ASSOCIATION OF AMERICAN UNIVERSITIES

February 3, 2015

The Honorable Arne Duncan
Secretary of Education
U.S. Department of Education
400 Maryland Ave, SW
Washington, DC 20202

Dear Secretary Duncan:

The Association of American Universities (AAU), a nonprofit association of 60 U.S. and two Canadian preeminent public and private research universities, is responding to the Department of Education's draft ratings framework, entitled "A New System of College Ratings — Invitation to Comment." We applaud President Obama's focus on college access and affordability at a time when Americans are concerned about the cost of a college education, the impact of student loan debt, and future employment prospects. AAU member institutions strive to provide the highest quality education possible and to make it affordable through a range of cost-savings measures, substantial institutional financial aid programs for low- and middle-income students, and creative use of technology.

AAU has submitted two previous sets of comments on the proposed rating system: on December 2, 2013 and January 24, 2014.[1] This response builds upon those comments. At the highest level, we continue to question whether the three stated objectives of the rating system — information for students and families, benchmarking for institutions, and accountability for federal student aid spending — can be met with a single tool, and whether a rating system is the best or most appropriate way to meet any of the objectives. We do not endorse a new rating system, and we are concerned that a rating system does not reflect the diversity of higher education institutions and their missions, and may confuse and mislead prospective students.

Providing students and families with clear, accurate, and useful information about higher education institutions is an appropriate federal role, given the significant investment the U.S. government makes in student financial aid. Likewise, the federal government has an interest in identifying bad actors among higher education institutions. But it is inappropriate for the federal government to rate institutions. Moreover, unlike the many college rankings and ratings that have proliferated in the media, this rating system would have the imprimatur of the federal government and the unbiased and authoritative image it conveys. As such, the potential for harm is considerable.

The Department's framework does not yet provide all the detail needed to evaluate a specific rating system proposal, and in this response we highlight some areas where tough choices must be made. We urge the Department to provide an opportunity for

[1] Available at http://www.aau.edu/WorkArea/DownloadAsset.aspx?id=14835 and http://www.aau.edu/WorkArea/DownloadAsset.aspx?id=14930, respectively.

public comment on a full and complete rating proposal before such a system goes live, and — given the timing — we do not believe it is wise to rush any rating system to completion before the start of the 2015-2016 academic year.

We would like to offer some comments and suggestions on the rating system generally, and then some comments and suggestions on data and metrics.

Comments on the rating system

We believe that the issues of rating categories, institutional groupings, and the available level of customization on the tool — all issues the Department has requested feedback on — are closely interrelated. The Department plans to divide institutions into three performance categories. It is not clear from the framework document how the Department plans to create these divisions. Will they be created relative to the total number of institutions, ranges in the metrics (individual or somehow combined), or some extrinsic factor? Will the divisions be equal or unequal?

Once divided, institutions will be categorized as high-performing, low-performing, and those falling in the middle. Such division and categorization may amplify small and arbitrary distinctions, while hiding most variation within each group.

Currently, the Department plans to separate four-year and two-year universities in presenting the ratings, and has asked for feedback on what other groupings might be used. Considering four-year universities alone reveals systematic variation in possible metrics (e.g., percent Pell, graduation rates, average net price) that are sector-dependent. It's not clear what will be achieved by grouping publics, private non-profits, and for-profits together and obscuring this variation. But dividing institutions into too-granular groupings, while maintaining the current proposed categorization, risks classifying high-quality institutions as "poorly-performing" (we presume, based on the framework, that institutions' ratings will vary depending upon the institutional grouping used).

The "best" institutional groupings for the Department to include depend upon the use to which the ratings will be put. Consumers, policymakers, and institutions themselves will all care about different groups of institutions. More customization of metrics and weights will likely be useful for consumers, but may change institutional results, making them relevant only in the context of that particular customization, and not necessarily otherwise descriptive of the institution. Chances are that any institution, with the right institutional grouping and metric weights, could be classified as "poorly-performing."

Consistent with our previous sets of comments, we urge more information be provided in any rating system. Providing disaggregated data would be more useful for consumers and institutions, and would avoid the tricky issues associated with weighting and combining metrics.

We support looking at institutional improvement over time, but it's unclear how this will work in the context of rating categories. An institution could exhibit a great deal of improvement without moving between categories. Alternatively, a small change may shift them. Even no change could move an institution when categorizations are based upon relative standing or performance. And top-performing institutions may have limited opportunity for improvement, for which they should not be penalized.

We appreciate the Department's recognition that annual fluctuations in the values used in any rating system may provide a misleading picture. Averaging values across years may help address these issues.

Finally, we would like to respond to the section of the framework about adjusting outcomes for student and institutional characteristics. We oppose such adjustments. We believe that adjusting outcomes is antithetical to the consumer information objective of the rating system. As noted in the framework itself, there are inherent dangers in adjusting outcomes for institution-level characteristics (such as state funding, endowment size, etc.). Rather than increasing transparency, adjusting outcomes risks turning the ratings system into a black box that neither consumers nor institutions fully understand. It is difficult to see how institutions would gauge improvement, or consumers understand their choices, if a complex regression model is first applied to student or institutional characteristics. (An adjustment that allowed an institution to compare itself to others while "factoring out" differences in student body socioeconomics, demographics, and academic characteristics is far beyond what existing data will allow and what the Department is considering, and such adjustment would not be appropriate for a consumer information tool.) Any student-level regression model, if based on FAFSA information, runs into the limitations we discuss in more detail in the next section. Such data may not be representative of the entire undergraduate student body, and may represent the student body at different institutions in different ways. It would be essential to understand and correct for biases in FAFSA data before using such a model. On the institutional side, including factors like endowment size in the regression would penalize schools who may be using that wealth for financial aid … this is especially problematic given the issues with the data sources like FAFSA, NSLDS, and IPEDS when it comes to representing students who do not receive federal aid. We also question whether the regression model would change each time data or metrics were updated. If so, results could be unstable and difficult to compare between years. But if not, older models may not be best-suited to current data.

Comments on data and metrics

The Department is considering using some existing metrics (e.g., percent Pell, graduation rates, and average net price) as well as some new metrics. The limitations in existing metrics are well-known. A significant portion of the variables in IPEDS are restricted to include only certain students, such as first-time full-time or those receiving Title IV financial aid. One concern is that the whole undergraduate student body is not represented in these variables. A deeper concern, especially relevant in the context of a rating system, is that not all institutions are represented equally. Institutions with a relatively small proportion of students receiving federal aid — even if the cause is a particularly generous institutional aid policy — may not be accurately reflected in these variables.

The Department's effort to think about how to use existing data sources, such as NSLDS and the FAFSA, to construct new and potentially useful metrics is laudable. Here, however, the concern is amplified. These data sources are also restricted to certain types of students. FAFSA completion rates vary by sector and level of institution,[2] and even for those who fill it out, FAFSA data are only tied back to particular institutions for Title IV aid recipients. Biases in the data from these sources are even less well-understood, to our knowledge, than those in IPEDS variables. The Department itself recognizes some of these limitations in the framework, which we appreciate. We understand the Department is on an expedited timeframe to establish a rating

[2] http://www.finaid.org/educators/20091014fafsacompletion.pdf

system, but putting new metrics to use before they have been adequately studied risks doing a disservice both to consumers and to institutions. Generally, we continue to encourage the Department to allow institutions to review their reported data before a rating system is published.

Below we include comments about specific metrics under consideration:

Percent Pell

Percent Pell is the most complete, accurate, and well-understood of the metrics the Department is considering. That being said, it is worth considering looking at both percentages and numbers of Pell grant recipients. We believe that these figures should be compared to the overall population of Pell-eligible students, not necessarily all undergraduates. Looking at Pell recipients also provides a narrow view of access, in that it potentially ignores a large population of low-income, middle-income, and first-generation students who should be part of any access equation. Many students who do not qualify for Pell grants still face considerable financial need. Information on institutional need-based aid should also be included, as it demonstrates institutional commitment to access and affordability. Some contextual information (e.g., comparisons to state or regional demographics) would give Pell grant figures more meaning. Finally, we would like to point out that the number of Pell recipients is limited, and by including this metric in a rating system, the Department is essentially encouraging institutions to compete for them.

EFC (Expected Family Contribution) gap and *Family income quartiles*

These are, as far as we can tell, new and untested metrics that rely on data from FAFSA and NSLDS. EFC, in particular, is difficult to understand. We have pointed out above that it would be desirable to study these possible metrics before including them in a rating system. Family income quartiles will be available only for students who fill out the FAFSA and receive Title IV aid. Likewise, the Department has stated in the framework that EFC gap “may misrepresent access for institutions with low participation in federal student aid programs.” This is likely to apply to many AAU universities because of their generous institutional aid policies. For example, 15 AAU universities, or one-quarter of our U.S. membership, have less than 25% of their undergraduate students receiving student loans.

First-generation college status

This metric again relies on FAFSA data with its concomitant limitations. Additionally, the way the question currently reads may not distinguish first-generation college students from those whose parents completed some college. The FAFSA question currently reads: “What is the highest school your father/mother completed?”, with the following potential answers: Middle School/Jr. High; High School; College or beyond; or Other/unknown. If a parent did not complete college but attended some college, the box “High School” would need to be checked since that is the highest level completed.

Average net price and *Net price by quintile*

Average price and net price by quintile are restricted to first-time, full-time students receiving grant or scholarship aid and those receiving Title IV federal aid, respectively. They are also restricted to in-state students at public universities. Net price by quintile provides more information than average net price, which we support, yet in the context of a rating may not accurately reflect those institutions with relatively low percentages of students receiving federal aid due to institutional aid policies.

Completion and transfer rates
As we pointed out in previous comments, we believe the Student Achievement Measure (SAM) provides more comprehensive information on completion/graduation rates than those provided by IPEDS,[3] and we strongly support the use of SAM by institutions. The Department is also considering developing a new rate based on NSLDS data, which would include part-time and transfer students but exclude students who do not receive federal aid. As mentioned earlier, this may exclude large numbers of students at AAU universities. We do not believe transfer rates should be used to rate four-year institutions, and we note that many institutions do not report transfer numbers on IPEDS. By reporting outcomes for transfer students (i.e. whether they are still enrolled or have graduated), SAM handles such students in the optimal way.

Labor market success
We applaud the Department for expressing a more sophisticated view toward labor market success than simply focusing on short-term earnings at the institutional level. As we and others have pointed out, earnings vary between programs and majors in a more meaningful way than they vary between institutions, and we continue to strongly discourage looking at earnings at the institutional level. A threshold measure of "substantial employment" could serve to identify bad actors without placing an undue emphasis on salaries, though such an approach would serve little additional purpose. Many decisions need to be made before such a measure could be used, including the appropriate level for benchmarking and the reliability of and biases in any data sources the Department would use to get the data. Use of the federal poverty line, or any other benchmark, would need to be adjusted for geographic cost-of-living variation. Short-term labor market outcomes may say more about economic conditions than about institutional performance, so we prefer taking a longer view in terms of the timing of looking at earnings.

The Department must be very cautious in not introducing perverse incentives through this measure. In our previous comments we emphasized the differences between disciplines, as well as the low earnings of individuals who are in the process of earning higher degrees in fields, like many STEM fields, that require long training periods. Even within disciplines, however, the potential for perverse incentives exists. Consider an education program that can place its graduates in well-paying suburban school districts or lower-paying urban school districts. An emphasis on earnings may be at odds with societal benefits.

Graduate school attendance
Graduate school attendance is an important piece of information for students and consumers. However, it's not practical to include this information in a rating system for two reasons. First, it's not clear what a "better" institutional outcome in terms of graduate school attendance is, and consumers are more likely to care about attendance at the program level than at the institutional level. Second, using data on loans from NSLDS to determine graduate school attendance would exclude most students pursuing doctorates, especially in the STEM fields. Data from the National Postsecondary Student Aid Study indicates that less than 25% of research doctoral students take Stafford loans, and in the sciences the figure is less than 10%. While we do not encourage its use in a rating system, we point out that the National Science Foundation's Survey of Earned Doctorates includes both baccalaureate institution and field of degree for those who earn doctorates at U.S. universities.

---

[3] http://www.studentachievementmeasure.org

The Department has asked for examples of online tools that are particularly helpful for students and the public, and in that regard we highlight the seekUT tool developed by the University of Texas system.[4] This tool allows current and prospective students to explore a variety of outcomes (including employment, earnings, and debt) by institution and program. seekUT is not a rating system and would not work as one, and its strengths are illustrative of some of the weaknesses of a rating approach. seekUT allows data to be accessed at a level of detail meaningful for students, includes appropriate nuance (e.g., earnings are provided one, five, and ten years after graduation and earnings percentiles are shown to give some idea of ranges), breaks data down at geographic levels that matter to students, and provides auxiliary data (e.g., job projection statistics from the Bureau of Labor Statistics) that may be helpful for students looking to the future. While some of the underlying data could potentially be used for purposes other than consumer information, seekUT does not attempt to simultaneously fulfill these purposes. One limitation of seekUT is that it includes earnings data only for those employed in Texas. For institutions sending a large number of graduates of out of state, and especially if the destination of students is not random, these kinds of data are limited. seekUT is very straightforward about the limitations of the data used.[5]

Thank you in advance for your consideration of AAU's views. We look forward to a conversation in the coming months about how we can enhance college access and promote college completion while protecting the integrity of federal student aid funding and encouraging states to reinvest in public higher education. There is an appropriate federal role in providing students and their families with clear, accurate, and useful information. We welcome the opportunity to continue to work with the Department to consolidate and streamline existing consumer disclosures.

Sincerely,

Hunter R. Rawlings

Hunter R. Rawlings III
President
Association of American Universities

---

[4] http://www.utsystem.edu/seekut/
[5] http://www.utsystem.edu/seekut/about_the_data.html