels, 2025).

Conversely, much research argues that organizations tend to create programs, subdivisions, or initiatives by copying existing designs, generating isomorphism (Boxenbaum & Jonsson, 2017; Ordanini et al., 2008). This occurs in part by default, as designers may not consider that alternatives to certain existing rules are possible. It also occurs because existing models may be perceived as successful or legitimate (DiMaggio & Powell, 1983). Given the

ubiquity and even inevitability of imitation, the relevant question involves not whether this is occurring but along what lines.

Finally, programs design will necessarily take available resources into account. All organizations are subject to finite resources. Organizations sponsoring college affordability programs may also face unpredictable revenues depending on the source, ranging from dedicated endowments (highly certain) to dependence on small private donations (highly unpredictable). To render programs sustainable, designers may aim to set expenditure in conformity with reliable revenues. Total expenditure is the product of average per-participant spending, the size of the eligible population, and the take-up rate among the eligible. Designers can impact the first of these through provision and applicability rules (given that college costs vary by institution), the second through eligibility rules, and the third through communication strategies.

Below, we will test three versions of the design-through-imitation thesis. We will also, via an inductive (clustering) analysis, present suggestive evidence that design is considerably influenced by the need to contain spending within available resources.

Methods

Data

We compiled a list of local[2] college affordability ("Promise" or "free college") programs in existence during the 2022-23 academic year, during which we completed this task. We included programs if they:

1) supply a monetary award or its equivalent (e.g., a tuition waiver) at least potentially equal to or greater than $500.

[2] We excluded fully state programs - those with statewide eligibility, within-sector statewide applicability, and state funding & operation

2) are distributed as an entitlement or until exhaustion of funds.

3) are applicable to at least one public college.

4) are usable by first-time college students, though not necessarily restricted to these students.

5) had a working website sufficiently informative about basic program features during the data collection period.

We first identified programs through databases compiled by the Upjohn Institute (Miller-Adams et al., 2023[3]) and the University of Pennsylvania's Alliance for Higher Education and Democracy (PennAHEAD) (Perna & Leigh, n.d.), College Promise (2021) and WestEd (n.d.), which contained 192, 425, 290, and 131 programs respectively. We removed duplicates and programs failing inclusion criteria, retaining 240 programs. We searched for additional programs online using search terms "Promise program", "college promise", "free college" and "free tuition", each in combination with the names of all fifty states and Washington DC. We identified an additional 74 eligible programs, bringing our total to 314.

We gathered data on program features from program websites, supplemented by information in press releases, news stories, and other sources. Through the data collection process, as we learned more about the ways programs vary, we repeatedly expanded and refined both our feature list and how we coded features, requiring iterative re-coding of previously coded programs. When websites and other sources did not provide explicit

[3] We used an earlier incarnation of this database sent to us upon request in February 2021

information about a feature[4], we assigned default values for that feature relying either on reasonable inference or category modes.

Variables

We sorted program rules into those relating to provision, applicability, and eligibility (Table 1). As eligibility is the largest category, we further divided it into rules relating to *initial* and *continuing* eligibility. The former condition ever receiving any award, and the latter condition continued receipt after the first installment into subsequent periods (i.e., later semesters or years).

We begin with our eleven provision variables[5]. *Relationship to other grants* measures whether a program reduces its award based on other grants received (e.g., Pell grants). Those that do not are "first-dollar" grants, and those that reduce based on all other grants are "last-dollar"[6]. *Award applicability* measures the items which program dollars may cover, fully or in part. *Minimum awards* are dollar amounts disbursed at minimum to each qualifying student. Without such stipulations, tuition guarantees provide no new funds to students whose eligible expenses are covered by need-based aid (Miller-Adams, 2015:45; Harnisch & Lebioda, 2016:10; Monaghan & Attewell, 2023)[7]. *Book/supply vouchers* may also be provided to all qualifiers. Programs vary in the maximum *number of years covered* (we convert credits and semesters to years presuming two yearly semesters of 12 credits each). *Maximum awards* cap awards, yearly or in total, for individual students. *Scaled awards* pay different amounts depending on some other status; most common is to pay different shares of tuition for students who spent different numbers of years in a targeted school district. Some awards *vary by college*

[4] For example, if no GPA eligibility threshold was referenced, we inferred that there is no such requirement.
[5] More detailed descriptions of all variables appear in Monaghan and Hawke 2025.
[6] In principle, programs can reduce using some but not all other awards. But in practice most so-called "middle-dollar" local programs only exempt private scholarships, which most students do not receive.
[7] This mostly affects low-income students at community colleges.

type, paying different set amounts depending on (for instance) private or public college attendance. *Tuition guarantees* ensure that a student pays nothing in tuition to attend at least one college; we measure this separately for *community colleges* and *four-year colleges*[8]. *Additional services* are extras provided to but not required of eligible students (e.g., priority registration, a bus or parking pass, or a designated advisor).

We identified six applicability rules. *Extensiveness* captures the number of colleges at which the program may be used. *Two-year applicability, four-year applicability*, and *private college applicability* measure to whether a program may be used to attend at least one two-year, four-year, or private college, respectively. *Program applicability* measures whether a grant is restricted to funding specified degree or certificate programs *within* applicable colleges.

(*Insert Table 1 about here*)

We measure eighteen eligibility features, beginning with nineteen that condition initial eligibility. *Availability* measures whether an award is guaranteed to all qualifying students (an entitlement), provided until funds are exhausted, or granted to a set number of recipients. *Geographic restrictions* condition eligibility based on residence within an area and/or attendance in a particular school or district. *Residence length* rules require recipients to have resided in the targeted area for more than a year prior to high school graduation. *Temporal eligibility* rules bound not the number of academic units covered, but the time frame in which one may access the scholarship (e.g., the first two years after high school graduation). *Citizenship* rules restrict eligibility, either explicitly or de facto[9], to those holding some citizenship statuses. Some programs restrict to *first-time, first-year* college-goers. *Income restrictions* condition eligibility on some measure of household resources. There are several ways in which programs restrict based on prior academics. These include requiring certain types of *high school completions*,

[8] To be coded positively, a program must guarantee tuition for at least one applicable college (per level). It doesn't need to guarantee tuition to *any* applicable college.
[9] E.g., requiring Pell eligibility disqualifies undocumented students.

reaching a minimum *high school grade point average (GPA)*, maintenance of a given *high school attendance* rate, exceeding a threshold on some *standardized test* (the SAT or a placement test like COMPASS), and taking some specified *high school curriculum*. Programs may restrict to those with *full time attendance* in their first college semester. They may require the completion of *community service* hours prior to college. They may condition eligibility on some *moral eligibility* criteria (e.g., do not be arrested). Then there are several procedural rules. The commonest is a *FAFSA requirement*, and many programs require the completion of a *program application*. *Early enrollment* programs require students to indicate interest several years prior to college. We also measure the number of *other initial procedural requirements* (e.g., completing a career plan, attending orientation, or meeting with an advisor).

We measure six rules relating to whether students retain an award after the first period (continuing eligibility). These include requiring *continuous enrollment*, completion of *community service* hours each year, and requiring the yearly completion of the *program application. Credit requirements* stipulate a minimum number of credits taken or completed over a period, and *GPA requirements* disqualify those falling beneath a GPA threshold. We also measure the *number of additional procedural requirements* to maintain a scholarship (e.g., attending workshops, meeting with advisors).

Finally, we measure two variables related to program infrastructure. We measure the source of *funding* for the program, and we identify the administering organization type (*administration*).

Analytical methods

We perform three sets of analyses below, with different methods for each. First, we study covariance in conceptually related features across programs. Crucially, we coded all feature within categories uniformly. Larger values (e.g., 1 rather than 0) for provision variables indicate greater provision (e.g., first-dollar rather than last-dollar). Larger values for applicability features indicate broader applicability. Larger values for eligibility variables indicate that a

program is more restrictive (i.e., less universal). Because we measured most rules as either dichotomous (yes/no) or ordinal, factor analysis is not appropriate. Instead, we calculate Chronbach's alpha using Stata's *alpha* command[10]. We require the program to accept signs as coded, rather than automatically incorporating negative relationships by reversing signs. This allows us to capture whether programs vary along spectrums in terms of greater or lesser provision, applicability, and universality. We examine four sets of rules: provision, applicability, initial eligibility, and continuing eligibility, adopting the prevailing wisdom (Spector, 1991) that a minimally acceptable alpha for internal consistency is 0.7. When this is not met, we proceed stepwise by dropping the least-consistent item and recalculating measures of consistency.

The second two analyses require us to measure dissimilarity between programs, which we do by calculating *Jaccard distance* (Jaccard, 1901). This is optimal given that most of our variables are nominal or ordinal. To prepare, we converted variables to asymmetric binary representation[11]. Jaccard distance is useful when a simple mismatch coefficient (SMC) would obscure program differentiation. For example, consider calculating dissimilarity between two shoppers' grocery carts (which we will label *A* and *B*) at a store selling 10,000 items. Given that any shopper will only buy a small share of available items, comparing carts across *all* items will result in extremely small measurements of dissimilarity, even if they contain mutually exclusive item sets. The SMC is calculated as

[10] This is an atypical use of this measurement. But we are essentially validating whether sets of items can be combined into an internally consistent scale, and this is precisely what Chronbach's alpha measures.
[11] For example, a field with three options such as vehicle = {car, truck, bicycle} would be converted into three fields: vehicle_car = {0, 1}, vehicle_truck = {0, 1}, and vehicle_bicycle = {0,1}. Numeric variables were converted into ordinal categories and then converted to asymmetric binaries.
We did not use variables in which most cases fell into a single category, as they would contribute little to differentiation. The variables excluded were: Availability, relationship to other grants, minimum award, vary by college type, program applicability, high school attendance, test scores, curriculum, community service prior, moral eligibility, early enrollment, FAFSA required, community service continuing, program application refiling.

$$SMC = \frac{M_{10} + M_{01}}{M_{11} + M_{00} + M_{10} + M_{01}}$$

Where:

M_{11} is the total number of attributes where A and B are both 1.
M_{01} is the total number of attributes where A is 0 and B is 1.
M_{10} is the total number of attributes where A is 1 and B is 0.
M_{00} is the total number of attributes where A and B are both 0.

In our example, the SMC is dominated by the large size of M_{00}. If each shopper bought 100 items, with only 5 in common, SMC would equal just 0.019. By contrast, Jaccard distance does not include the M_{00} term, focusing the metric on the differences between vectors (cases). It is calculated as:

$$J_d = \frac{M_{10} + M_{01}}{M_{11} + M_{10} + M_{01}}$$

For the grocery carts in question, Jaccard distance would be 0.974. Jaccard distances are symmetric, meaning that the distance from instance 1 to instance 2 is the same as the distance from instance 2 to instance 1.

The second analysis calculates correlation coefficients between program dissimilarity and other measures, which are discussed below. In our third analysis, we use Jaccard distances in hierarchical agglomerative clustering implemented in Python's sklearn.AgglormerativeClustering package (Pedregosa et al., 2011). In this technique, individual elements or clusters are repeatedly joined together until cases are in a singular cluster. We formed clusters using the linkage metric (Anderberg, 1973), which first measures the maximum distance between elements in distinct clusters and then joins the two clusters having the *smallest* maximum distance. We then analyzed clusters using partition trees (PART in Python), which partition the data into categories of a target variable (here, our selected number of derived clusters) using all available values of all available variables, identifying the one which does so optimally.

Results

How should we understand provision, applicability, and eligibility?

We claimed above that college affordability programs are frequently discussed as varying in terms of provision, applicability, and eligibility. There seems to be a prevalent assumption that these terms refer to *dimensions* of variance, and that related rules determine a program's relative placement in the resulting three-dimensional space. If so, then there should be substantial internal consistency among the rules in each of these three categories. If not, these categories are better thought of lists (we prefer "buckets") of rules that are only conceptually related. We measure internal consistency using Chronbach's *alpha*.

We begin in Table 2 by examining *provision*, which consists initially of eleven items. The covariance among these items is so poor that an alpha cannot be calculated, and average interitem covariance is negative. We proceed by dropping the indicator for community college tuition guarantee (arguably one of the most central components from a theoretical and policy perspective) and then the measure of the number of additional services provided, which permits us to finally calculate an alpha (0.09). We continue dropping items one by one in order of their maximizing subsequent consistency, and alpha rises accordingly. However, even with only two items remaining (number of years covered, and four-year college tuition guarantee) and chosen for maximal covariance, we never clear the 0.7 threshold. We conclude that while programs certainly vary in terms of what they provide, it does not make sense to discuss programs as characterized by *greater or lesser* provision (and certainly not more or less "generosity"). Provision, in this set of programs, is a bucket of conceptually related rules, not a cogent dimension of empirical variation.

(*Table 2 about here*)

Next, we turn to *applicability*. There are only five items to be analyzed, and together they generate an alpha of 0.5. We drop a dichotomous variable for two-year program applicability (again, this would seem to be theoretically important), which is negatively related to the remaining items. Alpha rises to the border of acceptability (0.64), and then above it when we drop another rule (program restriction). The remaining three items (extensiveness, four-year applicability, private college applicability) are related closely enough that we feel comfortable discussing this as a dimension of variation. However, we had to drop one theoretically and empirically important item (whether a scholarship is usable at a community college) to attain acceptable coherence. We have, for applicability, something of a dimension with sparsely populated bucket sitting next to it.

There are 19 items relating to *initial eligibility*, for which the initial alpha is 0.47. After dropping six of these items, we clear the 0.70 threshold. This leaves twelve items, or two-thirds of the initial set. However, some of the dropped items are utterly central from theoretical and policy perspectives. The very first item dropped (i.e., that is least consistent with others), measures income-based eligibility restrictions, or whether the grant is need-based. In early discussions of "universality" (e.g., Miller-Adams, 2015), this was (along with merit rules) the only form of restriction discussed. Other items dropped include highly prevalent rules like restriction to FTFY students and to full-time students. Thus, we achieve internal consistency, but at considerable cost in terms of theoretical robustness and policy relevance.

Finally, the six items relating to *continuing eligibility* have low initial consistency (0.45). Dropping items raises consistency, but never above the threshold of acceptability. This is a bucket of rules.

The findings of this analysis are quite clear. The three ways in which programs vary—provision, applicability, and eligibility—are not best discussed as coherent *dimensions* or clear

spectra of consistent variation. Or, at least, existing programs cannot be coherently placed relative to each other in terms of "universality" or "generosity". Among the programs we study, provision and eligibility (and, to a lesser extent, applicability) are buckets into which we sort program rules based on conceptual affinity – rules *relating to* provision, applicability and eligibility. Or, more precisely, we can create *some* of these dimensions, but only by doing considerable violence to initial constructs by eliminating conceptually and empirically important rules.

Do programs copy recent or nearby examples?

We determined in the prior analysis that there is little covariance between program rules measuring conceptually related constructs. That is, programs don't seem built to be more (or less) systematically generous, broadly applicable, or broadly available. Or, empirically, program structures don't covary in this manner in our sample.

We proceed to examine whether programs tend to copy their design from existing models. We examined three specific hypotheses regarding imitation. First, we hypothesized that programs would imitate *recently created* programs. Second, hypothesized that programs would imitate existing programs *in their geographic vicinity*. Third, we examined prototype effects—that programs copied the best-publicized "Promise" programs, the Kalamazoo Promise and the Tennessee Promise. We measured program dissimilarity through Jaccard distances (see above), temporal similarity in terms of years, and geographic proximity through the haversine formula, which uses the latitude and longitude of two locations to calculate the distance between them on a sphere. For each program, we examined relationships only with programs already existing at the time of its launch. We predicted a positive relationship between program dissimilarity and both temporal and geographic distance.

(*Figure 1 about here)*

(*Figure 2 about here*)

We display scatterplots depicting relationships between program similarity and temporal proximity in Figure 1, and between program similarity and geographic proximity in Figure 2. In both cases relationships are positive (as expected), but very weak. Pearson correlations are just 0.05 and 0.15. Correlations with logarithms of both spatial and temporal distance were similarly low, ruling out exponential relationships. This suggests that while designers might have drawn slightly more on recent or geographically proximate models than on more distal examples, these tendencies account for little program variation.

Next, we investigated evidence of prototype effects relating to the Kalamazoo Promise and Tennessee Promise. If these programs exerted strong influence on program design, programs created in the years immediately following their launch should be much more similar in design to them than programs created in prior years. We also expect that prototype effects will gradually fade over time. Operationally, then, we expect to see 1) mean Jaccard distance to the prototype program decreasing sharply in the year of the program's launch or in the subsequent years, and 2) this mean distance gradually rising again over time. We explore this graphically (Figure 3) and formally through a regression model (Table 3). Our formal test models distance to the prototype using an interaction between a dichotomous variable (*post*) equal to one for programs created after the prototype and a linear variable measuring *temporal distance* to the prototype. Prototype effects would be indicated by a negative coefficient for the *post* variable, and a positive coefficient for the interaction term (temporal distance when *post*=1). Since this is a nonrandom sample, we test hypotheses using permutation tests.

Graphically, the pattern for the KP seems to support the hypothesis that it served as a prototype. The Jaccard distance to it dropped slightly from 2005-2007, though there are few data points prior to 2005 for comparison. Still, this suggests that the programs launched immediately after KP's launch cleaved closer to its design than those created previously. Subsequently, except for 2011, the Jaccard distance mostly increased over time, indicating the

fading of prototype effects. However, our formal test shows these patterns to be very weak. Expected relationships are found, but coefficients are small and statistically nonsignificant.

Conversely, a formal test supports our hypotheses regarding the prototype effects of TP. There programs created after TP were considerably more like it than those created prior (the coefficient for *post* being a statistically significant -0.087), and this design similarity faded out over time (coefficient on the interaction being 0.015, also significant). However, graphical analysis complicates this conclusion. Programs created prior to 2010 were relatively dissimilar to TP, and more like KP. However, the tacking towards similarity to TP occurred over 2011-12 *preceding* the launch of TP. Programs remained relatively like TP, and more like TP than KP, through 2017. This suggests that rather than serving as prototype, TP was symptomatic of larger design trends.

Overall, none of our hypotheses regarding program imitation are strongly supported. This does not indicate that program imitation or influence did not occur, but that it did not clearly occur along the lines we specified.

What can we learn through an inductive process?

Local college affordability programs aren't well described as varying along continua of provision, applicability, and eligibility, nor are they strongly based on local or recent models. But there is likely *some* logic guiding program construction. At this point, we opted for an inductive, data-driven approach. Having used all the features in our data to determine inter-program (Jaccard) distances, we used hierarchical clustering to analyze the resulting distance matrix.

This result is depicted in Figure 4[12]. In such analyses, the usual strategy for determining the number of clusters is to examine the distances between clusters merged at each step,

[12] For visual clarity, in Figure 4 we label joins only once the distance between elements was greater than 0.65

looking for a discontinuity—an “elbow” in the resulting scree plot. Prior to and including the join creating three clusters, distance between cluster joined was between 0.02 and 0.03. Moving from three to two clusters crossed a distance of 0.072. For this reason we settled on three clusters.

Clusters alone are not particularly illuminating. What is of interest is how the clusters are differentiated, what distinguishes cases within from those without. That is, we need to understand the best determinants of the resulting classification. To do so, we used the PART classification algorithm in WEKA (Frank & Witten, 1998; Frank et al., 2016). This indicated the following set of rules:

1) **IF** program cannot be used at a four-year institution, **THEN** class = **1**
2) **IF** program cannot be used at a private institution **AND**
   program does not guarantee tuition at a community college, **THEN** class = **2**
3) **OTHERWISE,** class = 3

These rules correctly classified 96.8% (304 of 314) of the programs. These rules are concise and highly accurate, but not very intuitive. Next, we attempted the following alternative:

1) **IF** program can *only* be used at a two-year institution, **THEN** class = **1**
2) **IF** program can *only* be used at a four-year institution, **THEN** class = **2**
3) **IF** program can be used at *four or two-year* institution, **THEN** class = **3**

This rule set successfully classified 96.2% (302/314) of the programs—nearly as successful as that which we discovered through PART, and far more intuitive. In practice, it is nearly logically identical. Programs usable *only* at two-year colleges are also *not usable at* four-year (class 1 in both sets). Virtually all programs which *only* are usable at four-year colleges are funded and managed by individual public colleges and usable only at that college. Thus neither are they usable at any private college nor do they guarantee tuition to a community college (class 2). The remaining programs are usable at both levels (class 3).

This result is highly informative: the best way of sorting programs into maximally differentiated groups in terms of program rules is simply according to sectoral applicability. That is, where a program may be used is the most consequential determinant of the program's overall structure. Put another way, programs most closely resemble other programs usable at colleges in their sector. We quantify this differentiation in Table 4.

Programs usable at *both two-and four-year colleges* have some distinctive traits. To make sense of them, we suggest that the reader look to the bottom of Table 3. Such programs are nearly all privately financed and operated; they are philanthropic endeavors. That they finance *four-year* attendance suggests that they are *well-resourced* philanthropic endeavors (relative to privately financed programs restricted to two-year colleges). Thus, they *provide more* according to some measures than other programs; they are most likely to be first-dollar awards (though, still, most are not) and to cover something beyond tuition and fees (though, again most don't). They even sometimes permit students to enroll at *private* colleges. But since fully covering private college tuition wouldn't be sustainable, such programs introduce cost-containment measures rarely seen at other colleges. They impose maximum awards, vary awards by college type (private vs. public, usually), and are least likely to guarantee tuition coverage. The originating philanthropies tend to be linked to *places* (usually municipalities), seeking to benefit those places' populations (i.e., they are Miller-Adams' (2015) "place-based scholarships"). Thus, they involve tight residential restrictions. Why they are more likely to have merit-based requirements isn't quite so clear to us. It is possible that they conserve funds by restricting to applicants more likely to complete degrees.

The *four-year only* programs are nearly all single-institution programs financed and operated by the applicable college, and much of their design follows from this. Most are better understood as categorically determined tuition and fee waivers than as "scholarships", since they typically consist of a systematic redirection of existing institutional aid. They are mostly full tuition-and-fee waivers. Since they are run by *state* colleges, they usually are available to state

residents. They do not typically require program applications as they are granted automatically to qualifying admitted students by the college's financial aid office. Since these are four-year colleges, standard admissions procedures enact sufficient academic selection that *program* merit restrictions are redundant. We suspect that temporal restrictions are uncommon for similar reasons. To conserve funds and maximally leverage federal grants, these tend to be last-dollar programs restricted by income. They also tend to incentivize momentum towards completion, requiring full-time attendance, continuous enrollment, and meeting GPA and credit-accumulation thresholds. Such rules would be seen as boosting completion rates, but would also save money by dropping those not making progress.

Finally, *community college only* programs are the most heterogeneous in terms of financing. The heavy presence of government financing is owing mostly to programs in California and Michigan. California, through AB19 of 2017, provides modest funding enabling individual community colleges to set up their own "Promise" programs. Michigan, through the Promise Zone Authority Act (2009), establishes mechanisms to partially fund affordability programs through property tax recapture. These two states account for 85% of government-financed community college programs.

Perhaps the best way to grasp these programs is to suppose them to be financed by an entity with little to spend (or, in the case of states, wishing to spend little): a small philanthropy, a local government, or a community college. Community college tuition is relatively cheap, particularly for in-district students. Nearly all these programs (except those in California[13]) restrict to residents of areas smaller than states. Despite the low cost, this group is more likely than others to only cover tuition, and they are nearly all last-dollar programs. These rules enable the programs to guarantee tuition coverage through maximally leveraging Pell and state

[13] In California, most programs are funded by state funds (through AB19). Additionally, California community colleges do not distinguish between in-state and in-district students when determining charges.

grants, minimizing the amount of new award money required (Monaghan & Attewell 2023). As they finance attendance of low-cost and open-admission colleges, they are light on academic and income restrictions. They contain costs through temporal restrictions, requiring full-time attendance and continuous enrollment, and imposing GPA and credit accumulation minimums.

Having found these clusters of programs, we retested hypotheses regarding spatial and temporal imitation. Perhaps programs imitated recent or geographically proximate programs, but only those which were going to be used at similar types of colleges. That is, we tested for correlations between program dissimilarity (Jaccard distance) and temporal and geographic distance separately within clusters. Just as before, correlations were small. For two-year only programs, correlations were 0.31 (geographic) and 0.11 (temporal). For four-year only programs, these were 0.01 and -0.24. And for programs usable at both levels, they were 0.058 and -0.036. Thus there is no strong evidence of temporally- or geographically-structured imitation within program groups.

Discussion

The analyses in this paper shed light on the design of college affordability programs, building on earlier work (Monaghan & Hawke, 2025) detailing the many ways in which programs vary empirically. Our analyses are highly empirical and largely descriptive. Two are deductive, testing specific hypotheses testing, and one is data-driven and inductive. They represent to date the most detailed and systematic analysis of the design of college affordability programs.

The first analysis tested a prevalent assumption, which we treated as an implicit set of hypotheses. Researchers and policymakers often refer to programs as *more (or less) generous*, *more (or less) broadly applicable* and *more (or less) universal*. The assumption is that programs can be discussed coherently in these terms. That is, it is presumed that terms like “generosity” and “universality” refer to dimensions along which programs can be arranged relative to each other. Attempts at operationalization have utilized program rules. This presumes that program rules are valid to measure these latent constructs. It also seems to

assume that, in terms of such rules, programs really do vary in these ways—more or less generous, universal, and broadly applicable.

We converted these assumptions into hypotheses which could be tested. Compared to prior efforts, we use a much more comprehensive set of program features. For each feature set, we tested the research hypothesis that, across programs, there would be considerable covariance among features. Our "critical value" was the standard Chronbach's alpha of 0.7. We found that, with the partial exception of applicability, programs do not empirically vary in the manner that is broadly assumed. That is, programs aren't (for instance) so much *more or less* universally available as they make *different sets of decisions* about whom to include and to exclude. That is, program variation is much more complex than dimensional descriptions presume.

Part of the difficulty here, of course, has to do with measurement. By "generosity", commentators typically implicitly mean expenditure, both overall (extensive) and per recipient (intensive). And, of course, programs absolutely do vary dimensionally in these ways. However, directly measuring extensive and intensive spending requires access to program administrative data, which researchers usually lack. Provision rules, even a large set of them, are a poor proxy for spending. Indeed, given the large variation in pricing across colleges, we strongly suspect that spending is more directly determined by some applicability rules than by provision rules.

There are similar problems with universality. Programs bound eligibility in many, many ways. How even to determine the baseline program population isn't clear (i.e., the denominator for any measure of eligible share). Probably the only valid comparative metric would be the *number* of potentially eligible recipients, and this would depend not only on program rules but on the covariance of traits (e.g., income and high school GPA) within populations delimited by other rules (e.g., geographic and temporal restrictions).

There are two important takeaways here. First, researchers and others should use caution in comparing programs without sufficient data to do so. This obviously applies to claims about the relationship between program differentiation and program outcomes – that "more generous programs have stronger effects". Second, programs are put together in terms of rules in interesting ways, and how exactly this happens is a matter we know very little about yet. Program rules don't sort programs along dimensions of greater or lesser provision, applicability, or universality. But they are the result of specific decisions made by specific sets of people, and there are probably *reasons* for why these decisions are made

This brings us to our subsequent analyses. In our second set of analyses, we tested the hypothesis that maybe many decisions about how to structure programs are made by default, by simply copying other programs. There are, after all, many decisions to be made, and it is unlikely that individual program committees will consider *all* the different ways that they could structure their program. Rather than logically deriving all the choices they must make, designers begin by considering existing models and selectively making alterations. Organizational researchers have long pointed to such widespread mimetic practices.

We tested two specific hypotheses relating to program design through imitation. First, we supposed that programs would copy other recently designed programs. Our intuition here was that, over the history of "Promise" programs, new programs would copy earlier ones but selectively innovate. Subsequent programs would have an expanded set of models to copy and would innovate based on these, resulting in relative temporal covariance. Second, we supposed that programs would copy those in their geographic proximity. The intuition here was that programs would be most heavily influenced by organizations in their immediate vicinity, either through interorganizational networks or because they consider neighbors to be relevant comparisons.

We examined these by measuring the correlations between program dissimilarity (Jaccard distance) and measures of temporal and geographic distance. We did found very

small temporal or geographic dependence among the cases in our sample. This suggests that, at most, copying of the sorts we tested for was of minimal importance in program design. This does not rule out the importance of imitation overall, of course. It may be that programs copied well-known prototypes (e.g., in Kalamazoo or in Tennessee), or that inter-program links of influence were established through initiatives like PromiseNet.

In our final analysis, we switched from deductive hypothesis testing to the inductive study of patterns in data. Having measured the dissimilarity between programs, we clustered them into three groups. We then analyzed the determinants of cluster membership through a machine learning procedure (partition trees). This led us to the realization that in fact programs sorted into three classes very well simply based on the sorts of colleges at which they may be used: 1) only community colleges, 2) only four-year colleges, and 3) both two- and four-year colleges.

We then proposed that the reasons for this differentiation related to 1) the difference in tuition between community colleges and four-year colleges, and 2) the resources of the organizations sponsoring the programs. Four-year only programs are almost entirely sponsored by four-year colleges, and we suspect that these programs mostly constitute a re-allocation of existing pools of institutional aid. Four-year colleges, through these programs, regularize how they waive tuition (and fees) along specific predictable lines. They use eligibility and other rules to limit and target this spending. Programs usable at two- and four-year colleges are usually funded by wealthy private philanthropies tied to specific geographic communities. These funders are usually indifferent to precisely where a student attends, and they have the resources to finance (public) four-year attendance. But even these programs limit expenditure through eligibility and provision rules. Finally, two-year programs are adopted by community colleges themselves or entities (private or public) with fewer resources they can commit to the program. It is for these reasons that the sector where a program can be used can tell us a fair amount about other rules it adopts.

Conclusion

In short, having introduced considerable complication into how we think about the structure of college affordability programs, we move here towards greater simplicity. We suggest that program design is greatly influenced by available resources. Programs seek to limit spending to not exceed resources, and they do so by 1) controlling per-student expenses and 2) limiting the number of eligible potentially eligible recipients. Both are accomplished through program rules. Intensive spending is limited through provision and applicability rules, and extensive spending through eligibility rules. We postulate that these considerations are central to those planning any college affordability program, and indeed to any aid program more generally.

However, this does not mean that controlling spending is the only relevant consideration. Programs likely have explicit and implicit policy goals related to outcomes. They likely wish to promote the reputation and image of the sponsoring organization. They may wish to court, or at least avoid angering, well-resourced public and private actors. They may wish to limit organization time and effort by copying the design of an existing program.

In this regard, it is relevant that programs have numerous options for how to limit intensive and extensive spending. These options are the various rules of the program. Thus, eligibility rules may limit "target" spending in a manner that advances designers' other goals.

It is also relevant that decisions on some rules are highly impactful in terms of costs, triggering the need to employ countervailing mechanisms. This is the reason why four-year applicability and first-dollar design tend to be paired with award caps and income-based eligibility rules. And these compensatory tendencies are why it is so difficult to determine how "generous" or "universal" a program is through rules alone. Program designers strategically expand apparent generosity and universality with some rules only to retract them with others, making global comparisons nearly impossible. Programs really do spend different amounts per

student, and they fund a larger or smaller number of recipients. But these two “real” dimensions can only be wildly guessed at wildly through program rules.

There is much remaining to be learned. We suggest that this will be best accomplished by gathering as much data as possible for individual cases and for larger sets of cases--particularly direct measures of program spending—and by making as few assumptions as possible.

References

Anderson, C. (2021). Local-level, place-based scholarships: A review of the literature. *Educational Review*, 73(5), 638-661.

Anderberg, M. (1973). *Cluster analysis for applications.* New York: Wiley.

Bell, E. (2021). Does free community college improve student outcomes? Evidence from a regression discontinuity design. *Educational Evaluation and Policy Analysis*, *43*(2), 329-350.

Boxenbaum, E., & Jonsson, S. (2017). Isomorphism, diffusion and decoupling: Concept evolution and theoretical challenges. *The Sage handbook of organizational institutionalism*, *2*, 77-101.

College Promise (2020). *Evolution of Promise: The evidence behind the movement: College Promise reference guide*. College Promise.

College Promise (2023). *College Promise Catalogue of Local and State Programs.* College Promise.

Custer, B. D., & Akaeze, H. O. (2021). A typology of state financial aid grant programs using latent class analysis. *Research in Higher Education*, 62, 175-205.

de Mendoza y Ríos, J. (1801). *Memoria sobre algunos métodos nuevos de calcular la longitud por las distancias lunares: y aplicacion de su teórica á la solucion de otros problemas de navegacion (in Spanish).* Madrid, Spain: Imprenta Real.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American sociological review*, *48*(2), 147-160.

Frank, E., Hall, A. & Witten, I. (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques". (4th). Morgan Kaufmann.

Frank, E. & Witten, I. (1998). Generating Accurate Rule Sets Without Global Optimization. *Fifteenth International Conference on Machine Learning*, (pp. 144-151).

Harnisch, T.L., & Lebioda, K. (2016). *The promises and pitfalls of state free community college plans*. American Association of State Colleges and Universities.

Iriti, J., Page, L. C., & Bickel, W. E. (2018). Place-based scholarships: Catalysts for systems reform to improve postsecondary attainment. *International Journal of Educational Development*, 58, 137-148

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société vaudoise des sciences naturelles, 37*, 547–579. Retrieved from https://doi.org/10.5169/seals-266450

Jones, Tiffany, Ramirez-Mendoza, Jaime, and Victoria Johnson. (2020) *A Promise worth keeping: An Updated equity-based framework for free college programs.* EdTrust

March, J.G. & Simon, H.A. (1958). *Organizations*. Wiley.

Miller-Adams, M. (2015). *Promise nation: Transforming communities through place-based scholarships*. W.E. Upjohn Institute for Employment Research.

Miller-Adams, M. (2021). *The path to free college: In pursuit of access, equity, and prosperity*. Harvard Education Press.

Miller-Adams, M., Hershbein, B., Timmeny, B., McMullen, I., & Huisman, K. (2023). *Promise programs database*. W.E. Upjohn Institute for Employment Research.

Miller-Adams, M. & J. Iriti (2022). *The free college handbook: A practitioner's guide to promise research*. W.E. Upjohn Institute for Employment Research.

Millet, C.M., Saunders, S.R., Kanter, M.J., & Heistand, R. (2020). Setting the stage: A history and overview of the college Promise movement. In L.W. Perna & E.J. Smith (Eds.) *Improving research-based knowledge of college promise programs* (pp. 17-30). AERA.

Mishory, Jen. 2018. *The Future of Statewide College Promise Programs*. The Century Foundation. March 6.

Monaghan, D. B. (2025). How powerful are promises? A comprehensive review of the causal mechanisms and outcomes of 'free college' programs in the United States. *Policy Reviews in Higher Education*, *9*(1), 47-80.

Monaghan, D. B., Almanzar, C., Laughman, M., & Ritchey, A. (2024). *What Is a College" Promise" Program? The Creation and Transformation of a Concept, 2005-2022.* EdWorkingPaper No. 24-985. Annenberg Institute for School Reform at Brown University.

Monaghan, D. B., & Attewell, P. A. (2023). College promise programs as symbolic politics. *Sociology of Education*, *96*(4), 301-323.

Monaghan, D. B., & Hawke, E. A. *What are Promises Made of? The Design of Local College Affordability Programs.* EdWorkingPaper No. 25-1169. Annenberg Institute for School Reform at Brown University.

Monaghan, D. B., & Michaels, E. (2025). Empty Promises: Why and How Community Colleges Exclude Undocumented Students from "Free College" Promise Programs. *The Sociological Quarterly*, *66*(2), 411-432.

Ordanini, A., Rubera, G., & DeFillippi, R. (2008). The many moods of inter-organizational imitation: A critical review. *International Journal of Management Reviews*, *10*(4), 375-398.

Pedregosa, F., Varoquaux, G., Gamfort, A.,… & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830. Retrieved from https://dl.acm.org/doi/10.5555/1953048.2078195

Perna, L. W., & Leigh, E. W. (2018). Understanding the promise: A typology of state and local college promise programs. *Educational Researcher*, 47(3), 155-180.

Perna, L.W. & Leigh, E.W. (n.d.). *Database of College Promise Programs*. University of Pennsylvania Alliance for Higher Education and Democracy.

Perna, L. W., Rowan-Kenyon, H., Bell, A., Thomas, S. L., & Li, C. (2008). A typology of federal and state programs designed to promote college enrollment. *The Journal of Higher Education*, 79(3), 243-267.

Spector, P.E. (1991). *Summated rating scale construction: An introduction*. Sage

Swanson, E.., Watson, A.R., & Ritter, G.W. (2020). Promises fulfilled? A systematic review of the impact of college promise programs. In L.W. Perna & E.J. Smith (Eds.), *Improving research-based knowledge of college promise programs,* (pp. 33-68). AERA.

Upjohn Institute (2021). *Promise research bibliography*. W.E. Upjohn Institute for Employment Research.

U.S. Department of Education (2016). *America's college promise playbook: Expanding the promise of college education and economic opportunity for all students*. U.S. Department of Education.

WestEd (n.d.). *Active college promise programs in California.* https://californiacollegepromise.wested.org/programs/

Willard, J., Vasquez, A., & Lepe, M. (2019*). Designing for Success: The Early Implementation of College Promise Programs. MDRC*.

Figures

Figure 1: Jaccard distances between programs by years separating program launch

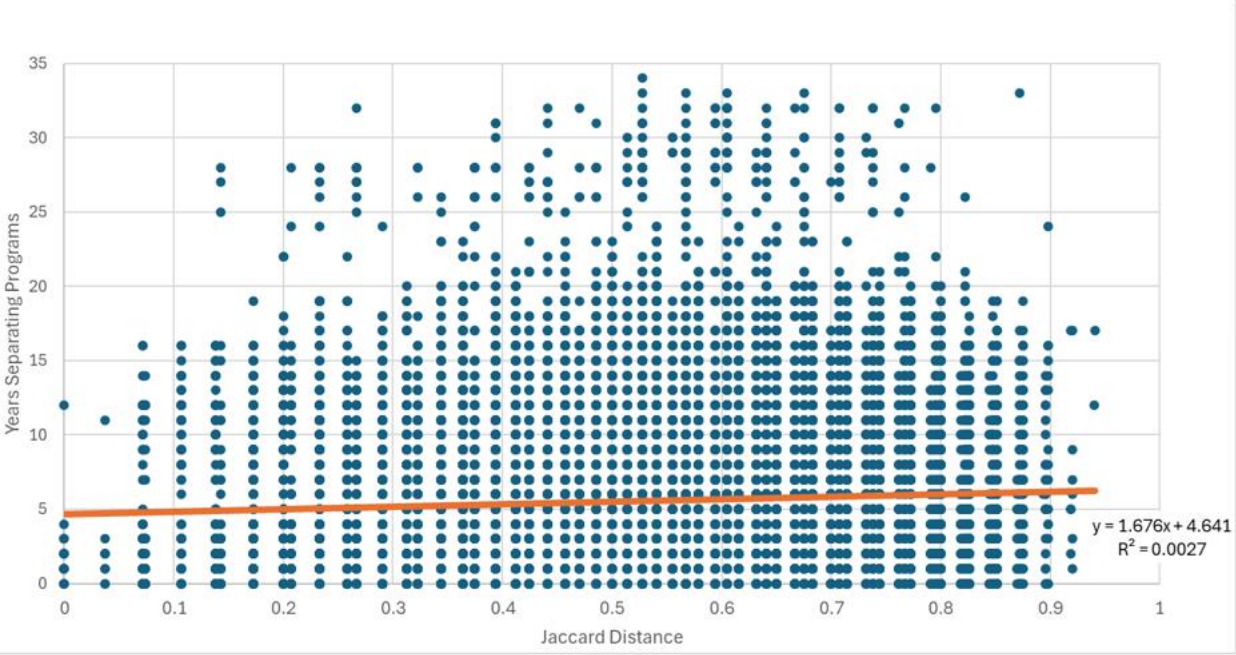


Figure 2: Jaccard distances between by physical distance between programs

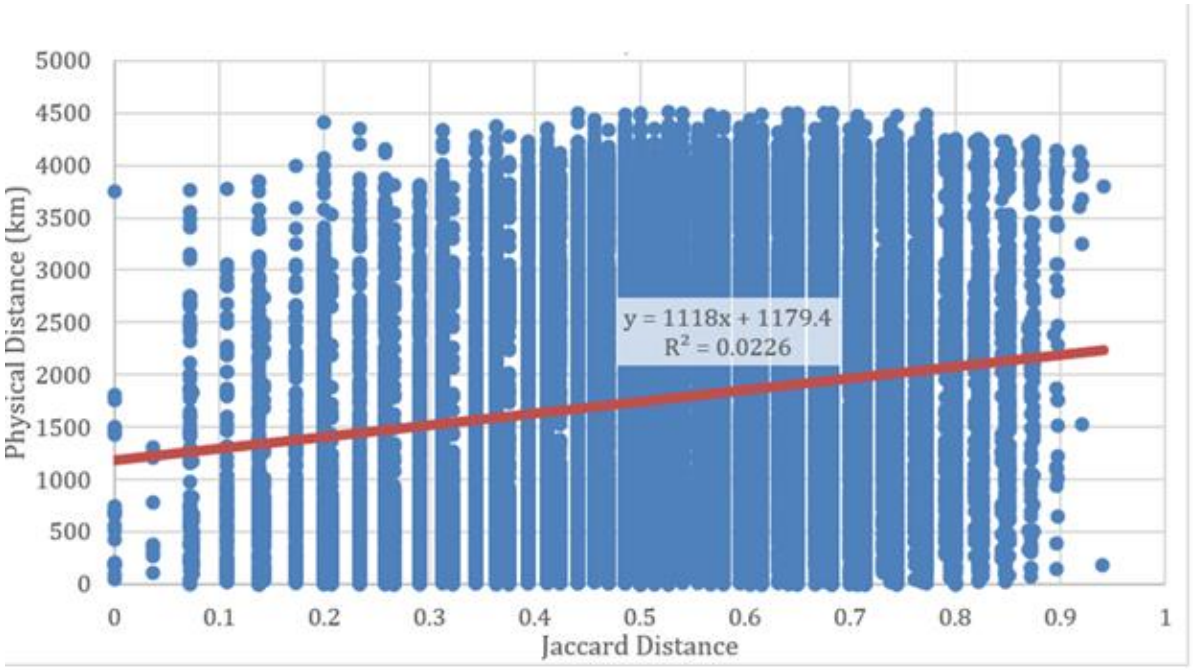


Figure 3: Mean Jaccard distance to the Kalamazoo Promise (KP) and the Tennessee Promise (TP), by year of program creation

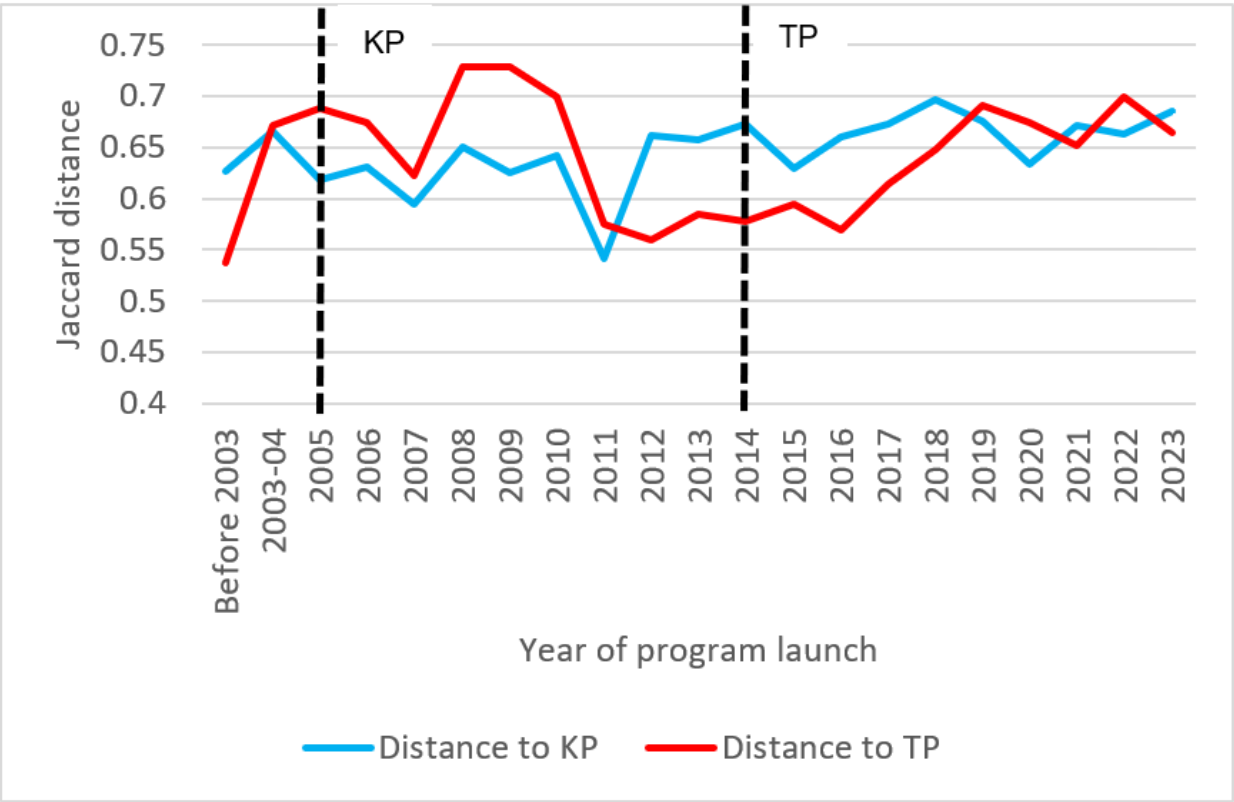
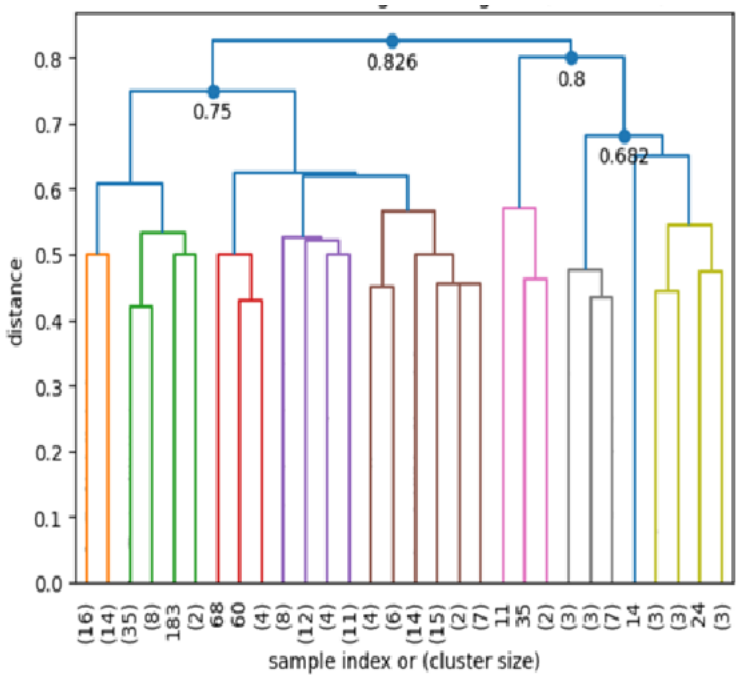


Figure 4: Hierarchical agglomerative clustering of Jaccard distances between programs



Tables

Table 1: Feature list with aspects of measurement

Feature	Measurement	Categories (ascending order)
Provision		
Relationship to other grants	Binary	Last-dollar, first dollar
Award applicability	Ordinal	Tuition, tuition & fees, beyond tuition & fees
Minimum award	Binary	No, yes
Book/supply voucher	Binary	No, yes
Number of years covered	Numeric	Equivalent half-year increments
Maximum award	Binary	Yes, no
Scaled award	Binary	Yes, no
Variance by college type	Binary	Yes, no
Tuition guarantee, community college	Binary	No, yes
Tuition guarantee, four-year college	Binary	No, yes
Number of aditional services	Numeric	Integer increments
Applicability		
Extensiveness	Ordinal	One college, more than one but less than all in-state, all in-state, more than in-state
Two-year college applicability	Binary	No, yes
Four-year college applicability	Binary	No, yes
Private college applicability	Binary	No, yes
Program applicability	Binary	Yes, no
Initial eligibility		
Availability	Ordinal	Entitlement, funds-limited, numerically restricted
Geographic restriction	Ordinal	Greater than in-state, state residency, locality restriction
Residence length restriction	Binary	No, yes
Temporal eligibility	Ordinal	Greater than five years, five years or less, immediate enrollment
Citizenship restriction	Ordinal	None, explicit DACA inclusion, implicit DACA inclusion, implicit restriction to citizen/LPR, explicit restriction to citizen/LPR
Restriction to first-time, first-year	Binary	No, yes
Income restriction	Ordinal	None, threshold greater than Pell, Pell or equivalent, threshold less than Pell

High school completion restriction	Ordinal	None, regular high school completion, on-time completion
High school GPA restriction	Ordinal	None, 2.0, between 2.0 and 3.0, 3.0 or greater
High school attendance restriction	Binary	No, yes
Test threshold restriction	Binary	No, yes
High school curriculum restriction	Binary	No, yes
Full-time attendance requirement	Binary	No, yes
Community Service requirement	Binary	No, yes
Moral eligibility criteria	Binary	No, yes
FAFSA requirement	Binary	No, yes
Program application	Binary	No, yes
Early enrollment	Binary	No, yes
Other required initial procedures	Numeric	Integer increments
Continuing eligibility		
Continuous enrollment requirement	Binary	No, yes
Community service requirement	Binary	No, yes
Program application requirement	Binary	No, yes
Credit requirements	Ordinal	None, less than 12, 12, more than 12
GPA Requirement	Ordinal	None, 2.0, greater than 2.0
Other required procedures to retain	Numeric	Integer increments
Infrastructure		
Funding	Nominal	Private, government, college
Administration	Nominal	Private, government, college

DACA=Deferred Action for Childhood Arrivals; GPA=Grade point average

Table 2: Analysis of feature covariance by category

#items	Feature dropped	Share excluded	Average covariance	Cronbach's Alpha
Provision				
11	None	0	-0.007	NC
10	Tuition guarantee (community college)	0.09	-0.001	NC
9	Additional services	0.18	0.003	0.09
8	Book/supply voucher	0.27	0.008	0.19
7	Maximum award	0.36	0.017	0.29
6	Scaled award	0.45	0.031	0.37
5	Varies by college type	0.55	0.053	0.43
4	Minimum award	0.64	0.083	0.45
3	Award applicability	0.73	0.119	0.45
2	Relation to other grants	0.82	0.119	0.64
	Applicability			
5	None	0	0.035	0.49
4	two-year applicability	0.2	0.070	0.64
3	Program restriction	0.4	0.137	0.71
2	four-year applicability	0.6	0.198	0.75
	Initial eligibility			
19	None	0.00	0.016	0.47
18	Income	0.05	0.02	0.53
17	FTFY only	0.11	0.026	0.58
16	Additional initial procedures	0.16	0.029	0.62
15	Full-time attendance	0.21	0.037	0.67
14	Citizenship	0.26	0.031	0.69
13	FAFSA	0.32	0.037	0.71
12	HS GPA	0.37	0.031	0.71
11	HS test	0.42	0.036	0.72
10	Availability	0.47	0.041	0.72
9	HS curriculum	0.53	0.5	0.73
8	Moral eligibility	0.58	0.5	0.73
	Continuing eligibility			
6	None	0	0.041	0.45
5	Additional retain procedures	0.17	0.047	0.49
4	Refile application	0.33	0.076	0.53
3	Community service	0.50	0.139	0.58
2	GPA threshold	0.67	0.198	0.60

FTFY=first time, first year; FAFSA-Free application for federal student aid; GPA-Grade point average; HS=high school

Table 3: OLS regression of Jaccard distance to Kalamazoo Promise (KP) and Tennessee Promise (TP)

	Distance to KP	Distance to TP
Post	-0.042	-0.086**
	(0.049)	(0.031)
Temporal distance (abs)	-0.002	0.001
	(0.008)	(0.003)
Post*Temporal distance	0.006	0.015**
	(0.007)	(0.005)
Constant	0.658	0.659
N	313	314
R2	0.037	0.071

Permutation *p* values **$p<0.01$

Table 4: Feature means by program applicability groups. Estimates in italics indicate significant difference from the two other applicability group means ($p<0.05$)

	Either two- or four-year (N=41)	Four-year only (N=79)	Two-year only (N=194)
Provision			
First-dollar	*21.95*	5	4.15
Tuition only	21.95	25	*40.93*
Tuition & fees	*34.15*	*63.75*	*43.01*
More than T&F	*43.9*	11.25	16.06
Maximum award	*58.54*	10	8.29
Scaled award	*43.9*	5	5.7
Vary by college type	*31.71*	1.25	0
Tuition guarantee: any college	*53.66*	88.75	91.19
Applicability			
Single institution	*0*	92.5	92.75
Any private college	*80.49*	1.25	1.55
Initial eligibility			
Local residence	*97.56*	*11.25*	*61.66*
Residence length	*58.54*	*6.25*	*20.73*
Temporal restriction	73.17	22.5	52.33
Citizenship restriction	63.41	58.75	*38.34*
FTFY	*2.44*	32.5	25.91
Income threshold	*24.39*	*87.5*	*10.88*
HS grad restriction	*80.49*	*30*	*43.52*
HS GPA requirement	*41.46*	25	26.42
Any academic requirement	*80.49*	37.5	48.7
Full-time attendance	63.41	*87.5*	61.14
Program application	*85.37*	*15*	*59.07*
Continuing eligibility			
Maintain GPA	73.17	*91.25*	68.39
Complete credits	*58.54*	*92.5*	*76.68*
Require continuous enrollment	*43.9*	*86.25*	*71.5*
Funding			
Private	*78.04*	*5.06*	*23.19*
College	*2.43*	*84.81*	*25.77*
Government	*19.51*	*10.12*	*51.03*
Operation			
Nonprofit	*82.92*	*0*	*10.31*
College	*7.31*	*94.93*	*84.53*
Government	*9.75*	5.06	5.15

T&F=Tuition & fees; FTFY=first-time, first-year student; HS=high school; GPA=grade point average