3

The Role of Generic Skills in Measuring Academic Quality

Roger Benjamin

Introduction

The goal of this chapter is to demonstrate why measuring student learning outcomes in undergraduate education is essential to understanding the concept of academic quality, including its critical role in measuring and improving academic quality. This aim is supported by addressing five topics: (1) the link between the generic-skills concept and the human-capital approach; (2) understanding the case for standardized assessment; (3) understanding the reliability and validity of generic-skills assessment (with evidence from the Collegiate Learning Assessment and the CLA+); (4) understanding the main barriers that need to be overcome; and (5) reframing the uses of generic-skills tests (i.e., understanding the current benefits of generic-skills assessment, and their important new uses).

The five topics are designed as building blocks that fit together. This chapter focuses solely on why the measurement of student learning is essential to understanding the larger topic of academic quality.

1. The primary focus of the human-capital approach (the knowledge, experience, and skills of a nation's citizenry) today is on generic skills, which justifies linking these two concepts together to form the premise for the full argument.

Assessing Quality in Postsecondary Education: International Perspectives, edited by Harvey P. Weingarten, Martin Hicks, and Amy Kaufman. Montréal and Kingston: McGill-Queen's University Press, Queen's Policy Studies Series.

2. While classroom-based assessments of undergraduates by instructors remain central, measurement scientists argue that any assessment with stakes attached requires comparisons. In turn, this also requires standardized assessments that are known to be reliable—given to students under the same conditions and within the same time period.
3. The reliability and validity evidence of the CLA+, a measure of generic skills, appears to be a positive case in point (ETS and ACT also field tests of generic skills seen as reliable and valid).
4. The fact that third-party standardized assessments are not welcomed by many department-based faculty should be a concern. This means reliable and valid comparisons are not possible. If standardized assessments were part of a larger suite of academic disciplines that embraced the value system of science, and were relevant to understanding academic quality, including student learning, that might change the situation. Faculty might then be more positive about third-party-based research, including standardized assessment, which is governed by the principles of transparency, peer review, and the ability to replicate results.
5. If these arguments for standardized assessments and, in particular, generic-skills tests make sense, there are important uses of this test within the university. Equally, there are also new interdisciplinary and global roles for the tests.

The Link Between the Human-Capital Approach and Generic Skills

Gary Becker and his colleagues in the economics department at the University of Chicago get credit for the human-capital approach (Becker 1993). These scholars define human capital as the stock of knowledge and skills present in a nation's population. Such capital accrues through education, training, and experience. As the human-capital field matured, economists began to mine its implications for education (Kreidl, Ganzeboom, and Treiman 2014). The analysis of the returns on the amount and quality of education achieved has become an important research program.

This body of research suggests that education should focus on the knowledge, skills, and experience required in the knowledge economy and society. This means focusing on the ability to access and structure information, and applying it to solve new problems (Simon 1996). Re-

cent theories of learning reflect this change in emphasis from specific content domains to a focus on being able to find and use information. Bransford, Brown, and Cocking (2000) agree that the goal now is to help students develop the intellectual tools and learning strategies needed to acquire the knowledge to think productively. All societies must ensure that their workforce can generate value-added ideas, which can be a foundation for sustained economic growth and prosperity. These skills are seen as requisites for success in the workplace by college graduates. Therefore, this means teaching critical-thinking skills and measuring students' progress toward desired attainment levels. In today's knowledge economy, this privileges the ability to access, structure, and use information, which, in turn, places the focus on generic skills.

The Case for Standardized Assessments

Measurement scientists who work in the education assessment space have developed criteria to evaluate assessment protocols. They are particularly concerned about reliability and validity. Validity is about the extent to which the assessment measures the knowledge, skills, and abilities it was designed to measure. Reliability refers to the degree of consistency of students' (or schools') scores across different assessors, and whether the assessments are given to students under the same conditions and over the same time period. The need for standardized assessments rests on the premise that decisions with stakes attached should be seen to be reliable and valid. If the assessment is not reliable and valid, how can stakeholders rely upon the test results when making decisions with consequences? For example, faculty understandably support student portfolios (Banta and Pike 2012; Rhodes 2012; AAC&U 2005). However, most measurement scientists are skeptical of the claim that portfolios are equal to, or better than, standardized assessments because they have doubts about the reliability of portfolio-based assessments. I share their view and argue that any decision with stakes attached should use a standardized test along with, or in addition to, formative assessments.

What, then, are the major differences between standardized and formative assessments? Perhaps the central distinction between the two groups concerns the different assumption about what unit and level of analysis is appropriate for educational assessment. Adherents of formative assessment privilege the classroom and individual universities as the unit and level of analysis to focus on. They do not believe com-

parisons between and among universities are possible or, in any event, necessary because they do not believe in standardized tests, and/or they do not believe it is possible to provide interuniversity comparisons. For example, one argument that is frequently expressed is that missions of colleges and universities are so different that it makes no sense to compare them. Furthermore, it is argued that research has shown no statistical differences between institutions on measures of critical thinking, which is the educational component measured most often (Banta and Pike 2012). The second often repeated argument is that variance is much higher within institutions than between institutions, so between-institution comparison is not worth doing (Kuh 2007, 32–3). There are two responses to these assertions.

The first is that most higher-education institutions commit to improving generic skills as a fundamental part of their compact with students. This commitment is enshrined in mission and vision statements of most colleges and universities. Second, the fact that there can be at least two standard deviations among similarly situated colleges and universities, including selective colleges on the CLA+ value-added protocol, means there is a substantial canvas of similar institutions where researchers may study best practices in teaching and learning (Benjamin 2008).

Finally, it has been argued for some time that including performance assessments would encourage greater coherence between instruction, curriculum, assessment, and the complex decision-making tasks faced in life beyond the classroom.

Reliability and Validity Evidence of One Generic-Skills Measure: CLA+

Reliability

The Cronbach's alpha measures the internal consistency of a set of test items. The reliability coefficients for two forms of the CLA+ assessment (one performance task and twenty-five selected-response questions) are .87 (form A) and .85 (form B). These scores are reliable enough for making decisions about grading, admissions, placement, scholarships, etc. (Zahner and James 2015). The reliability coefficient for CLA+ has been at or above .87 in four annual testing administrations, including 2017–18 and also in two test administrations of Teco, the Italian version of CLA+.

Second, CLA+ results can be compared within and between colleges. For example, value-added models can be used to estimate the growth in learning between freshmen and senior-year students. The average ef-

fect size reflecting differences in CLA+ performance between entering freshmen and graduating seniors has been .75 over several annual test administrations. There are significant intra- and inter-variations in the effectiveness of efforts to develop generic skills. Going to college matters a good deal and where students go to college is highly significant (Benjamin 2014).

Validity

Construct validity refers to the degree to which test scores may be interpreted as indicators of the particular skills (or construct). Construct validity is often evaluated by examining the patterns of correlations between (or among) a test and similar or different tests. In a technical validity study that carried out this kind of analysis by comparing the tests of critical-thinking skills fielded nationally (e.g., ETS, ACT, and CAE), construct validity for all three tests was demonstrated (Klein et al 2009).[1]

Are Generic Skills Independent?

In a summary of a number of studies, I find that generic skills are applicable over an array of academic disciplines, and can be both assessed and improved by teaching. CLA+ is based on the belief, supported by research, that learning is highly situated and context bound. However, through practice in one subject area, learned knowledge becomes sufficiently generalized to enable students to transfer it to the realm of enhanced reasoning, problem solving, and decision making that can be demonstrated across content domains.

One additional validity question concerns the test paradigm itself. Multiple-choice assessments remain the dominant testing regime. There is a significant education reform movement underway in the United States in both the kindergarten to Grade 12 and the postsecondary sectors. First, there is a shift from the long-standing lecture format to a student-centred approach. Second, there is a change in emphasis in text material, from a primary focus on content to a focus on case studies and problem-based materials. The significant changes underway in these two dimensions of education are ahead of the progress needed in creating assessments that measure the generic skills of students. As-

1 Also see Benjamin et al. (2012). Studies of the predictive validity of CLA+ include Steedle (2012), and Zahner and James (2015).

sessments that are better able to measure how well students are learning—and how well institutions are teaching—these skills have become necessary. If the human-capital school demonstrates the importance of education, the implications of the knowledge economy and recent theories of learning place the focus on improving the generic skills of the next generation of students. These developments create an urgent need to generate and implement a testing paradigm that measures and simulates these skills. That paradigm is performance-based assessment, such as that provided by CLA+. However, one issue inhibiting the introduction of this, or any external-based assessment to US universities, is the department-based governance model that is so critical to the success of the US higher-education system.

Department-Based Barriers to Overcome

Department-based governance means professionals in each field of inquiry organize themselves in departments based on the premise that those qualified in a field of knowledge are equipped to govern themselves and, in turn, to decide which fields of inquiry within their discipline should be covered, what subjects should be taught, who should be hired and promoted, and how students should be taught and assessed. No matter how great their knowledge, skills, and/or accomplishments, outsiders are perceived to lack the shared understanding needed to contribute to these decisions in a meaningful way. Faculty are, therefore, typically not interested in whether their instructional methods produce acceptable results based on independent, third-party assessments. Their interests do not often extend to research findings that question the premise of department-based governance (Benjamin and Carroll 1996).

As a result, department-based governance has led to a two-cultures split within the academy. Too many faculty resist science-based research of higher education. Thus, there is a paucity of empirical research supported by the value system of science. Scholarship that is not based on the value system of science lacks transparency and clear peer-review standards, and does not privilege the value of replication of research results. Without systematic empirically based evidence, it will not be possible to propose, develop, and implement effective remedies to the two-cultures division.[2]

2 Ostrom (1972) argued for transparent performance metrics about outcomes and key processes in non-profit institutions that are not clearly subject to the discipline of the

Researchers in cognitive science, macroeconomics and microeconomics, educational assessment, educational technology, and data analytics—to name a few—toil in independent silos, isolated from each other. However, they share a commitment to the logic and strategy of scientific inquiry. The premise of the value system of science, peer review, transparency, and the ability to replicate results are familiar to faculty and administrators. When paired with a coherent and compelling use-inspired basic research strategy, it is possible to imagine a more integrated, interdisciplinary, scientific approach to the challenges that higher education faces.

Framing the Uses of Generic-Skills Tests

The following statement from the "Revised Scoping Paper for an AHELO Main Study" (2015) provides the challenge to which we need to respond: "In a globalizing world, governments want to have more profound knowledge about the education and skills pool at the upper end of the distribution. Economic arguments relating to productivity, innovation, competitiveness and growth, and social arguments relating to social cohesion, trust and various other social outcomes of education create a need for governments to assess the learning outcomes of their new cohorts of tertiary graduates" (OECD 2015, 8).

Current Uses of Generic-Skills Assessment: The CLA+ Case

If leaders of colleges and universities are indeed at a tipping point—simultaneously facing rising costs, declining resources, and a decline in the quality of student-learning outcomes—new decision-making tools that assist college leaders in responding to this challenge would be useful and welcome. CLA+ attempts to provide one decision-making tool for this purpose. The following practical uses of the CLA+ generic-skills assessment, in the form of reports and data analytics offered to all test takers, are designed to assist the higher-education sector improve the quality of student learning, and anchor interdisciplinary research conducted by researchers from the disciplines noted above. Because researchers in all these fields of inquiry share a commitment to the value system of science, which privileges peer review, the ability to

market. Simply putting the spotlight on performance indicators causes changes in attitudes and behaviour of the participants, in this case higher-education institutions. This is an example of what Mayo (1949, pp. 60–77) called the Hawthorne Effect.

replicate research results, and public transparency of those results, department-based faculty should be reassured that the results are not controlled or manipulated by policy makers or administrators privately.[3]

Participants in CLA+ receive test results with student, institution, region, and country-based reports with:

- Value-added results
- Certificates/badges with test results for qualifying students (students at the proficient-to-accomplished end of the distribution of generic skills) to showcase their levels of mastery to employers (CLA+ CareerConnect)
- Online results analysis tool, CLA+ Analytics, part of CLA+ Data Miner
- Online videos and interactive exercises to help students improve their generic skills
- Professional development seminars to train professors on the techniques to improve their students' performance

These applications are designed to do the following:

- Permit employers to more easily identify students of high ability who warrant interviews for high, value-added jobs
- Permit graduating seniors to distinguish the level of generic skills they have attained from that of other students
- Permit universities to identify departments and programs that contribute the most to the growth and attainment of generic skills
- Permit ministries of education to identify universities that produce the most value-added growth and/or the highest level of attainment of generic skills
- Permit graduating secondary-school students and their parents to know the level of value-added growth their potential choice of university produces
- Provide the basis for research to understand the impact of resources on the value-added growth, and the highest attainment levels universities provide their students

3 Because of the importance of human capital, local, state, and national public leaders are likely to increase their interest in holding institutions accountable for the student-learning results they achieve. Of course, there is considerable debate about whether assessment results should be used for accountability purposes versus improvement. If the faculty and institutions do not have a credible voice in this debate, the department-based barriers will likely continue (Benjamin and Klein 2006, 19).

- Permit researchers to evaluate which academic disciplines contribute the most to student learning-outcomes success
- Provide diagnostic information about the generic skills of entering students, and the retention and graduation rates of students from various demographic backgrounds, in particular under-represented groups
- Provide the basis for cross-national comparisons of similarly situated students, universities, and national systems

Potential Uses and Interdisciplinary Roles of the Generic-Skills Measure

We know from the history of the development of empirical, evidence-based research in agricultural- and health-policy research that there can be an evolution from specific evidence-based results to major new research programs.[4] The question is whether the time is right for a similar transition to occur in higher-education policy research. Two immediate possibilities present themselves.

The decline of productivity growth is an important puzzle to solve. Productivity, defined as the output per hour worked, adjusts for the contribution of capital and materials, and provides a measure of the pace of technological change by tracking productivity growth year over year. From 1948 to 1973 the annual average growth in US productivity declined from 2.5 percent to a stabilized rate of about 1.01 percent in the past decade.[5]

Why does this productivity slowdown, which appears to be similar for other advanced economies, matter? Little or no productivity growth for one or two years is not especially noteworthy. However, annualized year-over-year productivity growth is essential to a national economy and society. Lower economic growth accelerates the rise in social and economic inequality, which appears to be a growing problem in advanced economies today (Piketty 2014). To examine the possible explanations of the slowdown in productivity growth, two subfields of eco-

4 Hayami and Ruttan (1987) showed that as a result of the progress in scientific research on agriculture, agricultural economists were able to measure agricultural productivity growth. Following the Flexner Report (Flexner 1910), leaders of medical education decided to change medicine from a clinical to a science-based field.

5 The US permits the most extensive post–World War II historical period to measure productivity growth in advanced economies. Other advanced economies in Western Europe and Asia did not fully recover economically until the mid-1950s. However, OECD-based measures of productivity growth for these countries now presents similar trends to the US.

nomics appear most relevant for our purposes. They are both focused on the service sector.

The first subfield, returns to tertiary education, shows a net decline in the benefits of a BA degree. Kreidl et al. (2014),[6] reviewing occupational trends in education over labour force entry cohorts in forty-two nations over most of the twentieth century and the beginning of the twenty-first century, find that occupational returns to education have been steadily decreasing (Abel and Deitz 2014) .

The second possible explanation is an error in measuring the productivity of the service sector. If we could more accurately measure the service sector (e.g., health, social, and education), productivity growth would look much better. The OECD defines the service economy as "…a diverse group of economic activities, not directly associated with manufacture of goods, mining or agriculture. They typically involve the provision of human value added in the form of labor, advice, managerial skill, entertainment, training, intermediation and the like" (OECD 2000, 7).

Powell and Snellman state that the rise of the service economy involves "…a shift in focus from the principal production and consumption of physical goods to today's principal focus on the production and consumption of services, in particular … knowledge intensive activities" (Powell and Snellman 2004, 199). Nordhaus writes, "… the structural shift from high to low productivity growth sectors, from manufacturing to services)" is the most important contributor to slowing productivity growth, which requires further careful examination (Nordhaus 2016, 3).

The service sector now accounts for over 80 percent of the GDP of the United States and more than 70 percent in OECD member countries overall. Baily and Montalbano argue "…if productivity growth were more accurately measured, particularly in health, education and other services, the growth rate would look better than [it does] currently" (Baily and Montalbano 2016; see also Sprague 2017). This a reasonable position, which leads to the need to re-conceptualize the way we measure GDP. A generic-skills measure could be used to track the productivity growth year over year.

6 Other explanations of declining productivity growth include a) mismanagement of new information-technology advances, b) impact of artificial intelligence on investment, c) decline in investment, and d) slowdown in global trade due to national populism.

The OECD Findings Regarding the Skills Mismatch Problem

A generic-skills measure could also illuminate trends in the current debate over a skills mismatch. Is there evidence of a skill mismatch? The OECD finds that more than 40 percent of European workers feel their skill levels do not correspond to those required to do their job. In parallel, many employers report that they face recruitment problems due to skill shortages. "The costs of persistent mismatches and shortages are substantial. For instance, skill shortages can constrain the ability of firms to innovate and adopt new technologies while skill mismatches reduce labour productivity due to the misallocation of workers to jobs. Individuals are also affected as skills mismatch can bring about a higher risk of unemployment, lower wages, lower job satisfaction and poorer career prospects" (OECD 2016a, 7; see also Bol 2015).

In the language of economics, this description is labelled as a maldistribution of human capital at the national level. This statement also describes the impact of a skills mismatch on individuals, which translates as a problem of unequal opportunities for individuals. Since the equality of individual opportunity is a fundamental tenet of liberal democracies, this is also a major policy issue that most, if not all, countries must be concerned with (Benjamin 2016).

The OECD divides skills valued in every job, occupation, and sector into (a) cognitive and non-cognitive skills, and (b) job-specific skills such as technical knowledge associated with a job or occupation (i.e., practical competencies). Key cognitive skills are critical thinking, problem solving, qualitative and quantitative reasoning, and writing mechanics and persuasiveness. Non-cognitive skills refer to persistence, teamwork, entrepreneurial ability, and moral or ethical reasoning ability. While it is recognized that non-cognitive skills are important, it is also recognized that they are not yet measured in a reliable and valid way. Therefore, the immediate focus is on cognitive skills (OECD 2016b).

Experts at the OECD offer two alternatives for moving the cognitive generic-skills agenda forward:

1. Create a comprehensive qualifications framework that would cover all jobs. The challenge here is the need to constantly update changes in each occupation, and make sure that the occupations compared across countries are defined in the same way. The OECD experts recognize this alternative is complex, requires

intensive, large-scale labour, and is cumbersome.[7]

2. The second alternative is to "focus on developing generic or general skills in the education and training system so that workers can more readily adapt to different working environments and allow them to learn field- or job-specific skills in the job" (Montt 2015, 40).

This second alternative is aligned with, and supports, the rationale for focusing on generic skills. The problem, then, is whether we can advance our understanding of both the productivity growth issue and the skills mismatch problem by exploring the potential linkages between them.

The first task is to see whether declining productivity growth supports a closer look at the skills mismatch problem. Second, is there a compelling rationale to introduce a generic-skills assessment to assist efforts to improve measurement of productivity, and aid research that attempts to explain economic growth, the skills mismatch problem, and the increasing inequality of individual opportunity in the workplace? If the generic-skills measurement supports the argument that there is a connection between productivity growth and the skills mismatch issue, this would also demonstrate an absence of a level playing field for many students at the high end of the distribution of generic skills.

Conclusion

The logic of the human-capital approach in today's knowledge economy privileges critical-thinking skills. The focus on the importance of generic skills in today's knowledge economy, which favours the service sector, provides the fundamental rationale for exploration of a new interdisciplinary role for measuring generic skills. First, a generic-skills measure may be used as additional information to track the growth of productivity year over year. Second, the measure may illuminate the trends in the skills mismatch space.

In addition to the current and potential uses of generic-skills measures such as CLA+, quality-assurance agencies are interested in student learning-outcome measures that clearly demonstrate whether the quality of student learning is improving. Because of the problematic history of efforts to develop accountability systems that compel univer-

7 A precedent for this in the United States is ACT's Work Keys, a comprehensive map of thousands of vocations.

sities and colleges to demonstrate the level of student learning attained by their graduating seniors, at least in the United States, it may be preferable to focus on measuring the generic skills of a representative sample of graduating seniors within a state, region, or country. This method is more likely to obtain a more statistically accurate picture of what the level of student learning for graduating seniors has reached on an annual basis. The National Assessment of Educational Progress (NAEP) represents one model for such an approach.

The premise of this approach differs from that of the current uses of CLA+ noted above. Instead of using the institution of higher education as the unit of analysis, the student is the focus. The advantage of this approach is that it bypasses the institution altogether. See "An Overview of the NAEP" for a description of this approach, which has become the gold standard for the assessment of the quality of elementary and secondary education in the United States (NAEP 2017). However, since individual universities would not be initially involved in this approach, quality-assurance groups would need to engage the colleges and universities to review the generic-skills assessment results for their state or region to understand the diagnostic results of the assessment, and provide compelling evidence about the level of student learning skills reached by graduating seniors at their institution. The quality-assurance agency might then request that the institution propose changes to its curriculum and pedagogy to improve its student learning results to meet the state or regional requirements set by the quality-assurance agency. In such a model, the quality-assurance agency might recommend the best practices to improve writing, quantitative and qualitative reasoning, and problem solving, (which are core sub-components of generic skills) to levels negotiated between the institutional leaders and the representatives of the quality-assurance agency. The quality-assurance agency might also provide various positive incentives or negative sanctions to encourage the desired improvements.

Finally, the only way to find out whether the generic-skills measure proves to be a useful additional indicator of the productivity growth and skills mismatch issues is to try it out using a number of pilot programs. The proposition put forward here is that a collegiate measure of educational progress, C-NAEP, could serve as the generic-skills measure used to evaluate changes in the skills mismatch problem and the production growth issue.

References

AAC&U (Association of American Colleges & Universities). 2005. *Liberal Education Outcomes: A Preliminary Report on Student Achievement in College*. Washington, DC: Association of American Colleges and Universities.

Abel, Jaison R., and Richard Deitz. 2014. *Do the Benefits of College Still Outweigh the Cost?* Current Issues in Economics and Finance, Vol 20, No. 3. New York: Federal Reserve Bank of New York. https://www.newyorkfed.org/medialibrary/media/research/current_issues/ci20-3.pdf

Baily, Martin, and Nicholas Montalbano. 2016. *Why Is U.S. Productivity Growth So Slow? Possible Explanations and Policy Responses*. Hutchins Center Working Paper No. 22. Washington, DC: The Brookings Institution.

Banta, Trudy W., and Gary Pike. 2012. "Making the Case Against—One More Time." *The Seven Red Herrings about Standardized Assessments in Higher Education.* National Institute of Learning Outcomes. Occasional Paper 15. September: 24–30. http://cae.org/images/uploads/pdf/The_Seven_Red_Herrings_About_Standardized_Assessment_in_Higher_Education.pdf

Becker, Gary. 1993. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education,* 2nd ed. Chicago: University of Chicago Press.

Benjamin, Roger. 2008. "The Case for Comparative Institutional Assessment of Higher-Order Thinking Skills." *Change: The Magazine of Higher Learning* 40 (6): 15–21.

———. 2014. "Two Questions about Critical-Thinking Tests in Higher Education." *Change: The Magazine of Higher Learning* 46 (2): 32–9.

———. 2016. *Leveling the Playing Field from College to Career*. New York: Council for Aid to Education. http://cae.org/images/uploads/pdf/Leveling_the_Playing_Field_From_College_To_Career.pdf

Benjamin, Roger, and Stephen J. Carroll. 1996. "Impediments and Imperatives in Restructuring Higher Education." *Educational Administration Quarterly* 32 (1): 705–19.

Benjamin, Roger, and Stephen Klein. 2006. *Assessment versus Accountability: Notes for Reconciliation.* UNESCO Occasional Paper No. 2. Paris: UNESCO.

Benjamin, Roger, Stephen Klein, Jeffrey Steedle, Doris Zahner, Scott Elliot, and Julie Patterson. 2012. *The Case for Generic Skills in the United States and International Settings*. New York: Council for Aid to Edu-

cation. http://cae.org/images/uploads/pdf/The_Case_for_Generic_Skills_and_Performance_Assessment.pdf

Bol, Thijs. 2015. "Has Education Become More Positional? Educational Expansion and Labour Market Outcomes, 1985–2007." *Acta Sociologica*, 58 (2): 105–20.

Bransford, John, Anthony Brown, and Robert Cocking. 2000. *How People Learn: Brain, Mind, Experience, and School.* Washington, DC: National Academy Press.

Flexner, Abraham. 1910. *Medical Education in the United States and Canada: A Report to the Carnegie Foundation for the Advancement of Teaching.* New York: The Carnegie Foundation for the Advancement of Teaching, Bulletin No. 4.

Hayami, Yujiro, and Vernon W. Ruttan. 1987. *Agricultural Development: An International Perspective.* 2nd ed. Baltimore: Johns Hopkins University Press.

Klein, Stephen, Lydia Liu, James Sconing, Roger Bolus, Brent Bridgeman, Brent, Heather Kugelmass, Alexander Nemeth, Steven Robbins, and Jeffrey Steedle. 2009. *Test Validity Study Report*. Fund for the Improvement of Postsecondary Education, U.S. Department of Education. https://cp-files.s3.amazonaws.com/26/TVSReport_Final.pdf

Kreidl, Martin, Harry B. G. Ganzeboom, and Donald J. Treiman. 2014. *How Did Occupational Returns To Education Change Over Time?* Los Angeles: California Center for Population Research, University of California, Los Angeles. http://citeseerx.ist.psu.edu/viewdoc/download?-doi=10.1.1.661.8511&rep=rep1&type=pdf

Kuh, George. 2007 "Risky Business: Promises and Pitfalls of Institutional Transparency." *Change* 29 (5): 30–5.

Mayo, Elton. 1949. *Hawthorne and the Western Electric Company. The Social Problems of Industrial Civilization.* London: Routledge.

Montt, Guillermo. 2015. *The Causes and Consequences of Field-of-Study Mismatch: An Analysis Using PIAAC*. OECD Social, Employment and Migration Working Papers, No. 167. Paris: OECD Publishing.

NAEP (National Assessment of Educational Progress). 2017. *An Overview of NAEP*. Washington, DC: Center for Education Statistics, United States Department of Education. https://nces.ed.gov/nationsreportcard/subject/_commonobjects/pdf/2013455.pdf

Nordhaus, William D. 2016. "Why Growth Will Fall." *The New York Review of Books*. 18 August: 1–3. http://www.nybooks.com/articles/2016/08/18/why-economic-growth-will-fall/

Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Translated by

Arthur Goldhammer. Cambridge: Harvard University Press.
Powell, Walter W., and Kaisa Snellman. 2004. "The Knowledge Economy." *Annual Review of Sociology* 30 (1): 199–220.
OECD (Organisation for Economic Co-operation and Development). 2000. *The Service Economy*. Business and Industry Policy Forum Series. Paris: OECD Publishing. https://www.oecd.org/sti/ind/2090561.pdf
———. 2015. *Revised Scoping Paper For An AHELO Main Study*. Directorate for Education and Skills. Paris: OECD Publishing.
———. 2016a. *Getting Skills Right: Assessing and Anticipating Changing Skill Needs*. Paris: OECD Publishing. http://www.keepeek.com/Digital-Asset-Management/oecd/employment/getting-skills-right-assessing-and-anticipating-changing-skill-needs_9789264252073-en#.WeTQAFtSxhE
———. 2016b. *Skills Matter: Further Results from the Survey of Adult Skills*. Paris: OECD Publishing. http://www.keepeek.com/Digital-Asset-Management/oecd/education/skills-matter_9789264258051-en#.WeTQQFtSxhE
Ostrom, Vincent. 1972. "Polycentricity." Paper presented at the annual meeting of the American Political Science Association, Washington, DC, September.
Rhodes, Terrel. 2012. "Getting Serious About Assessing Authentic Student Learning." *The Seven Red Herrings about Standardized Assessments in Higher Education*. National Institute of Learning Outcomes. Occasional Paper 15. September: 19–23. http://www.learningoutcomesassessment.org/documents/HerringPaperFINAL1.pdf
Simon, Herbert. 1996. *The Sciences of the Artificial*. Boston: MIT Press.
Sprague, Shawn. 2017. "Below Trend: The U.S. Productivity Slowdown Since the Great Recession." *Beyond the Numbers: Productivity*, Vol. 6, No. 2. Bureau of Labor Statistics. January. https://www.bls.gov/opub/btn/volume-6/below-trend-the-us-productivity-slowdown-since-the-great-recession.htm
Steedle, Jeffrey T. 2012. "Selecting Value-Added Models for Postsecondary Institutional Assessment." *Assessment & Evaluation in Higher Education*, 37 (6): 637–52.
Zahner, Doris, and Jessalynn K. James. 2015. *Predictive Validity of a Critical Thinking Assessment for Post-College Outcomes*. New York: Council for Aid to Education. 30 September.