EXECUTIVE SUMMARY

College completion rates in the U.S. are stubbornly low despite the large and rising returns to a college degree. Efforts to increase student success in college have largely ignored a potentially key factor: the instruction that students receive in the sequence of courses that add up to a college education. Little evidence exists about how well students learn the material taught in these courses, largely because student performance is assessed using exams developed by instructors and thus cannot be compared to students at other institutions or even in other sections of the same course at the same college.

The lack of direct measures of student learning in higher education severely hampers efforts to measure the quality of instruction delivered in different classrooms. Improving the quality of instruction may represent a promising path to increasing the number of students who earn high-quality degrees by decreasing frustration and failure, and improving the skills of college graduates. But it is nearly impossible to improve instructional quality without being able to measure it.

This report describes a sophisticated set of common final exams implemented in two developmental algebra courses at Glendale Community College in California. These common finals enable instructors and administrators to compare student performance across different sections, and have earned broad faculty support by being implemented in a way that strikes a balance between standardization and the preservation of faculty autonomy.

I show how data from common finals can be used to measure how much students learn in sections of the same course taught by different instructors, and how instructor characteristics such as education and full-time status are related to student mastery of algebra. These results are limited in scope to the two courses at a single institution represented in my data, but serve as a “proof of concept” of the kind of analyses that are made possible by the adoption of common final exams.

I conclude with four policy recommendations aimed at moving forward efforts to assess and improve the quality of postsecondary instruction and ultimately increase the number of students who earn high-quality credentials:

- First, more departments at more postsecondary institutions should adopt common final exams in their large, multi-section, introductory courses. The exams should be developed by faculty and reflect a consensus among professors
about what students ought to be able to do after completing these introductory courses.

• Second, campus administrators should encourage and provide support for these efforts, such as financial support to cover the modest costs of developing and implementing common finals as well as financial incentives to departments that undertake these efforts. Public university systems and higher education associations such as the American Council on Education and the Association of Public and Land-grant Universities could help coordinate efforts across member institutions.

• Third, administrators should directly address concerns that common finals will be used to evaluate faculty. Some faculty may worry that test-score data will be used in ways that are unfair, and others may be resistant to any form of evaluation that represents a departure from business as usual. But some faculty may support learning-based measures as an alternative to sole reliance on student course evaluations.

• Finally, higher education researchers and practitioners should work to continuously improve common finals. Pre-tests could be developed and administered at the beginning of the semester so that student learning is measured as growth over the course of the semester. Ways to assess student learning in courses other than large, multi-section courses also need to be developed for use in settings such as introductory lecture courses taught by a single instructor.
COMMON SENSE:
Using Common Finals to Measure Postsecondary Student Learning

Introduction

Popular discussions of higher education feature two competing narratives. Critics have argued that higher education is a risky bet that leaves many students deeply in debt, without a job, and often without even a degree. Others, including President Obama, have argued that too few students obtain postsecondary credentials and have set ambitious goals for the U.S. to increase its level of educational attainment.

There is certainly some truth to both of these narratives. Many students take too long to finish college or never graduate at all, and face rising costs that have long outpaced inflation. But the stubborn fact remains that the economic return to a college degree for the average student is substantial and larger than it has been in decades. In 2011, college graduates ages 23-25 earned $12,000 more per year on average than high school graduates, and had employment rates 20 percentage points higher. Over the last 30 years, the increase in lifetime earnings brought by a college degree has increased by 75 percent, whereas costs have increased by 50 percent.

The truth is that a college degree yields substantial benefits on average, but of course “on average” is not the same as “for everyone.” College drop-outs reap fewer benefits than college graduates, and post-graduation outcomes vary from major to major and institution to institution. And costs can vary dramatically across different types of institutions. The end result is that the average return to a college degree is the center-point of a wide distribution, with some students realizing few benefits and even ending up worse off than if they hadn’t gone to college at all, and other students faring even better than the average.

Obtaining more education, like most investments, entails some risk. There will always be students who choose a postsecondary path that turns out to be a bad bet for them. The challenge for policymakers and the
American system of higher education is to reduce that level of risk by improving the quality of education students can expect to receive and reigning in the ever-increasing tuitions they are expected to pay. Much has been written about how to improve educational productivity, and the bottom line is that we do not know nearly as much as we need to about how to produce more high-quality degrees at lower costs.7

An underappreciated path to improving the quality of postsecondary education is to improve the quality of instruction delivered to undergraduates. Higher quality instruction could increase persistence to degrees by decreasing frustration and failure, particularly at institutions that have notoriously low completion rates. The problem is that it is very difficult to improve the quality of instruction without being able to measure it. Some colleges administer tests of general skills such as critical thinking and writing to their students. A recent study of student performance on such tests came to troubling conclusions about the lack of learning that takes place on many college campuses.8

But students don’t receive a standardized body of knowledge called “college” the way most students in a given state do in fourth grade.9 They take individual courses, with some forming part of a coherent program of study, that add up to what we call “college.” The problem is that there is little evidence on how well students in math courses are learning math, and how well writing courses teach writing, and so on. Courses are generally taught by instructors who select the content for the course, and write the exams and other assignments used to assess students’ mastery of the content. The only outcome reported outside of the classroom is the final grade assigned by the instructor. These grades are often based on a curve that ranks students based on their performance but results in the same number of As, Bs, and so on, regardless of how much students actually learned.

Without direct measures of student learning, the quality of instruction can only be assessed using indirect measures such as course completion rates and student evaluations of their instructors.10 The obvious solution to this problem is for instructors to use measures of
student learning that allow for comparisons to be made outside their classroom. The most straightforward settings in which to do this are large, multi-section courses where students learn the same material from different instructors. Such courses lend themselves well to using common exams, such as a common final exam that all students take in the course regardless of which section they are in.

This report describes a handful of common final exam systems currently in use in community colleges in the U.S. The focus of this report is the sophisticated system of common finals used in two developmental algebra courses at a community college in California. Below I describe this system in detail, including its origin and perceived benefits, and then describe the kind of data analysis that it allows. I conclude with policy recommendations for how these kinds of assessments of student learning can be adopted and put to good use in more courses at more institutions.

Algebra Common Finals at Glendale Community College

Glendale Community College (GCC) is a large, diverse campus with a college-credit enrollment of about 25,000 students located in Glendale, California, a city near Los Angeles. GCC has in place a sophisticated system of final exams in its developmental algebra courses that was initiated in 2001 at the suggestion of faculty in the Math Department; this bottom-up initiative had the support of the Division Chair. Faculty members were concerned about the preparation of students in their classes who had taken previous classes that were graded too leniently or did not cover all of the expected material. For example, instructors of intermediate algebra were concerned about students who had passed elementary algebra (the previous course in the sequence) but did not learn the material well enough to succeed in the next course. This grade inflation resulted partly from instructors, many of them adjuncts, who found it easier to give higher grades than to deal with student complaints.

GCC’s algebra exams have evolved over time, particularly with regard to how they are graded. Initially, instructors graded the exams of
their own students, but this created concerns about comparability of grading across sections. Next, instructors were assigned one problem to grade on a partial-credit basis. However, this type of grading was very time-consuming. In the current iteration, which has been in place for about five years, all of the exam questions are graded on a right/wrong basis. There are usually 25 questions, all of which are open-ended (not multiple-choice). Instructors grade a stack of exams that is proportional in size to the number of students they teach, which takes about one hour for the typical instructor. The number of questions answered correctly is recorded, and then the exams are returned to students’ own instructors. The instructors are free to re-grade the exam using whatever standards they prefer (such as partial credit instead of right/wrong grading), and it is that score that is factored into a student’s course grade. Instructors are also able to supplement the exam with their own questions if they so choose.

Test security is taken very seriously. The exams are developed by algebra instructors who are not currently teaching a section covered by the common final, and are not seen by the other instructors until 30 minutes before the exam is administered. Instructors can suggest questions and receive a list of topics covered on the exam. Instructors do not proctor the exams of their own students, and two different forms of each exam are used to mitigate cheating. Many instructors appreciate that much of the work surrounding the final is done for them—all they have to do is pick up a stack of exams, a seating chart, and proctor the exam.

Administrators at GCC indicate that the common final system has gained acceptance among instructors, especially as it has been modified in response to faculty concerns, most notably the reduction in the burden on faculty that came with the move to right/wrong grading. The few instructors who do not wish to have a common final simply choose to teach another course. The common final system has likely been successful because it strikes a careful balance between standardization and instructor autonomy. Administrators report that the standardization has increased accountability for adjuncts—most now cover most of the expected
material—and has reduced grade inflation (with the unintended consequence of reducing pass rates somewhat). At the same time, instructors write their own syllabi and have the option of adding to and re-grading their students’ final exams in addition to assigning course grades.13

The common final exam scores are aggregated by instructor and made available both to the individual instructors, and for all instructors on an anonymous basis. The aggregated data include measures of how students in each section did on the final compared to the rest of the course, and similar information on the grades assigned by instructors. The data are calculated both overall and by test items covering specific topics, so that instructors can identify which topics their students struggled with. The data are discussed at staff workshops, and are occasionally used to identify individual instructors whose grading standards are too lax (or too strict) or whose students performed particularly poorly on the exam. The exam results are not used as part of any formal evaluation process.

The algebra common final system is a significant undertaking that presents logistical challenges and entails financial costs. The primary logistical challenge is reserving enough classrooms so that the exam can be given at the same time to all students in courses with enrollments of about 1,000. The financial costs are mostly comprised of compensation for the exam coordinator and an administrative assistant who digitizes the exam data. Earlier in the history of the common finals, the total cost was about $24,000 per year for the two algebra courses. After recent budget cuts, the cost has been reduced to about half the original amount. The costs of the algebra common finals were covered internally initially and later received support from a Carnegie Foundation grant. Currently, the costs are covered by state money targeted at developmental education.
Other Common Finals

It is difficult to assess the extent to which common finals are used in American higher education. No comprehensive data sources exist to even approximate the extent to which multi-section courses tend to have the same or different final exams across sections. Given the logistical challenges and costs of coordinating exams, especially at institutions where classrooms tend to be small, it seems likely that common finals are the exception rather than the rule. However, they are not exceedingly rare.

Common finals have been used in two developmental English courses at GCC for roughly 20-25 years. They were created at the suggestion of faculty members. The current version of the final, which has been in place for about 15 years, consists of an essay that students write based on long excerpts from 6-7 articles that are provided one week before the final. In order to mitigate cheating, 3-4 different forms of the exam are created that have slightly different questions, but draw from the same sources. The exam is graded by two instructors who do not teach the student; a third reader is used if the first two disagree about whether the student should receive a passing grade.

Instructors can receive data on the relative performance of their students on request, and the exam data are sometimes used to spot problems with part-time faculty. Instructors retain responsibility for determining their students’ course grades; the goal of the common final is achieving consistency in material covered and grading standards. The common final is not mandatory in one of the two courses in which it is used; in that course, 2-3 out of about 17-18 instructors opt out. In general, faculty have come to accept the common final over time, especially as it has evolved.

Common finals are not specific to GCC. Nearby Pasadena City College (PCC) also uses them in multiple courses, including intermediate algebra and chemistry. The intermediate algebra common final is a relatively recent, voluntary effort begun by a professor in the math department. Some instructors choose not to participate because they want...
the freedom to write their own final or are afraid of being evaluated (even though the exam results are not used for that purpose). Participation varies by semester; of 35-40 sections of intermediate algebra, 6-7 used the common final in one semester and 21 used it in another.

Common finals have been used in all major-track classes in the chemistry department at PCC for more than 25 years. The common finals, which are written in multiple versions for test security, must be used by adjuncts and are typically used by full-time faculty as well. In their current iteration, instructors grade their own students’ exams (grading was coordinated in previous iterations). Many of the exams are multiple-choice. The exam results are monitored informally, and patterns of low scores attract attention, but do not factor into any formal evaluation process. The common exams were originally adopted for consistency and to make it easier to bring in part-time faculty and monitor their performance.

A theme of this handful of examples is that common finals have been successful where they have faculty support. Indeed, in most of these cases the idea for common finals originated with the faculty and their support was maintained by modifying the common finals in response to their concerns. Future research should aim to systematically measure the extent to which common finals are used in large, multi-section courses at both community colleges and four-year colleges and universities across the United States. Further in-depth case studies could provide additional evidence on the circumstances under which common final exam systems can be implemented and maintained successfully.

**Putting Common Final Data to Use**

High-quality systems of common final exams such as the one used in GCC’s algebra courses create data that can be used to calculate direct measures of student learning and instructional quality. Whereas most prior studies of student learning in higher education have been forced to use indirect measures of learning such as course completion rates, student course evaluations, and performance in follow-on courses, a well-
designed algebra exam directly measures how well students have learned algebra. Student performance on the common exams can then be aggregated to the section level to examine how much learning varies across classrooms taught by different instructors.

I obtained eight semesters of data from GCC’s elementary and intermediate algebra courses, covering 14,220 observations of 8,654 unique students in 281 sections taught by 76 unique instructors between the spring 2008 and fall 2011 terms. Student performance on the common final exams was linked to administrative records containing background information on students and instructors. Background data on students include their math placement level, race/ethnicity, gender, whether they received a fee waiver (a proxy for financial need), birth year and month, units (credits) completed, units attempted, and cumulative GPA. The instructor data include education level (master’s, doctorate, or unknown), full-time status, birth year and month, gender, ethnicity, years of experience teaching at GCC, and years of experience teaching the indicated course.

These data enable me to measure how much student learning varies across different sections of the same algebra course, controlling for students’ background characteristics. I examine both whether certain instructor characteristics, such as education, are associated with student performance as well as whether some instructors appear to consistently produce better outcomes than others. This analysis is described in detail in the accompanying technical paper, “Instructional Quality and Student Learning in Higher Education: Evidence from Developmental Algebra Courses.” Below I briefly summarize some key results and methodological challenges.

The GCC data indicate that student mastery of algebra is associated with both observed instructor characteristics as well as the identity of individual instructors. Instructor education, full-time status, and experience teaching at GCC are all associated with student performance on the final exam. Students whose instructors have a doctoral degree perform about 0.17 standard deviations worse, on average, than students...
who instructors have a master’s degree. It is not clear why more educated instructors appear to produce worse outcomes, but one potential explanation is that many instructors with doctorates teach at community colleges not because they want to but because they do not have any other options. Instructors with a master’s degree may be more committed to educating undergraduates, on average, than instructors with a doctoral degree.

The students of full-time instructors perform 0.25 standard deviations better, on average, than the students of part-time instructors. Potential explanations for this finding include the hiring of better instructors as full-time faculty and the ability of better paid full-time faculty to focus on teaching at a single institution than their worse-paid part-time counterparts, many of whom have to cobble together a living by teaching courses at multiple institutions. Students also perform better, on average, by 0.21-0.28 standard deviations if they are taught by an instructor with at least one year of prior teaching experience at GCC. This difference appears mostly after the first year, with much smaller differences after additional years of experience.

These differences are large in magnitude but explain a relatively small amount of variation in student learning because the observed instructor characteristics do not vary much across students. Most students are taught by a part-time instructor with a master’s degree who has at least one year of experience teaching at GCC. The identity of a student’s instructor is a much stronger predictor of exam scores than the three observed characteristics of that instructor. Whereas the observed characteristics explain one percent of the differences in exam scores, the identity of the instructor explains eight percent of the differences.

Another way to measure this relationship is to calculate how much student performance on the final exam varies across the classrooms of different instructors. A one-standard-deviation change in instructor quality—for example, moving from an instructor at the median to one at the 84th percentile—is associated with an increase in exam scores of 0.21 standard deviations. To put this number in context, it is similar in size to
increase in student performance predicted by an increase in cumulative GPA (prior to taking the algebra course) of more than half a grade point (e.g., a 3.5 GPA vs. a 3.0 GPA).

Using common final exam data to measure the quality of postsecondary instruction raises potentially important methodological concerns. First, the results may be biased if better students consistently choose some instructors over others in a way that cannot be accounted for by the student characteristics factored into the analysis. Second, final exam scores are not available for students who drop the course before the final exam. Consequently, an instructor who does a good job at getting students to complete the course may have an unfairly lower average score among her students than an ineffective instructor who has many students drop the course. These issues, which are discussed in detail in the technical paper, do not appear to significantly bias the analysis of the GCC data, but they should continue to be addressed in future work on student learning in postsecondary education.

The results of this analysis should not be extrapolated beyond the limited context from which the data are drawn: two courses at a single community college. In other words, the findings may not apply to other courses at other colleges. Instead, the results should be interpreted as preliminary and the analysis regarded as a “proof of concept” of the kind of work that is made possible by the use of common final exams.

Conclusions and Policy Recommendations

There is a widespread consensus in the research and policy communities that increasing the number of high-quality postsecondary credentials earned by American students would be good for both individuals and society. However, there is disappointingly little evidence on how to get more students to complete college successfully, and to learn more while they are there. The research that does exist has largely focused on policies and practices at the periphery of the educational experience, such as student aid, financial incentives, and student support services.17 What goes on in the classroom—the heart of a student’s
educational experience—has been largely ignored. Presumably exposing students to higher quality instruction could increase both completion rates and degree quality (how much students have learned), but little research exists that attempts to measure student learning and instructional quality at the course level.

The unfortunate dearth of research in this area results largely from a lack of data on student mastery of course content, a problem that is solvable in many contexts through the use of common final exams. The case studies of common finals documented in this report show not just how the resulting data can be used, but provide anecdotal evidence that the use of common exams can address other problems such as grade inflation and the lack of consistency in material covered by different instructors responsible for the same course, many of whom do not have a full-time appointment at the institution. Common finals also have potential uses not discussed in this report such as the evaluation of instructional materials, including the interactive, technology-enhanced products that are quickly becoming popular in many postsecondary institutions.

The use of common final exams to measure the quality of instruction is clearly at an early stage of development. Systems of common finals exist, but very few have been used to carry out the kinds of analyses described in this report. I propose four recommendations that, if adopted, could help move forward efforts to assess and improve the quality of postsecondary instruction. First, more departments at more postsecondary institutions should adopt common final exams in their large, multi-section, introductory courses. These exams should not be imposed externally or from on high by campus administrators, but rather should be developed by faculty in the department. The exams should reflect a consensus among professors about what students ought to know and be able to do after completing the course. Faculty autonomy can be preserved by allowing professors to supplement and regrade the common final, but the scores from the common portion should be recorded in the institution’s administrative records.
Second, campus administrators should encourage and provide support for these efforts. It is important that administrators not unilaterally require common finals if they are to obtain faculty support. But administrators can provide financial support to cover the modest costs of developing and implementing common finals, as well as financial incentives to departments that undertake these efforts. Public university systems could also provide support for these efforts on multiple campuses, and coordinate the development of exams that are common across multiple institutions. There would be clear advantages to knowing that successful completion of a course conveys the same level of mastery at multiple campuses, especially across the two- and four-year sectors. Higher education associations such as the American Council on Education and the Association of Public and Land-grant Universities could also help coordinate efforts across member institutions.

Third, administrators should directly address concerns that common finals will be used to evaluate faculty. Some faculty may worry that test-score data will be used in ways that are unfair, and others may be resistant to any form of evaluation that represents a departure from business as usual. These concerns may be relatively easy to alleviate in the short-run given the job protections enjoyed by tenured faculty. Such faculty may be willing to support the use of exam data as one element in an evaluation system, especially if evaluations have the highest stakes for part-time faculty. Some instructors may also be more willing to support the use of exam data if existing evaluation systems are based largely on student evaluations.

Finally, higher education researchers and practitioners should work to continuously improve common finals. Professional psychometricians could be retained to develop assessments that are more reliable and valid, in some cases by adding items that require students to complete more complex tasks than those assessed by short answer or multiple-choice questions. Pre-tests could be developed and administered at the beginning of the semester so that student learning is measured as growth over the course of the semester. And ways to assess student
learning in courses other than large, multi-section courses need to be developed. For example, a large introductory course taught by a single lecturer would likely use a single exam for all students, but student performance on the exam couldn’t be compared to students of a different instructor because there is only one instructor. A potential solution would be to mimic common finals by drawing from a large bank of test items to create final exams in each semester of a given course. With the right design features, student performance on different administrations of the exam (e.g., in different semesters or at different campuses) would be comparable. Such an approach would also solve many of the logistical challenges of having a large number of students taking an identical exam at the same time for security reasons.

At many institutions, it will not be easy to move away from the traditional model in which individual instructors are responsible for assessing the performance of their own students independent of any agreed-upon standard. Common finals offer clear benefits to instructors, such as removing the work associated with writing a final exam every semester, but may draw the ire of those who see it as an infringement on their independence. But the political challenges are likely to be surmountable. The existence of common final exams shows that it is possible for such systems to be adopted if they have faculty support. Convincing higher education practitioners to develop and adopt common finals is therefore a crucial first step in efforts to implement common yardsticks that institutions can use to assess how much their students are learning, and ultimately increase the number of students who earn high-quality credentials.
Endnotes


3 Michael Greenstone and Adam Looney, “Regardless of the Cost, College Still Matters,” Brookings on Job Numbers, Brookings Institution, 2012. See also Figure 1.6 of Education Pays 2010, College Board, 2010.

4 Figure 1.6 of Education Pays 2010, College Board, 2010.

5 See, for example, the State Council of Higher Education for Virginia’s Post-Completion Wages of Graduates, available at http://research.schev.edu/apps/cms/Post-Completion-Wages-of-Graduates.aspx.


9 There is surely variation in the content delivered across fourth-grade classrooms in a state, but it is much less than the variation produced by the course-taking decisions of college students.

10 For a discussion of previous research, see the technical paper that accompanies this report.


12 This section is largely based on interviews conducted in April 2012 with the following GCC staff: Kathy Holmes (current Math Division Chair), Peter Stathis (Math Division Chair from 1998 to 2008), and Yvette Hassakoursian (common final exam coordinator).

13 In addition to standardizing the teaching of algebra at GCC, the common final system has also created more of a sense of camaraderie among instructors, especially adjuncts. Many instructors find the friendly competition both challenging and fun.

14 The description of the English common finals is based on an interview conducted in April 2012 with Steve Taylor (English professor at GCC).

15 The descriptions of these common finals are based on interviews conducted in April 2012 with Yu-Chung Chang (math professor) and Kerin Huber (chemistry professor).

16 The results discussed are those that control for student characteristics (reported in Table 3 of the technical paper that accompanies this report). Results that do not control for student characteristics are qualitatively similar.

About the Brown Center on Education Policy
Established in 1992, the Brown Center on Education Policy conducts research and provides policy recommendations on topics in American education. The Brown Center is part of The Brookings Institution, a private nonprofit organization devoted to independent research and innovative policy solutions. For more than 90 years, Brookings has analyzed current and emerging issues and produced new ideas that matter - for the nation and the world.

Brown Center on Education Policy
The Brookings Institution
1775 Massachusetts Ave. NW
Washington DC, 20036
202.797.6090
202.797.6144 (f)
http://www.brookings.edu/brown.aspx

Email your comments to
browncenter@brookings.edu

The Brookings Institution is a private non-profit organization. Its mission is to conduct high-quality, independent research and, based on that research, to provide innovative, practical recommendations for policymakers and the public. The conclusions and recommendations of any Brookings publication are solely those of its author(s), and do not reflect the views of the Institution, its management, or its other scholars.

Brookings recognizes that the value it provides is in its absolute commitment to quality, independence and impact. Activities supported by its donors reflect this commitment and the analysis and recommendations are not determined or influenced by any donation.