Journal for the Study of Postsecondary and Tertiary Education

An Official Publication of the Informing Science Institute InformingScience.org

JSPTE.org

Volume 2, 2017

BEYOND RESEARCH PRODUCTIVITY: MATCHING PRODUCTIVITY MEASURES TO INSTITUTIONAL MISSION

Patricia Bartholomew — Economic Modeling Specialists, Moscow, ID, USA — pcbartholomew@Hotmail.com

ABSTRACT

Aim/Purpose	The aim of this paper is to develop a unified methodology inclusive of the three primary areas of faculty responsibility (teaching, research, and service) to calculate departmental productivity that fills the gap in methodological benchmarking tools for overall faculty productivity.
Background	A disproportionate number of departmental and faculty productivity indices in higher education rely solely on research. Productivity in other areas of faculty workload areas, like teaching and institutional and community service, are either measured separately or ignored all together – even when those activities are institutionally mandated. This does a disservice to those who work in those institutions and skews incentives.
Methodology	This paper utilizes a unified methodology inclusive of the three primary areas of faculty responsibility (teaching, research, and service) to calculate departmental productivity in five disparate departments (English, Biology, Mathematics, Sociology, and Computer Science) common to two universities with differing missions (teaching and service).
Findings	The results reveal the bias inherent in relying solely on research as a proxy for overall productivity in institutions that have differing missions.
Recommendations for Practitioners	Utilizing better metrics informs higher education administrators, promotes better decision-making, and allows incentives to re-align with desired outcomes.
Recommendation for Researchers	This paper recommends combing all aspects of faculty workload into a single benchmark index to better measure departmental productivity.
Future Research	Further research into improving this simple index is warranted and would include how to account for quality and other facets of productivity.
Keywords	higher education departmental productivity, faculty productivity, productivity management, methodology, university mission, accurate benchmarking

Accepted as a Peer Blind Reviewed Research Article by Editor Sydney Freeman Jr. | Received: May 4, 2016 | Revised: September 3, 2016 | Accepted: September 4, 2016.

Cite as: Bartholomew. P. (2017). Beyond research productivity: Matching productivity measures to institutional mission. *Journal for the Study of Postsecondary and Tertiary Education, 2,* 1-11. Retrieved from http://www.informingscience.org/Publications/3559

INTRODUCTION

As Paul Krugman, Nobel Prize economist, famously commented, "Productivity isn't everything, but in the long run it is almost everything" (1997, p. 11). But what is departmental productivity? This paper argues that the currently accepted understanding of departmental productivity in higher educational settings as a sole function of research productivity needs to be widened to encompass more aspects of the generally accepted faculty workload. A disproportionate number of departmental and faculty productivity indices in higher education rely solely on research. Productivity in other areas of faculty workload areas, like teaching and institutional and community service, are either measured separately or ignored all together – even when those activities are institutionally mandated. This does a disservice to those who work in those institutions and skews incentives. This paper utilizes a unified methodology inclusive of the three primary areas of faculty responsibility (teaching, research, and service) to calculate departmental productivity in five disparate departments (English, Biology, Mathematics, Sociology, and Computer Science) common to two universities with differing missions (teaching and service). The results reveal the bias inherent in relying solely on research as a proxy for overall productivity in institutions that have differing missions. Utilizing better metrics informs higher education administrators, promotes better promotes better decision-making, and allows incentives to re-align with desired outcomes.

A typical problem that higher education administrators have is one of comparison with outside institutions. This is due to complications introduced by variant missions and department size. Several surveys of faculty workload allocation show that faculty at institutions with differing missions allocate their time accordingly, which results in variant workload allocation (Blackburn, Bieber, Lawrence, & Trautvetter, 1991; Jordan, 1994; Serpe, Large, Brown-Large, Newton, Kilpatrick, & Mason, 2002). Selecting peer institutions in some way moderates this problem, but does not solve it completely. Therefore this study contributes to the tools available to higher education administrators to better understand productivity in their own institutions and how they compare to others. Further, the utility of knowing what the impact of hiring decisions would have on the department's overall current productivity when making decisions regarding resource allocation by being able to use the equation to measure the impact of prospective faculty on the current departmental index would be enormously useful.

Not having a methodology to combine diverse productivity metrics results in not being able to compare productivity of departments across their own campus or with outside universities (Cohen & Kisker, 2010, pp. 359-360). Webber (2011b) writes that many scholars, including Baird (1986) and Becher (1994), point out discipline as an important variable in discussing academic productivity. Citation levels are typically lower in the English and liberal arts disciplines than in say biology or other scientific fields, social science lower than life and physical scientists (Webber, 2011b, p. 111). Therefore, straightforward comparisons using productivity measures such as H-indices introduce unfair bias.

Many surveys of faculty workload allocation show that faculty at institutions with differing missions allocate their time differently (Blackburn et al., 1991; Jordan, 1994; Serpe et al., 2002).

Massy and Zemsky (1994) coined the popular catch-phrase "academic rachet" to mean that emphasis on research and teaching productivity causes faculty to move "away from their traditional teaching and student advising responsibilities to focus on research and scholarly activity directed at meeting their own needs rather than those of the institution" (Middaugh, 1992, p. 61). Part of the fault, of course, is of the metric by which faculty are measured. A unified index that reflects the value of traditional teaching and student advising as well as other faculty activities would discourage the distortion of faculty emphasis solely toward publication.

So it is surprising that so little research has been done to include faculty activity other than research in the productivity calculus. A better method would extend the methodology in a way to account for the three aspects of departmental faculty workload in a way that allows for inter and intra-

comparability across US campuses. This paper uses the methodology in Bartholomew (2015) to calculate and apply to five diverse departments (English, Biology, Mathematics, Sociology, and Computer Science) common to two universities with differing missions. One of the universities has a teaching mission and one has a service mission. The results are compared to the results obtained when only research is considered as a dimension of departmental workload.

BACKGROUND

What few modern *academic* attempts to measure faculty productivity there are have focused primarily on research productivity as the sole proxy for overall productivity (Baird, 1986; Creswell, 1986; Webber, 2011a) at the expense of other important faculty activities like teaching and service (although classroom hours and student contact hours have been also been used as a "crude index of gross productivity" according to Cohen & Kisker, 2010).

More recently Jorge Hirsch's H-index (2005), which measures how many times an author has been cited relative to the number of publications, improved on the research-as-proxy-for-total productivity methodology by introducing a way to measure the quality dimension of research.

While this improvement is useful, it does not address the fact that most of the solutions and research into quantitative measures of faculty productivity, at worst, ignore the teaching and service aspects completely. At the least, the current solutions calculate individual, distinct indices for only two of the three aspects, not all three aspects and certainly not in a single unified index, which leaves the administrator without guidance on how to weigh the indices in decision making. (See Cohen and Kisker, 2010, p. 359–360, for an enlightening explanation as to why this is so.) Alonso, Cabrerizo, Herrera-Viedma, and Herrera observe that even the H-index cannot be used to compare research across disciplines. They cite lack of normalized practices across academic disciplines as the primary reason.

Private companies have also gotten into the game of measuring faculty research productivity, but again is only limited to research as a proxy for all types of productivity. Academic Analytics releases a "Faculty Scholarly Productivity Index" on a subscription basis. The product ranks both individuals and departments on important research variables and then compares the department with the national averages in a radar plot to give department heads and administrators a quick visual on where departments might be weak or strong in comparison with the nation or other peer or aspirant institutions. Further information on Academic Analytics can be found at <academicanalytics.com>.

In terms of measuring teaching productivity, the groundbreaking researcher has been Michael Middaugh (2001). He tackled teaching productivity from the perspective of generated student credit hours. His work, known colloquially as the Delaware Study, is considered the gold standard on teaching productivity and is used widely as a benchmarking tool for institutional researchers and higher education administrators. Middaugh's work builds on the earlier work of the Joint Commission for Accountability Reporting (Mortimer, 1997, as cited in Middaugh, 2001, pp. 37-38), one of whose major contributions to the field in Middaugh's view was to specifically define the three aspects of faculty workload and define that the sum should total 100% (Middaugh, 2001 p. 37-38). Middaugh's contribution introduces the managerial aspect of "cost per instructional hour" in an attempt to add an efficiency aspect to productivity. The most surprising outcome of the study was finding that direct instruction costs vary more by discipline than Carnegie classification, a system of classification established by the Carnegie foundation in 1973 to identify groups of roughly comparable institutions. "While Carnegie institutional classification could be expected to account for some of the variance, its explanatory power does not approach that of the disciplinary mix within the institutional curriculum" (Middaugh, Graham, Shahid, & Carroll. 2003, p. 18– 20).

A weighted average method of evaluating teaching productivity at medical schools was proposed using a relative value scale in teaching by Bardes, Hayes, Falcone, Hajjar, and Alsonso (1998). Weights

are increased by difficulty of courses taught (new courses have a higher weight than old, problem based learning courses have a higher weight than ordinary courses).

Crosta and Packman (2005) modeled teaching productivity based on the number of doctoral students supervised at Cornell and found departmental prestige and longevity were significant predictors of productivity. As found in research productivity, they also found unequal distribution of productivity among discipline lines, with faculty members in the Humanities being the most productive when judged by how many committees they had chaired Weighting differing level classes properly remains controversial, as there are as many ideas on proper weighting as there are faculty members. It is generally acknowledged that upper division classes should carry more weight; however, how much more weight is a contentious issue. The same controversy exists also for class size and type of class.

Although there is not much research into teaching productivity, even less exists on service. The exception is Blackburn (1974) who divides service into three components: public (non-academic), professional (peer-review), and campus (committee). Blackburn attempts to find significant determinants of service hours like professorial rank (full professor was significant), self-reported self-efficacy, and career age. Consistent with some of the literature, finding time spent on teaching reduces research productivity, more time spent on service was also found to decrease research productivity.

Townsend and Rosser (2007) studied workload allocation and its effects on productivity. They limited themselves to research and teaching productivity as proxies for total productivity, ignoring the service component for tractability's sake. Their key finding is that institutional type was a key determinant of faculty workload allocation (faculty at teaching colleges spend more time teaching, for example).

Looking into faculty salary research (as perhaps its own proxy of faculty productivity), Fairweather (2005) found "the relationship between hours spent in the classroom per week and pay was substantially more negative in 1998–99 than 1992–93" even in teaching-oriented institutions (p. 412). Toutkoushian, Bellas, and Moore (2007) found more demographic reasons ("white, married, male") for higher pay while controlling for publications and discipline, which were also significant.

Tangentially, Dundar and Lewis (1995) use cost data to predict economies of scale and scope (diversity of disciplines) in cost and productivity in teaching and research structures. The productivity variables modeled include annual student hours per department (separated by undergraduate, master, and PhD) and number of publications by department. Their results confirmed the conventional wisdom that research universities have the higher cost structure, and advanced education is the most costly from a teaching perspective (except for social sciences where masters students cost more to educate than doctoral students).

A review of the literature would not be complete without a review of the sub-genre of literature that argues the futility of measuring faculty productivity. Blackburn (1974) wrote one of the earliest and most persuasive arguments against applying productivity indices to faculty. He posits several arguments, but the main thrust is that free time and work time are indistinguishable in the life of an academic. Academia is a way of life for faculty and thus measuring hours becomes a fruitless exercise in splitting hairs about such topics as to whether thinking about one's work while walking about campus counts as scholarship. He uses four arguments against research productivity in particular: 1) discipline differences in publishing rates, 2) gender differences favor males, 3) age of earning the PhD, and 4) type of institution.

This prior research points to the importance of utilizing a benchmark metric that reflects more than one aspect of the departmental faculty activity. This paper proposes a solution that fills the gap in methodological benchmarking tools for overall faculty productivity.

DATA AND METHODOLOGY

A unified index methodology was created in line with Bartholomew's P-score methodology developed in 2015. The composite index is a weighted average Z-score-like calculation with the weights

taken from a department's estimated workload allocation. A Z-score is a statistical method that transforms individual scores into unit-less measures, thus allowing comparison across a variety of measures. The P-score, however, differs from a typical Z-score in that it is not used in the cardinal sense of the term, but for use in an ordinal sense. Hence it is called a P-score. The formula for the composite index is found below:

$$Composite\ P-score = \sum_{i=1}^{3} p_i \frac{x_{i,j} - \overline{x}_{i,j}}{\sigma_{i,j}}$$

Where: i = workload allocation area (1= research, 2=teaching, 3 = service)

j= discipline or department in which the faculty member is working

p_i= proportion of time spent on faculty activity

$\overline{x}$= activity (research, teaching, or service) mean

$\sigma_{i,j}$ = standard deviation of activity measure

The research measure (i=1) includes average publications per faculty member, average citations, average grant dollars, average number of grants, average number of books published, average number of conference proceedings, and average number of awards over the active academic lifetime of the faculty. (Therefore, departments with longer tenured faculty may see higher productivity in their index. Restricting the data to the last five years would reduce this bias.)

The teaching measure (i=2) is student credit hours per academic year divided by full time equivalent faculty. The service measure is service hours per academic year.

The data are at the national level for the United States. The data for the research portion of the calculation were from an extract from the Academic Analytics (2014) data set. Total journal publications, total citations, total dollar amounts of grants, total number of grants, total awards, total conference proceedings, and total books variables were each divided by the number of faculty to arrive at average numbers for the seven measures by department. The database includes all faculty members listed on the departmental roster, and as such combines all tenure track faculty, adjunct faculty, research/teaching faculty who are listed on the departmental roster. Means and standard deviations were then calculated by discipline. The seven p- scores for research were then averaged (this assumes that each institution and department values the variables equally, i.e., number of publications is just as important as citations. As a caveat, it would be straightforward to change that assumption by assigning weights by institutional and/or department value. Also note this method averaged seven p-scores into a single number p-score to avoid the trap of Simpson's Paradox, a statistical effect in which aggregation of data before taking averages results in the opposite than expected effect. (See Pearl, 2014, for a more thorough discussion.)

For teaching hours, data was taken from the 2014 Delaware study for undergraduate teaching hours. Due to the difficulties in measuring learning outcomes for students, classroom hours as measured by the Delaware Study will be used as a proxy for teaching productivity. Further, only undergraduate teaching hours were used due to the controversy in weighting higher-level courses.

For community service, the P-statistics were estimated based on the mission statement of the universities. The mean and standard deviations by department were extrapolated from the Higher Education Research Institute (HERI) annual Faculty Survey (Eagan & Aragon, 2014).

These datasets are independently collected with the commonality in that the all five of the same departments at the two selected universities were present in the data for teaching and research. They all have the commonality of collecting data from the population of four-year educational institutions within the United States. However, they all depend on voluntary submission of data and as such may suffer from selection bias and differing populations of participating four-year institutions. Currently

there is no common dataset attempting to measure all three variables with the same degree of granularity as these three combined datasets.

Weights were estimated from range data. The calculations took the number of hours from the most frequent responses, multiplied by the mid-point, then summed and divided by the total. When frequencies were nearly equal, ranges were combined.

Five departments commonly found in most universities were chosen for analysis: English, Biology, Mathematics, Sociology, and Computer Science. They were chosen primarily because they were well represented in the data samples at the institution level and range from soft to hard skill, liberal arts, social sciences, and hard sciences. Thus they represent the disciplinary differences most commonly cited as troublesome for inter-departmental comparison according to the literature. Observations by dataset can be found in Table 1.

Table 1: Number of Observations (Departments*) by Index Component

Department	N-Research	N-Teaching	N-Service
English	216	219	289
Biology	215	219	289
Mathematics	230	219	289
Sociology	200	219	289
Computer Science	237	219	289

*As the data spans multiple data sets, N here represents the number of departments analyzed

A national average statistic would be equal to zero on a P-score. Any score above zero (positive) reflects a department preforming above national averages. Any score in negative represents a department preforming below the national standard. Additionally, the higher the p- score means the more productive the department.

RESULTS

Two universities with data available for the three-workload areas were selected to compare. Both universities are classified has RVH (Research-Very High) in the Carnegie classification. One is in an urban area (top quartile of the metropolitan statistical areas by population) and one is in a rural community. Both are public institutions. University One was established as a normal (teaching) school. University Two has land grant (community service) status. University One has in its mission statement a clear emphasis on its institutional heritage as being established as Normal School and a clear commitment to teaching. University Two emphasizes its heritage as a Land Grant establishment and in the second sentence of the mission statement states that service to the state and community is central to its mission.

Weighting research at 100% (and thus ignoring teaching and service components as current criteria does) gives the reader an idea of the bias of benchmarking using research as a proxy for overall productivity. The first university is under national averages in almost all departments, except Sociology where they hover slightly over the national mean (zero). The second university is below national averages in mathematics, sociology, and computer science but excels in English. Neither of these measures, however, accurately reflects the benchmark performance of the departments or reflects

the efforts of the departments on behalf of teaching and commitment to institutional and community service (see Figure 1).

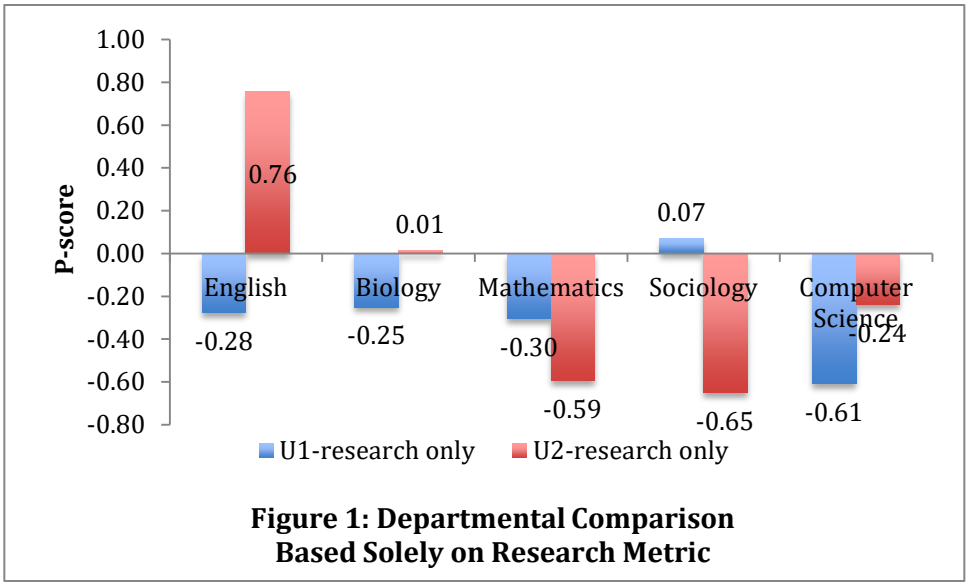

Figure 1: Departmental Comparison Based Solely on Research Metric

Adding in teaching corrects some of the bias and drives most of the results over average. Mathematics and Sociology in University Two, however, remain below average. Not including the service metric in this calculation disadvantages University Two, which is a land grant and as such has service mandated in its mission (see Figure 2). In the following section, service is added as a component in the benchmark metric.

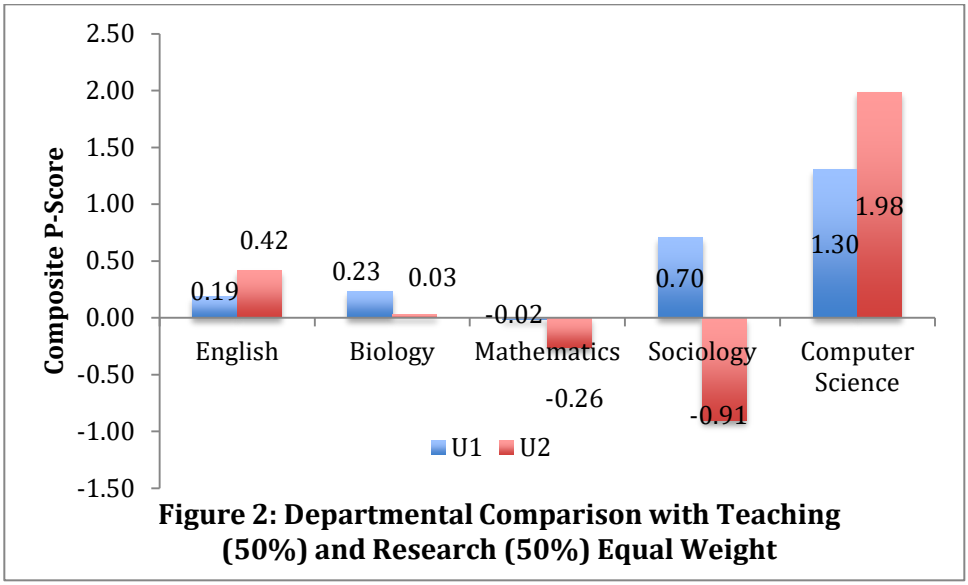

Figure 2: Departmental Comparison with Teaching (50%) and Research (50%) Equal Weight

Figure 3 shows the calculations not only adding a service component, but also assigning the weights to reflect the varying university missions. University One's percentages are teaching 60%, research 30%, service 10%; University Two's percentages are set at teaching 50%, research 20%, and service 30%. Seven of the ten resulting composite P- scores remain above, but hover near, national averages Computer Science is the exception. Those scores increase dramatically to over 2 for both universities.

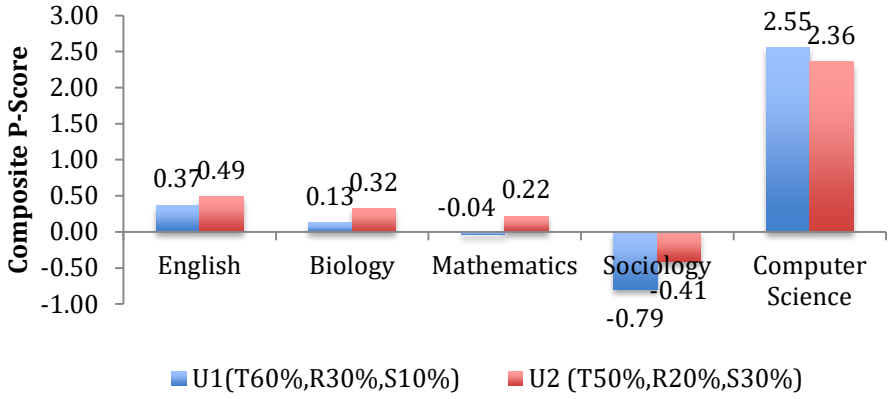

Figure 3: P-scores using weights tailored to university mission

The three permutations of the index reveal the importance of good benchmarking. The first case, where only research productivity was benchmarked, causes both universities to be seen as performing below national averages in almost all departments. Metrics improve for both universities when the workload allocations are adjusted to be more in line with university mission. The metrics improve in eight out of ten of the departments in both universities, showing greater productivity than when judged only based on a sole research metric. This lends weight to the premise that a sole research metric biases productivity estimates and comparisons for universities with missions inclusive of other goals.

Assumptions, Limitations, and Areas for Future Research

No one measure would be able to adequately adjust for all the peculiarities of comparing faculty productivity. Decisions about what to include and exclude, quality concerns, and the like will ultimately affect outcomes. Limiting the index to include the most common three metrics of faculty workload (research, teaching, and service) by definition excludes other facets of productivity like professional development and administrative work (although some administrative work is logged as university service). The method could be easily extended to include those facets if and when the data are included in the faculty work allocation statistics and available databases. Confusion between self-reported service and extension and outreach activities and double reporting may also be limiting factors (although all attempts have been made to exclude instances of double reporting).

A critical assumption is that each university values every department equally. However, if this is not the case, the bias could be easily remediated by assigning qualitative weights reflecting the relative importance of individual departments. This requires close consultation with university decision makers and assumes there will be general agreement on this topic among administrators.

A further assumption of the study is that input teaching hours are a good proxy for learning outcomes. Given the complications in measuring student learning outcomes, it is difficult to know how much of an assumption this is and how much variability there is across departments that is not reflected in the data. This assumption also holds true for service hours, i.e., that more service hours reflect higher productivity. Although intuitive, this does carry the implicit assumption that the input hours are all uniformly productive, which, of course, is somewhat tautological.

If data at the individual level were available, it would be possible (and more methodologically sound) to calculate the departmental index as the average of its individual constituents; however, to date that data set at the national level does not exist.

Quality is subjective and therefore a difficult dimension to measure. While some research productivity indices now include a quality aspect (Hirsch's H-index), assessing teaching and service quality remains problematic. It is an issue that continues to challenge researchers.

Due to the difficulties in measuring learning outcomes for students, undergraduate classroom hours as measured by the Delaware Study will be used as a proxy for teaching productivity. Weights will vary by university mission only; despite the fact that certain research (Middaugh, 2001) has suggested workload allocation also varies by discipline. This limitation affects only the application of the unified index, as the application of the index itself posits only weights as fixed. The weights actually selected are limited only by user's access to data that points to accurate weighting for each individual's situation. It is assumed that the department head will know how she or he would like the faculty in the department to be spending their time.

Data is by department. No attempt has been made to tease out the number of levels of faculty (i.e., assistant professor, full professor, tenured, tenured track, full time, part time, temporary, teaching assistants, etc.).

The Delaware Study and Academic Analytics rely on voluntary participation for their data and such, the data is likely skewed upward to represent this selection bias in the underlying data.

This paper is limited to institutions of higher education in the United States. An interesting area of further research could look into how the index could be implemented in a global context and what insights could be gleaned from a comparison of US and other global institutions. This would undoubtedly lead to interesting insights into international productivity.

Further research into improving this simple index is more than warranted. Useful research into how to improve the index would include how to account for quality (especially in the teaching metric); inclusion of more output-oriented metrics into the model; inclusion of other facets of productivity, like professional development and administrative work (although some administrative work is logged as university service); and survey work to establish the benchmark weights not only by university or college, but by department. Further, it would make sense to calculate the P-scores based not only on departmental discipline, but break the population into mission subsets for better and more precise comparison.

CONCLUSION

The results of this paper show that departments at institutions with mandated missions separate from research are benchmarked inadequately when only research productivity is taken into account (and used as a proxy for overall productivity). Five departments common to two universities with differing missions (service and teaching) are compared under various weighting scenarios of departmental activity (100% research, 50% research/50% teaching, addition of service as a variable). It is found that when only research is counted, the supermajority of departments at both universities are found to be below national average; however, when the weightings are adjusted to more accurately reflect university missions, those results invert and show the supermajority of departments performing at or above national averages.

Benchmark metrics across the five departments showed improvement when the additional metrics of teaching and service were included. In the first case, where only research productivity was benchmarked, both universities were seen as performing below national averages in almost all departments. When workload allocations are adjusted to be more in line with university mission, the metrics improved in eight out of ten of the departments in both universities-showing greater productivity than when judged based only on a sole research metric.

The results highlight the importance of an adequate benchmarking procedure. Massey and Zemsky (1994) coined the popular catch-phrase "academic rachet" to mean that emphasis on research and teaching productivity causes faculty to move "away from their traditional teaching and student advis-

ing responsibilities to focus on research and scholarly activity directed at meeting their own needs rather than those of the institution" and associated costs increases. Part of the fault, of course, is with the metric by which faculty are measured. A unified index that reflects the value of traditional teaching and student advising as well as other faculty activities would discourage the distortion of faculty emphasis solely toward publication.

Better metrics inform higher education administrators by providing accurate information that accounts for institutional mission when benchmarking against peer and aspirant institutions, resulting in better decision-making, better resource allocation, and aligning incentives with desired outcomes.

REFERENCES

Academic Analytics. (2014). Retrieved from http://www.academicanalytics.com

Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). h-Index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, *3*(4), 273–289. doi:10.1016/j.joi.2009.04.001

Baird, L. L. (1986). What characterizes a productive research department? *Research in Higher Education*, *25*(3), 211–225. Retrieved from http://spot.lib.auburn.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ349450&site=ehost-live

Bardes, C., Hayes, J., Falcone, D., Hajjar, D., & Alsonso, D. (1998). Measuring teaching: A relative value scale in teaching. *Teaching and Learning in Medicine*, *10*(1), 40–43.

Bartholomew, P. (2015). *Measuring faculty productivity: Towards a unified methodology at the departmental level.* Unpublished Dissertation. Auburn University.

Becher, T. (1994). The significance of disciplinary differences. *Studies in Higher Education*, *19*, 151–161. doi:10.1080/03075079412331382007

Blackburn, R. (1974). The meaning of work in academia. *New Directions for Institutional Research*, *1974*(2), 75–99.

Blackburn, R., Bieber, J., Lawrence, J., & Trautvetter, L. (1991). Faculty at work: Focus on research, scholarship, and service. *Research in Higher Education*, *32*(4), 385–412.

Cohen, A., & Kisker, C. (2010). *The shaping of American higher education* (2nd ed.). San Francisco: Jossey Bass.

Creswell, J. (1986). Concluding thoughts: Observing, promoting, evaluating, and reviewing research performance. In J. Creswell (Ed.), *Measuring faculty research performance* (pp. 87–102). San Francisco: Jossey Bass.

Crosta, P. M., & Packman, I. G. (2005). Faculty productivity in supervising doctoral students' dissertations at Cornell University. *Economics of Education Review*, *24*, 55–65. doi:10.1016/j.econedurev.2004.03.011

Delaware Study. (2014). The National Study for Instructional Cost & Productivity < http://ire.udel.edu/hec/cost/>

Dundar, H., & Lewis, D. (1995). Departmental productivity in American universities: Economies of scale and scope. *Economics of Education Review, 14*(2), 119-144.

Eagen, K., & Aragon, M. (2014). *Undergraduate teaching faculty: The 2013-2014 HERI Faculty Survey*. Higher Education Research Institute, UCLA. Accessed from http://heri.ucla.edu/pr-display.php?prQry=151

Fairweather, J. S. (2005). Beyond the rhetoric: Trends in the relative value of teaching and research in faculty salaries. *The Journal of Higher Education, 76*(4), 401-422. doi:10.1353/jhe.2005.0027

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569–16572. doi:10.1073/pnas.0507655102

Jordan, S. (1994). What we have learned about faculty workload. In J. Wergin (Ed.), *Analyzing faculty workload*: *New Directions for Institutional Research*, *Number 83* (pp. 15–23) San Francisco, CA: Jossey-Bass.

Krugman, P. (1997). *The age of diminished expectations*. Boston, MA: MIT Press.

Massy, W., & Zemsky, R. (1994). Faculty discretionary time: Departments and the academic ratchet. *Journal of Higher Education*, *65*(1), 1–22.

Middaugh, M. F. (1992). Examining academic and administrative productivity measures. *New Directions for Institutional Research*, *75*, 61–75.

Middaugh, M. F. (2001). *Understanding faculty productivity*. San Francisco: Jossey Bass.

Middaugh, M. F., Graham, R., Shahid, A., & Carroll, C. D. (2003). A study of higher education instructional expenditures: The Delaware Study of instructional costs. *Higher Education.* National Center for Education Statistics Publication 2003-161. US Department of Education.

Pearl, J. (2014). Understanding Simpson's paradox. *The American Statistician*, (December 2013), 1–12. doi:10.1080/00031305.2014.876829

Serpe, R., Large, M., Brown-Large, L., Newton, R., Kilpatrick, K., & Mason, R. (2002). *CSU faculty workload report*. Retrieved from https://www.calstate.edu/acadres/docs/csu_facwrkldrpt.pdf

Toutkoushian, R. K., Bellas, M. L., & Moore, J. V. (2007). The interaction effects of gender, race, and marital status on faculty salaries. *The Journal of Higher Education*, *78*(5), 572–601. doi:10.1353/jhe.2007.0031

Townsend, B. K., & Rosser, V. J. (2007). Workload issues and measures of faculty productivity. *Thought and Action*, *Fall*, 7–20. Retrieved from http://www.nea.org/assets/img/PubThoughtAndAction/TAA_07_02.pdf

Webber, K. L. (2011a). Factors related to faculty research productivity and implications for academic planners. *Planning for Higher Education*, (September), 32–44.

Webber, K. L. (2011b). Measuring faculty productivity. In J. C. Shin, R. K.Toutkoushian, & U. Teichler (Eds.), *University rankings* (pp. 105–121). Springer Verlag. doi:10.11007/978-94-007-1116-7_6

BIOGRAPHY

Patricia Bartholomew serves as an economist in the consulting division of Emsi. Her focus includes, but is not limited to, gap analysis, economic scans, cost-benefit analysis, regional development strategies, and economic impact studies. Pricing higher education, bench-marking faculty productivity, educational research methodology, and education to career transition are some of her research interests. Prior to joining Emsi, Dr. Bartholomew taught at several universities and headed economic research efforts in central Europe. She holds a PhD in higher education administration from Auburn University, where she previously worked in institutional research. She also has a master's degree in economics from Duke University and a bachelor's degree in economics from the University of Miami (Coral Gables). She is fluent in German.