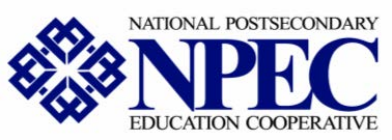


Studying Student Learning in Postsecondary Populations

A Deliberative Paper Commissioned by the

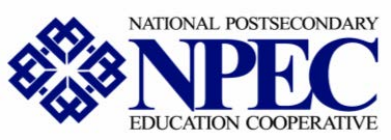
National Postsecondary Education Cooperative – Sample Surveys

Principle Author:

Dary Erwin

Consultant

Coffey Consulting, LLC



October 2012

This project has been funded, either wholly or in part, with federal funds from the U.S. Department of Education under Coffey Consulting, LLC's Contract No ED-04-CO-0044-0001. The contents of this publication do not necessarily reflect the views or policies of the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. Government.

About The National Postsecondary Education Cooperative

The National Postsecondary Education Cooperative (NPEC) was established by NCES in 1995 as a voluntary organization that encompasses all sectors of the postsecondary education community including federal agencies, postsecondary institutions, associations and other organizations with a major interest in postsecondary education data collection. NPEC's mission is to "promote the quality, comparability and utility of postsecondary data and information that support policy development at the federal, state, and institution levels." It is composed of two panels: NPEC-IPEDS (NPEC-I) and NPEC-Sample Surveys (NPEC-S).

NPEC Panels

NCES has assigned NPEC-I the specific responsibility for developing a research and development agenda for the Integrated Postsecondary Education Data System (IPEDS). IPEDS is the core postsecondary education data collection program for NCES. NPEC also intermittently produces advisory publications for use by postsecondary data providers, users, and institutional representatives. In contrast, NPEC-S is designed to provide high level guidance on the evolution of a suite of studies that includes the National Postsecondary Student Aid Study (NPSAS), the Beginning Postsecondary Students Longitudinal Study (BPS), the Baccalaureate and Beyond Longitudinal Study (B&B), and other survey and administrative data collections.

NPEC Publications

NPEC publications do not undergo the formal review required for standard NCES products. The information and opinions published in them are the products of NPEC and do not necessarily represent the policy or views of the U.S. Department of Education or NCES.

Contents

The Need for Data on Learning Outcomes in Postsecondary Education

The absence of nationally-representative data on the learning outcomes associated with college attendance is noted in virtually every report about postsecondary education from the 1980's to the present. The meaning and measurement of college student learning has continued to attract attention from higher education organizations (Dwyer, Millett, & Payne, 2006; NCPPHE, 2008), faculty (Shavelson, 2007, 2009; Zimmerman, 2012), state policymakers (State of Tennessee, 2010), federal officials (U.S. Department of Education, 2006, 2011), employers (Schneider, 2012), and the media (de Vise, 2012; Keeling & Hersh, 2011). Despite this perennial interest—and despite assessment models in elementary and secondary education such as the National Assessment of Educational Progress—incremental efforts to build a base of evidence about the relationship between college attendance and student learning at the national level, such as the broad-based adoption of a set of specific learning outcomes or the systematic use of a new assessment instrument, can evoke controversy.

In the absence of such evidence, our collective understanding of the outcomes associated with postsecondary education is murky. Pascarella and Terenzini's (1991, 2005) reviews of thirty years of outcomes research suggest that differences in students' collegiate experiences may be related to variation in a broad range of outcomes, such as: (a) cognitive, moral, and psychosocial development; (b) attitude change; (c) occupational and economic benefits; and (d) post-college quality of life. However, the authors critique the studies they consider by noting prior scholarship has often drawn conclusions on the basis of data drawn from only a single institution, or by using of samples or measures of convenience.

In an effort to develop more robust evidence of student learning, recent scholarship has sought to address these methodological concerns. Arum and Roksa (2011) provide a recent example that has garnered the attention of both the media and policymakers. Using a multi-institutional sample and a well-known assessment instrument, they conclude that, by the end of their second year, undergraduates do experience growth on measures of higher-order thinking skills, but that the magnitude of such growth appears to be "moderate" (p. 6), approximately .18 standard deviations (Arum, Roksa, & Velez, 2008).

Not surprisingly, ways to build upon Arum and Roksa's scholarship have been offered in its wake, and at least two caveats related to external validity are important to note. First, although their sample was multi-institutional in nature, it is characterized as consisting of "traditional age freshmen at four-year institutions[1]" (Arum, Roksa, & Cho, 2011, pg. 16). Second, in contrasting their sample with participants in NCES's *Beginning Postsecondary Longitudinal Study*, they find "our sample did have fewer men as well as a smaller number of students of lower scholastic ability as measured by standardized tests (e.g., students' combined scores at the 25th percentile of the SAT …)," (pg. 17), and argue that as a result, they may overestimating students' learning gains.

[1] Because Arum and Roksa's work relied upon multiple waves of data, the number of institutions and student at each wave of collection varies. Arum, Roksa and Cho (2011) note that analyses based on the first two years of data are based on 2,322 students at 24 institutions, but that, in total, 29 institutions and more than 3,000 students participated in their project.

Without nationally-representative data on student learning at the postsecondary level, institutional, state, and federal decision-makers face difficulties in addressing a wide range of policy and pedagogical questions. These questions include:

- What growth is observed, if any, in students' capacities by virtue of their participation in postsecondary education;
- How are elements of students' educational experiences related to growth on measures of key outcomes of interest; and
- How well, at their exit from baccalaureate or sub-baccalaureate education, students' capacities are aligned with the needs of the labor market and, if applicable, future study at the post-baccalaureate level?

While the specific questions a given study could hope to address would vary based upon its specific design (see discussion below), virtually *any* effort to produce systematic measures of student learning will improve our understanding of the results of the collegiate enterprise.

To that end, the National Postsecondary Education Cooperative—Sample Surveys (NPEC-S) recommends that NCES engage the higher education community in a deliberative process that explores the development of data on student learning through a nationally-representative sample survey that can generalize to undergraduate students enrolled in all institutional sectors. Below, a subset of issues considered in the development of that recommendation is summarized. The summary is *not* meant to constrain the many discussions that NPEC-S hopes will occur in the months ahead. Instead, it is offered as a way of providing insight into the process that gave rise to the panel's recommendation.

What Might be Measured?

The generally-acknowledged first step in the assessment process is agreement upon the objectives to be measured. Objectives may involve factual knowledge, demonstrable skills, or developmental tasks, and may be wholly framed within a specific domain of knowledge (e.g., a licensing examination) or more broadly (e.g., an admissions examination). While the former may entice interest, assessment within a discipline necessarily reflects the experience of only a subset of students. As a result, most prior large-scale efforts at quantifying the outcomes associated with students' participation in postsecondary education have taken a more generic approach.

The National Education Goals of 2000 project, which spanned both the Bush and Clinton administrations, is a notable example. That project delineated three skill areas: (a) oral and written communication; (b) critical thinking; and (c) problem solving. Because the meaning or definitions of these skill areas was seen to be lacking, steps were taken to define them more specifically (Jones, Hoffman, Moore, Ratcliff, Tibbetts, & Click, 1995). Definitions, however, were about as far as the postsecondary component of the National Education Goals (US Department of Education, 1994) project progressed.

The Lumina Foundation's Degree Qualifications Profile (DQP), inspired by the European Union's Bologna Process, is a more recent example (Adelman, Ewell, Gaston, and Schneider, 2011). The DQP, which attempts to define the substance of an undergraduate degree, includes five broad areas: (a) integrative knowledge, (b) specialized knowledge, (c) intellectual skills, (d) applied learning, and (e) civic learning. While the profile offers some general characterizations of these five areas, the specific definition of each DQP dimension has not yet been worked out. And, to date, no assessment instrument is part of this DQP process.

Although we consider growth on the learning and developmental constructs mentioned above as valuable ends in and of themselves, such growth can also be seen as a stepping stone to other important outcomes. Those related to workforce participation (e.g., employment, earnings) are key elements of contemporary conversations about the benefits of postsecondary education. Pascarella and Terenzini (2005) note that prior literature has drawn myriad connections between attainment and other important adult behaviors, including voting, volunteerism, and healthy living. As such, there may be merit in assessing not only collegiate learning, but also distal gains that represent both public and private goods.

What Measurement Instruments Already Exist?

Although dozens of published instruments have sought to measure outcomes similar to those identified above, only a subset demonstrates wide-spread diffusion. We review them briefly below, distinguishing between two test types: general education and adult literacy. The general education tests we consider include measures that, today, are being used to support inferences about institutional performance for the purpose of accreditation and policymaking, as well as consumer information. In contrast, the literacy tests we consider have primarily been used to develop descriptive portraits of adult skills, particularly for the purpose of cross-national comparisons.

General Education Tests

Three popular, proprietary instruments purport to measure the extent of student learning within traditional collegiate institutions (Bridgeman, Klein, Sconing, & Erwin, 2008): the *Collegiate Assessment of Academic Proficiency*® (CAAP), the *Collegiate Learning Assessment*®

(CLA), and the *Proficiency Profile*® (ETS-PP). The CAAP, CLA, and ETS-PP serve as the learning instruments of choice for the Voluntary System of Accountability (APLU, 2007), and the CLA is being used by the Organisation for Economic Co-operation and Development (2010) in their forthcoming Assessment of Higher Education Learning Outcomes (AHELO) study.

The CAAP of the American College Testing program offers five modules: reading, writing from a selective response format, writing from a constructed response format, mathematics, science, and critical thinking. Modules, which can be administered separately or in combination, are 40 minutes each in length. Sample items are located on-line at http://www.act.org/caap/sample/q.html.

The Council for Aid to Education's CLA produces institutional, rather than student-level, scores based upon a computer-administered performance task, a critique-an-argument task, and an analytic writing task. Individual respondents are assigned one of the three tasks, and, within a task, a specific "subset" of items (i.e., matrix sampling). Each individual's assessment takes between 75 and 90 minutes to complete. Those assessments are then scored and combined to form institution-level metrics. Sample items are available on-line at http://www.cae.org/content/pro_collegiate_sample_measures.htm and additional details about the CLA can be found at http://www.collegiatelearningassessment.org/files/CLA_Technical_FAQs.pdf.

Finally, the Educational Testing Service offers the *Proficiency Profile*® (ETS-PP), previously the Measure of Academic Proficiency and Progress and Academic Profile. ETS-PP produces both student and group scores in reading/critical thinking, writing, and mathematics over the humanities, social sciences, and natural sciences. ETS offers its standard version of the

Proficiency Profile®, which takes approximately two hours to complete, as well as an abbreviated test form that can be completed in approximately 40 minutes. A choice of paper-and-pencil or online delivery is also available. Sample items may be viewed at: http://www.ets.org/s/proficiencyprofile/pdf/sampleques.pdf

Adult Literacy Tests

Adult literacy tests seek to measure skills necessary for effective functioning in society and the workplace in populations that are typically no longer engaged in formal schooling. NCES has a long history of administering nationally-representative surveys of adult literacy, including the National Adult Literacy Survey (NALS; Kirsch, Jungeblut, Jenkins, & Kolstad, 1993) and its successor, the National Assessment of Adult Literacy (NAAL; U.S. Department of Education, 2012).

More recently, NCES has joined international partners in the Program for the International Assessment of Adult Competencies (PIAAC). Designed to permit international comparisons, the PIAAC assesses three domains, described below[2].

[2] PIAAC also administers a Reading Components module assessing vocabulary, sentence comprehension, and basic passage comprehension for adults with the lowest levels of literacy. NPEC-S does not recommend its inclusion in a study of postsecondary learning outcomes. For more information on the Reading Components module, see http://nces.ed.gov/surveys/piaac/reading-components.asp and http://dx.doi.org/10.1787/220367414132.

- Literacy is defined as "understanding, evaluating, using and engaging with written text to participate in the society, to achieve one's goals and to develop one's knowledge and potential" (see also http://nces.ed.gov/surveys/piaac/literacy.asp and http://dx.doi.org/10.1787/220348414075).

- Numeracy is defined as "the ability to access, use, interpret, and communicate mathematical information and ideas, to engage in and manage mathematical demands of a range of situations in adult life" (see also http://nces.ed.gov/surveys/piaac/numeracy.asp and http://dx.doi.org/10.1787/220337421165).

- Problem solving in technology-rich environments is defined as "using digital technology, communication tools, and networks to acquire and evaluate information, communicate with others, and perform practical tasks," includes such tasks as purchasing goods over the web, locating health information, and managing one's personal finances electronically. Simulations of email, spreadsheets, and web pages are also posed to participants. For more information, see http://nces.ed.gov/surveys/piaac/problem-solving.asp and http://dx.doi.org/10.1787/220262483674.

PIAAC's instrumentation is adaptive in nature. That is, test items are administered to participants based upon the accuracy of their prior responses, continually matching test difficulty with estimates of a person's ability.

To oversimplify, tests of general education demand a demonstration of *knowledge acquisition* while tests of adult literacy demand a demonstration of *knowledge application*. Some tasks—such as skill in critical thinking—may be demanded by both types of tests. While both

knowledge acquisition and knowledge application are critical outcomes of postsecondary education writ large, NPEC-S recognizes that not all programs of study are designed to promote them equally. In general, educational programs that are highly specialized (e.g., short-cycle certificates that prepare students for specific occupations) may be less likely to promote broad-based general education gains than programs that, historically, have included a "liberal arts" component (e.g., programs leading to a bachelor's of arts degree). However, as Arum and Roksa's (2011) work demonstrates, it is far from evident that participation in a baccalaureate program inevitably leads to substantive gains in critical thinking.

Constructs for any Future Study

NPEC-S acknowledges that its preference for fielding a study that is applicable to undergraduate students at all institutional sectors—that is, all combinations of institutional controls (i.e., public, private non-profit, and private for-profit) and levels (i.e., less-than two-year, two-year, and four-year)—constrains the constructs that can tenably be measured.

We believe that subject- or field-specific assessment is inappropriate for a study that seeks to generalize to the entire undergraduate population, and offer a brief example. It might well be reasonable to expect that all bachelor's degree-seeking students would evidence growth in their college-level mathematics proficiency. Indeed, in programs that include a substantial focus on science, technology, engineering, or mathematics, we might expect any such growth to be substantial. However, we might not expect that students enrolled in a less-than one-year certificate program in medical records administration would develop substantial proficiency in college-level mathematics: limited instructional time seems better spent on developing the specific skills needed for employment.

This leads us to conclude that the construct or constructs for assessment must instead focus on the competencies needed for general functioning and success in today's world. This is generally consistent with the adult literacy framework employed by PIAAC and its predecessors. Note that we do not advocate for the wholesale adoption of the PIAAC instrument *per se*, if for no other reason than we do not believe it to be technically feasible due to its length, usage restrictions, and design. Instead, we use the notion of PIAAC as *shorthand* for the general sort of instrumentation envisioned. NCES, in conjunction with stakeholder groups, assessment experts, and psychometricians will be left with the substantial challenge of refining (or, should they choose to, wholly redefining) the broad vision set forth by NPEC-S.

A Focus on Applied Knowledge

Our affinity for PIAAC is driven by the factors identified below, which we would hope to see evidenced in the instrumentation developed for any future study:

- Its components are "curriculum-neutral," instead aligning with the competencies needed for general life functioning. We believe the concepts embodied in the literacy, numeracy, and problem-solving subscales are likely to be most useful in studying collegiate impact.

- It uses both selective response (e.g., multiple choice test items) and constructed response formats (e.g., performance tasks) that are administered via a laptop computer to individuals. Using a combination of selective and constructed response formats allows the methodological advantages of each format to be available across the study.

- It is adaptive. That is, item response theory has been used to develop a set (and sequence) of items that minimizes respondent burden by quickly honing in a respondent's

measured ability, rather than administering test items well below or above the individual's capacity.

Supplemental Constructs

NPEC-S believes that the addition of one construct, absent from the PIAAC framework, might add value to this effort and urges NCES consider its inclusion in any future study: scientific literacy. Scientific literacy can be defined as: "the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity" (OECD, 2003, pp. 132–33). The National Science Foundation (NSF) is currently sponsoring projects that define and attempt to define understanding and awareness of science . Appendix A contains a sample of scientific literacy test items from a recent NSF sponsored workshop (Guterbock et al., 2011).

Finally, NPEC-S recommends NCES consider the inclusion of so-called "non-cognitive constructs" that research has suggested may be related to success in a broad range of contexts and, importantly, may be malleable and sensitive to institutional efforts to improve or enhance them. Related to a robust prior literature in higher education pioneered by Sedlacek (n.b., Sedlacek, 2004), recent work by Duckworth and colleagues (e.g., Duckworth, Grant, Loew, Oettingen & Gollwitzer, 2011; Duckworth, Peterson, Matthews, & Kelly, 2007; Duckworth, Tsukayama & May, 2010; Duckworth & Quinn, 2009) suggests personality characteristics including consistency of interest, perseverance of effort, and self-control are positively related to both academic (e.g., GPA) and non-academic outcomes (e.g., stability of employment).

Study Design

In addition to making recommendations vis-à-vis instrumentation, NPEC-S considered several other issues related to a future study's research design. These issues, which include providing additional specificity about the population to which the study sample should generalize, clarifying the research design, identifying additional analytic variables to be collected alongside outcome assessment data, and clearly stating the limitations of the proposed design, are summarized below.

Population and Sampling Strata

NCES longitudinal studies are designed to generalize to either cohorts of secondary school students of the same grade (e.g., ninth graders in the High School Longitudinal Study of 2009) or cohorts of college entrants in the same year (e.g., the Beginning Postsecondary Students Longitudinal Study). Both possibilities were considered by NPEC-S.

A secondary school grade cohort has the advantage of providing a measure of the "value added" of college attendance compared to alternative pathways, such as work or the military, and offering the opportunity to administer a true pre-test prior to college entry. In contrast, a college-entrance cohort can only include students who self-select for college attendance, and the administration of a pre-test prior to entry is, by definition, impossible.

Despite those challenges, the use of an entrance cohort has at least one distinct advantage: it is increasingly the case that secondary age cohorts are no longer capable of generating a sample that is reflective of today's postsecondary population. NCES reports that 40% of the students who were first-time, beginning college students in 2003-04 were not 2003

high school graduates. While NPEC-S believes there may be merit in exploring the use of a secondary cohort that is "freshened" with an entrance cohort to surmount this problem, it acknowledges that such an approach is likely to be costly and may not be technically feasible. Barring input from NCES that such an approach can be made workable, NPEC-S is not inclined to support research that reifies a view of postsecondary participation that, while still widely held, is simply not borne out by reality.

As noted above, NPEC-S recommends students enrolled in all institutional sectors be included in this research. More specifically, it envisions a two-stage sampling design typical of NCES studies. At the first stage of sampling, institutions will be selected from among the more than seven thousand primarily postsecondary, Title IV-participating institutions contained in IPEDS. NCES should stratify the institutional sample by level (i.e., 4-year, 2-year, and less-than 2-year) and control (i.e., public, private not-for-profit, private for-profit), in addition to any other potentially relevant characteristics.

Then, within institutions, NCES should sample first-time, beginning postsecondary students enrolled during a given academic year. Within that sample, potential strata include initial degree type program (i.e., bachelor's degree, associate's degree, certificate, and non-degree-seeking) or other student characteristics hypothesized to be related to differential learning gains (e.g., enrollment in programs that are conducted entirely on-line). NPEC-S also recommends NCES consider that possibility that one or more states might wish to partner on this effort.

Design

Given the goal of this study—the identification of change over time—NPEC-S recommends a true longitudinal design. Although there are status quo examples of research projects that attempt to infer change through cross-sectional comparisons of entering and completing cohorts, the challenges associated with such designs (e.g., suitable matching on observables, bias introduced through attrition due to drop-out) make them undesirable. Unfortunately, a longitudinal design based upon the sampling strategy identified above poses its own problem.

Specifically, if students are sampled on the basis of institutional enrollment lists—a process already used in existing NCES postsecondary studies—this means that a true pre-test before exposure to the educational "treatment" is not possible. Because enrollment lists are used as an institution's sampling frame, sampling cannot begin until an institution's enrollment is "known" for a given year. In an institution following a semester system, this will be some number of weeks into the Spring term, suggesting sampling could begin as early as February. However, in a continuous enrollment institution (or even one that follows the quarter system), enrollment may not be definitively known until early Summer. As such, for some types of institution, sampling might not begin until mid-July.

The net result of an enrollment list-based sampling strategy is that, for some students, the pre-test measurement of student learning occurs well after their entry to college. Indeed, for some students—most notably students in very short-cycle degree programs (e.g., less-than one-year certificates)—sampling for their institution may occur *after* an award has already been conferred. To be sure, the timing of the pre-test for students in longer degree programs is less problematic. Nonetheless, this notable challenge should give NCES cause to fully explore ways

in which interviewing can be begun as soon as possible after a student has entered postsecondary education.

After resolving issues related to the timing of the pre-test, NCES must determine when it is appropriate to implement one or more post-tests. NCES's current study of first-time beginning students' persistence, the Beginning Postsecondary Students Longitudinal Study (BPS), currently interviews students at the end of their first, third, and sixth years. While NPEC-S does not recommend attempting to graft an assessment of learning on to the existing BPS study due to respondent burden, the timing of follow-ups used by BPS may be instructive. Because they are nominally tied to the familiar "150% of normal time to completion" of common degrees (e.g., three years for a 2-year Associate's degree program), they are likely to capture significant proportions of students who have recently transitioned out of postsecondary education.

Importantly, NPEC-S recommends following *all* students for a period of at least six years, irrespective of their completion/persistence status. Doing so would provide important information about the trajectory of learning outcomes after the conclusion of formal education and would perhaps capture additional variation due to exposure to additional formal and informal training after college. Prior NCES studies using IPEDS data have suggested there is relatively little increase in institutional graduation rates after six years (Horn, 2010), both because IPEDS graduation rates are based on only students who complete at their first institution and also because NCES's Baccalaureate and Beyond study suggests a substantial proportion of baccalaureate students complete beyond the sixth year (24%; Cataldi et al., 2011). Therefore, NCES should consider whether an even longer follow-up period is warranted.

Other Analytic Variables

As is the case with most NCES studies, additional analytic variables will be gathered alongside assessment results. These fall in to three categories: (a) institutional characteristics, (b) student background characteristics, and (c) questions about the student experience. Each category is described below.

Institutional Characteristics

Given the sampling strategy suggested above, NPEC-S presumes that the most important institutional characteristic for analysis of any assessment data is sector (that is, the combination of institutional level and institutional control). Other strata characteristics, such as status as a minority serving institution or region of the country, may also be of interest to the analyst. Caution should be used, however, when introducing other variables that might be erroneously used as the basis of generalization when doing so would be statistically inappropriate (e.g., state in a non-representative study). Finally, and perhaps most importantly, NCES must caution any user that results *must not* be generalized to specific institutions; indeed, NCES may elect to perturb IPEDS UNITIDs in any data release and, instead, release data with coarsened institutional characteristics.

Student Background Characteristics

In addition to any characteristics used in the development of student sampling strata, other relevant background characteristics include those that could covary with student learning, such as: sex/gender, race/ethnicity, age, parental education, financial aid dependency status, income measures, household composition, and labor market behaviors. Sources for potential items include existing NCES studies such as PIAAC's extensive background questionnaire.

NCES may also be able to leverage matches to other extant data sources. Examples include data related to students' K-12 schooling, perhaps from a state student-level longitudinal data system or college admissions tests (e.g., SAT or ACT).

Student Experiences

Several facets of a student's collegiate experience may be useful in describing the variance observed in measurements of learning. This includes type of undergraduate degree program (e.g., bachelor's degree, associate's degree, certificate, or non-degree seeking), major or field of study, depth or breadth of academic coursework (e.g., course "clusters"), high impact practices (e.g., undergraduate research), pre-college work and learning experiences (e.g., military experience, professional certifications), working while enrolled, and engagement with college academic and social systems.

A Cautionary Note About "Treatments"

NPEC-S strongly advises analysts not to attempt to use data on the variables above to make causal claims. As noted earlier, these covariates should be used to help contextualize what is found in any study of student learning, not attempt to "explain" it. While many issues preclude these variables use in that way, one is the notion of *treatment fidelity*, or how accurately an intervention is applied or modeled across (and even within) institutions (Cordray & Pion, 2006; Hulleman & Cordray, 2009). For example, not all biology programs, work experiences, enrichment programs, or less-than two-year programs are alike.

Limitations

As with all research endeavors, any study of student learning that comports with the rough design identified above has its limitations. These include:

- An inability to generalize at the institutional level, only at a national level;
- A parsimonious set of learning outcomes that may not be appropriate for all students in all programs of study, particularly those that are more vocationally-focused;
- An imperfectly timed pre-test;
- An inability to make comparisons against the non-college-going population.

Conclusion

NPEC-S acknowledges this paper is but a next step in the long national conversation about the measurement of student learning. However, we believe researchers and policymakers have gone too long without the data this study hopes to develop. Although this work will be complex, and although it presents political and technical challenges, NCES has the opportunity to build relationships with a wide range of institutional, organizational, state, and national partners to advance the current state of assessment and to model good practice for others to follow.

While adult literacy assessments already underway at NCES may "jump start" a future study of postsecondary learning outcomes, the rewards of any such study are years away. Because the questions facing education policymakers are weighty and numerous, NPEC-S strongly believes the conversations, collaborations, and efforts needed to make this effort a reality are too important to delay any longer. The panel hopes this paper encourages NCES and

the broader higher education community to consider the issues raised within it, engage in a process of thoughtful deliberation and careful design, and take action to field a study of learning in postsecondary populations.

References

Adelman, C., Ewell, P., Gaston, P., & Schneider, C. G. (2011). *The degree qualifications profile.* Indianapolis, IN: Lumina Foundation for Education. Retrieved from http://www.luminafoundation.org/ publications/The_Degree_Qualifications_Profile.pdf

ACT, Inc. (n.d.) *Collegiate assessment of academic proficiency (CAAP*). Iowa City, IA: Authors. Retrieved from http://www.act.org/caap/

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago, IL: University of Chicago Press.

Arum, R., Roksa, J., & Cho, E. (2011). *Improving undergraduate learning: Findings and policy recommendations from the SSRC-CLA Longitudinal Project.* Brooklyn, NY: Social Science Research Council. Retrieved from http://highered.ssrc.org/wp-content/uploads/2012/01/Improving-Undergraduate-Learning-2011.pdf

Arum R., Roska, J., & Velez, M. (2008). *Learning to reason and communicate in college: Initial report of findings from the CLA Longitudinal Study*. Brooklyn, NY: Social Science Research Council. Retrieved from http://highered.ssrc.org/wp-content/uploads/2010/10/Arum-and-Roksa-2008-Learning-in-Higher-Ed.pdf

Association of Public and Land-Grant Universities and the American Association of State Colleges and Universities. (n.d.). Voluntary System of Accountability. Washington, DC: Authors. Retrieved from http://www.voluntarysystem.org/index.cfm

Bridgeman, B., Klein, S., Sconing, J., & Erwin, T. D. (2008). *Research on alternative methods to measure accountability in postsecondary education*. Unpublished paper presented at the Third Annual Research Conference of the Institute of Education Sciences, Washington, DC.

Cataldi, E.F., Green, C., Henke, R., Lew, T., Woo, J., Shepherd, B., & Siegel, P. (2011). *2008–09 Baccalaureate and Beyond Longitudinal Study (B&B:08/09): First Look* (NCES 2011-236). Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Cordray, D.S., & Pion, G.M. (2006). Treatment strength and integrity: Models and methods. In R. Bootzin & P. McKnight (Eds.), *Strengthening Research Methodology: Psychological Measurement and Evaluation* (pp. 103–124). Washington, DC: American Psychological Association.

Council for Aid to Education. *Collegiate Learning Assessment (CLA)*. New York, NY: Authors. Retrieved from http://www.collegiatelearningassessment.org/

Council for Aid to Education. *Collegiate Learning Assessment (CLA): Frequently asked technical questions: 2007-2008*. New York, NY: Authors. Retrieved from http://www.collegiatelearningassessment.org/files/CLA_Technical_FAQs.pdf

de Vise, D. (March 14, 2012). Trying to assess learning gives colleges their own test anxiety. *Washington Post.*

Duckworth, A. L., Grant, H., Loew, B., Oettingen, G., & Gollwitzer, P. M. (2011). Self-regulation strategies improve self-discipline in adolescents: Benefits of mental contrasting and implementation intentions. *Educational Psychology*, 31, 17-26.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087-1101.

Duckworth, A. L., Tsukayama, E., & May, H. (2010). Establishing causality using longitudinal hierarchical linear modeling: An illustration predicting academic achievement from self-control. *Social Psychological and Personality Science, 1*, 311-317.

Duckworth, A., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (Grit-S). *Journal of Personality Assessment*, *91*(2), 166-174.

Dwyer, C. A., Millett, C. M., & Payne, D. G. (2006). *A culture of evidence: Postsecondary assessment and learning outcomes.* Princeton, NJ: ETS.

Educational Testing Service (n.d.). ETS® Proficiency Profile Content. Princeton, NJ: Authors. Retrieved from http://www.ets.org/proficiencyprofile/about/content/

Guterbock, T. M., Allum, N., Besley, J., Conrad, F., Holbrook, A., Keeter, S., Losh, S., Mondak, J., Reeve, R., Rexrode, D., Sikkink, D., Stares, S., Tourangeau, R., & Tourney, C. (2011). *Measurement and operationalization of the "Science in the Service of Citizens and Consumers" framework.* Charlottesville, VA: University of Virginia Center for Survey Research. Retrieved from

http://www.lgava.org/sites/default/files/publications/Instrumentation%20Workshop%20Final%20Report.2011-01-31.pdf

Horn, L. (2010). *Tracking students to 200 percent of normal time: Effect on institutional graduation rates* (NCES 2011-221). Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*(1), 88-110.

Jones, E.A., Hoffman, S., Moore, L.M., Ratcliff, G., Tibbetts, S., & Click, B.A. (1995). *National assessment of college student learning: Identifying college graduates' essential skills in writing, speech and listening, and critical thinking* (NCES 95-001). Washington, DC: U.S. Government Printing Office.

Keeling, R. P., & Hersh, R. H. (2011). *We're losing our minds: Rethinking American higher education*. New York, NY: Macmillan.

Kirsch, I. S., Jungeblut, A., Jenkins, L., & Kolstad, A. (1993). *Adult Literacy in America: A First Look at the Findings of the National Adult Literacy Survey* (NCES 1993-275). Washington, DC: U.S. Department of Education.

Klein, S., Liu, O.L., Sconing, J., Bolus, R., Bridgeman, B., Kugelmass, H., Nemeth, A., Robbins, S., & Steedle, J. (2009). *Test validity study report.* Retrieved from http://www.voluntarysystem.org/index.cfm?page=research

National Center for Public Policy and Higher Education (2008). *Measuring Up 2008.* Retrieved from http://measuringup2008.highereducation.org/index.php

Organisation for Economic Co-operation and Development (2003). *The PISA 2003 assessment framework.* Paris, France: Authors.

Organisation for Economic Co-operation and Development and Statistics Canada. (2005). *Learning a Living: First Results of the Adult Literacy and Life Skills Survey.* Paris, FR: Authors.

Organisation for Economic Co-operation and Development. (2010). *Testing student and university performance globally: OECD's AHELO*. Paris, FR: Retrieved from http://www.oecd.org/document/22/0,3746,en_2649_35961291_40624662_1_1_1_1,00.html

Pascarella, E., & Terenzini, P. (1991). *How college affects students: Findings and insights from twenty years of research.* San Francisco, CA: Jossey-Bass.

Pascarella, E. T., & Terenzini, P. T. (2005). *How College Affects Students: A Third Decade of Research* (2nd ed.). San Francisco, CA: Jossey-Bass.

PIAAC Literacy Expert Group. (2009). *PIAAC literacy: A conceptual framework* (OECD Education Working Papers, No. 34). Paris, FR: Authors. Retrieved from http://dx.doi.org/10.1787/220348414075

PIAAC Numeracy Expert Group. (2009). *PIAAC numeracy: A conceptual framework* (OECD Education Working Papers, No. 35). Paris, FR: Authors. Retrieved from http://dx.doi.org/10.1787/220337421165

PIAAC Expert Group in Problem Solving in Technology-Rich Environments. (2009). *PIAAC problem solving in technology-rich environments: A conceptual framework* (OECD Education Working Papers, No. 36). Paris, FR: Authors. Retrieved from http://dx.doi.org/10.1787/220262483674

Schneider, M. (February 22, 2012). A better approach to 'gainful employment.' *Inside Higher Education.*

Sedlacek, W. E. (2004). *Beyond the Big Test: Noncognitive Assessment in Higher Education*. San Francisco, CA: Jossey-Bass.

Shavelson, R. J. (2007). *A brief history of student learning assessment.* Washington, DC: Association of American Colleges and Universities.

Shavelson, R. J. (2009). *Measuring college learning responsibly*. Palo Alto, CA: Stanford University Press.

State of Tennessee. (2010). *The public agenda for Tennessee higher education 2010-2015.* Retrieved from *http://www.tn.gov/thec/complete_college_tn/ccta_files/master_plan/The%20Public%20Agenda%20with%20Appendices%20Jan2011.PDF*

National Center for Education Statistics. (n.d.). *Beginning postsecondary students*. Retrieved from http://nces.ed.gov/surveys/bps/

National Center for Education Statistics. (n.d.). *National assessment of adult literacy* Retrieved from http://nces.ed.gov/naal/

National Center for Education Statistics. (n.d.). *Program for the international assessment of adult competencies*. Retrieved from http://nces.ed.gov/surveys/piaac/

National Education Goals Panel. (1994). *National education goals report: Building a nation of learners*. Washington, DC: Authors. Retrieved from http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED360394

U.S. Department of Education. (2006). *A test of leadership: Charting the future of U.S. higher education.* Washington, DC: Authors.

U. S. Department of Education. (2011). *Committee on measures of student success: A report to Secretary of Education Arne Duncan*. Washington, DC: Authors. Retrieved from http://www2.ed.gov/about/bdscomm/list/cmss-committee-report-final.pdf

Zimmerman, J. (January 31, 2012). Are college students learning? *Los Angeles Times.*

Appendix A

Sample Test Items of Scientific Literacy (Guterbock et al., 2011)

Process Items

- When you read or hear the term scientific study, do you know what it means?
- What does it mean to study something scientifically?

Factual Items

- All radioactivity is man-made.
- Whose gene decides whether the baby is a boy or girl.
- Lasers work by focusing sound waves.
- Electrons are smaller than atoms.
- Antibiotics kill viruses as well as bacteria.
- How do most fish get the oxygen they need to survive?
- Why do people experience shortness of breath at the top of a mountain?

Graphical Literacy Items

- Looking at a graph, in which time periods did the most errors occur?
- Which combination of bodily features is BEST suited to a small animal that needs to minimize heat loss?
- Which is the BEST method to report the weight of the leaf?