3

This chapter examines the validity of college student self-reported gains by exploring the correspondence between self-reported gains and longitudinally assessed gains for diverse groups of students.

Assessing Learning and Development Among Diverse College Students

Nicholas A. Bowman

To determine the impact of college experiences, it is critical to have measures of student learning and development that are valid, reliable, and affordable. Data on student outcomes are useful not only to institutions that want to assess their students' growth but also to external constituencies, such as federal and state governments, that demand evidence about the effects of the college experience (for example, U.S. Department of Education, 2006). Moreover, having accurate and affordable measures of student development enables researchers to identify aspects of the college experience that are most educationally effective.

To measure student growth, many researchers have asked students to report how much they have gained on a variety of cognitive, attitudinal, and behavioral outcomes (Gonyea, 2005). Because these assessments are cross-sectional and rely on one item per outcome, they require far fewer resources to administer than do longitudinal assessments. Unfortunately, though, the validity of college student self-reported gains is questionable at best (Bowman, 2008; Dumont and Troelstrup, 1980; Pascarella, 2001; Pike, 1993, 1999). Within large samples of students, the correspondence between self-reported gains and longitudinal gains is quite low, and systematic and substantial biases in self-reported gains are evident. However, no study has examined whether certain groups of students (such as first-generation or female students) are more accurate in making these self-assessments than are other groups. If substantial group differences exist, then colleges and

NEW DIRECTIONS FOR INSTITUTIONAL RESEARCH, no. 145, Spring 2010 © Wiley Periodicals, Inc.
Published online in Wiley InterScience (www.interscience.wiley.com) • DOI: 10.1002/ir.322

universities with particular student populations may be justified in using self-reported gains as valid indicators of student learning and development.

Estimating Self-Reported Gains

For any form of self-reported information, Kuh and colleagues (Kuh and others, 2001; Umbach and Kuh, 2006) proposed five conditions that should be met for participants' responses to be valid:

> (a) the information requested is known to the respondents, (b) the questions are phrased clearly and unambiguously, (c) the questions refer to recent activities, (d) the respondents think the questions merit a serious and thoughtful response, and (e) answering the questions does not threaten, embarrass, or violate the privacy of the respondent or encourage the respondent to respond in socially desirable ways [Umbach and Kuh, 2006, p. 173].

Unfortunately, some of these conditions are unlikely to be met among college students providing self-reported gains. The discussion here focuses on three of these areas, giving special attention to whether these conditions are more likely to be satisfied for some students than for others.

Understanding the Questions. Many items that assess self-reported gains contain concepts that are commonly used in higher education and would be generally understood by the majority of college students. However, this familiarity does not imply that all students would interpret the same item in the same way. For example, what exactly does "critical thinking" mean? The definition of this concept may be a source of contention or disagreement among scholars who study critical thinking, let alone among college students. Moreover, what higher education scholars might mean by this term could differ substantially from what college students might think. Similar critiques could be leveled at the interpretive ambiguity of a variety of items, such as tolerance for people from diverse backgrounds, general knowledge, spirituality, and leadership abilities.

Systematic between-group differences may also be apparent in students' definitions. For example, East Asians and Caucasian Americans have substantially divergent views on what constitutes complex thinking, interpersonal relationships, and even how a person defines oneself (for reviews, see Markus and Kitayama, 1991; Nisbett, 2003). That is, East Asians tend to think holistically and define themselves in terms of their relationships with close social others, whereas Caucasian Americans tend to think analytically and define themselves as unique individuals (that is, distinct from social others). These differences in cognitive style and conceptions of self are also apparent among Asian Americans and European Americans. Given this divergence, students from diverse cultural backgrounds are likely to interpret the same broad concept differently in attempting to determine their

development, which then contributes to lower correspondence between self-reported and longitudinal gains.

Social Desirability. An important issue in administering surveys is the potential impact of social desirability on responses (overreporting positive traits, behaviors, or attitudes, while underreporting negative attributes). Within the context of college student self-reports, it seems reasonable to assume that gaining on positive traits or skills is seen as desirable, and some college students may report gains that are greater than what they actually believe. Social desirability is most easily observed when a single, objective answer exists. For example, Cole and Gonyea (2008) found that many college students inaccurately report their SAT and ACT scores, and these students are much more likely to overreport their scores (in other words, report scores that are higher than what they actually received) than to underreport them. This topic has received virtually no empirical attention in regard to student self-reported gains, even though social desirability may have substantial effects on these responses.

To the degree that social desirability does play a role in shaping student responses, these effects may vary across student groups. Substantial cross-cultural and within-culture differences exist in the concern for authority and hierarchy (Kohn, 1969; Triandis, 1989). Students who have greater attention to and respect of authority could plausibly be motivated to give more accurate responses, because they do not want to deceive the researchers or their institutions. On the other hand, these students might actually overreport their gains in an effort to "look good" in the eyes of the researchers or institutions. In a large institutional sample, students from several groups that are traditionally underrepresented in American higher education—Latinos, low-income students, and nontraditional-age students—provided greater self-reported gains than their peers, even when controlling for longitudinal gains on the same measures (Bowman, 2008). However, it is unclear whether these students simply reported genuinely held perceptions that were more positive than those of their peers or inflated their perceptions because of social desirability pressures.

Students' Knowledge of Their Learning and Development. The largest problem with self-reported gains is that students have little idea how much they have grown while in college. Interestingly, students are quite accurate in reporting their *current* attributes; according to multiple studies (Berdie, 1971; Pohlmann and Beggs, 1974), correlations between students' perceptions of their knowledge and skills and objective measures of the same constructs are moderate to high, ranging from .47 to .74. However, these relationships are much lower when comparing students' perceptions of *gains* to longitudinally measured gains. Bowman (2008) found that the median correlation between self-reported gains and longitudinal gains was .07, and these correlations for cognitive outcomes were virtually zero. Self-reported gains and self-reports of current attributes are both subject to social desirability and potential difficulties with understanding the question; therefore the

most plausible account for the low correspondence between gain measures is that students are very poor at estimating their learning and development.

The surprising inability to make accurate judgments about oneself has been demonstrated in a variety of contexts. For instance, Gilbert (2007) showed that people are not very good at predicting what makes them happy. Intuitively, it seems quite easy to predict which experiences will lead to happiness, but people's predictions do not align well with their own subsequent ratings of happiness in real-world situations. In general, people are confident about all kinds of self-knowledge, such as knowledge of their own traits and characteristics and their ability to identify factors that influenced their decision making, but their views are often erroneous (Nisbett and Wilson, 1977; Pronin and Kugler, 2007; Wilson, 2002).

In the context of self-reported gains, students probably feel that they have a great deal of insight into their own learning and development, even if they are inaccurate in their judgments. According to several theorists (Nisbett and Wilson, 1977; Ross, 1989; Wilson, 2002), people's introspections reflect causal theories that make intuitive sense to them, and these introspections are correct to the degree that these implicit causal theories align with reality. This phenomenon seems to be apparent within students' perceived gains. For example, college students report greater knowledge gains in their major than in other disciplines (Pace, 1984); this subjective assessment, which is consistent with general "common sense," seems exceedingly likely to be confirmed by any objective measure of subject area knowledge. However, experiences that would logically appear to influence students' skills and attributes may not always do so, which can lead to errors in judgment. Conway and Ross (1984) randomly assigned college students into either a series of study skills workshops or a waitlist for the workshops. Relative to the control group, students who participated in the workshops expected to receive higher grades in their major, and they reported receiving higher grades when subsequently asked about the semester in which they took the workshops. However, students who took the workshops overestimated their grades in both instances; that is, they expected higher grades than they later received, and they falsely recalled having received higher grades than they had actually been given. These faulty self-reports were consistent with many students' causal theory, which is that study skills workshops should lead to better study skills and higher grades. In contrast, students who were on the waitlist had no reason to expect their grades to improve, and they did not overestimate their predicted or actual grades.

In another form of potential bias, students may respond to specific self-reported gain items in a way that reflects their overall perception of gains, which would diminish the validity of self-report responses. In other words, students form an impression of how much they have grown while in college, and that global perception unduly affects their judgments about gains in specific areas. This tendency to respond to specific items on the basis of general perceptions of a subject is known as the halo effect (Wells, 1907).

These effects have been well documented in a variety of domains over time and observed in people's judgments about themselves and about others (Cooper, 1981; Symonds, 1925; Thorndike, 1920; Wells, 1907). Within the college context, Pike (1999) has found that halo effects in student self-reported gains can be massive, accounting for more than half of the explained variance in self-reported gains among first-year students and at least one-quarter of explained variance among seniors. Moreover, Pike (1993) found that halo error is more strongly related to students' perceptions of gains in general personal development than to other perceptions of development or overall college satisfaction.

It is possible that certain groups might have better knowledge of their own learning and development than do others. For instance, middle-class European Americans are highly concerned with developing and maintaining a unique, individualized sense of self (Kusserow, 1999, 2004; Markus and Kitayama, 1991; Stephens, Markus, and Townsend, 2007). This focus on the individual would seem to promote self-knowledge, and people who often reflect on their unique qualities and attributes may be more likely to notice changes in themselves over time. Simultaneously, though, European Americans are also likely to have an overly high opinion of their own attributes (for a review, see Heine and Hamamura, 2007). In contrast, other cultural groups (such as East Asians) have a relatively greater emphasis on self-criticism and self-improvement (Kitayama, Markus, Matsumoto, and Norasakkunkit, 1997). Successful attempts to engage in self-improvement must start with accurate knowledge of one's strengths and weaknesses, and they must involve knowledge of how one has grown. Moreover, East Asians and Asian Americans are more likely than European Americans to take an "outsider" approach on self-perception, which involves viewing oneself and one's own attributes from the view of an observer (Cohen, Hoshino-Browne, and Leung, 2007). This outsider perspective and attention to self-improvement may lead to more objective and unbiased self-assessments among Asian Americans than among European Americans.

Present Study

In sum, college students have a limited understanding of their own learning and development. However, it seems likely that certain groups of students have more accurate perceptions of their growth. To examine this possibility, the current study focused on differences in the correspondence between self-reported and longitudinal gains across race and ethnicity, gender, and social class. In addition, because some evidence suggests that students with greater cognitive ability are more accurate in assessing their own test performance (Truxello, Seitz, and Bauer, 2008), potential differences across academic achievement were also explored. This study included both cognitive and noncognitive outcomes; moreover, some of the longitudinal gains were assessed with objective measures (exams), whereas others

used subjective measures (students' perceptions of their attributes). The breadth of the types of outcomes and the forms of longitudinal assessments enhance the generalizability of these findings.

Method

Data Source and Participants. Data from the Wabash National Study of Liberal Arts Education (WNSLAE) were used for this study. Nineteen colleges and universities (eleven liberal arts colleges, three research universities, three regional universities, and two community colleges) were included in the sample on the basis of their strong commitment to liberal arts education. The study sample contained both private and public institutions, along with religiously affiliated, single-sex, and minority-serving schools. These institutions exhibited a range of selectivity, tuition costs, and geographic diversity.

Students who were beginning their first year of college in fall 2006 were invited to participate in a longitudinal study. Before classes began or during their first two or three weeks on campus (Time 1), students completed a registration form that included demographic information; a questionnaire of various high school experiences, interests, attitudes, and values; and a battery of five assessments. All students completed all assessments, but half of the students completed a critical thinking measure (the critical thinking module of the College Assessment of Academic Proficiency), whereas the other half completed a measure of moral reasoning (the Defining Issues Test 2). Students received $50 for their participation, and a total of 4,501 students completed this first wave. At the end of their first year (Time 2), students who took part in the initial assessment were invited to participate in a second round of data collection. They completed the same battery of assessments, along with questionnaires that asked about their college experiences, interests, attitudes, values, and self-reported gains during their first year. Once again, students who completed all measures received $50 as compensation. A total of 3,081 students participated in the second wave, yielding a retest response rate of 68 percent. Among students who responded to both waves of the survey, 65.4 percent were female, 79.6 percent were White non-Hispanic, 7.3 percent were Asian/Pacific Islander, 5.6 percent were Hispanic, 5.0 percent were black non-Hispanic, 0.3 percent were American Indian/Alaska Native, and 1.7 percent did not report their race or ethnicity. Of these, 3,072 had valid data on the assessments; 1,569 students completed the CAAP critical thinking module, and 1,503 completed the DIT2.

A sample weighting algorithm was developed to make the sample more representative of the incoming students at these colleges and universities. Specifically, follow-up participants were weighted according to each institution's first-year undergraduate population by sex, race (white, African American/black, Hispanic/Latino, Asian/Pacific Islander, or other), and ACT

quartile (or equivalent score). Applying weights in this manner has the effect of making the overall sample more similar to the population from which it was drawn, but it cannot adjust for nonresponse bias.

Several outliers were removed from the analyses. One student who had an extremely high weight was eliminated so that his results would not excessively influence the subgroup analyses. Moreover, students who were outliers on a given variable were removed for analyses that involved that variable. Because the overall sample is quite large, if the traditional outlier cutoff of plus or minus three standard deviations from the mean were used, then ten students would be eliminated from a perfectly normal distribution. Therefore, cases that were greater than four standard deviations from the mean were eliminated. By this criterion, some students were eliminated from analyses of critical thinking (n = 19), understanding others from different racial backgrounds (n = 57), and understanding oneself (n = 37). Preliminary analyses showed that the removal of outliers had little impact on the substantive findings.

Measures

Dependent Variables. Four items that asked students to estimate self-reported gains were included. These items were "thinking critically and analytically," "developing a personal code of values and ethics," "understanding people of other racial and ethnic backgrounds," and "understanding yourself." All items used a four-point scale (1 = gained "very little," to 4 = gained "very much").

Four corresponding longitudinal measures were used. The College Assessment of Academic Proficiency (CAAP) critical thinking module was used to gauge critical and analytical thinking over time. Developed by ACT, the forty-minute, thirty-two-item CAAP module measures a student's ability to clarify, analyze, evaluate, and extend arguments. Students receive an overall score for the number of questions they answer correctly. The internal consistency reliabilities for the CAAP critical thinking test range between .81 and .82 (ACT, 1991), and CAAP critical thinking scores correlate very highly (r = .75) with the Watson-Glaser Critical Thinking Appraisal (Pascarella, Bohr, Nora, and Terenzini, 1995).

Developing a personal code of values and ethics was measured longitudinally with the Defining Issues Test 2 (DIT2; Bebeau and Thoma, 2003; Rest, Narvaez, Thoma, and Bebeau, 1999). The DIT2 and its predecessor, the Defining Issues Test, are designed to measure students' moral judgments. Application of one's own complex framework for considering moral problems is associated with a high level of reasoning; in other words, students who have high N2 scores on the DIT2 have developed and are employing their own code of values and ethics. Previous evidence has shown that the N2 score is reliable (α = .77 to .81; Rest et al., 1999; University of Minnesota,

n.d.), and the N2 and its predecessor (the P-score) predict a variety of forms of moral thinking and behavior (for a synthesis of this literature, see Pascarella and Terenzini, 1991, 2005).

The two longitudinal measures described here—the CAAP critical thinking module and DIT2—are objective in nature. The other two longitudinal measures are primarily subjective; that is, they are based on participants' self-perceptions. Understanding people from other racial and ethnic backgrounds was measured longitudinally with the relativistic appreciation subscale of the Miville-Guzman Universality-Diversity scale (M-GUDS; Fuertes and others, 2000; Miville and others, 1999). Basing their conception on the work of Vontress (1986, 1988), Yalom (1985), Miville and colleagues (1999) describe a relativistic appreciation of diversity as a recognition and valuation of the similarities and differences among diverse people. Clearly, recognizing commonalities and similarities among diverse groups is a necessary component of understanding people from diverse racial backgrounds. The overall score for the M-GUDS scale is positively associated with empathy and a more complex racial identity, whereas it is negatively associated with dogmatism and homophobia (Miville and others, 1999). The current study used the short form of the M-GUDS (Fuertes and others, 2000); the relativistic appreciation subscale contains five items that are each rated on a five-point scale (1 = strongly disagree, to 5 = strongly agree). In the present sample, the internal reliability of the M-GUDS relativistic appreciation scale is .77 at Time 1 and .79 at Time 2, and this measure is highly correlated with Pascarella and colleagues' openness to diversity and challenge scale (r = .50 at Time 1, and r = .56 at Time 2; Pascarella and others, 1996).

Understanding oneself was measured by the consciousness of self subscale of the Socially Responsible Leadership Scale (SRLS; Tyree, 1998), which is based on Astin and colleagues' social change model of leadership development (Astin and others, 1996). In this model, knowing who you are is seen as a critical component for exhibiting effective leadership skills. This subscale directly asks participants to rate their level of self-knowledge with items such as, "I know myself pretty well." Thus the consciousness of scale represents a purely subjective assessment that is designed to gauge the same construct as the corresponding self-reported gain measure. The scale included nine items, and the internal reliabilities were .78 at Time 1 and .80 at Time 2. In the present sample, consciousness of self is highly correlated with Ryff's self-acceptance scale (r = .64 at Time 1, and r = .68 at Time 2; Ryff, 1989).

To compare changes in longitudinal measures with self-reported gains, the differences between pretest and posttest scores for the four longitudinal measures were computed.

Categorical Variables. Several variables were used to divide students into subgroups. Social class was delineated by first-generation status; first-generation college students were defined as students whose parents had not attended any postsecondary education. Race and ethnicity was divided into

four categories: black non-Hispanic, Asian/Pacific Islander, Latino/Hispanic, and white non-Hispanic. Gender was also included. Finally, students were placed into one of two categories of high school grade point average: those who had an A average (A+ to A–) and those who had a B+ average or less. High school GPA was used as the best possible measure of academic ability; grading stringency in college varies dramatically by major, even within the same institution (Johnson, 2003), and standardized test scores were available for only some students. The weighted sample sizes for each group are seen in Table 3.1.

Analyses

To determine the correspondence between self-reported gains and longitudinal gains, Pearson correlations between self-reported gains and pre-post differences in longitudinal assessments were computed. High correspondence between self-report and longitudinal gains suggests that students' self-reported gains are accurate indicators of learning and development (Bowman, 2008; Dumont and Troelstrup, 1980). Fisher *r*-to-*z* transformations were used to compare the size of correlations among subgroups. For race and ethnicity, the correlations for all three groups of students of color were compared to those of white/Caucasian students. When significant differences were found, Levene's test of equality of variances was used to determine whether these differences can be attributed, at least in part, to greater variances among one of the groups. That is, for the variables X and Y, group

Table 3.1. Weighted Sample Sizes of Subgroups in the Longitudinal Sample

Subgroup	*Number of Students*	*Percentage of Sample*
First-generation students	316	10.5
Nonfirst-generation students	2656	88.3
Missing data	34	1.1
Female	1687	56.1
Male	1320	43.9
Missing data	0	0.0
High school GPA of B+ or less	1209	40.2
High school GPA of at least A–	1798	59.8
Missing data	0	0.0
Black/African American	123	4.1
Latino/Hispanic	157	5.2
Asian/Pacific Islander	226	7.5
White/Caucasian	2438	81.1
Other or missing data	63	2.1

Note: Percentages may not add up to 100.0% because of rounding.

A might exhibit a larger correlation than group B. However, this may be attributed to the fact that the dispersion of X or Y is larger in group A than in group B. Therefore, when applicable the equality of variances between groups was tested for both self-reported gains and longitudinal gains.

Limitations. Some limitations should be noted. First, the sample includes only first-year college students. Because halo effects in self-reported gains tend to be greater in first-year students than in seniors (Pike, 1999), the patterns of correspondence between longitudinal measures and self-reported gains may differ for more advanced students. Second, the content of the longitudinal measures may not perfectly match that of self-reported gains, particularly in the case of developing a personal code of ethics and understanding people from other racial backgrounds. Although the constructs measured by DIT2 and the M-GUDS relativistic appreciation scale are closely related to the concepts in the self-reported gains, they are not synonymous. Fortunately, the intended content of self-report and longitudinal measures is virtually identical for critical thinking and understanding oneself. In addition, the overall correspondence between self-report and longitudinal gains for critical thinking and understanding oneself is quite similar to that for personal code of values and understanding people from other racial backgrounds (Bowman, 2008), which suggests that this is not a serious concern. Third, the sample sizes for students of color are fairly small; thus strong conclusions should not be drawn from any single correlation, because one or two students may be driving a particular effect. However, the pattern of correlations across all outcomes may indicate meaningful group differences.

Results and Discussion. By far, the most pronounced differences in self-reporting ability occur across social class groups. As shown in Table 3.2, correlations between self-reported and longitudinal gains are significantly higher for first-generation students than for nonfirst-generation students in critical thinking ($r = .26$ and $.07$, respectively; $p < .03$), moral reasoning ($r = .17$ and $-.03$; $p < .02$), and understanding people from different racial backgrounds ($r = .31$ and $.07$; $p < .001$). Overall, the correlations for first-generation students are still generally small, but the correlations for nonfirst-generation students are virtually zero.

Levene's tests of equality of variance were conducted to determine whether these group differences can be attributed in part to greater dispersion in the distributions for first-generation students. First-generation students do have greater variance than other students on self-reported gains in critical thinking ($p < .04$), so this correlation difference should be interpreted with caution. However, the variance for first-generation students is significantly *smaller* than that of nonfirst-generation students for self-reported gains in personal code of ethics ($p < .01$) and for longitudinal gains in understanding people from different racial backgrounds ($p < .005$). In other words, the variance differences between groups for these two variables make the observed correlation patterns *less* likely to occur. No other group

Table 3.2. Correlations Between Self-Reported and Longitudinal Gains on the Same Outcome by First-Generation Status

Outcome	*First-Generation Student*	*Nonfirst-Generation*	*Z-value of Difference*
Critical thinking	.26***	.07*	2.51*
Personal code of ethics	.17*	–.03	2.30*
Understanding people from other races	.31***	.07**	4.10***
Understanding oneself	.13*	.07***	0.98
Median correlation	.22	.07	

Note: Fisher *r*-to-*z* transformations cannot be used to compare median correlations. * $p < .05$; ** $p < .01$; *** $p < .001$.

differences in variance are significant (*p*s > .15). In sum, correlations between longitudinal and self-reported gains are higher for first-generation students than for nonfirst-generation students, and this general pattern of results cannot be explained by variance differences between groups. Thus it seems that first-generation students may have a better understanding of how they have developed than do other students.

To some degree, this finding may come as a surprise; first-generation students often have little knowledge of what to expect in college and often feel out of place in their early college experiences (Terenzini and others, 1994; Zwerling and London, 1992). However, this unfamiliarity, along with a relatively difficult adjustment period, may actually contribute to a heightened sense of self-awareness regarding their development. First-generation students are probably quite cognizant of their own traits and skills, because the college environment is novel and unfamiliar, and they generally differ importantly from their middle- and upper-class peers. As a result, first-generation students are likely attuned to their current attributes and how these attributes change over time. Such an interpretation is consistent with the particular outcomes for which significant differences are observed (critical thinking, personal code of values and ethics, and understanding people from other racial backgrounds). First-generation students typically have poorer academic preparation than their college peers (Warburton, Bugarin, and Nunez, 2001), so critical thinking and other academic skills would likely be a concern. Moreover, relative to middle-class Americans, working-class Americans place a very high value on the persistence of personal and moral integrity (Lamont, 2000; Snibbe and Markus, 2005), and norms and values are often defined through external factors such as authority figures and one's family (Kohn, 1969; Stephens, Markus, and Townsend, 2007). As a result, first-generation students' personal code of ethics is likely to be salient during college, in which the prevailing norms may be substantively different from their earlier environment. Finally, because first-generation students often have more difficulty establishing social ties with

their college peers, understanding others from all racial and ethnic backgrounds is likely a key concern.

Relatively few significant differences in the correlations between self-reported and longitudinal gains are apparent across the other student groups. The only observed difference for academic achievement is that lower-achieving students exhibit a higher correlation between critical thinking measures than do higher-achieving students (r = .17 and .03, respectively; see Table 3.3). This significant finding cannot be accounted for by group differences in the dispersion of these variables; the variances on the critical thinking measures for these two groups' gains do not differ significantly (ps > .20). The most likely explanation for the correlation disparity is similar to that of first-generation students: because the academic skills of lower-achieving students are more often challenged or questioned than are those of high-achieving students (and are therefore more salient), lower-achieving students may be more attentive to their own cognitive development.

The only gender difference that was close to significant (p < .12) dealt with interpersonal concerns, which are widely considered to be more salient for women than for men (see, for example, Deaux and LaFrance, 1998). Specifically, the correlation between self-reported and longitudinal gains for understanding people from other backgrounds is twice as high for women as for men (r = .12 and .06, respectively; see Table 3.4). Although at the trend level, this pattern supports the interpretation that students are better at estimating gains in domains that are highly relevant to them.

Significant differences across racial and ethnic groups are infrequent, which may be primarily the product of small sample sizes for the three groups of students of color. However, the patterns of results were generally consistent with the foregoing explanation. For instance, Asian/Pacific Islander students, whose cultures often place a strong emphasis on interpersonal dynamics and attunement, have a marginally higher correlation (p < .08) for understanding people from other racial groups than do

Table 3.3. Correlations Between Self-Reported and Longitudinal Gains on the Same Outcome by High School Academic Achievement

Outcome	*B Average or Lower*	*A Average*	*Z-value of Difference*
Critical thinking	.17***	.03	2.62**
Personl code of ethics	.04	–.04	1.50
Understanding people from other races	.08**	.10***	–0.56
Understanding oneself	.04	.09***	–1.31
Median correlation	.06	.06	

Note: Fisher r-to-z transformations cannot be used to compare median correlations. * p < .05; ** p < .01; *** p < .001.

Table 3.4. Correlations Between Self-Reported and Longitudinal Gains on the Same Outcome by Gender

Outcome	*Female Students*	*Male Students*	*Z-value of Difference*
Critical thinking	.10**	.08*	0.31
Personal code of ethics	.00	–.02	0.40
Understanding people from other races	.12***	.06*	1.57
Understanding oneself	.07**	.06*	0.27
Median correlation	.09	.06	

Note: Fisher *r*-to-*z* transformations cannot be used to compare median correlations. * $p < .05$; ** $p < .01$; *** $p < .001$.

white/Caucasian students ($r = .21$ and .08, respectively; see Table 3.5). There are no group differences in variance for the relevant self-report and longitudinal gains (ps > .15). The correlation among Latino/Hispanic students for understanding people from other racial groups ($r = .20$) is very similar to that of Asian students, but the small sample size for Latino students yields a nonsignificant difference between Latino/Hispanic and white/Caucasian students ($p = .16$; see Table 3.6). Latino/Hispanic students are marginally more accurate at estimating their own critical thinking gains than are white/Caucasian students ($r = .27$ and .09, respectively; $p < .09$), even though Latino students have a marginally *smaller* variance on self-reported gains than do white students ($p < .08$). Given the general skepticism regarding the academic ability of Latino college students (for example, Aronson and Salinas, 1997), it seems reasonable to assume that these students would be more attentive to their cognitive development than white students.

Table 3.5. Correlations Between Self-Reported and Longitudinal Gains on the Same Outcome by Race and Ethnicity (Asian American/Pacific Islander vs. White/Caucasian)

Outcome	*Asian American/ Pacific Islander*	*White/ Caucasian*	*Z-value of Difference*
Critical thinking	.16	.09**	0.64
Personal code of ethics	.03	.00	0.32
Understanding people from other races	.21**	.08***	1.78+
Understanding oneself	.10	.06**	0.55
Median correlation	.13	.07	

Note: Fisher *r*-to-*z* transformations cannot be used to compare median correlations. + $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 3.6. Correlations Between Self-Reported and Longitudinal Gains on the Same Outcome by Race and Ethnicity (Latino/Hispanic vs. White/Caucasian)

Outcome	*Latino/ Hispanic*	*White/ Caucasian*	*Z-value of Difference*
Critical thinking	.27**	.09**	1.73+
Personal code of ethics	.13	.00	0.97
Understanding people from other races	.20*	.08***	1.39
Understanding oneself	.06	.06**	0.05
Median correlation	.17	.06	

Note: Fisher *r*-to-*z* transformations cannot be used to compare median correlations. + $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

One significant (and unexpected) correlation difference is apparent for black students: as shown in Table 3.7, Black students are significantly worse than white students at estimating their gains in personal code of ethics ($r = -.36$ and .00, respectively; $p < .005$). There are no significant variance differences across groups ($ps > .15$), so this difference is not the result of range restriction within one of the subsamples. The significant negative correlation among black students may be the product of one or two outliers within this subsample. As noted earlier, only about half the students in the entire sample took the DIT2, so there were a very small number of black students for this particular correlation (n = 64). Overall, black students have the lowest median correlation of any group in the study ($r = -.03$). It is unclear why this might be the case, but this pattern suggests that self-reported gains may be particularly ineffective for assessing the college outcomes of black students.

Table 3.7. Correlations Between Self-Reported and Longitudinal Gains on the Same Outcome by Race and Ethnicity (Black/African American vs. White/Caucasian)

Outcome	*Black/African American*	*White/ Caucasian*	*Z-value of Difference*
Critical thinking	–.05	.09**	–1.04
Personal code of ethics	–.36**	.00	–2.84**
Understanding people from other races	–.01	.08***	–0.99
Understanding oneself	.12	.06**	0.68
Median correlation	–.03	.07	

Note: Fisher *r*-to-*z* transformations cannot be used to compare median correlations. * $p < .05$; ** $p < .01$; *** $p < .001$.

Conclusions and Implications

The findings of this study support three broad conclusions, each of which has important implications for institutional research. First, regardless of race, gender, social class, and academic achievement, students are largely unable to estimate how much they have learned and developed during the first year of college. The median correlation between self-reported gains and longitudinal gains for the various groups ranges from .22 to –.03; in other words, self-reported gains explain less than 5 percent of the variance in longitudinal gains for all groups of students. Thus it seems that self-reported gains do not constitute reasonable proxies for longitudinal measures of student learning and development in the first year of college. This conclusion may not extend to self-reported gains for all four years of college; Pike (1999) finds that one substantial source of potential bias in self-reported gains (halo effects) is considerably stronger among first-year students than among seniors. Given the importance of assessing student outcomes over the entire length of the college experience, this constitutes a critical topic for future research.

Second, first-generation students—and to a lesser degree, Latino/Hispanic students—are more accurate in reporting their gains than are their peers. These two groups had the highest median correlations between longitudinal and self-reported gains (.22 for first-generation students, and .17 for Latino/Hispanic students). As with all of the analyses, the correlations noted here are depressed to some degree by the unreliability of the longitudinal and self-report measures; that is, the true relationship between self-reported and longitudinal gains would be higher if these constructs were measured with perfect reliability. At institutions that serve a very high proportion of first-generation students, then, using self-reported gains may constitute a reasonably valid technique for assessing student learning and development. Because the sample of Latino students was quite small in this sample, more support is needed to make a strong recommendation.

Third, students seem to be more accurate at assessing their growth when the outcome is salient and important to them. For instance, lower-achieving students are better than higher-achieving students at estimating their gains in critical thinking, Asian/Pacific Islander students are somewhat more accurate than white/Caucasian students in assessing changes in their understanding of racial groups, and first-generation students are more effective than other students at estimating their cognitive and interpersonal development. For each group of students, these respective topics likely constitute an area of concern upon entering college and throughout their first year. If this interpretation is correct, then accurate self-assessments are made possible by continuous reflection during the college experience, not by post-hoc consideration of an outcome after attending an institution for an entire year. Asking students to reflect on their learning and development may improve their objective ability to assess their own growth. Through appropriate training and practice,

students may become reasonably skilled at understanding and monitoring their own learning and development (see Peet, 2006). Fostering a nuanced and accurate self-understanding may significantly improve the quality of institutional research on student growth. From a student development perspective, students' keeping track of their own progress would help them know where and how they still need to improve, which could in turn foster greater improvements on these outcomes.

Given the strong reliance on self-report measures in institutional research, more research on their validity—and on how to improve their validity—is certainly needed. For example, Pascarella (2001) suggested that some biases in self-reported gains may be diminished or eliminated by asking college students to report how much they gained in high school and then statistically controlling for these high school gains. Despite the fact that this technique may yield substantial benefits, higher education researchers virtually never ask students about their gains in high school, and no empirical study has examined whether controlling for self-reported high school gains would in fact improve the validity of self-reported gains in college.

Another intriguing possibility is that students may be able to furnish accurate estimates of their learning and development in making judgments about more specific skills and tendencies. For instance, instead of asking students how much they have gained in "critical thinking," institutional researchers might ask about students' development in identifying assumptions, strengthening or weakening arguments, comparing and contrasting multiple viewpoints, and so on. This increased specificity could potentially reduce ambiguity in the meaning of items, diminish tendencies toward socially desirable responding, and offer concrete domains in which students have relatively accurate knowledge of their own learning and development. In this time of increasing accountability and decreasing financial resources in higher education, more inquiry is needed to identify methods for assessing student growth that are valid, reliable, and cost-effective.

References

American College Testing Program (ACT). "CAAP Technical Handbook." Iowa City: Author, 1991.

Aronson, J., and Salinas, M. F. *Stereotype Threat, Attributional Ambiguity, and Latino Underperformance*. University of Texas, 1997.

Astin, A., and others. "A Social Change Model of Leadership Development: Guidebook (Version III)." Los Angeles: University of California at Los Angeles, Higher Education Research Institute, 1996.

Bebeau, M., and Thoma, S. "Guide for the DIT-2." Minneapolis: Center for the Study of Ethical Development, University of Minnesota, 2003.

Berdie, R. F. "Self-Claimed and Tested Knowledge." *Educational and Psychological Measurement*, 1971, *31*, 629–636.

Bowman, N. A. "Can First-Year College Students Accurately Report Their Learning and Development?" Paper presented at the Association for the Study of Higher Education Conference, Jacksonville, Fla., Nov. 7, 2008.

Cohen, D., Hoshino-Browne, E., and Leung, A. K.-Y. "Culture and the Structure of Personal Experience: Insider and Outsider Phenomenologies of the Self and Social World." In M. Zanna (ed.), *Advances in Experimental Social Psychology* (vol. 37). San Diego: Elsevier, 2007.

Cole, J., and Gonyea, R. "Accuracy of Self-Reported SAT and ACT Scores: Implications for Research." Paper presented at the Association for the Study of Higher Education Conference, Jacksonville, Fla., Nov. 7, 2008.

Conway, M., and Ross, M. "Getting What You Want by Revising What You Had." *Journal of Personality and Social Psychology*, 1984, *47*, 738–748.

Cooper, W. "Ubiquitous Halo." *Psychological Bulletin*, 1981, *90*, 218–244.

Deaux, K., and LaFrance, M. "Gender." In D. T. Gilbert, S. T. Fiske, and G. Lindzey (eds.), *The Handbook of Social Psychology* (4th ed.). New York: McGraw-Hill, 1998.

Dumont, R. G., and Troelstrup, R. L. "Exploring Relationships Between Objective and Subjective Measures of Instructional Outcomes." *Research in Higher Education*, 1980, *12*, 37–51.

Fuertes, J., Miville, M., Mohr, J., Sedlacek, W., Gretchen, D., and others. "Factor Structure and Short Form of the Miville-Guzman Universality-Diversity Scale." *Measurement and Evaluation in Counseling and Development*, 2000, *33*, 157–169.

Gilbert, D. *Stumbling on Happiness*. New York: Vintage, 2007.

Gonyea, R. M. "Self-Reported Data in Institutional Research: Review and Recommendations." In P. D. Umbach (ed.), *Survey Research: Emerging Issues*. New Directions for Institutional Research, no. 105. San Francisco: Jossey-Bass, 2005.

Heine, S. J., and Hamamura, T. "In Search of East Asian Self-Enhancement." *Personality and Social Psychology Review*, 2007, *11*, 1–24.

Johnson, V. E. *Grade Inflation: A Crisis in College Education*. New York: Springer, 2003.

Kitayama, S., Markus, H. R., Matsumoto, H., and Norasakkunkit, V. "Individual and Collective Processes in the Construction of the Self: Self-Enhancement in the United States and Self-Criticism in Japan." *Journal of Personality and Social Psychology*, 1997, *72*, 1245–1267.

Kohn, M. L. *Class and Conformity*. Homewood, Ill.: Dorsey, 1969.

Kuh, G. D., and others. "NSSE Technical and Norms Report." Bloomington: Indiana University, Center for Postsecondary Research and Planning, 2001.

Kusserow, A. S. "De-Homogenizing American Individualism: Socializing Hard and Soft Individualism in Manhattan and Queens." *Ethos*, 1999, *27*, 210–234.

Kusserow, A. S. *American Individualisms*. New York: Palgrave Macmillan, 2004.

Lamont, M. *The Dignity of Working Men*. New York: Sage, 2000.

Markus, H. R., and Kitayama, S. "Culture and the Self: Implications for Cognition, Emotion, and Motivation." *Psychological Review*, 1991, *98*, 224–253.

Miville, M., and others. "Appreciating Similarities and Valuing Differences: The Miville-Guzman Universality-Diversity Scale." *Journal of Counseling Psychology*, 1999, *46*, 291–307.

Nisbett, R. E. *The Geography of Thought*. New York: Free Press, 2003.

Nisbett, R. E., and Wilson, T. D. "Telling More Than We Know: Verbal Reports on Mental Processes." *Psychological Review*, 1977, *84*, 231–259.

Pace, C. *Measuring the Quality of College Student Experiences: An Account of the Development and Use of the College Student Experiences Questionnaire*. Los Angeles: Higher Education Research Institute, 1984.

Pascarella, E., Bohr, L., Nora, A., and Terenzini, P. "Cognitive Effects of 2-Year and 4-Year Colleges: New Evidence." *Educational Evaluation and Policy Analysis*, 1995, *17*, 83–96.

Pascarella, E. T. "Using Student Self-Reported Gains to Estimate College Impact: A Cautionary Tale." *Journal of College Student Development*, 2001, *42*, 488–492.

Pascarella, E. T., and others. "Influences on Students' Openness to Diversity and Challenge in the First Year of College." *Journal of Higher Education*, 1996, *67*, 174–195.

Pascarella, E. T., and Terenzini, P. T. *How College Affects Students*. San Francisco: Jossey-Bass, 1991.

Pascarella, E. T., and Terenzini, P. T. *How College Affects Students*. Vol. 2: *A Third Decade of Research*. San Francisco: Jossey-Bass, 2005.

Peet, M. R. "We Make the Road by Walking It: Critical Consciousness, Structuration, and Social Change." Unpublished doctoral dissertation, School of Education, University of Michigan, 2006.

Pike, G. R. "The Relationship Between Perceived Learning and Satisfaction with College: An Alternative View." *Research in Higher Education*, 1993, *34*, 23–40.

Pike, G. R. "The Constant Error of the Halo in Educational Outcomes Research." *Research in Higher Education*, 1999, *40*, 61–86.

Pohlmann, J., and Beggs, D. "A Study of the Validity of Self-Reported Measures of Academic Growth." *Journal of Educational Measurement*, 1974, *11*, 115–119.

Pronin, E., and Kugler, M. B. "Valuing Thoughts, Ignoring Behavior: The Introspection Illusion as a Source of the Bias Blind Spot." *Journal of Experimental Social Psychology*, 2007, *43*, 565–578.

Rest, J., Narvaez, D., Thoma, S., and Bebeau, M. "DIT2: Devising and Testing a Revised Instrument of Moral Judgment." *Journal of Educational Psychology*, 1999, *91*, 644–659.

Ross, M. "Relation of Implicit Theories to the Construction of Personal Histories." *Psychological Review*, 1989, *96*, 341–357.

Ryff, C. D. "Happiness Is Everything, or Is It? Explorations on the Meaning of Psychological Well-Being." *Journal of Personality and Social Psychology*, 1989, *57*, 1069–1081.

Snibbe, A. C., and Markus, H. R. "You Can't Always Get What You Want: Educational Attainment, Agency, and Choice." *Journal of Personality and Social Psychology*, 2005, *88*, 703–720.

Stephens, N., Markus, H. R., and Townsend, S. "Choice as Acts of Meaning: The Case of Social Class." *Journal of Personality and Social Psychology*, 2007, *93*, 814–830.

Symonds, P. "Notes on Rating." *Journal of Applied Psychology*, 1925, *9*, 456–461.

Terenzini, P. T., and others. "The Transition to College: Diverse Students, Diverse Stories." *Research in Higher Education*, 1994, *35*, 57–73.

Thorndike, E. "A Constant Error in Psychological Ratings." *Journal of Applied Psychology*, 1920, *4*, 25–29.

Triandis, H. C. "The Self and Social Behavior in Differing Social Contexts." *Psychological Review*, 1989, *96*, 506–520.

Truxello, D. M., Seitz, R., and Bauer, T. N. "The Role of Cognitive Ability in Self-Efficacy and Self-Assessed Test Performance." *Journal of Applied Social Psychology*, 2008, *38*, 903–918.

Tyree, T. "Designing an Instrument to Measure Socially Responsible Leadership Using the Social Change Model of Leadership Development." Unpublished doctoral dissertation, College of Education, University of Maryland-College Park, 1998.

Umbach, P. D., and Kuh, G. D. "Student Experiences with Diversity at Liberal Arts Colleges: Another Claim for Distinctiveness." *Journal of Higher Education*, 2006, *77*, 169–192.

University of Minnesota, Center for the Study of Ethical Development. "New index (N2)," n.d. Retrieved Jan. 14, 2008, from http://www.centerforthestudyofethical development.net/New percent20Index.htm.

U.S. Department of Education. "A Test of Leadership: Charting the Future of U.S. Higher Education." Washington D.C.: Author, 2006.

Vontress, C. E. "Social and Cultural Foundations." In M. D. Lewis, R. Hayes, and J. A. Lewis (eds.), *An Introduction to the Counseling Profession*. Itasca, Ill.: Peacock, 1986.

Vontress, C. E. "An Existential Approach to Cross-Cultural Counseling." *Journal of Multicultural Counseling and Development*, 1988, *16*, 78–83.

Warburton, E. C., Bugarin, R., and Nunez, A. "Bridging the Gap (NCES 2001–153)." Washington, D.C.: U.S. Department of Education, National Center for Education Statistics, 2001.
Wells, F. "A Statistical Study of Literary Merit." *Archives of Psychology*, 1907, *7*, 5–30.
Wilson, T. D. *Strangers to Ourselves*. Cambridge, Mass.: Belknap Press of Harvard University Press, 2002.
Yalom, I. D. *The Theory and Practice of Group Psychotherapy* (2nd ed.). New York: Basic Books, 1985.
Zwerling, L. S., and London, H. B. (eds.). *First-Generation Students*. New Directions for Community Colleges Series, no. 80. San Francisco: Jossey-Bass, 1992.

NICHOLAS A. BOWMAN is a postdoctoral research associate in the Center for Social Concerns at the University of Notre Dame.