Teacher-Scholar: The Journal of the State Comprehensive University

Volume 11 | Issue 1 Article 2

January 2022

A State University’s Assessment of ACUE: Feasible Model for Evaluating the Impact of a Faculty Instruction Quality Program

Jeffrey Budziak
Western Kentucky University, jeffrey.budziak@wku.edu

Daniel Super
Western Kentucky University, daniel.super@wku.edu

Thomas Gross
Western Kentucky University, thomas.gross@wku.edu

Douglas McElroy
Western Kentucky University, doug.mcelroy@wku.edu

A State University's Assessment of ACUE: Feasible Model for Evaluating the Impact of a Faculty Instruction Quality Program

Cover Page Footnote

The authors would like to acknowledge Dr. Pam Petty for her contributions to the evaluation design, Angela Byrd for her assistance in data collection, and the Center for Innovative Teaching and Learning at Western Kentucky University for sponsoring and facilitating the ACUE training program.

A State University's Assessment of ACUE: Feasible Model for Evaluating the Impact of a Faculty Instruction Quality Program

Jeffrey Budziak
Western Kentucky University
Daniel Super
Western Kentucky University
Thomas Gross
Western Kentucky University
Douglas McElroy
Western Kentucky University

Nationally, university administrators and faculties face an ever-increasing pressure to recruit, retain, and graduate students. The looming demographic cliff—the steep decline in traditional-aged college bound students born immediately following the 2008 financial crisis—is rapidly approaching (Campion, 2020). This, coupled with stagnating rates for traditional college-going students (NCES, 2021) and increased competition for attracting incoming students, places a greater emphasis on retaining current students rather than recruiting increasingly larger incoming classes. In the U.S., the undergraduate dropout rate is 40% overall, with approximately 30% of freshmen failing to enroll for a sophomore year (NCES, 2019). This has driven retention-focused programs to target first-year undergraduate students. Some of these programs focus specifically on student engagement, such as collaborative learning and other high-impact practices. Others focus on improving teaching practices, theorizing that quality teaching will lead to greater learning and greater learning will lead to better grades, meaning students are likelier to be retained. Scholarship supports this expectation; higher college grade point averages (GPAs) are associated with college completion (Denning et al., 2021). It is therefore unsurprising that institutional strategic plans tend to emphasize investment in faculty teaching capabilities (Minter, 2009).

High quality instruction is a goal for most institutions and is particularly prioritized by state comprehensive universities. These institutions are often beacons of educational and economic development in their respective service regions, many of which trace their genesis to land grant and normal school origins. State comprehensives commit to providing an accessible, quality education through faculty who prioritize teaching and student support. However, the financial implications of shepherding incoming classes of students from admission to graduation are salient to faculty and administration in these institutions. In sum, student retention is key to fulfilling the mission of many public universities missions as well as their budgetary needs to maintain university operations.

Assessing Faculty Development Program Outcomes. Universities frequently launch initiatives to improve the quality of classroom instruction because teaching quality is one of the most important factors related to increased student achievement (Darling-Hammond, 2003). Further, the scholarship of teaching and learning (SoTL) permeates higher education nationwide (Condon et al., 2016), and investigations of programs can be used to help instructors improve course delivery and responsiveness to student needs. Unfortunately, there is often more consideration paid to investigating the degree of faculty satisfaction with programming than the program influence on changes in teaching or student learning (Hines, 2009). One explanation for the relative dearth of program assessment in this area is that faculty development programs pose a particular assessment challenge. Many programs are replete with confounding variables, lack longitudinal data, and are frequently orchestrated by individuals who lack training in rigorous evaluation techniques (Kucsera & Svinicki, 2010). Such programs also frequently suffer from low response rates in post-evaluation surveys. As a result, what assessments are done are frequently limited to short-term, small-scale evaluations of changes in process or broad-brush measurements of student or instructor satisfaction, and the outcomes of these assessments may not be representative of the program broadly. This limits understanding of the success of faculty development initiatives. The addition of multiple independent data sources (e.g., Hewson et al., 2001), longitudinal data elements (LaFleur et al., 2009), and statistical data techniques designed to draw appropriate causal claims could help improve assessment of these initiatives.

The methodological challenges to outcomes-focused assessment of faculty development programs make it difficult to have confidence in conclusions concerning the success of the programs. University decision makers understand that there are many fixed (e.g., academic preparedness for college, family income, first-generation status) but relatively few variable (e.g., faculty instruction development and ongoing support) predictors of student success. Investing in pedagogical training that is relevant for and accessible to a large number of instructors is a possible step toward improved student experiences and academic outcomes across degree areas. However, with the absence of quality outcomes data, administrators and faculty development leaders are often ill-equipped to adjust programming and offer compelling justifications for continued or increased funding. Just as assessment both closes and opens the loop of quality instruction, so it is an integral component of program design and improvement.

Administrators and program directors could consider descriptive statistical trends of aggregated course outcome metrics such as pass rates (often presented as the inverse rate of D’s, F’s, and W’s) or class GPAs. A descriptive approach or a univariate analysis intending to draw inferential conclusions related to the influence of pedagogical adjustments lacks the rigor of design necessary to instill confidence in the findings. An important methodological consideration in the assessment design is to, whenever possible, identify a group of nonparticipants as a comparison group to which the participants in the intervention group can be compared (Fink, 2013; Devlin, 2008). While this is a preferable design approach, it is accompanied by its own logistic and methodological challenges.

Here we offer a case study of an assessment strategy for a faculty training program directed at improving DFW rates via improved faculty teaching methods. In 2019 Western Kentucky University, a regional comprehensive university, implemented a faculty development program from The American Council on Undergraduate Education (ACUE). This program is a faculty training course focused on incorporating best practices for quality instruction in the university classroom. We employ a propensity score matching approach to assess the effect of this program on authentic measures of student success across a wide range of courses and disciplines. As we detail below, this approach avoids many of the pitfalls commonly associated with less rigorous assessment strategies and allows us to address the question of whether the program was successful in changing student outcomes. The results demonstrate that it was; students in sections taught by faculty who had previously completed the ACUE program earned fewer D’s, F’s and W’s than matched students in courses taught by faculty who had not completed the program. In conclusion, we discuss the potential value of faculty development programs and the importance of proper assessment.

The ACUE Program. ACUE’s course addresses over 200 evidence-based teaching practices that promote student success. Dozens of colleges and universities have adopted this professional development program for their faculty, and initial data suggest that the course has a positive impact on a variety of institutional outcomes (Lawner & Snow, 2019; Lawner & Snow, 2020). The course is divided into 25 online learning modules that are organized into five major units of study: (1) Designing an Effective Course and Class; (2) Establishing a Productive Learning Environment; (3) Using Active Learning Techniques; (4) Promoting Higher Order Thinking; (5) Assessing to Inform Instruction and Promote Learning. While the program is delivered asynchronously, the design for implementation is one of self-paced autonomy within a sequence and timeframe negotiated between the university partner and ACUE staff.

Unique to ACUE is the sustained and supported development design. The “workshop” models of professional development characterized by episodic offerings with little or no obligation for participants to implement that which they learned have been demonstrated to be relatively ineffective (Darling-Hammond & Richardson, 2009; Knapp, 2003). The ACUE model mimics the structure of a university course in both its intensity and the feedback provided. Participants are expected to not only engage in the asynchronous delivery of modularized content but to apply and reflect on their pedagogical adjustments as they implement what they have learned. Professional staff employed by ACUE then provide feedback to faculty incorporating these changes. Upon completion of the course, participating faculty are awarded the ACUE Certificate in Effective College Instruction signifying their accomplishment.

The ACUE course is one example of an intensive faculty training program designed to improve the effectiveness of classroom instruction. Certainly, other similar commercialized products are available and there are innumerable "in-house" faculty development initiatives being implemented at institutions of higher learning across the country. These offerings may differ in size, scope, target outcomes, and design, but they each share a need for sound, systematic assessment designed to inform stakeholders regarding impact and value.

Implementation

Beginning in late November, participating faculty were divided into four cohorts of either seven or eight members akin to the model of professional learning communities (PLCs) often employed by educational entities endeavoring to engage faculty in long-term professional development. These PLCs were organized and facilitated by the university's center for faculty development and were mobilized with the intention of fostering a sense of group camaraderie while providing the structured support that is often necessary to maintain progress through self-paced programs. The ACUE course spanned 25 weeks across the spring and summer semesters, with faculty "graduating" from the experience in time to implement pedagogical and course structure adjustments beginning in the fall semester.

In total, 30 faculty members representing more than a dozen disciplines across all five university colleges, enrolled in the program. All of the "targeted" courses were relatively high enrollment and lower-level (100 and 200 level) offerings within the university's general education curriculum. Selecting courses of this nature aligned with the general hypothesis that enhancing the level of quality instruction by incorporating ACUE's evidence-based practices should have a positive impact on university metrics of student success and retention.

The ACUE course included faculty who taught disparate disciplines, employed differing course delivery modalities, and targeted multiple levels of the undergraduate curriculum. Large scale faculty development programs, whether contracted or designed locally, often adopt a similar approach in an effort to scale the effects of the program. These design constraints call for assessment methodologies that account for myriad factors while endeavoring to isolate the effects of program participation. As such, we offer an approach to systematic assessment of faculty development programs that helps to minimize the noise in the data by employing propensity score matching (PSM) techniques.

Data and Analytical Methods

Data were collected from fall and spring semesters across a three academic-year period. This period began in the fall semester of 2018 (prior to the implementation of the program) and covered through the spring semester of 2020 (after the completion of the program). The sample included 1682 course sections taught by 280 unique instructors from 26 different academic subjects. Most sections (1266, or 75%) were delivered through an in-person modality on the university's main campus. The overwhelming majority (1466 sections, or 88%) were lower-level courses. Some sections (1219 sections, or 72%) were taught by instructors who never participated in the program or had yet to complete the program. The rest (463 sections, or 28%) were taught by instructors who had already completed the program. The structure of the data is well-suited to assess the effectiveness of the program because it allows the analysis to account for both inter-instructor differences (i.e. instructors who never participated in the ACUE program compared to those who completed the program) as well as intra-instructor differences (how instructors performed before and after completing the program).

A number of variables were collected for each course section. Among them were the days of the week on which the class was offered, the time the class began and ended, the enrollment in the course section, among many others. Like many other assessment strategies, we examined student outcomes data. The primary variable of interest was the percentage of students in the course who earned a grade of D or F or who withdrew from the course, thus receiving a grade of "W" ("DFW rate"). While DFW rate does not perfectly measure of student performance, one of the primary goals of the program was to increase the academic performance of students in the section, and assigned grades generally reflect that performance. At the university, grades of D, F, and W typically indicate deficient student performance and prevent students from progressing toward graduation. For many reasons discussed above, failure to make sufficient academic progress undermines the mission of most institutions of higher learning. For all these reasons, we interpret a reduction in DFW rates as support for the effectiveness of the intervention.

To assess whether the intervention reduced DFW rates, we utilized propensity score matching. Matching refers to a class of statistical techniques commonly used to assess the effect on an intervention (commonly described as a "treatment") when the treatment cannot be randomly assigned to the observed units (see Morgan & Winship 2014). The lack of random assignment threatens the ability of researchers to draw appropriate causal inferences concerning the effect of the treatment because of the potential for bias caused by unobserved confounding factors (Rubin, 1973). In matching, units that received the treatment are "matched" to units that share many observable characteristics that did not receive the treatment prior to estimating the effect of the treatment. This reduces the potential for bias, meaning scholars can have greater confidence they are observing a true causal relationship between the treatment and the outcome of interest even if they cannot implement a true randomized experiment. Assessments that fail to account for the possibility of unobserved confounding factors run the risk of drawing improper inferences about the effectiveness of their intervention.

Propensity score matching (PSM) is a commonly used matching method. To implement PSM, each observation receives a probability (“propensity”) score that it was assigned to receive the treatment (Rosenbaum & Rubin, 1983). This score is commonly estimated using a logistic regression model, where the probability of receiving the treatment is predicted by a set of observed potential confounders that affect the likelihood of receiving the treatment. Observations are then matched based on the similarity of their propensity score. Once this matching is complete, one can estimate the average treatment effect by comparing how the matched treated and non-treated observations compare on a quantity of interest. In our analysis, the outcome variable of interest is DFW rate (measured as a percentage), and the treatment is whether the instructor was a participant in the ACUE program (1 = yes, 0 = no).

Prior to assessing the effect of the treatment, it is important to determine whether the propensity score matching has achieved appropriate covariate balance between the treated and non-treated group. Covariate balancing can be understood as the degree to which the distributions of relevant covariates are similar across treated and non-treated units. While the literature recommends a variety of tools to assess balance (see Ho et al., 2007 for details), standardized mean differences (Stuart et al., 2013) and variance ratios (Austin, 2009) are two commonly deployed tools. In our analysis, the data were matched based on four variables. Sections were matched on whether it was an honors section (1) or not (0), whether the class was a lower-level class (1) or not (0), the enrollment in the section (count of students), and whether the section occurred after the completion of the program (1) or not (0). The final covariate accounts for the fact that program instructors should be indistinguishable from non-program instructors prior to completion of the program. This, in effect, permits a comparison of the DFW rates of all faculty in the sample prior to the ACUE program with the DFW rates of ACUE faculty and non-ACUE faculty after the completion of the program.

Both tools suggest the data are well balanced after matching. The standardized mean differences, variance ratios, and average treatment effect were calculated using the “teffects” command in Stata 15.0. For all four covariates, the standardized mean differences are less than 0.1 after matching. In addition, three of the four covariates had variance ratios very close to 1 (with the final being moderately greater than one) after matching. Each of these results are well within the suggested ranges for achieving appropriate balance.

Results and Discussion

When sufficient balance is achieved, the estimated treatment effect (ATE) is less sensitive to things like model specification, meaning the researcher can be more confident the results capture a true causal effect. The ATE is the difference in the average outcomes between the treated and non-treated units on the outcome variable of interest. In our case, this constitutes the average difference in DFW rate between instructors who have and have not completed the ACUE program. The analysis offers support for the efficacy of the intervention. The average treatment effect is -0.037 ($p < .001$), meaning that sections taught by ACUE instructors had, on average 3.7% fewer Ds, Fs, and Ws than those sections taught by non-ACUE instructors. To provide some context of the magnitude of this effect, the average enrollment of the course sections in sampled data was just above 30. A 3.7 percentage-point reduction in Ds, Fs, and Ws suggests that, on average, one additional student earned at least a C in the section as a result of the program intervention. To understand the magnitude in a different way, there were 171 course sections in the data that were taught by instructors who had completed the ACUE program at the time they taught the course. The result suggests there were potentially 171 fewer DFW grades. Given the central role that grades play in student retention and degree completion, it is reasonable to speculate that a substantial number of additional credit hours were completed at the institution as a result of the intervention.

Many institutions struggle with student success, which is central to the mission of all institutions of higher learning. For the purposes of this study, student success can be understood as students successfully achieving the learning outcomes of a course, persisting at their institution, and progressing toward graduation. Given that the sections in the analysis were most commonly lower-level courses that disproportionately serve early career students, the observed improvements in DFW rate are likely to implicate student success in a number of ways. Students earning a "C" or better are being judged as having satisfactorily met the course learning outcomes. As a result of the improved grade, students are more likely to persist at the university, and will make better progress toward graduation. Therefore, while hardly a panacea for all student success challenges, interventions like the one assessed here can play a key role in achieving student success, particularly at public institutions.

The results demonstrate that students earned fewer DFW grades in classes taught by instructors who completed the ACUE program. This suggests the intervention was successful. However, it is important to note that the results do not conclusively demonstrate that the success is exclusively a function of the ACUE program itself. As noted above, ACUE is one of many commercial products available for faculty development, and many institutions generate in-house programs that have similar goals. It is possible that the ACUE program offers unique insights, and the results observed here would not occur with any other intervention, commercial or otherwise. However, it may also be the case that any comprehensive faculty development program, in the form of ACUE or something comparable, which asks faculty to think intentionally about their teaching strategies, would produce a similar result. Because the analysis lacked a group of faculty members who experienced a different intervention (rather than just ACUE or no intervention), it is not possible to distinguish between these explanations. What is clear is that this effort to develop faculty teaching appears to improved student grades.

Conclusion

As the demographic profile of students seeking higher education continues to change, institutions, should continually evaluate their strategies for helping students succeed. Too often decisions are made based on expectations about what will help students, rather than evidence-based practices that have demonstrated success in improving measurable outcomes. Even when institutions do seek to measure the effectiveness of student success initiatives, faculty satisfaction is frequently prioritized over rigorous analysis of student outcomes. Faculty satisfaction is an important factor to consider; faculty who are satisfied with the program are more likely to be engaged. However, faculty satisfaction does not guarantee student success—the primary goal of any such program. And even if institutions successfully transcend faculty satisfaction as a measure of success and investigate student outcomes, reliance on descriptive statistics that fail to account for potential confounding factors leaves administrators vulnerable to drawing improper inferences about the effectiveness of their intervention.

In this analysis, we have demonstrated how one intervention—a faculty development program—can help promote student success. We have also demonstrated how proper assessment techniques can promote confidence in the conclusion that the intervention was effective. While our analysis cannot distinguish the effectiveness of this specific intervention from other types of interventions, it does demonstrate that successful interventions are possible. However, only proper assessment will allow an institution to be confident in that conclusion. Many institutions are quick to consider and implement interventions but fail to make a systematic effort to assess their effectiveness. This is particularly troubling for public institutions, where (at least in many places) resources continue to erode.

We believe matching methodologies are particularly well-suited for these types of assessments. Institutions are rarely afforded the opportunity to randomly assign students to treatment and control groups. Instead, they must rely on observational data from observations where students self-select into courses, professors, course times, etc. Statistical matching techniques, when implemented correctly, reduce the potential for unobserved confounders that frequently lead to improper inferences about the effectiveness of an intervention. Efforts at assessment are critical, but poorly designed practices have the potential to lead institutions to draw the incorrect conclusions about their interventions. Institutions undoubtedly wish to implement effective programs and discontinue ineffective ones; only appropriate, rigorous assessment can help them achieve that goal.

References

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine, 28*(25), 3083–3107. https://doi.org/10.1002/sim.3697

Campion, L. L. (2020). Leading through the enrollment cliff of 2026. *TechTrends, 64*, 542-544. https://doi.org/10.1007/s11528-020-00492-6

Condon, W., Iverson, E. R., Manduca, C.A., Rutz, C., & Willett, G. (2016). *Faculty development and student learning: Assessing the connections.* Indiana University Press.

Darling-Hammond, L. (2003). Keeping good teachers: Why it matters and what leaders can do. *Educational Leadership, 60*(8), 6-13.

Darling-Hammond, L., & Richardson, N. (2009). Teacher learning: What matters? *Educational Leadership, 66*(5), 46-53.

Denning, J. T., Eide, E. R., Mumford, K., Patterson, R. W., & Warnick, M. (2021). Why have college completion rates increased? An analysis of rising grades. *National Bureau of Economic Research*. Working Paper. https://doi.org/10.3386/w28710

Devlin, M. (2008). Research challenges inherent in determining improvement in university teaching. *Issues in Educational Research, 18*(1), 12-25.

Fink, L.D. (2013). Innovative ways of assessing faculty development. *New Directions for Teaching & Learning, 2013*(133), 47-59. https://doi.org/10.1002/tl.20045

Hewson, M. G., Copeland, H. L., & Fishleder, A. J. (2001). What's the use of faculty development? Program evaluation using retrospective self-assessments and independent performance ratings. *Teaching and Learning in Medicine, 13*(3), 153-160. https://doi.org/10.1207/S15328015TLM1303_4

Hines, S. R. (2009). Investigating faculty development program assessment practices: What's being done and how can it be improved? *Journal of Faculty Development, 23*(3), 5-19.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15*(3), 199–236. https://doi.org/10.1093/pan/mpl013

Knapp, M. S. (2003). Professional development as policy pathway. *Review of Research in Education, 27*(1), 109–157.

Kucsera, J. V., & Svinicki, M. (2010). Rigorous evaluations of faculty development programs. *Journal of Faculty Development, 24*(2), 5-18.

LaFleur, E. K., Babin, L. A., & Lopez, T. B. (2009). Assurance of learning for principles of marketing students: A longitudinal study of a course-embedded direct assessment. *Journal of Marketing Education, 31*(2), 131-141. https://doi.org/10.1177/0273475309335242

Lawner, E. K., & Snow, M. (2019). *Improved teaching at democracy's college: Findings from Miami Dade College.* Association or College and University Educators. https://acue.org/wp-content/uploads/2019/03/ACUE_MDC-Tech-Report_100120.pdf

Lawner, E. K., & Snow, M. (2020). *Advancing academic equity at Broward College: Improved course completion and passing, particularly among pell-eligible and black students.* Association or College and University Educators. https://acue.org/wp-content/uploads/2020/06/ACUE_Broward-Tech-Report_100120.pdf

Minter, R. L. (2009). The paradox of faculty development. *Contemporary Issues in Education Research, 2*(4), 65-70. https://doi.org/10.19030/cier.v2i4.1073

Morgan, S. L., & Winship, C. (2014). *Counterfactuals and causal inference: Methods and principles for social research* (2nd ed.). Cambridge University Press.

National Center for Education Statistics (2019). Yearly success and progress rates - 2019. https://nscresearchcenter.org/snapshot-report-yearly-success-and-progress-rates-2019/

National Center for Education Statistics (2021). Immediate college going rate. https://nces.ed.gov/programs/coe/pdf/2021/cpa_508c.pdf

Rosenbaum, P. R. & Rubin, D. R. (1983). The central role of propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55. https://doi.org/10.1093/biomet/70.1.41

Rubin, D. B. (1973). Matching to remove bias in observational studies. *Biometrics*, *29*(1), 159-183. https://doi.org/10.2307/2529684

Stuart, E. A., Lee, B. K., & Finbarr L. P. (2013). Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of Clinical Epidemiology, 66*(8), 84-90. https://doi.org/10.1016/jclinepi.2013.01.013