

# Ten Challenges and Recommendations for Advancing Research on the Effects of College on Students

**Tricia A. Seifert**

*Montana State University*

**Nicholas A. Bowman**

*University of Iowa*

**Gregory C. Wolniak**

*New York University*

**Alyssa N. Rockenbach**

*North Carolina State University*

**Matthew J. Mayhew**

*The Ohio State University*

*The first two volumes of How College Affects Students constitute the most-cited publications in higher education research. Since these volumes were published, the literature on college impact has expanded greatly, which is at least partially the result of the nearly 500 journals focusing on the scholarship of teaching and learning internationally. In addition to the increase in the quantity of college impact research, methodological advancements and the changing notion of the higher education enterprise have further shaped the rigor and tenor of the literature. We draw on our experience and lessons learned from conducting the most recent comprehensive synthesis of college impact literature to discuss 10 challenges and corresponding recommendations to advance future research of the effects of college on students.*

**Keywords:** college effects, higher education, student learning outcomes

SINCE at least the early 1960s with the publication of *The American College* (Sanford, 1962), higher education has occupied a distinct and well-defined area of research within the social and behavioral sciences. For nearly 50 years, researchers, policymakers, practitioners, and other stakeholders have benefited from several comprehensive syntheses of research on the influence of college on students. Feldman and Newcomb's (1969) *The Impact of College on Students* and the two volumes of *How College Affects Students* (Pascarella & Terenzini, 1991, 2005) have contributed valuable up-to-date empirical information on the complex and expanding literature on the effects of college on students' learning and development as well as their career and life trajectories. These large-scale reviews provide the opportunity to identify key developments in the higher education literature and highlight critical areas for future research.

Several notable trends have occurred in the area of college impact research, including the sizable and steady increase in the quantity of published research, the

sophistication of research designs and analytic methods, and the evolving nature of what constitutes a college education. For example, whereas Feldman and Newcomb (1969) identified and reviewed roughly 1,500 studies that spanned a 40-year period, Pascarella and Terenzini's (1991) first volume of *How College Affects Students* was based on approximately 2,600 studies published in the 1970s and 1980s. The second volume of *How College Affects Students* identified and provided a synthesis of 2,400 studies occurring primarily within the single decade of the 1990s (Pascarella & Terenzini, 2005). The trend of proliferating college impact research has continued in dramatic fashion. The current synthesis (Mayhew, Rockenbach, Bowman, Seifert, & Wolniak, 2016) collected information from over 10,000 sources that appeared over a 10-year review period, which roughly corresponds to the first decade of the 21st century. Such an increase in literature necessitated focusing primarily on research appearing in peer-reviewed journals, thereby generally excluding unpublished conference papers and dissertations. Even with these criteria in place, over



1,800 rigorous peer-reviewed articles were reviewed and synthesized (Mayhew et al., 2016). With publications frequently appearing both within and outside of mainstream higher education outlets, it is difficult for anyone to stay up to date with this rapidly increasing literature.

Accompanying the rise in the quantity of college impact research, notable developments have occurred in the use of sophisticated methods for assessing the effects of college experiences and conditions for learning on student outcomes. Higher education studies are increasingly employing experimental, quasiexperimental, and nonexperimental designs with thorough analytic controls; drawing from longitudinal data; obtaining multi-institutional samples; and using validated measures of both student outcomes and experiences. The greater use of these methods, along with the heightened scrutiny of self-reported gains versus directly measured longitudinal change, is improving the state of knowledge on the effects of college attendance on student learning by enabling stronger causal conclusions. Over the past decade, studies on the effects of postsecondary experiences and conditions for learning are providing credible and reliable evidence as well as aligning with the standards put forth by the What Works Clearinghouse (2014) and American Educational Research Association (2006).

In addition to research design and methodological considerations, syntheses of research across decades highlight the evolving and dynamic nature of the postsecondary system. The profile of the average college student in the United States is different today than it was previously, with women outnumbering men; an increase in adult learners; and a more racially, culturally, and religiously diverse student body (Eagan et al., 2016; National Center of Education Statistics, 2010). Even the meaning of “a college education” is changing, amid rapid expansion of online degree programs, subsequent or simultaneous attendance at multiple institutions (2-year and 4-year), and the increasingly borderless nature of college (Adelman, 2006; Bell & Federman, 2013; Institute of International Education, 2015; Snyder, de Brey, & Dillow, 2016). Research on the effects of college on students over numerous decades highlights the changes in the sociodemographic characteristics of college-going students and evolving nature of the college student experience.

Altogether, the sheer volume of research, sophistication of methods, and evolving nature of the higher education enterprise present challenges, and hold the keys to informing the next wave of research and improving our understanding of how college affects student learning and personal growth. Consistent with the theme of this special issue, we draw on our experience and lessons learned conducting the most recent comprehensive synthesis of college impact literature (Mayhew et al., 2016) to discuss 10 challenges and provide a series of recommendations and directions for future inquiry. We organize the discussion according to four main themes: reconciling definitions, determining the “weight of

the evidence,” generalizing research findings, and expanding inquiry in underexamined areas.

## Reconciling Definitions

As Descartes (1644/2017) notes, “If something exists, it exists in some amount. If it exists in some amount, then it is capable of being measured.” Most scholars and practitioners interested in assessing student learning outcomes would agree—at least in part—with the ontological, epistemic, and axiologic assumptions embedded in Descartes’ message. Applied to college impact work, rather than assuming exposure to and participation in a set of educational conditions and experiences result in student learning, it is necessary to measure such learning. Even if we can agree with these ideas or at least their embedded tenets, how one defines and measures a construct is socially constructed and inherently context based (Reale, 2014). Although the problem of achieving definitional consensus is neither new (see Batey, 2012) nor idiosyncratic to college impact work, scholars interested in learning outcomes face a series of distinctive challenges, due to working with the vast number of definitions offered for different constructs, terms, and methodologies used to understand college and its effects on students. We turn to a brief overview of the definitional challenges, which we reconciled and conclude with recommendations for future research.

### *Defining Constructs Across International Contexts*

Since Pascarella and Terenzini’s 2005 volume, which reviewed articles from 1990 to 2002, higher education research has taken on an increasing international comparative perspective (Kosmützky & Krücken, 2014). Perhaps the most fundamental matter of definitions, a construct in one context may be called something quite different in another. Ulrich Teichler (1996) described the practical issue of “acquiring sufficient field knowledge” (p. 453) for conducting comparative higher education research. Although not an international comparative research project per se, reviewing research conducted in multiple contexts required developing a lexicon of names and meanings for common constructs. It is worth noting that the key construct underpinning our synthesis, “college,” is subject to this precise challenge, having multiple definitions within and across contexts. Reflecting on the historic notion of college, Judith Eaton (2014) stated, “‘College’ meant experiencing a fixed curriculum in a fixed place (a campus) on a fixed timetable” (p. 223). Even in a single national context like the United States, *college* refers simultaneously to a 4-year institution that awards baccalaureate and graduate degrees, a 2-year institution that may be accredited to award 2-year (associate’s) degrees and a host of certificates, and a collection of disciplines that compose a specific organizational unit (e.g., College of Engineering). In

Canada, *college* has referred typically to a postsecondary institution that awards subbaccalaureate credentials, although the distinction between the university and non-university sector is blurring (Jones, 2007). To add to the challenge of definitions, *college* in the Canadian context may also refer to one of the many residential colleges within the university (e.g., Trinity College at the University of Toronto), whereas the organizational unit for a collection of disciplines is often called a *faculty*.

Researchers undertaking this kind of project must develop an awareness of the “national or cultural relativity of terms and concepts” (Teichler, 2014, p. 399), as such field knowledge is critical for interpretation. Drawing on this knowledge, the opportunity exists to confront these definitional challenges by discussing findings in comparative fashion and highlighting the historical, social, cultural, and political differences across international contexts. The presence of one or more researchers who are familiar with these specific contexts can be instrumental for understanding linguistic variations.

#### *Defining Experiences, Environments, and Outcomes*

In addition to the definitional challenges inherent when including research from different higher education contexts, constructs are often defined and measured in a multitude of ways. This is an issue for not only experiences and environments but also student learning outcomes. For example, student–faculty interaction has been researched extensively but has differed considerably in its definition and measurement. Studies drawing from the National Survey of Student Engagement (Carini, Kuh, & Klein, 2006; Lundberg, 2012; Pike, Smart, & Ethington, 2012) tended to define student–faculty interaction by a composite index of the frequency with which students reported interacting with faculty. Such an operationalization yielded mixed effects with respect to learning about subject matter competence (Mayhew et al., 2016). On the other hand, when student–faculty interaction was defined as students’ perceptions of the quality of the relationship, the association with intellectual ability was positive (e.g., Kim, Chang, & Park, 2009). Like the qualitative details that differentiate an everyday practice from one that has the potential for “high impact” (Kuh, 2008), defining student–faculty interaction as students’ perception of the quality of the interaction may measure the nature of the interaction better than how often students interact with faculty.

The challenge of multiple definitions also extends to learning outcomes. Outcomes, like the indicators described by Reale (2014), “are social constructs based on conceptual frameworks, providing definitions and normative understanding of the underlying reality” but are yet “a synthetic representation, not a complete and objective description of reality; in this sense they are proxies of the phenomenon

they want to represent” (p. 418). Critical thinking, as an indicator of student learning, exemplifies Reale’s point of the varied social constructions used to define a commonly identified learning outcome in that no theoretical framework has consistently been used to discuss it as a student learning outcome (Mayhew et al., 2016). Noting the multiple ways researchers defined *critical thinking*, Pascarella and Terenzini (1991) offered a guiding definition that included (among other aspects) processing and utilizing new information, reasoning objectively and drawing objective conclusions from various types of data, and evaluating arguments and claims critically. Such a definition may result in outcome ambiguity: Does a measure need all of these dimensions or a select few? In reviewing the literature on critical thinking, should researchers view certain operationalizations as more valid?

Layered onto the challenge presented by multiple and/or ambiguous definitions is the distinction between measuring critical thinking or any learning outcome through an authentic learning task or a standardized assessment versus students’ self-reported gains on the outcome. For example, we found that diversity course taking was more positively associated with self-reported measures of cognitive gains than with objective measures of critical thinking (Mayhew et al., 2016). Different definitions of experiences, environments, and outcomes challenge researchers when synthesizing and summarizing across studies.

We do not feel that a single definition of experiences or outcomes is necessary or, in some cases, even desirable for college impact research. However, researchers must be careful not to conflate findings that use different definitions of the same construct, because these studies may be examining qualitatively different phenomena. Moreover, sufficient similarity in experiences and outcomes must exist to draw meaningful conclusions within and across contexts. In our work (Mayhew et al., 2016), to balance the desire for comparability and reduce linguistic challenges while also conducting a large-scale international review, we examined higher/tertiary/postsecondary systems that primarily teach courses and disseminate research findings in English and are largely grounded in the Oxford-Cambridge residential colleges model. This approach allowed us to extend our review beyond a single nation and provide a meaningful synthesis across nations.

#### *Defining Student Learning Outcomes*

Beyond issues of operationalization, some deeper challenges remain for “defining” learning outcomes: What should be the outcomes of a college education? What are the general competencies and/or generic/transferable skills? What should learners be “expected to know, understand and be able to do at the end of a period of learning” (Bologna Working Group, 2005, p. 29)? The statements that answer

these questions are learning outcomes. Determining and agreeing to the substance of such statements is at the heart of the student learning outcome definitional challenge and for researchers examining the extent to which colleges and universities succeed in promoting their desired outcomes.

Building on the Humboldtian foundation of *lehrfreiheit*, defining what and determining how to teach within one's area of expertise are quintessential components of academic freedom (Altbach, 2013). Faculty often perceive external bodies' definition and assessment of student learning outcomes as a public accountability lever and an encroachment on academic freedom (Eaton, 2014; Gold, Rhoades, Smith, & Kuh, 2011; Hutchings, 2010). Gary Rhoades, former president of the American Association of University Professors, acknowledged the organization's position of local control and concluded "nationally standardized outcomes and assessments . . . are inappropriate for higher education, particularly when they get beyond the level of the discipline or professional field" (Gold et al., 2011, pp. 7–8). To date, college student learning outcomes have not been defined by a set of standardized assessments, as is commonplace in K–12 education (see the Program for International Student Assessment; Organisation for Economic Co-operation and Development, n.d.). Rather, student learning outcomes are defined commonly by international/national quality assurance bodies at the degree or credential level (Adelman, Ewell, Gaston, & Schneider, 2014; Australian Qualifications Framework Council, 2013; Bologna Working Group, 2005; Council of Ministers of Education, Canada, 2007; European Consortium for Accreditation, n.d.) or by professional accrediting bodies (see Council for the Accreditation of Educator Preparation for the teaching field or ABET for engineering). For example, the Ministerial Statement on Quality Assurance of Degree Education in Canada (Council of Ministers of Education, Canada, 2007) specifies six expectations: depth and breadth of knowledge, knowledge of methodologies and research, application of knowledge, communication skills, awareness of limits of knowledge, and professional capacity/autonomy. Degree-level distinctions address the change in standards for each expectation, from a basic understanding of disciplinary knowledge at the bachelor's level to a systematic understanding at the master's level. The qualification frameworks from other contexts also distinguish degree levels through more advanced standards for each criterion.

Historically, quality has been assured through a third-party audit of institutional processes. The logic was that if the process was sound, then the product—in this case, student learning—would manifest. The new norm for quality assurance mechanisms, however, focuses on the product, specifically, institutional performance in the form of student learning and achievement (Eaton, 2014). From this perspective, faculty have a tremendous role to play in defining

student learning outcomes and designing curricula and assessments to gauge the realization of such outcomes. The definitional issues become a challenge, however, when articulating what constitutes the breadth of knowledge for a particular discipline. If faculty wish to define, plan, design, implement, and evaluate student learning outcomes, then it will be incumbent that faculty are present to define and develop outcomes and standards not only locally but nationally. For researchers synthesizing research not only within but across disciplines, degree levels, and international contexts, keeping abreast of how student learning is defined and measured is essential.

### Determining the “Weight of the Evidence”

After these definitional issues have been settled (at least to some extent), studies that explore the same constructs must be synthesized to draw meaningful and well-supported conclusions about the effects on student learning outcomes. However, this task of combining results also comes with a set of challenges that must be resolved or at least addressed. Below, we discuss three salient concerns when attempting to determine the overall impact of a particular student experience or intervention.

#### *Combining Results Into an Overall Effect Size*

The practical significance of findings is a critical issue. A study of thousands of students may yield a statistically significant link between an experience and student learning, but is this result large enough to have any real-world consequence? And when synthesizing a variety of studies, does the “weight of the evidence” suggest that there is a practically meaningful effect? Ideally, the results of these studies would be combined into a single effect size through a quantitative meta-analysis. The overall result could then be used to determine whether the magnitude of the effect is practically meaningful as well as the conditions under which the effect is strongest (see Borenstein, Hedges, Higgins, & Rothstein, 2009; Cooper, 2016; Cooper, Hedges, & Valentine, 2009).

However, quantitative syntheses are complicated by a variety of factors. First, a common effect size metric must be obtained across studies. Given that a great deal of research on student learning outcomes uses multiple regression, standardized regression coefficients are a promising option, but these can also have problems. For instance, when determining the impact of service-learning coursework, studies might operationalize coursework with a variable that is binary (0 = no courses, 1 = at least one course), count (number of courses), or ordinal (1 = no courses have a service-learning component to 4 = most courses have a service-learning component). Using standardized regression coefficients helps account for these

differences better than using unstandardized coefficients, but interpretation of a one-standard-deviation unit for the binary predictor is difficult. This issue is largely avoided if the vast majority of studies on a topic consistently use a similar type of variable to indicate a particular experience, because these can then be readily combined into an effect size metric that has a similar meaning across studies.

In a second concern, studies that use the same effect size metric (or for which that metric can be computed) may differ in numerous other ways that may affect the results. For instance, in a meta-analysis of college diversity experiences and civic engagement (Bowman, 2011), research that used self-reported gains to indicate student growth had an average effect size that was almost 3 times as large as research that used longitudinal assessments. In this same review, studies that included multiple diversity experiences within a single statistical model had smaller effect sizes than studies that had only one diversity experience within the model. This issue can be addressed to some extent by conducting subgroup analyses on studies with different methodological characteristics or by including these characteristics within a meta-regression analysis. Additional effect size measures using partial and semipartial correlations have also been created that account for some features of multivariate analyses, including the number of predictors in the model (e.g., Aloe & Becker, 2011, 2012).

Third, many studies report the results of multiple statistical models with different sets of predictors, so a synthesis that involves these studies must decide which of these models should be used. This choice is somewhat subjective, but an important issue is whether one or more predictors in the analysis might mediate the relationship between the key experience and outcome. For instance, in our review (Mayhew et al., 2016), numerous studies examined the link between living on campus and student retention, and the findings were mixed between positive and nonsignificant relationships. However, studies that included social involvement or adjustment in the statistical model almost always obtained nonsignificant results, whereas studies that did not contain such control variables almost always yielded positive results. It therefore seems that living on campus may improve students' social lives, which then promotes college retention. In this particular example, no single study reported separate analyses with and without social involvement/adjustment in the model, but the synthesis of this research highlights the fact that studies that include mediators can lead to misleading results. Therefore, when reviewing literature on a topic, we recommend using the most fully identified model (i.e., with the most covariates) that does not contain potential mediators of the relationship of interest. Researchers should also attend to the presence or absence of different control variables when synthesizing these studies.

### *Weighing the Results of Studies With Different Methodological Rigor*

The benefits of quantitative meta-analysis are that the results of various studies can be combined and that one can examine whether the results differ significantly depending upon methodology and other characteristics of the study and sample. However, this broaches an important interpretive issue that occurs regardless of whether a quantitative or qualitative synthesis is used: What happens when the results of a small number of high-quality studies diverge notably from those of many lower-quality studies? Should the synthesis favor a handful of studies—or perhaps even just one—because they permit stronger conclusions, or should the findings from the vast majority of studies outweigh these?

We faced exactly this issue when examining the literature on learning communities and student attrition (Mayhew et al., 2016). The vast majority of recent studies that used observational data found positive relationships between learning communities and postsecondary retention, persistence, and graduation (e.g., W. Hill & Woodward, 2013; Hotchkiss, Moore, & Pitts, 2006; Jones-White, Radcliffe, Huesman, & Kellogg, 2010; Mangold, Bean, Adams, Schwab, & Lynch, 2002; Popiolek, Fine, & Eilman, 2013; Stassen, 2003). However, a large-scale experimental study of learning communities at six 2-year colleges found virtually no meaningful effects on student success, except for positive results at only one of the institutions (Visher, Weiss, Weissman, Rudd, & Wathington, 2012). Given these findings, do we believe that learning communities reduce student attrition? A quantitative meta-analysis of all results would likely reach this conclusion, but the only study that completely eliminated self-selection into learning communities suggests otherwise.

Unfortunately, no obvious answer exists for reconciling all instances of divergent findings. In the learning communities-student attrition example, the multisite design and large sample size of the randomized controlled trial certainly increased the generalizability of the findings, but all institutions were 2-year colleges, all but one was located in a major metropolitan area, and all learning communities involved linking a developmental (remedial) course with other course work. These characteristics diverged from the nonexperimental studies, which generally occurred at 4-year universities and did not include developmental course work. Ultimately, after considering these studies as well as research on other student experiences and findings from previous reviews (Pascarella & Terenzini, 1991, 2005), we concluded that learning communities are primarily effective when they integrate student services and/or other resources. Thus, our conclusion focused more on the conditions under which learning communities are effective rather than some studies being considerably more accurate than others.

This challenge with differential study quality will become increasingly important given the growing demand for research that uses experimental and quasiexperimental designs. Most research on student learning outcomes has been correlational, so the question of how much those findings should be used to inform current understandings of these outcomes is still up for debate. Some research questions are clearly more amenable to randomized controlled trials than others; for instance, exploring the impact of classroom pedagogy experimentally is much easier than randomly assigning students to participate (or not) in a social fraternity or sorority. The importance placed on the most rigorous studies should depend upon various factors, including the potential for the less rigorous research to rule out alternative explanations for their results as well as the generalizability of the most rigorous studies. As with the learning community example, research syntheses should consider whether divergent results occur because some results are “right” and others are “wrong” or because this divergence illustrates the conditions under which an experience may be most effective.

#### *Examining the Overall Impact of a Treatment*

Many studies of postsecondary student learning outcomes examine the extent to which an experience predicts one outcome or a narrow range of outcomes. This focus makes sense from a research perspective, because (a) researchers may have access to only a small number of outcomes, (b) the hypothesized relationships may not be applicable to other outcomes, and (c) providing too many outcomes could create a confusing story (especially if the results are contradictory). However, from a practical perspective, administrators, practitioners, and policymakers likely care about implementing experiences that promote a broad array of desired outcomes; this desire is likely bolstered when financial resources are limited.

Therefore, when attempting to synthesize the available literature, researchers may consider whether and how to integrate multiple outcomes into the review. Quantitative meta-analyses can conduct moderator analyses to determine whether the results differ by the nature of outcome or how it is measured. For instance, two meta-analyses of problem-based learning have explored whether the outcome measured learning concepts, principles, or application, and both reviews found substantial differences in this relationship across outcome type (Gijbels, Dochy, Van den Bossche, & Segers, 2005; Walker & Leary, 2009). Similarly, using active learning rather than lecture in science, technology, engineering, and mathematics course work appeared to have larger effects when the learning outcomes were assessed through concept inventories than through exams (Freeman et al., 2014). Sometimes these investigations find no differences

across outcome types, which is also important. For instance, a meta-analytic comparison of learning in online, blended, and face-to-face courses found no difference regardless of whether the assessment tested declarative, procedural, or strategic knowledge (Means, Tomaya, Murphy, & Baki, 2013). In all cases, these reviews provided both the magnitude of the link between an experience and a range of outcomes as well as the variation by outcome type or measurement, thereby conveying a more complete picture of the potential overall impact.

#### **Generalizing Research Findings**

A critical feature of quantitative research is its potential to yield results that can be generalized beyond the participants within a particular study. Even if very high-quality research evidence is available, considerable challenges exist for determining where, when, and for whom these results are applicable. Some choices apply to decisions that individual researchers might make, whereas others pertain to making sense of the existing research literature. Two key issues for generalizability are discussed here.

#### *Providing Multisite Evidence Versus In-Depth Understanding*

A crucial trade-off regarding the scope of the study occurs in many cases. A large-scale national study can be used to examine the extent to which a particular student experience may promote college learning. Depending on the data set, the students and/or institutions may come from a (potentially diverse) convenience sample, or they could have been sampled to be nationally representative. By using large, representative data sets, researchers can obtain results that provide a more accurate picture of the link between experiences and outcomes at many institutions. Such studies could also explore whether the strength of these relationships are moderated—or are conditional—by student attributes (e.g., demographics, precollege academic preparation) and institutional characteristics (e.g., selectivity, public/private control).

This approach contrasts with studying an intervention or experience at a single institution. By focusing on a single college or university, the researcher can provide rich detail about the intervention and what exactly it entailed, along with key attributes of the institution and participating students. As a result, consumers of the research clearly know what happened, so they can assess whether a given intervention might yield desired results in similar contexts. This understanding is important, because many practices designed to improve postsecondary student learning can vary considerably in their design and implementation, such as first-year seminars and service learning. For instance, although a large-scale study may suggest that diversity course work is

effective at promoting learning, important details could be overlooked, such as the content and pedagogy of these courses and whether the course was required or optional.

In addition, small-scale studies (as well as a nuanced review of numerous smaller studies) may provide a greater opportunity for exploring mechanisms and moderators that shape these results. For instance, an overall finding that study abroad is seemingly beneficial does not necessarily provide insights for practitioners, other than suggesting that they should try to encourage students to engage in this practice. The finding begs the question: What occurs during study-abroad trips that contributes to student learning? Does visiting a country in which residents do not speak the students' native language lead to greater growth? Which aspects of classroom experiences and direct engagement with the new country are more effective and therefore should be encouraged? And can these outcomes be achieved in a short-term study-abroad experience, or is a full semester or year necessary? In theory, a large-scale, multi-institutional data set could address these questions, but these data sets (in practice) tend to contain general information about many experiences and outcomes rather than in-depth information about specific college experiences.

When conducting a research synthesis, these various study attributes must be weighed carefully. If the results of the research are consistent across a variety of students, institutions, and implementations of the intervention, then one can confidently draw conclusions about the far-reaching effects of the experience (or lack thereof). However, this ideal situation is quite rare, because the findings are often mixed across studies (Mayhew et al., 2016). Therefore, most syntheses require close attention to the processes that promote student learning as well as the conditions under which these are most likely to occur. Such information is highly useful for informing practice and designing future studies.

### *Assessing and Interpreting Conditional Effects*

Reflecting on future directions for research soon after the second volume of *How College Affects Students* was released in 2005, Pascarella (2006) urged scholars studying college impact to take stock of whether and how collegiate experiences differentially influence students of diverse backgrounds. In short, studies of *conditional effects* ascertain the extent to which a given educational intervention has similar, stronger, or weaker effects among certain students depending on their personal characteristics (often defined in terms of student identity or precollege academic ability). Inquiries of this sort have the potential to generate evidence to transform policies and practices so that students of different identities or abilities derive similar benefits from their institutions. Yet the college impact literature prior to the 2000s rarely addressed such questions and focused largely on general

effects. In so doing, any differences in effects between various subpopulations were lost in the aggregated data.

College impact researchers through the 2000s made some progress—including a number of significant advances with respect to certain outcomes and subpopulations—in assessing conditional collegiate effects. Notably, research detailing conditional effects by gender, race/ethnicity, first-generation status, and religious/worldview identity grew substantially, particularly in psychosocial, attitudinal, and some cognitive domains. For example, Sax (2008) identified the unique ways that faculty influence women and men across numerous outcomes. The pattern involving well-being was especially striking: Women who receive honest feedback from faculty find their physical health improves; however, feeling their comments are not taken seriously by faculty reduces women's sense of physical health. Such effects were not apparent for men, whose physical health was shaped in more pronounced ways than women's by their academic major choice and course-taking patterns. Standing as another robust example, a number of scholars examined the effects of myriad curricular, cocurricular, and interactional diversity experiences on diversity attitudes. In some instances, these studies noted stronger positive effects among White students (e.g., Bowman, 2010; Gurin, Dey, Hurtado, & Gurin, 2002; Gurin, Nagda, & Lopez, 2004; Hu & Kuh, 2003), but some evidence illuminated further nuances about campus climate conditions and exchanges with diverse peers that furthered positive diversity attitudes among students of color more so than among White students (e.g., Bowman, 2013; Cabrera, 2011; Harper & Yeung, 2013). All told, these bodies of research yielded useful findings to aid in dismantling assumptions that college effects are uniform and necessarily generalizable to diverse subpopulations.

Nonetheless, two distinctive challenges remain when it comes to advances in attention to conditional effects in the college impact literature. First, many studies had key weaknesses in their analytical strategies. To accurately estimate conditional effects, one must construct variables indicating the interaction between the student characteristic (e.g., gender) and the educational experience or intervention (e.g., faculty interaction, diversity course) and subsequently assess the strength and direction of the interaction term's relationship to the outcome (e.g., well-being, diversity attitude) after controlling for the main effects. Alternatively, one may run parallel subgroup models and statistically compare the resultant regression coefficients. For instance, is the coefficient representing the relationship between faculty feedback and physical health statistically significantly stronger, weaker, or the same for women relative to men? We found a number of studies conducted in the past 10 years that claimed to interrogate conditional effects, but they did not use either of these two analytical approaches and thereby produced less convincing evidence. Typically, these studies reported results of

subgroup models and indicated the collegiate variables that were “significant” only for one group (often for the one with the larger sample size) rather than directly comparing the coefficients.

The second major challenge pertains to translating conditional effects into actionable strategies to address inequities in higher education. Despite the expansion of the college impact literature attentive to conditional effects, practice and policy implications were challenging for researchers—and by extension, their audiences—to articulate. Thus, the literature provides a much better understanding of the nuanced relationships between experiences and outcomes for diverse subpopulations but with limited insight as to their practical meaning or consequence. How are educators to respond when a particular educational intervention is “good” for certain groups but results in poorer outcomes or limited gains for others? In our review (Mayhew et al., 2016), we encountered many instances in which a complex array of conditional effects was reported without sufficient direction on how to make sense of the findings and implement meaningful changes on campus. The opportunity exists for future research to examine conditional effects and to discuss the findings in ways that improve existing theory and advance policy and practice recommendations.

### Expanding Inquiry in Underexamined Areas

The existing evidence on college impact highlights not only challenges for conducting and synthesizing evidence across studies but also areas in need of attention among researchers where severe knowledge gaps exist. Just as Pascarella (2006) recommended that researchers investigate “rational myths” of higher education—policies or programs that seem like they should be beneficial escape the scrutiny of researchers, and thus have no evidence of efficacy—we highlight two key areas in need of research we believe to be essential for improving our understanding of the full effects of college on students.

#### *Estimating Net Effects of College*

*Net effects* refers to the influence of college attendance over and above other factors that induce individual change, such as normal maturation or sociohistorical and episodic factors. Appraising net effects necessitates research designs that can be costly or logistically difficult to implement, but the concept is simple: Compare people who did not attend college to those who did. Because such data sets are few and far between (especially for measuring student learning outcomes), other approaches have materialized in the college impact literature over the years. For instance, some studies prior to the 2000s (e.g., Astin, 1993) examined the impact of the *degree* of college exposure, such as the length of time in college (“extensity”) and the depth of engagement in college

experiences (“intensity”). Importantly, models that were designed to approximate net effects included controls for a pretest measure of the outcome, age (to account for maturation), and other demographic variables. In addition, to assess whether sociohistorical factors were at work, the degree of change within a cohort was typically compared to changes across different cohorts.

In our review (Mayhew et al., 2016), we identified few studies of college net effects overall, and the results did not substantially add to the conclusions drawn in previous reviews (Pascarella & Terenzini, 1991, 2005). Little emphasis was placed on comparing college attenders to nonattenders. When such studies surfaced, they were often couched within particular disciplinary contexts. For example, one national sociological study of 2,500 college students and their noncollege counterparts revealed sharper religious declines among young adults who did not attend college (Smith & Snell, 2009), challenging the assumption that college has liberalizing effects and demonstrating that modest changes in religious identity over time cannot be solely attributed to attending college. Meanwhile, political scientists used sophisticated propensity score matching techniques to approximate the net effects of educational attainment on civic engagement and political participation. This series of studies debated conventional wisdom about whether attending college increases commitment to civic and sociopolitical involvement (see Henderson & Chatfield, 2011; Kam & Palmer, 2008, 2011; Mayer, 2011). This highlights the value of consecutive rigorous studies—all in conversation with one another—designed to investigate net effects, advance emergent methodological innovations, and address prevailing disciplinary- or field-specific questions.

In the end, longitudinal studies of college-going versus non-college-going students were uncommon through the 2000s, so we were unable to draw many definitive conclusions about the net effects of college. Recognizing that endeavors to follow cohorts of non-college goers introduces a host of challenges, this design is the most precise and effective way to assess whether individual change and development is rooted in college attendance or caused by other factors. It enables researchers to answer questions such as the following: Does postsecondary participation deepen knowledge, strengthen cognitive proficiencies, facilitate psychosocial growth, and cultivate other outcomes? In an era when the value of higher education is often called into question, it is of utmost importance to determine whether those who attend college make unique gains compared to those who do not.

#### *Conducting Studies of the Long-Term Effects of Student Learning and Growth*

For all the attention and concern focused on the impact of college on student learning and development, little direct empirical evidence explores the extent to which the effects

of college persist after graduation. Specifically, if students learn and develop more during college, then do they perform better in their careers, contribute more to their postcollege communities as engaged citizens, or continue to learn and develop more in years following college? Although a robust literature demonstrates the effects of college experiences and educational attainment on postcollege outcomes—including measures related to careers, socioeconomic status, quality of life, and a myriad of cognitive and attitudinal outcomes—virtually no research examines the relationship between students' learning gains during college and their postcollege outcomes.

As examples of more “traditional” studies of postcollege outcomes, Bowman, Brandenberger, Hill, and Lapsley (2011) explored the effects of racial/cultural workshop participation and other diversity interactions on personal growth and engaged purpose 13 years after college, and Jayakumar (2008) examined the link between college diversity experiences and pluralistic orientation several years after college. Both studies yielded evidence that specific college experiences affect graduates' attitudes years after college. In addition, parents' educational attainment, particularly having at least one parent who has attained at least a bachelor's degree, reduces the likelihood of their children dropping out of college (Chen & DesJardins, 2008, 2010) and increases educational attainment (e.g., Clotfelter, Ladd, Muschkin, & Vigdor, 2013; Ishitani, 2003, 2006; Niu & Tienda, 2013; Roksa, 2011). Although addressing important research questions and contributing valuable evidence, these studies do not explore the associations between learning and growth that occurs during college and postcollege outcomes.

One explanation for why essentially no prior research has sought to evaluate the long-term effects of college student learning and growth is that few, if any, data sets contain pretest-posttest measures capturing both change during college and postcollege outcomes. Data sets that longitudinally follow students beyond college, such as those provided by the National Center for Education Statistics in the United States (e.g., Educational Longitudinal Study, Baccalaureate and Beyond, High School and Beyond), lack direct measures of student learning or cognitive development that enables one to measure change over the course of an individual's college years. Alternatively, data sets that contain direct and comprehensive measures of college student learning or cognitive development during college (e.g., Wabash National Study, National Study of Student Learning) have not followed students beyond college graduation. Two notable exceptions are Arum and Roksa (2014) and Hill, Jackson, Roberts, Lapsley, and Brandenberger (2011). For example, Arum and Roksa studied students' social and academic learning based primarily on measures of critical thinking and complex reasoning captured by the Collegiate Learning Assessment (CLA). The study longitudinally followed a sample of U.S. college freshmen at 4-year institutions into their senior year

as well as up to 2 years later. Although the CLA was measured at multiple time points, the authors reported postcollege outcomes (e.g., probability of unemployment, working in an unskilled occupation) in relation to senior-year CLA score rather than change over time. The results support the notion that college seniors who scored higher on the CLA (one standard deviation above the mean) experience better employment outcomes. Although the analyses did not examine outcomes in relation to change in CLA scores during college, the study design serves as an important example.

An additional explanation for the lack of research on student learning in relation to postcollege outcomes is the complexity and cost associated with tracking students beyond their college years. In conjunction with this issue, one of the largest sources of major funding for research on postsecondary and adult education in the United States—the Institute of Education Sciences—has prioritized studies of access, persistence, completion, and achievement in mathematics, reading, writing, and English language proficiency. Absent from these priority areas are topics that seek to examine student learning during college in relation to outcomes later in life. With relatively few opportunities for federal funding, researchers must turn to private funders, who often lack the resources needed to sponsor a multiyear, complex longitudinal study of students from college entry to years after graduation. The circumstances surrounding postsecondary funding sources severely limit researchers' ability to identify the long-term influence of students' learning and growth during college.

Researchers have responded to Pascarella's (2006) admonition that future research examine the relationship between a specific college experience and its potential influence on students' postcollege lives. If we are to fully capture the effects of college, however, we need to extend and expand inquiry on the long-term effects of student learning during college on educational, career, and civic outcomes beyond college graduation. Inquiry in this area would make a significant contribution to the college impact literature and provide important counterpoint to increasing scrutiny and skepticism aimed at postsecondary education. Without such evidence, stakeholders will continue to question if the higher education system is honoring its implicit social contract.

## Conclusion

This article has outlined 10 salient challenges and corresponding recommendations for achieving reliable, valid, important, and ultimately useful evidence on the effects of college on students. These fall into broad categories around reconciling definitions, determining the weight of the evidence, generalizing research findings, and expanding inquiry in underexamined areas. Obtaining useful evidence is a product of both the studies that are conducted and how research is synthesized into a broader understanding of a

particular topic. We have attempted to offer some recommendations, informed in part through our experience conducting a large-scale synthesis of college impact literature in several countries. We invite readers to seize the opportunity to act on these recommendations so to strengthen the research base and substantively contribute to higher education policy and practice.

## References

- Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school to college*. Washington, DC: U.S. Department of Education.
- Adelman, C., Ewell, P., Gaston, P., & Schneider, C. G. (2014). *The degree qualifications profile*. Indianapolis, IN: Lumina Foundation for Education.
- Aloe, A. M., & Becker, B. J. (2011). Advances in combining regression results in meta-analysis. In M. Williams & W. P. Vogt (Eds.), *The Sage handbook of innovation in social research methods* (pp. 331–352). Thousand Oaks, CA: Sage.
- Aloe, A. M., & Becker, B. J. (2012). An effect size for regression predictors in meta-analysis. *Journal of Educational and Behavioral Statistics*, 37, 278–297.
- Altbach, P. (2013). *The international imperative in higher education*. Rotterdam, Netherlands: Sense.
- American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*, 35, 33–40.
- Arum, R., & Roksa, J. (2014). *Aspiring Adults Adrift: Tentative Transitions of College Graduates*. Chicago, IL: University of Chicago Press.
- Astin, A. W. (1993). *What matters in college? Four critical years revisited*. San Francisco, CA: Jossey-Bass.
- Australian Qualifications Framework Council. (2013). Australian Qualifications Framework (2nd ed.). Retrieved from <http://www.aqf.edu.au/wp-content/uploads/2013/05/AQF-2nd-Edition-January-2013.pdf>
- Batey, M. (2012). The measurement of creativity: From definitional consensus to the introduction of a new heuristic framework. *Creativity Research Journal*, 24(1), 55–65.
- Bell, B. S., & Federman, J. E. (2013). E-learning in postsecondary education. *Future of Children*, 23, 165–185.
- Bologna Working Group. (2005). *A framework for qualifications of the European higher education area*. Bologna Working Group report on qualifications frameworks. Copenhagen, Denmark: Danish Ministry of Science, Technology, and Innovation.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, UK: Wiley.
- Bowman, N. A. (2010). Disequilibrium and resolution: The nonlinear effects of diversity courses on well-being and orientations toward diversity. *Review of Higher Education*, 33(4), 543–568.
- Bowman, N. A. (2011). Promoting participation in a diverse democracy: A meta-analysis of college diversity experiences and civic engagement. *Review of Educational Research*, 81(1), 29–68.
- Bowman, N. A. (2013). The conditional effects of interracial interactions on college student outcomes. *Journal of College Student Development*, 54(3), 322–328.
- Bowman, N. A., Brandenberger, J. W., Hill, P. L., & Lapsley, D. K. (2011). The long-term effects of college diversity experiences: Well-being and social concerns 13 years after graduation. *Journal of College Student Development*, 52, 729–739.
- Cabrera, N. L. (2011). Using a sequential exploratory mixed-method design to examine racial hyperprivilege in higher education. *New Directions for Institutional Research*, 151, 77–91.
- Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student engagement and student learning. *Research in Higher Education*, 47(1), 1–32.
- Chen, R., & DesJardins, S. (2008). Exploring the effects of financial aid on the gap in student dropout risks by income level. *Research in Higher Education*, 49, 1–18.
- Chen, R., & DesJardins, S. L. (2010). Investigating the impact of financial aid on student dropout risks: Racial and ethnic differences. *Journal of Higher Education*, 81, 179–208.
- Clotfelter, C., Ladd, H., Muschkin, C., & Vigdor, J. (2013). Success in community college: Do institutions differ? *Research in Higher Education*, 54(7), 805–824.
- Cooper, H. M. (2016). *Research synthesis and meta-analysis: A step-by-step approach* (5th ed.). Thousand Oaks, CA: Sage.
- Cooper, H. M., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY: Russell Sage Foundation.
- Council of Ministers of Education, Canada. (2007). *Ministerial statement on quality assurance of degree education in Canada*. Retrieved from <http://www.cicic.ca/docs/cmec/QA-Statement-2007.en.pdf>
- Descartes, R. (2017). The principles of philosophy [ebook]. Retrieved March 17, 2017 from [https://books.google.com/books?id=GMWzAwAAQBAJ&printsec=frontcover&dq=rene+descartes+principles+of+philosophy&hl=en&sa=X&ved=0ahUKEwjhkJNjv\\_93SAhXB1CYKHsR4Bc4Q6AEIITAB#v=onepage&q=rene%20descartes%20principles%20of%20philosophy&f=false](https://books.google.com/books?id=GMWzAwAAQBAJ&printsec=frontcover&dq=rene+descartes+principles+of+philosophy&hl=en&sa=X&ved=0ahUKEwjhkJNjv_93SAhXB1CYKHsR4Bc4Q6AEIITAB#v=onepage&q=rene%20descartes%20principles%20of%20philosophy&f=false) (Original work published 1644).
- Eagan, M. K., Stolzenberg, E. B., Ramirez, J. J., Aragon, M. C., Suchard, M. R., & Rios Aguilar, C. (2016). *The American freshman: Fifty-year trends, 1966–2015*. Los Angeles: Higher Education Research Institute, University of California, Los Angeles.
- Eaton, J. S. (2014). International quality assurance: Where have we been and where are we going. In A. Maldonado-Maldonado & R. M. Bassett (Eds.), *The forefront of international higher education: A festschrift in honor of Philip G. Altbach* (pp. 223–233). Dordrecht, Netherlands: Springer.
- European Consortium for Accreditation. (n.d.). *Learning outcomes*. Retrieved from [http://ecahe.eu/w/index.php/Learning\\_outcomes](http://ecahe.eu/w/index.php/Learning_outcomes)
- Feldman, K., & Newcomb, T. (1969). *The impact of college on students*. San Francisco, CA: Jossey-Bass.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410–8415.
- Gijbels, D., Dochy, F., Van den Bossche, P., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the

- angle of assessment. *Review of Educational Research*, 75(1), 27–61.
- Gold, L., Rhoades, G., Smith, M., & Kuh, G. (2011). *What faculty unions say about student learning outcomes assessment* (NILOA Occasional Paper No. 9). Urbana: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.
- Gurin, P., Dey, E. L., Hurtado, S., & Gurin, G. (2002). Diversity and higher education: Theory and impact on educational outcomes. *Harvard Educational Review*, 72(3), 33–66.
- Gurin, P., Nagda, B. R. A., & Lopez, G. E. (2004). The benefits of diversity in education for democratic citizenship. *Journal of Social Issues*, 60(1), 17–34.
- Harper, C. E., & Yeung, F. (2013). Perceptions of institutional commitment to diversity as a predictor of college students' openness to diverse perspectives. *Review of Higher Education*, 37(1), 25–44.
- Henderson, J., & Chatfield, S. (2011). Who matches? Propensity scores and bias in the causal effects of education on participation. *Journal of Politics*, 73(3), 646–658.
- Hill, P. L., Jackson, J. J., Roberts, B. W., Lapsley, D. K., & Brandenberger, J. W. (2011). Change you can believe in: Changes in goal setting during emerging and young adulthood predict later adult well-being. *Social Psychological and Personality Science*, 2(2), 123–131.
- Hill, W., & Woodward, L. S. (2013). Examining the impact learning communities have on College of Education students on an urban campus. *Journal of College Student Development*, 54(6), 643–648.
- Hotchkiss, J. L., Moore, R. E., & Pitts, M. M. (2006). Freshman learning communities, college performance, and retention. *Education Economics*, 14(2), 197–210.
- Hu, S., & Kuh, G. D. (2003). Diversity experiences and college student learning and personal development. *Journal of College Student Development*, 44(3), 320–334.
- Hutchings, P. (2010). *Opening doors to faculty involvement in assessment* (NILOA Occasional Paper No. 4). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.
- Institute of International Education. (2015). *2015 fast facts*. Retrieved from <http://www.iie.org/~media/Files/Corporate/Open-Doors/Fast-Facts/Fast-Facts-2015.pdf?la=en>
- Ishitani, T. T. (2003). A longitudinal approach to assessing attrition behavior among first-generation students: Time-varying effects of pre-college characteristics. *Research in Higher Education*, 44(4), 433–449.
- Ishitani, T. T. (2006). Studying attrition and degree completion behavior among first-generation college students in the United States. *Journal of Higher Education*, 77(5), 861–885.
- Jayakumar, U. M. (2008). Can higher education meet the needs of an increasingly diverse and global society? Campus diversity and cross-cultural workforce competencies. *Harvard Educational Review*, 78, 615–651.
- Jones, G. (2007). Canada. In J. F. Forest & P. G. Altbach (Eds.), *International handbook of higher education* (pp. 627–645). Dordrecht, Netherlands: Springer.
- Jones-White, D. R., Radcliffe, P. M., Huesman, R. L., Jr., & Kellogg, J. P. (2010). Redefining student success: Applying different multinomial regression techniques for the study of student graduation across institutions of higher education. *Research in Higher Education*, 51(2), 154–174.
- Kam, C. D., & Palmer, C. L. (2008). Reconsidering the effects of education on political participation. *Journal of Politics*, 70(3), 612–631.
- Kam, C. D., & Palmer, C. L. (2011). Rejoinder: Reinvestigating the causal relationship between higher education and political participation. *Journal of Politics*, 73(3), 659–663.
- Kim, Y. K., Chang, M. J., & Park, J. J. (2009). Engaging with faculty: Examining rates, predictors, and educational effects for Asian American undergraduates. *Journal of Diversity in Higher Education*, 2(4), 206–218.
- Kosmützky, A., & Krücken, G. (2014). Growth or steady state? A bibliometric focus on international comparative higher education research. *Higher Education*, 67(4), 457–472.
- Kuh, G. D. (2008). *High-impact educational practices: What they are, who has access to them, and why they matter*. Washington, DC: Association of American Colleges and Universities.
- Lundberg, C. A. (2012). Predictors of learning for students from five different racial/ethnic groups. *Journal of College Student Development*, 53(5), 636–655.
- Mangold, W. D., Bean, L. G., Adams, D. J., Schwab, W. A., & Lynch, S. M. (2003). Who goes who stays: An assessment of the effect of a freshman mentoring and unit registration program on college persistence. *Journal of College Student Retention*, 4(2), 95–122.
- Mayer, A. K. (2011). Does education increase political participation? *Journal of Politics*, 73(3), 633–645.
- Mayhew, M. J., Rockenbach, A. N., Bowman, N. A., Seifert, T. A., & Wolniak, G. C., with Pascarella, E. T., & Terenzini, P. T. (2016). *How college affects students: Vol. 3. 21st century evidence that higher education works*. San Francisco, CA: Wiley.
- Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record*, 115, 147.
- National Center for Education Statistics. (2010). *Digest of education statistics*. Washington, DC: U.S. Department of Education. Retrieved from [http://nces.ed.gov/programs/digest/2010menu\\_tables.asp](http://nces.ed.gov/programs/digest/2010menu_tables.asp)
- Niu, S. X., & Tienda, M. (2013). High school economic composition and college persistence. *Research in Higher Education*, 54(1), 30–62.
- Organization for Economic Co-operation and Development (n.d.). Programme for International Student Assessment. Retrieved from <http://www.oecd.org/pisa/>
- Pascarella, E. T. (2006). How college affects students: Ten directions for future research. *Journal of College Student Development*, 47, 508–520.
- Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students*. San Francisco, CA: Jossey-Bass.
- Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research* (Vol. 2). San Francisco, CA: Jossey-Bass.
- Pike, G. R., Smart, J. C., & Ethington, C. A. (2012). The mediating effects of student engagement on the relationships between academic disciplines and learning outcomes: An extension of Holland's theory. *Research in Higher Education*, 53(5), 550–575.

- Popiolek, G., Fine, R., & Eilman, V. (2013). Learning communities, academic performance, attrition, and retention: A four-year study. *Community College Journal of Research and Practice*, 37(11), 828–838.
- Reale, E. (2014). Challenges in higher education research: The use of quantitative tools in comparative analyses. *Higher Education*, 67, 409–422.
- Roksa, J. (2011). Differentiation and work: Inequality in degree attainment in U.S. higher education. *Higher Education*, 61(3), 293–308.
- Sanford, N. (1962). *The American college: A psychological and social interpretation of the higher learning*. New York, NY: Wiley.
- Sax, L. J. (2008). *The gender gap in college: Maximizing the developmental potential of women and men*. San Francisco, CA: Jossey-Bass.
- Smith, C., & Snell, P. (2009). *Souls in transition: The religious and spiritual lives of emerging adults*. New York, NY: Oxford University Press.
- Snyder, T. D., de Brey, C., & Dillow, S. A. (2016). *Digest of education statistics 2014* (NCES 2016-006). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Stassen, M. L. A. (2003). Student outcomes: The impact of varying living-learning community models. *Research in Higher Education*, 44(5), 581–613.
- Teichler, U. (1996). Comparative higher education: Potentials and limits. *Higher Education*, 32(4), 431–465.
- Teichler, U. (2014). Opportunities and problems of comparative higher education research: The daily life of research. *Higher Education*, 67(4), 393–408.
- Visher, M. G., Weiss, M. J., Weissman, E., Rudd, T., & Wathington, H. D. (2012). *The effects of learning communities for students in developmental education: A synthesis of findings from six community colleges*. New York, NY: MDRC.
- Walker, A., & Leary, H. (2009). A problem-based learning meta-analysis: Differences across problem types, implementation types, disciplines, and assessment levels. *Interdisciplinary Journal of Problem-Based Learning*, 3(1), 6–28.
- What Works Clearinghouse. (2014). *Procedures and standards handbook, Version 3.0*. Retrieved from [http://ies.ed.gov/ncee/wwc/pdf/reference\\_resources/wwc\\_procedures\\_v3\\_0\\_standards\\_handbook.pdf](http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf)

## Authors

TRICIA A. SEIFERT is an associate professor in Adult and Higher Education at Montana State University and maintains a faculty appointment at the Ontario Institute for Studies in Education, University of Toronto. Her research examines postsecondary organizational structures and cultures as well as student experiences associated with learning and success.

NICHOLAS A. BOWMAN is an associate professor of higher education and student affairs as well as the director of the Center for Research on Undergraduate Education at the University of Iowa. He studies issues of college diversity, religion/worldview outcome assessment, college rankings and prestige, and student success.

GREGORY C. WOLNIAK is a director of the Center for Research on Higher Education Outcomes and an associate professor of higher education at New York University. His research examines the socioeconomic effects of college, focusing on the factors that influence students' pathways into college and the career and economic effects of college.

ALYSSA N. ROCKENBACH is a professor of higher education at North Carolina State University. She studies students' spiritual development, the religious/worldview diversity of postsecondary institutions, and campus climate.

MATTHEW J. MAYHEW is the William Ray and Marie Adamson Flesher Professor of Educational Administration with a focus on higher education and student affairs at The Ohio State University. His research focuses on the relationship between college attendance and student learning, with a particular emphasis on democratic outcomes.