

Assessing Long-Term Effects of Inquiry-Based Learning: A Case Study from College Mathematics

Marina Kogan · Sandra L. Laursen

Published online: 17 August 2013

© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract As student-centered approaches to teaching and learning are more widely applied, researchers must assess the outcomes of these interventions across a range of courses and institutions. As an example of such assessment, this study examined the impact of inquiry-based learning (IBL) in college mathematics on undergraduates' subsequent grades and course selection at two institutions. Insight is gained upon disaggregating results by course type (IBL vs. non-IBL), by gender, and by prior mathematics achievement level. In particular, the impact of IBL on previously low-achieving students' grades is sizable and persistent. The authors offer some methodological advice to guide future such studies.

Keywords Mathematics · Inquiry-based learning · Academic records · Grades · Course selection

Student-centered or “active” forms of instruction have been shown to improve student learning and affective outcomes in the sciences and in other fields (Ambrose et al., 2010; Froyd, 2008; Hake, 1998; Prince & Felder, 2007; Springer, Stanne & Donovan, 1999). Yet these proven, “high-impact” educational practices are not typical of what students experience in college (Kuh, 2008). Especially crucial are mathematics courses, key prerequisites that may regulate students' access to many majors and careers, or to any college degree at all (Carlson et al., 2011; Seymour & Hewitt, 1997; Stigler, Givvin & Thompson, 2010). Thus the use of active learning methods in college mathematics may help to attract and retain students, including students of diverse backgrounds (Ellis, Rasmussen & Duncan, 2013; Fullilove & Treisman, 1990; Watkins & Mazur, 2013).

Marina Kogan has bachelors' degrees in sociology/anthropology and computer science from City University of New York and an M.A. in sociology from the University of Illinois Urbana-Champaign. She conducted this work as a researcher at Ethnography & Evaluation Research at the University of Colorado Boulder and is now pursuing a doctorate in computer science from the University of Colorado Boulder, with research interests in the application of computer modeling and data mining methods to social problems.

Sandra L. Laursen has undergraduate degrees in chemistry and French from Grinnell College and a Ph.D. in chemistry from the University of California-Berkeley. She is the Co-director of Ethnography & Evaluation Research at the University of Colorado Boulder and leads evaluation and research studies of higher education and career paths in science, engineering, and mathematics. She can be reached at Sandra.laursen@colorado.edu.

M. Kogan · S. L. Laursen (✉)

Ethnography & Evaluation Research, University of Colorado Boulder, 580 UCB, Boulder,
CO 80309-0580, USA

e-mail: sandra.laursen@colorado.edu

To date, the most persuasive studies of active learning have examined student outcomes within a single course at one or several institutions (e.g., Deslauriers et al., 2011; Kwon, Rasmussen & Allen, 2005; and studies analyzed by Froyd, 2008, and Ruiz-Primo et al., 2011). However, as active learning approaches are applied more broadly, evaluating their outcomes presents new methodological challenges. Measures to assess effectiveness must be general enough to apply across different classrooms and institutions. A common test—the most direct method of evaluating classroom learning—may not be available or applicable.

Students' course grades and course-taking patterns—their choices to pursue (or not) subsequent courses in a discipline—offer broad and arguably objective measures for evaluating the effects of an educational intervention. While grading standards differ across instructors, courses, and campuses, grades have a fairly stable social meaning (Pattison, Grodsky & Muller, 2013). As part of students' academic transcripts, grades become lasting records of achievement. Like grades, course-taking patterns apply to varied academic contexts; they may reflect students' sustained or lost interest in a discipline following an initial experience. Several recent studies have used various grade- and course-taking measures to evaluate the success of an educational intervention, including final grades and pass/fail rates (e.g., Dubetz et al., 2008; Tai, Sadler & Mintzes, 2006; Tien, Roth & Kampmeier, 2002), the next grade in a course sequence (e.g., Farrell, Moog, & Spencer, 1999; Gafney & Varma-Nelson, 2008), grades in multiple subsequent courses (e.g., De Paola, 2009; Weinberg, Hashimoto, & Fleisher, 2009), and enrollment in higher level electives (Carrell & West, 2010).

Mostrom and Blumberg (2012) suggested that student-centered courses are subject to accusations of grade inflation because the course has lost content or rigor or because different assessment methods enable students to do better. They also argued that grade improvement may in fact measure real improvement in learning. Measures that focus on *subsequent* courses avoid this issue because students who did and did not experience the intervention all take the same later courses. Moreover, such measures can detect valued and lasting impact on students' learning, academic success, or academic choices (Derting & Ebert-May, 2010).

This study examined undergraduates' grades and course-taking following an inquiry-based learning (IBL) experience in college mathematics. In the context of mathematics, IBL approaches engage students in exploring mathematical problems, proposing and testing conjectures, developing proofs or solutions, and explaining their ideas. As students learn new concepts through argumentation, they also come to see mathematics as a creative human endeavor to which they can contribute (Rasmussen & Kwon, 2007). Consistent with current socio-constructivist views of learning, IBL methods emphasize individual knowledge construction supported by peer social interactions (Ambrose et al., 2010; Cobb, Yackel & McCain, 2000; Davis, Maher & Noddings, 1990).

In this article we report our analysis of student academic records for patterns in grades and course-taking among students who had earlier taken an IBL mathematics course or comparative, "non-IBL" course taught with other methods. We focus on results for two groups often under-served by traditionally taught college mathematics courses, women and low-achieving students.

The Study

The academic records study was one element of a large, mixed-methods study of IBL mathematics as implemented at four universities hosting IBL Math Centers (Laursen, Hassi & Hough, 2013; Laursen, Hassi, Kogan, Hunter & Weston, 2011; Laursen, Hassi, Kogan & Weston, 2013).

Observation, survey, interview, and test data were gathered from over 100 sections of 40 courses aimed at varied levels and audiences. First we describe results of classroom observations which establish that IBL was a student-centered, educational intervention. We then outline the methods used to study subsequent grades and course-taking for students who had completed an IBL course or its non-IBL counterpart (for details see Laursen et al., 2011).

Setting and Courses

Each of the four institutions selected and developed its IBL courses independently and labeled them as IBL or non-IBL based on instructor participation in their grant-funded IBL Center. The courses were well established, having been taught several times prior to our data collection in 2009. To check these labels and to establish whether observed differences in student outcomes were meaningful, we carried out over 300 hours of classroom observation of 42 course sections, having received human subjects approval from our University's Institutional Review Board and that of each study site where required. The results showed that, despite variation among courses and instructors, several key characteristics differentiated the IBL courses from the non-IBL courses. On average, about 60% of class time in IBL courses was spent on student-centered activities such as small group work, student presentation of problems at the board, or whole-class discussion, while in non-IBL courses over 85% of class time consisted of the instructor talking. In IBL courses, students more often took on leadership roles and asked more questions. Trained observers rated IBL courses higher for creating a supportive classroom atmosphere, eliciting student intellectual input, and providing feedback to students on their work. Overall, the data clearly show that students who took IBL sections experienced a different instructional approach than those in lecture-based, non-IBL sections (Laursen et al., 2011, 2013b).

The academic records study focused on two Centers that offered high-enrollment courses taught in both IBL and non-IBL formats. We chose three target courses based on the following:

- Placement early enough in a typical course sequence to allow for variation in subsequent course choices and grades,
- Target sections taught in prior years early enough that subsequent course-taking was near complete at the time of data collection in 2009, and
- Adequate numbers of students enrolled in both IBL and non-IBL sections.

The two courses at Center L were a middle-level introduction to proof course (designated L1) and an advanced proof-based course (L2). L1 aimed to help students shift from the problem-solving of calculus to the rigorous proof-based approach of advanced courses. It met degree requirements for mathematics, some science and engineering fields, and secondary math teaching. Course L2 was not required but counted toward the math major. Both L1 and L2 were taught in sections of 20–30 students. Course sections were not institutionally labeled as IBL or non-IBL; self-selection occurred but was not extensive. Therefore we used statistical methods to control for differences among entering students that might affect their later academic outcomes.

The third course, G1, was the first course in a three-term sequence including multivariable calculus, linear algebra, and differential equations; all three were offered in IBL and non-IBL formats. Both institutional selection and student self-selection operated heavily in this course. Students were invited to join the IBL “honors” section based on past mathematics performance, thus populating these sections with high-achieving, self-motivated students. Non-IBL sections included students of all prior achievement levels taught in large lectures with recitations led by graduate teaching assistants. On average, IBL students had

higher SAT scores and high school GPAs than non-IBL students, took G1 earlier (often in their first college term), and pursued mathematics majors in higher numbers. To compare these groups fairly, we constructed a matched sample, which is detailed below.

Variables

We considered several measures by which to assess academic outcomes, beginning with anonymized raw data from standard institutional records. DFW rates, the proportion of students who fail (earn D or F grades) or withdraw (W) from a class (Dubetz et al., 2008), were not useful since they are low in honors and upper-level courses. Instructors argued that grades and exam scores could not be compared across IBL and non-IBL sections, given differences in emphasis and assessment. Instead, we developed standardized approaches to counting and averaging grades in courses taken *after* the target course. Because students who took IBL or non-IBL sections of the target later co-enrolled in other courses, their later grades can be compared directly to each other, albeit not across courses or institutions. Here we outline the standardized variables.

Variables describing course counts focus on new courses completed. Repeated courses were counted if the initial attempt ended in withdrawal or a failing grade. Because many students had not graduated by the time of data collection, we analyzed only courses taken within two years after the target course. Course-counting variables included the following:

- Number of *prior* math courses—before the target course, a control for math background;
- Number of *subsequent* math courses—all courses taken after the target course;
- Number of subsequent *elective* courses—elective courses taken after the target course, other than core courses required for the mathematics major; and
- Number of subsequent *IBL* courses—IBL-method courses taken after the target course.

The number of subsequent required courses is largely determined by students' progress in the major and not a useful measure of student choice. Major-switching in or out of mathematics was minimal for all groups in these courses.

Variables to describe math grades focus on courses for which a letter grade was received. Differently from the count variables, all letter grades were included in calculating averages, including repeated courses. Grade variables included the following:

- Average *prior* grade—in courses before the target course, a control for math achievement;
- *Next* term average grade; and
- Average grade in subsequent *elective*, *required*, and *IBL* courses.

For the average grade variables, sample sizes differed, since not all students took all types of courses. For example, most non-IBL students took no IBL courses. Statistical control procedures further reduced sample sizes as not all student records contained the data used to control for incoming differences. Table 1 shows the maximum sample sizes for each course; actual sample sizes for each grade variable are included with the detailed results in Tables 2 and 3.

To account for students' prior achievement or ability entering the first-year course G1, we created an index combining students' high school GPA with college admissions test scores. Concordance tables were used to convert ACT to SAT mathematics scores and ACT English and reading scores to SAT verbal scores (Dorans et al., 1997; Dorans, 1999, 2004). Because high school grades and admissions test scores are comparably important predictors of college success (Hoffman & Lowitzki, 2005; Noble, 1991), the index weighted high school GPA, math SAT score, and verbal SAT score approximately equally. The new index was divided into seven equal brackets.

Table 1 Study Samples for Academic Records Analyses

Course	Sample size			Major demographic sub-groups			
	Total	IBL	Non-IBL	Class year	Gender	Race & ethnicity	Major
L1: Mid-level statistical	1341	211	1130	23% soph 28% jr 39% sr	71% M 29% F	52% white 21% Asian 12% Hispanic	60% math 30% S&E
L2: Upper-level statistical	909	123	786	26% jr 52% sr	65% M 35% F	51% white 19% Asian 16% Hispanic 10% foreign	71% math 18% S&E
G1: Introductory sampling, statistical	197 (of 962)	49	98	72% first 24% soph	62% M 38% F	64% white 18% Asian 10% Hispanic	27% math 55% S&E

Sampling and Analysis

Table 1 summarizes the demographic information for samples from the three courses, totaling 3,212 students. Non-IBL samples were larger because IBL section offerings were limited. To develop a non-IBL sample comparable to the selective population in IBL sections of G1, we used the pre-college index plus demographic variables to match students. For each IBL student we selected two non-IBL students matched by index bracket, academic major (math, science, non-STEM, undeclared), academic status (freshman-senior), gender, and race/ethnicity—in that order of priority. Overall, this process yielded highly similar IBL and non-IBL samples.

SPSS (version 18) was used for statistical analyses. To compare means for IBL and non-IBL students we used primarily non-parametric tests (Mann-Whitney, Kruskal-Wallis, Chi-square), as most of the data were not normally distributed. We found some incoming differences between IBL and non-IBL student groups in the number of math courses and average prior math grade. For L1, IBL students had taken fewer prior math courses and earned higher average math grades prior to the target course. For L2, these differences were not significant. For G1, even after our close-match sampling, there was still a significant difference in the number of prior math courses, with IBL students taking fewer. Thus all reported results are based on applying the General Linear Model (GLM) procedure in SPSS to control for these incoming differences, using as covariates the number of math courses and average prior math grade, or for G1, the pre-college index. We report estimated marginal means, which are intended to offset the effect of the covariates as intervening variables.

Effect sizes for the IBL intervention were computed from estimated marginal means and pooled standard deviations for all students and by gender. A different approach was required for effect sizes by prior achievement group. The GLM procedure adjusts the post-intervention student outcomes by controlling for prior math GPA. Because the achievement subgroups are based on prior math GPA, using these adjusted outcome measures to calculate effect sizes by achievement group would obfuscate precisely the group differences of interest. Instead, we used Morris’ (2008) Pretest-Posttest-Control group design. This method controls for preexisting differences even when treatment and control groups are nonequivalent, by allowing “each individual to be used as his or own control, which typically increases the power and precision of statistical tests” (p. 365). This design is only appropriate for the grade variables, as the number of math courses taken prior to the intervention does not have a pretest relationship to subsequent course counts; it is a measure of student preparedness instead.

Table 2 Estimated marginal means for average grade and courses taken subsequent to an IBL or non-IBL mathematics course, for all students and by gender, for three courses

<i>Course L1, 2001-2008</i>	non-IBL			IBL		
	men	women	all	men	women	all
Sample size (overall & all course counts)	755	322	1077	147	57	204
<i>Average grade in the next term</i>						
N	351	175	526	66	23	89
Mean	2.816	2.726	2.786	2.970	2.922	2.957
standard deviation	0.918	0.926	0.917	0.926	0.921	0.925
<i>Average grade in subsequent required courses</i>						
N	336	163	499	67	37	104
Mean	2.626	2.635	2.629	2.846	2.706	2.796
standard deviation	0.917	0.945	0.916	0.925	0.918	0.918
<i>Average grade in subsequent elective courses</i>						
N	507	218	725	89	41	130
Mean	2.790	2.909	2.826	2.936	2.963	2.945
standard deviation	0.878	0.886	0.862	0.877	0.871	0.878
<i>Average grade in subsequent IBL courses</i>						
N	51	29	80	22	8	30
Mean,	2.523	2.534	2.528	3.033	2.826	2.975
standard deviation	1.000	1.007	0.975	1.008	0.973	0.997
Sig., IBL vs. non-IBL			*			*
Effect size for IBL intervention						0.456
<i>Number of subsequent required courses</i>						
Mean	0.473	0.586	0.507	0.529	0.792	0.604
standard deviation	0.632	0.646	0.624	0.630	0.627	0.628
Sig., IBL vs. non-IBL			*			*
Sig., IBL vs. non-IBL, within gender		*			*	
Effect size for IBL intervention					0.320	0.155
Sig., men vs. women	**	**		*	*	
<i>Number of subsequent elective courses</i>						
Mean	1.768	1.887	1.803	1.675	1.551	1.641
standard deviation	1.841	1.848	1.805	1.831	1.820	1.828
<i>Number of subsequent IBL courses</i>						
Mean	0.059	0.075	0.064	0.137	0.142	0.138
standard deviation	0.275	0.269	0.263	0.267	0.264	0.271
Sig., IBL vs. non-IBL			***			***
Sig., IBL vs. non-IBL, within gender	**			**		
Effect size for IBL intervention				0.285		0.280
<i>Course L2, 2002-2008</i>						
	non-IBL			IBL		
	men	women	all	men	women	all
Sample size (overall & all course counts)	477	270	747	77	40	117
<i>Average grade in the next term</i>						
N	204	122	326	30	11	41
Mean	2.498	2.736	2.588	2.716	3.039	2.797
standard deviation	1.000	0.994	0.993	0.997	1.002	0.999

Table 2 (continued)

<i>Average grade in subsequent required courses</i>						
N	92	38	130	17	5	22
Mean	2.214	2.418	2.274	2.434	3.014	2.564
standard deviation	1.045	1.042	1.038	1.047	1.042	1.046
<i>Average grade in subsequent elective courses</i>						
N	289	157	446	48	23	71
Mean	2.642	2.708	2.665	2.579	2.924	2.690
standard deviation	0.918	0.915	0.908	0.908	0.911	0.910
<i>Average grade in subsequent IBL courses</i>						
N	15	8	23	6	1	7
Mean,	2.194	3.286	2.564	2.205	4.574	2.568
standard deviation	0.918	0.939	1.093	0.926	0.928	1.101
<i>Number of subsequent required courses</i>						
Mean	0.051	0.022	0.041	0.019	0.031	0.023
standard deviation	0.197	0.197	0.191	0.193	0.196	0.195
<i>Number of subsequent elective courses</i>						
Mean	1.399	1.482	1.429	1.337	1.338	1.338
standard deviation	1.594	1.610	1.585	1.588	1.600	1.590
<i>Number of subsequent IBL courses</i>						
Mean	0.020	0.010	0.016	0.038	-0.001	0.024
standard deviation	0.131	0.148	0.137	0.140	0.139	0.141
Course G1, 2004-2006	non-IBL (matched sample)			IBL		
	men	women	all	men	women	all
Sample size (overall & all course counts)	61	37	98	28	19	47
<i>Average grade in the next term</i>						
N	47	21	68	24	15	39
Mean	2.935	3.132	2.995	3.477	3.279	3.402
standard deviation	0.912	0.926	0.907	0.916	0.906	1.001
Sig., IBL vs. non-IBL			*			*
Sig., IBL vs. non IBL within gender	*			*		
Effect size for IBL intervention				0.593		0.430
<i>Average grade in subsequent required courses</i>						
N	46	24	70	24	16	40
Mean	3.022	3.095	3.046	3.209	3.026	3.137
standard deviation	0.794	0.813	0.786	0.794	0.788	0.791
<i>Average grade in subsequent elective courses</i>						
N	17	7	20	9	7	16
Mean	3.052	2.975	3.029	3.498	2.325	2.979
standard deviation	1.064	0.958	0.997	0.945	0.937	1.000
Sig., men vs. women				*	*	
<i>Average grade in subsequent IBL courses</i>						
N	3	0	3	21	14	35
Mean	3.419	-	3.416	3.602	3.317	3.469
standard deviation	0.461	-	0.476	0.454	0.456	0.467

Table 2 (continued)

<i>Number of subsequent required courses</i>						
Mean	2.040	1.815	1.958	2.300	1.788	2.088
standard deviation	1.453	1.490	1.445	1.460	1.447	1.460
<i>Number of subsequent elective courses</i>						
Mean	0.771	0.475	0.661	0.996	1.500	1.197
standard deviation	1.953	1.995	1.841	1.958	1.940	1.954
<i>Number of subsequent IBL courses</i>						
Mean	0.042	0.019	0.034	1.203	0.898	1.206
standard deviation	0.461	0.474	0.465	0.466	0.462	0.466
Sig., IBL vs. non-IBL			***			***
Sig., IBL vs. non-IBL within gender	***	***		***	***	
Effect size for IBL intervention				2.51	1.87	2.52

Notes on Table 2: Average grades are reported on a 0–4 point scale. Statistically significant differences between subgroups are indicated in the marked columns: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Rows are included only for comparisons where at least one statistically significant difference was detected. Effect sizes are reported for the IBL intervention, only for groups or sub-groups where a significant difference was detected

Results

The IBL status of the target course is the primary independent variable by which grades and course-taking are compared, using the defined variables. We first describe the results for all IBL vs. non-IBL students, then disaggregate results by gender and by prior achievement level.

All Students

Table 2 and Figure 1 show the grades and course-taking patterns of all students who had taken IBL or non-IBL sections of the target courses. In all three courses, L1, L2, and G1, IBL students' grades were as good or better than their non-IBL peers (Figure 1a). In two cases, IBL students' grades were statistically significantly better.

When number of courses is examined (Figure 1b), little difference is seen for the more advanced students in L1 and L2. Both IBL groups took modestly (but not significantly) fewer elective courses. Students who took L1 in IBL format were more likely to opt for a second IBL course, while students in L2 had little opportunity to take additional IBL courses.

Among the first-year students in course G1, however, IBL students pursued more math courses, especially IBL courses (of which three more were available). The difference in elective course count is large, though not significant due to the small sample and high variance. Because the IBL and non-IBL groups were well matched by major, the difference in elective choice is not due to differing requirements. The difference in students' pursuit of IBL courses is significant.

By Gender

Prior results from immediate post-course surveys had flagged gender as salient in students' response to IBL vs. non-IBL classes (Laursen et al., 2013b). Therefore we examined whether gender differences persisted into later courses (Figure 2, Table 2). For grades

Table 3 Estimated marginal means for grades and course-taking subsequent to an IBL or non-IBL mathematics course, by prior achievement level, for one course

<i>Course L1, 2001-2008</i>	Non-IBL			IBL		
	Low	Medium	High	Low	Medium	High
Sample size (overall & all course counts)	360	353	364	49	76	79
<i>Average grade in the next term</i>						
N	186	184	156	18	36	35
Mean	2.064	2.885	3.438	2.427	2.924	3.680
std. deviation	0.941	0.922	0.937	0.925	0.924	0.935
Sig., L vs. M	***	***				
Sig., M vs. H		***	***		***	***
Sig., L vs. H	***		***	**		**
<i>Average grade in subsequent required courses</i>						
N	180	162	157	24	39	41
Mean	1.959	2.584	3.360	2.429	2.748	3.377
std. deviation	0.939	0.929	0.940	0.931	0.931	0.941
Sig., IBL vs. non-IBL	*			*		
Sig., L vs. M	***	***				
Sig., M vs. H		***	***		**	**
Sig., L vs. H	***		***	**		**
<i>Average grade in subsequent elective courses</i>						
N	235	253	237	30	45	55
Mean	2.195	2.778	3.442	2.344	2.963	3.521
std. deviation	0.889	0.891	0.893	0.887	0.885	0.897
Sig., L vs. M	***	***		**	**	
Sig., M vs. H		***	***		**	**
Sig., L vs. H	***		***	***		***
<i>Average grade in subsequent IBL courses</i>						
N	37	21	22	7	8	15
Mean	1.571	3.053	3.224	2.992	3.263	3.420
std. deviation	1.010	0.999	0.999	0.997	0.998	1.026
Sig., IBL vs. non-IBL	*			*		
Sig., L vs. M	***	***				
Sig., L vs. H	***		***			
<i>Number of subsequent required courses</i>						
Mean	0.495	0.531	0.494	0.513	0.622	0.648
std. deviation	0.626	0.639	0.630	0.630	0.628	0.631
<i>Number of subsequent elective courses</i>						
Mean	1.572	1.914	1.909	1.379	1.651	1.868
std. deviation	1.840	1.822	1.832	1.820	1.822	1.840
Sig., L vs. M	**	**				
Sig., L vs. H	*		*			
<i>Number of subsequent IBL courses</i>						
Mean	0.077	0.054	0.061	0.122	0.105	0.179
std. deviation	0.266	0.263	0.267	0.266	0.270	0.267

Table 3 (continued)

Course L1, 2001-2008	Non-IBL			IBL		
	Low	Medium	High	Low	Medium	High
Sig., IBL vs. non-IBL			***			***

Notes on Table 3: Average grades are reported on a 0-4 point scale. Categories for prior achievement are based on mathematics grades for courses prior to Course L1: Low = $GPA \leq 2.5$, Medium = $2.5 < GPA \leq 3.4$; High = $GPA > 3.4$. Statistically significant differences between subgroups are indicated in the marked columns: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Rows are included only for comparisons where at least one significant difference was detected

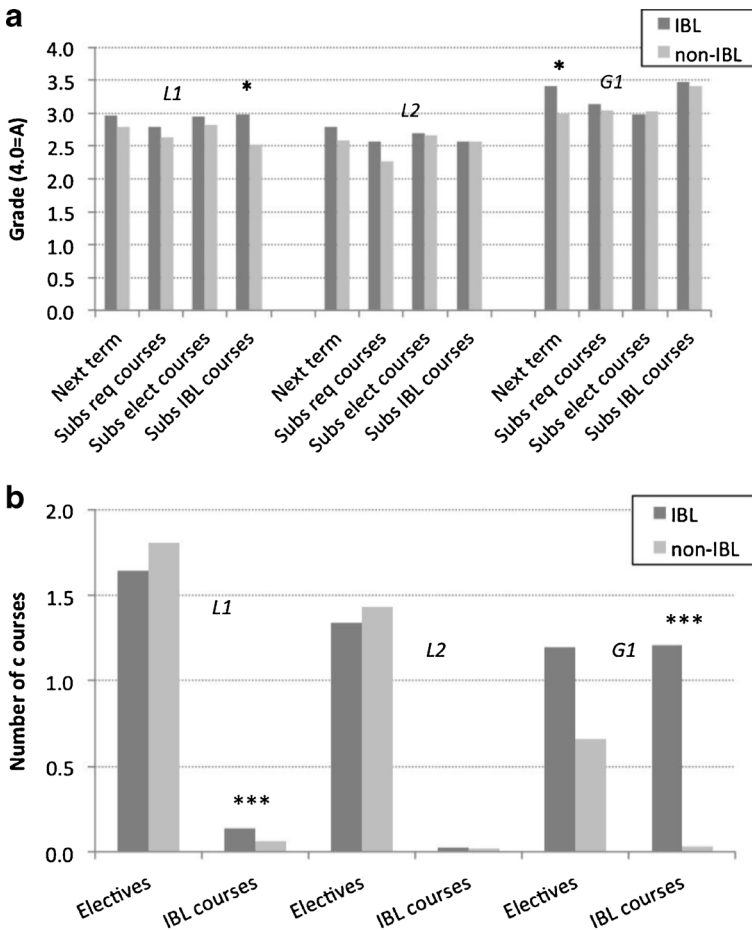


Fig. 1 Grades and number of subsequent courses for all students after taking IBL and non-IBL sections of courses L1, L2, and G1. (a) Average grades for next term and all subsequent required, elective and IBL courses. (b) Average number of courses taken, elective and IBL courses. Group difference is statistically significant: * at $p < 0.05$, ** at $p < 0.01$, *** at $p < 0.001$

subsequent to the target course (Figure 2a), few significant differences were detected. In the upper-division courses L1 and L2, there was a minor but not significant pattern of IBL students (both men and women) out-performing non-IBL students of their own gender. For course G1, this pattern continued and is significant in one case, men’s next-term grades.

Comparing student grades by gender, some differences are large, at several tenths of a grade point, though not statistically significant. In the advanced course L2, women in both IBL and non-IBL groups tended to outperform their male classmates. For the mid-level course L1, there was little difference in men’s and women’s subsequent grades, while men in the early course G1 tended to outperform women in their own (IBL or non-IBL) sections.

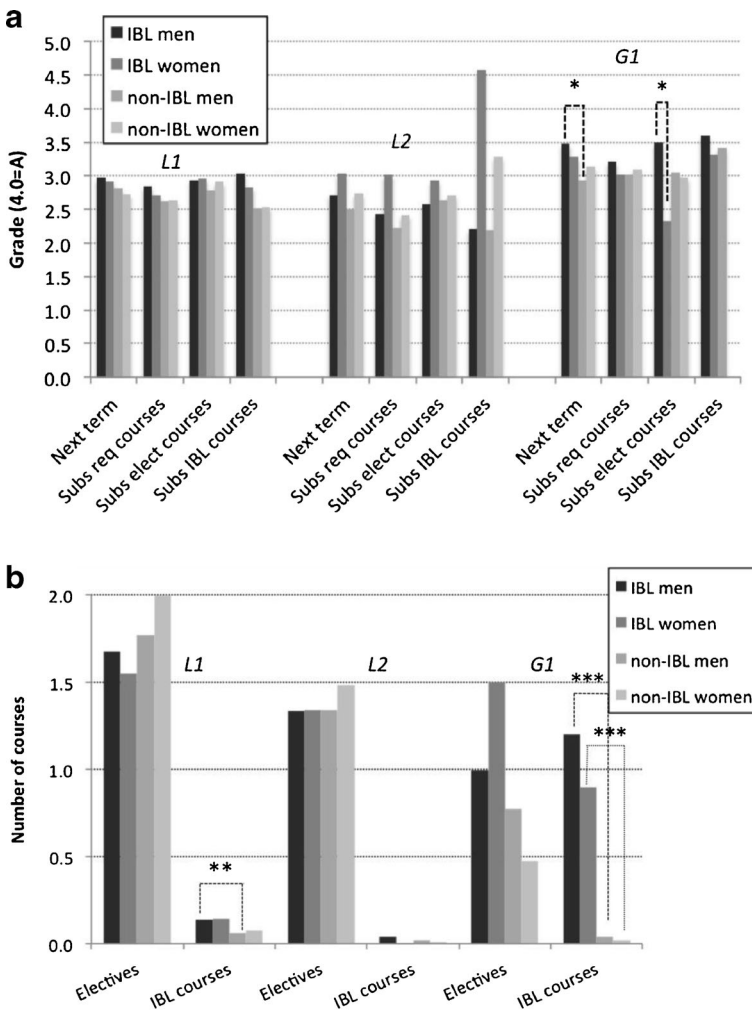


Fig. 2 Grades and number of subsequent courses for men and women after taking IBL and non-IBL sections of courses L1, L2, and G1. (a) Average grades for next term and all subsequent required, elective and IBL courses. (b) Average number of courses taken, elective and IBL courses. Subgroup difference is statistically significant: * at $p < 0.05$, ** at $p < 0.01$, *** at $p < 0.001$

Portraying numbers of courses by gender and IBL status, Figure 2b shows that both male and female IBL students pursued further IBL courses at higher rates. Significant differences versus non-IBL students are found for both L1 (men) and G1 (men and women). IBL women from G1 persisted significantly longer, taking on average a full elective course beyond their non-IBL female peers. This pattern is not mirrored in the more advanced courses L1 and L2.

By Achievement Group

In interviews conducted with instructors as part of the broader study, instructors proposed that IBL would particularly benefit students with weaker academic backgrounds. The strongest students, they felt, would enjoy the challenge of IBL but would succeed in both IBL and non-IBL settings. Based on these hypotheses, we disaggregated the data for IBL and non-IBL students by prior mathematics achievement level. We present results from L1 only. L2 students took too few subsequent courses to support further division of the sample, and this analysis held no meaning for G1 where all students were high achievers. For L1, we empirically divided students into three achievement subgroups: low, GPA < 2.5; medium, GPA 2.5 to 3.4; and high, GPA > 3.4, taking care to match the underlying distributions for IBL and non-IBL samples.

Table 3 and Figure 3 show the average subsequent grades of L1 students, disaggregated by IBL status and achievement level. Stairstep patterns in Figure 3a show that students' prior grades tend to predict their later grades. Medium and high achievers who took L1 as an IBL course earned subsequent grades similar to those of their non-IBL peers, but low achievers from IBL sections earned consistently *higher* grades than their low-achieving peers in non-IBL sections. These differences were statistically significant for required and IBL courses, and large enough to hold academic meaning. For example, IBL low achievers averaged a high C+ (2.41) for subsequent required courses, while non-IBL low achievers averaged below a C (1.95).

Figure 3b shows the L1 data plotted as the average difference in grades before and after the target course. Most students' grades fell by several tenths of a grade point as they moved from computationally oriented, lower-division courses to more abstract, proof-based courses. Especially students with previously high grades did not maintain an A average as courses became more challenging. The contrasting pattern for IBL low achievers is thus striking: the grades of these students *improved* in later courses, unlike both their non-IBL peers and their higher-achieving classmates. The improvements were sizable and persistent: for instance, grades improved by about 0.3 grade points in the next term, and even more in further IBL courses. Only two of the group differences were statistically significant, but, as Table 4 shows, some of the effect sizes for the IBL intervention are substantial. Medium to large effect sizes (Cohen, 1988) are observed not only for the increase in IBL low achievers' grades, but also for the lesser pre-to-post decline in grades of IBL medium- and high-achievers versus non-IBL peers.

We found few differences in course-taking by achievement level (Table 3). There was one statistically significant difference: high achievers who had taken an IBL section of L1 took more IBL courses than did their non-IBL peers. This finding matches instructors' expectations that high achievers would find the IBL method stimulating.

Discussion

Overall, the effect of IBL on students' subsequent grades and course-taking was modest when comparing IBL and non-IBL students in their entirety. Certainly no harm was done;

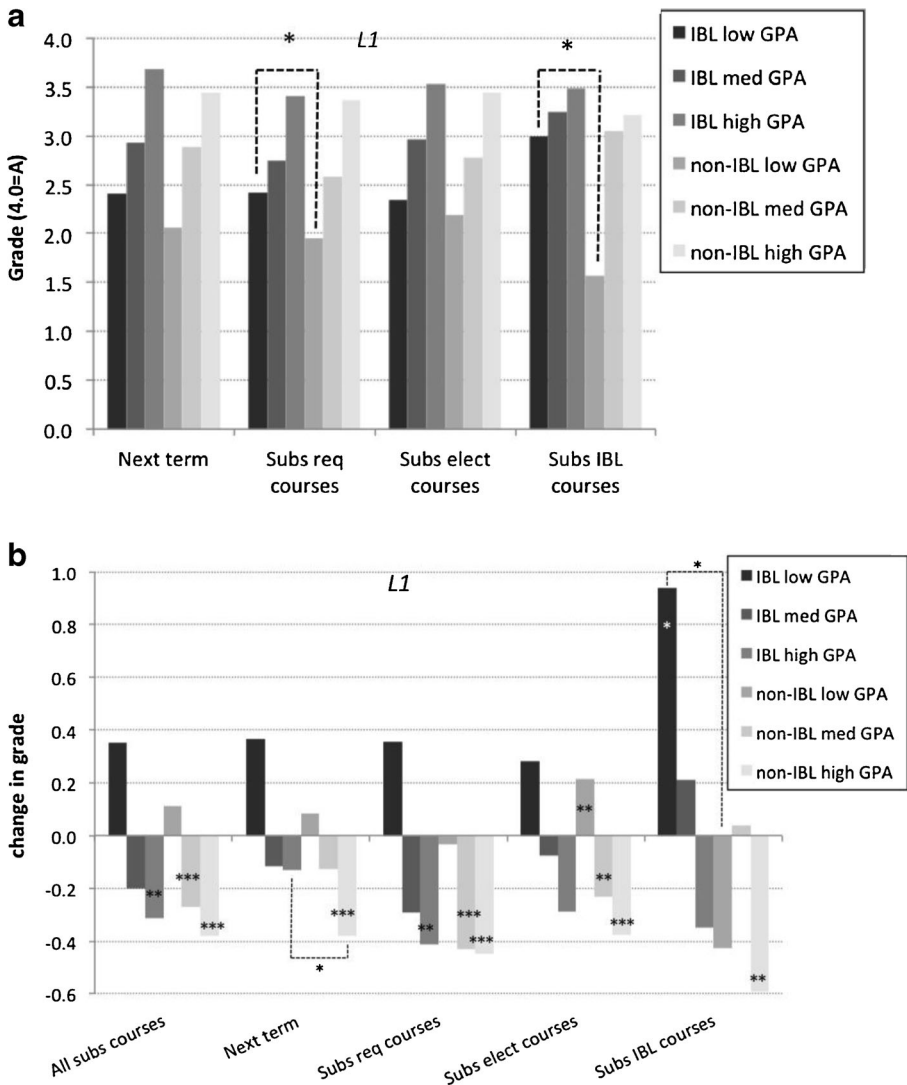


Fig. 3 Grades for previously low, medium and high-achieving students after taking IBL and non-IBL sections of course L1. (a) Average grades for next term and all subsequent required, elective and IBL courses. (b) Average change in grades from all previous math courses to next term and all subsequent required, elective and IBL courses. Subgroup difference is statistically significant: * at $p < 0.05$, ** at $p < 0.01$, *** at $p < 0.001$. In (b), significance of the pre-to-post-target grade change is marked on the bar; subgroup differences are marked outside the bars with brackets

IBL students succeeded at least as well as their peers in later courses. This result challenges instructors’ common concern that material omitted to accommodate the slower pace of IBL courses may hinder student success in later courses (Yoshinobu & Jones, 2012).

IBL students also tended to take additional IBL courses if available. While this study controlled for differences in prior mathematics background and achievement, other factors also affect students’ choice of an IBL or non-IBL section: learning beliefs, professor choice, peer influences, and even the time of day the section is offered. Our analyses may not fully

Table 4 Effect size for IBL intervention on grades and course-taking subsequent to an IBL or non-IBL mathematics course

<i>Course L1, 2001-2008</i>	Prior achievement group		
	Low	Medium	High
<i>Grade variable</i>			
Average of all subsequent math grades	0.56	0.29	0.35
Average grade in the next term	0.65	0.04	1.28
Average grade in subsequent required courses	0.90	0.55	0.18
Average grade in subsequent elective courses	0.16	0.63	0.45
Average grade in subsequent IBL courses	3.11	0.68	1.19

Notes on Table 4: Categories for prior achievement are based on mathematics grades for courses prior to Course L1: Low = $GPA \leq 2.5$, Medium = $2.5 < GPA \leq 3.4$; High = $GPA > 3.4$. Effect sizes are calculated using the pretest/posttest/control design of Morris (2008)

separate pre-selection from causal effects linking pursuit of further IBL courses to a good IBL experience.

Positive effects of IBL on students' pursuit of further mathematics courses were general to both men and women. These effects are detected among courses where student choice may be most apparent, electives and courses taught with IBL methods. The results by gender are particularly interesting given our findings from immediate post-course survey data: in non-IBL courses, women reported significantly lower learning gains than did men (Laursen et al., 2011, 2013a, b). This gender gap persisted across several types of intellectual (e.g., conceptual learning, problem-solving) and affective (confidence, interest) gains, though there were no actual differences in men's and women's grades. That is, women in non-IBL courses succeeded at similar rates to men, but reported less mastery and lower confidence at the end of the course. The present analysis shows that non-IBL women also persisted in mathematics at lower rates.

In IBL courses, however, women reported similar intellectual and affective gains to men on surveys (Laursen et al., 2011), and their grades were no different. This analysis indicates that IBL women were also more likely to persist in mathematics. Enhanced persistence was apparent following G1, a course early in the curriculum, while after courses L1 and L2, such effects were less detectable as students had fewer terms left in which to adjust their major or course choices. Moreover, IBL experiences may matter more earlier in undergraduates' careers (Watkins & Mazur, 2013) as also suggested by the higher gains reported by first- and second-year IBL students vs. upperclassmen (Laursen et al., 2011). Women's apparent grade improvement relative to their male peers from lower-division to advanced courses may suggest that women who persist to advanced courses are high achievers who also have high tolerance for their minority status.

Disaggregated by prior achievement, differences in students' grades and course-taking patterns became apparent. Taking an IBL course did not erase achievement differences among students, but did flatten them. In non-IBL courses, initial patterns of achievement difference were preserved; previously low-achieving students gained no ground.

Figure 3a compares students to each other, while Figure 3b compares students to their own prior performance. Low achievers' performance was boosted after taking an IBL course, relative both to their own previous performance and to non-IBL peers. Differences of 0.3-0.5 grade points are meaningful to students' future academic options.

The differing impact of IBL on women and low achievers shows that the intervention functions differently for these two groups. For women, the impact of IBL appears to be primarily

affective; it is not permanent. IBL courses offer features that are known to be effective for women, including collaborative work (Springer, Stanne & Donovan, 1999), problem-solving, and communication (Du & Kolmos, 2009) and that may enhance women's sense of belonging to the discipline (Good, Rattan & Dweck, 2012). Public sharing and critique of student work may serve as vicarious experiences that enhance self-efficacy (van Dinther, Dochy & Segers, 2011) and link effort, rather than innate talent, to mathematical success (Good, Rattan & Dweck, 2012). For low-achieving students, however, the effect is longer-lasting. We propose that IBL experiences promote what one student called "fruitful struggle," thereby strengthening transferable problem-solving strategies and study habits. For students who do not already have these skills, this is a powerful and lasting impact (Hassi & Laursen, 2013).

This study also yields some methodological insight. Overall, grades and course-taking choices are blunt instruments for detecting the impact of an educational intervention. Because these outcomes and their meaning varied importantly by student sub-group, disaggregating results was essential—but necessitated large samples. Comparing patterns in results across three courses yielded insight about the relative impact of IBL experiences on students at different academic stages. The methods used are entirely general and not specific to mathematics courses.

The utility of academic records analysis was understandably sensitive to the nature and timing of the target course. Effects on subsequent grades and course-taking were most easily detected in courses earlier in the curriculum. Self- and institutional selection required the use of stringent controls that in turn required large samples. Variation by prior achievement could not be studied in the G1 course because all students were strongly prepared. Results could be rigorously compared only within a single course, not across courses or institutions. Finally, the analysis required substantial up-front work to gather institutional records, transform data, and define and compute standardized variables. In sum, academic records analysis is not a tool to be applied lightly, yet techniques like these may yield insight for studies of multi-course or multi-site educational reform in cases where these design constraints can be accommodated.

Conclusion

College instructors using student-centered methods in the classroom are often called upon to provide evidence in support of the educational benefits of their approach—an irony, given that traditional lecture approaches have seldom undergone similar evidence-based scrutiny. Our study indicates that the benefits of active learning experiences may be lasting and significant for some student groups, with no harm done to others. Importantly, "covering" less material in inquiry-based sections had no negative effect on students' later performance in the major. Evidence for increased persistence is seen among the high-achieving students whom many faculty members would most like to recruit and retain in their department. Thus these results should be useful to instructors seeking evidence to persuade colleagues and students of the value of their approach.

This work raises many interesting questions for future studies. The differential benefits of inquiry learning experiences for low-achieving students highlights their potential to help overcome historical inequities for other groups, such as students of color and first-generation college students—groups we could not examine in this study. Comparison of longitudinal effects will be especially interesting in cases where inquiry courses are offered to first-year students and where multiple inquiry experiences are offered within a single program or institution. While the quantitative approach reported here establishes patterns of student achievement and persistence after an active-learning course, only mixed-methods approaches can both document such effects and reveal the reasons for them.

Acknowledgments The Educational Advancement Foundation supported this work. We thank Clint Coburn and Tim Archie for research assistance and Andy Cameron, Michele Keeler, and Steven Velasco for access to the data.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching*. San Francisco, CA: Jossey-Bass.
- Carlson, M., Rasmussen, C., Bressoud, D., Pearson, M., Jacobs, S., Ellis, J., et al. (2011). Surveying mathematics departments to identify characteristics of successful programs in college calculus. In S. Brown, S. Larsen, K. Marrongelle, and M. Oehrtman (Eds.), *Proceedings of the 14th Annual Conference on Research in Undergraduate Mathematics Education*, Vol. 3, pp. 3–33–3–38. Portland, OR. Retrieved from http://sigmaa.maa.org/rume/RUME_XIV_Proceedings_Volume_3.pdf
- Carrell, S. E., & West, J. E. (2010). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy*, *118*, 409–432.
- Cobb, P., Yackel, E., & McCain, K. (Eds.). (2000). *Symbolizing and communicating in mathematics classrooms: Perspectives on discourse, tools, and instructional design*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cohen, J. (1988). *Statistical power for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Davis, R. B., Maher, C. A., & Noddings, N. (Eds.) (1990). Constructivist views of the teaching and learning of mathematics. *Journal for Research in Mathematics Education, Monograph No. 4*. Reston, VA: National Council of Teachers of Mathematics.
- De Paola, M. (2009). Does teacher quality affect student performance? Evidence from an Italian university. *Bulletin of Economic Research*, *61*, 353–377.
- Derting, T. L., & Ebert-May, D. (2010). Learner-centered inquiry in undergraduate biology: Positive relationships with long-term student achievement. *CBE-Life Sciences Education*, *9*, 462–472.
- Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science*, *332*, 862–864.
- Dorans, N. J. (1999). Correspondence between ACT and SAT I scores. *College Board Research Report 99-1*. New York, NY: The College Board.
- Dorans, N. J. (2004). Equating, concordance, and expectation. *Applied Psychological Measurement*, *28*, 227–246.
- Dorans, N. J., Lyu, C. F., Pommerich, M., & Houston, W. M. (1997). Concordance between ACT Assessment and recentered SAT I Sum scores. *College and University*, *73*, 24–35.
- Du, X., & Kolmos, A. (2009). Increasing the diversity of engineering education—a gender analysis in a PBL context. *European Journal of Engineering Education*, *34*, 425–437.
- Dubetz, T., Barreto, J. C., Deiros, D., Kakareka, J., Brow, D. W., & Ewald, C. (2008). Multiple pedagogical reforms implemented in a university science class to address diverse learning styles. *Journal of College Science Teaching*, *38*(2), 39–43.
- Ellis, J., Rasmussen, C., & Duncan, K. (2013). *Switcher and persister experiences in Calculus I*. Sixteenth Annual Conference on Research in Undergraduate Mathematics Education. Denver, CO. Retrieved from http://pzacad.pitzer.edu/~dbachman/RUME_XVI_Linked_Schedule/rume16_submission_93.pdf
- Farrell, J. J., Moog, R. S., & Spencer, J. N. (1999). A guided-inquiry general chemistry course. *Journal of Chemical Education*, *76*, 570–574.
- Froyd, J. E. (2008). *White paper on promising practices in undergraduate STEM education*. Commissioned paper, Board on Science Education, National Academies. Retrieved from http://sites.nationalacademies.org/DBASSE/BOSE/DBASSE_080106#UUoV5hngJ8g
- Fullilove, R. E., & Treisman, P. U. (1990). Mathematics achievement among African American undergraduates at the University of California, Berkeley: An evaluation of the Mathematics Workshop Program. *Journal of Negro Education*, *59*, 463–478.

- Gafney, L., & Varma-Nelson, P. (2008). *Peer-Led Team Learning: Evaluation, dissemination and institutionalization of a college-level initiative*. New York, NY: Springer.
- Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology*, *102*, 700–717.
- Hake, R. R. (1998). Interactive-engagement vs. traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*, 64–74.
- Hassi, M.-L., & Laursen, S. L. (2013). Transformative learning: Personal empowerment in learning mathematics. Manuscript in review.
- Hoffman, J. L., & Lowitzki, K. E. (2005). Predicting college success with high school grades and test scores: Limitations for minority students. *The Review of Higher Education*, *28*, 455–474.
- Kuh, G. (2008). *High-impact educational practices: What they are, who has access to them, and why they matter*. Washington, DC: American Association of Colleges and Universities.
- Kwon, O. N., Rasmussen, C., & Allen, K. (2005). Students' retention of mathematical knowledge and skills in differential equations. *School Science and Mathematics*, *105*, 1–13.
- Laursen, S., Hassi, M.-L., Kogan, M., Hunter, A.-B., & Weston, T. (2011). *Evaluation of the IBL Mathematics Project: Student and Instructor Outcomes of Inquiry-Based Learning in College Mathematics*. (Report to the Educational Advancement Foundation and the IBL Mathematics Centers) Boulder, CO: University of Colorado, Ethnography & Evaluation Research. Available at <http://www.colorado.edu/eer/research/steminquiry.html>
- Laursen, S. L., Hassi, M.-L., & Hough, S. (2013a). Inquiry-based learning in mathematics content courses for pre-service teachers. Manuscript submitted for publication.
- Laursen, S. L., Hassi, M.-L., Kogan, M., & Weston, T. J. (2013b). From innovation to implementation: Multi-institution pedagogical reform in undergraduate mathematics. Manuscript submitted for publication.
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*, 364–386.
- Mostrom, A. M., & Blumberg, P. (2012). Does learning-centered teaching promote grade improvement? *Innovative Higher Education*, *37*, 397–405.
- Noble, J. P. (1991). *Predicting college grades from ACT assessment scores and high school course work and grade information*. ACT Research Report Series 91-3. American College Testing Program. Retrieved from http://www.act.org/research/researchers/reports/pdf/ACT_RR91-03.pdf
- Pattison, E., Grodsky, E., & Muller, C. (2013). Is the sky falling? Grade inflation and the signaling power of grades. *Educational Researcher*, *42*, 259–265.
- Prince, M., & Felder, R. (2007). The many facets of inductive teaching and learning. *Journal of College Science Teaching*, *36*(5), 14–20.
- Rasmussen, C., & Kwon, O. (2007). An inquiry oriented approach to undergraduate mathematics. *Journal of Mathematical Behavior*, *26*, 189–194.
- Ruiz-Primo, M. A., Briggs, D., Iverson, H., Talbot, R., & Shepard, L. A. (2011). Impact of undergraduate science course innovations on learning. *Science*, *331*, 1269–1270.
- Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences*. Boulder, CO: Westview Press.
- Springer, L., Stanne, M. E., & Donovan, S. (1999). Measuring the success of small-group learning in college-level SMET teaching: A meta-analysis. *Review of Educational Research*, *69*, 21–51.
- Stigler, J. W., Givvin, K. B., & Thompson, B. J. (2010). What community college developmental mathematics students understand about mathematics. *MathAMATYC Educator*, *1*(3), 4–16.
- Tai, R. H., Sadler, P. M., & Mintzes, J. J. (2006). Factors influencing college science success. *Journal of College Science Teaching*, *36*(1), 52–56.
- Tien, L. T., Roth, V., & Kampmeier, J. A. (2002). Implementation of a peer-led team learning instructional approach in an undergraduate organic chemistry course. *Journal of Research in Science Teaching*, *39*, 606–632.
- van Dintter, M., Dochy, F., & Segers, M. (2011). Factors affecting students' self-efficacy in higher education. *Educational Research Review*, *6*, 95–108.
- Watkins, J., & Mazur, E. (2013). Retaining students in science, technology, engineering and mathematics (STEM) majors. *Journal of College Science Teaching*, *42*(5), 36–41.
- Weinberg, B. A., Hashimoto, M., & Fleisher, B. M. (2009). Evaluating teaching in higher education. *The Journal of Economic Education*, *40*, 227–261.
- Yoshinobu, S., & Jones, M. G. (2012). The coverage issue. *PRIMUS*, *22*, 303–316.