

Measures for a College and Career Indicator: Final Report



Prepared for the California Department of Education
by the Educational Policy Improvement Center

David T. Conley

with

Paul Beach

Michael Thier

Sarah Collins Lench

Kristine L. Chadwick



November 2014



Table of Contents

Acknowledgments	iv
Executive Summary	v
I. Introduction	1
Project Overview	1
Project History	2
Senate Bill 1458	3
The Academic Performance Index	3
The College and Career Indicator	4
Methodology	5
II. College and Career Preparedness Measures	8
College Admission Exams	8
The SAT	8
The ACT	9
Advanced Coursework	9
Advanced Placement and International Baccalaureate	9
Innovative Measures	10
Metacognitive Assessments.....	10
Performance Assessments	10
California State Seal of Biliteracy.....	11
Course-Taking Behavior	11
The a–g Subject Requirements.....	12
Career and Technical Education Course Pathway	13
Integrated Course Pathway.....	13
Career Preparedness Assessments	13
ACT WorkKeys.....	13
Armed Services Vocational Aptitude Battery	14
National Occupational Competency Testing Institute	14
Industry Certifications.....	15
Comparison of Measures	15
Across Measures	15
Within Measures	17
III. Constructing a College and Career Indicator	24

Technical Considerations	24
Setting Benchmarks	24
Complementary (or Compensatory), Matrix, and Conjunctive Methods	25
Aggregating Student-Level Indicators to Measure School Effectiveness.....	27
Equity Considerations	28
Biases of Race, Ethnicity, and Socioeconomic Status	28
Access and Opportunity to Learn	29
Pathway Choices Between College and Career	29
Additional Possible Contributors to the College and Career Indicator	32
Dual/Concurrent Enrollment.....	33
Culminating Projects	33
Languages Other Than English Coursework	33
Lab Science Coursework	34
College Remediation Rate	34
IV. Recommended College and Career Indicator.....	35
V. Considering a Multiple-Measure System	37
Recognition Approaches	40
VI. Creating a Coherent Accountability System	40
Interacting Accountability Components	40
Local Control Funding Formula, Local Control Accountability Plans, and the California Collaborative for Excellence in Education	41
Multiple Levels of Accountability and the Challenge of Compliance Versus Goal- Oriented Mindsets.....	42
A State/Local Model.....	43
VII. Conclusion.....	45
References	47

Acknowledgments

EPIC's researchers were ably assisted by two members of the Public Schools Accountability Act Advisory Committee, who offered invaluable comments and insights that helped shape the final report. Committee cochair Kenn Young and member Ed Haertl provided their views on earlier versions of this report that were helpful in refining its focus and content.

The staff at the California Department of Education offered steady and consistent guidance, leadership, and attention to detail. EPIC is grateful for the support and flexibility of Keric Ashley, director of the Analysis, Measurement, and Accountability Report Division, and Jenny Singh, who headed the project team, which also included Ryan Lam.

The authors wish to acknowledge the important contributions of EPIC staff members, most notably Barbara Hewick in her role as copyeditor for this report and the six white papers that preceded it, and Dr. Matt Coleman for his guidance and insights throughout this project.

Executive Summary

In 2012, California Senate Bill 1458 added a measure of college and career preparedness to the Academic Performance Index (API). The Public Schools Accountability Act Advisory Committee was charged with making recommendations to the State Superintendent of Public Instruction and the State Board of Education regarding measures that could serve as indicators of college and career preparedness at the high school level.

Nature of Evaluation

The Educational Policy Improvement Center (EPIC) was commissioned to evaluate potential measures identified by the Committee. To do so, EPIC employed a criterion-based evaluation framework that focused on the technical quality, stakeholder relevance, and system utility of each potential measure as represented in Table 1.

Table 1. Evaluative Criteria for Potential College and Career Preparedness Measures

Dimension	Criterion
Technical quality	has a <i>research base</i> demonstrating a relationship with postsecondary success
	allows for <i>fair comparisons</i>
	has <i>stability</i>
Stakeholder relevance	has <i>value for students</i>
	is <i>publicly understandable</i>
	has <i>instructional sensitivity</i>
	emphasizes <i>student performance</i> , not educational processes
System utility	<i>minimizes burden</i>
	provides as much <i>student coverage</i> as possible
	recognizes <i>various postsecondary pathways</i>

The Measures

Five potential categories of measures were evaluated and reported in a series of white papers (and a sixth white paper examined multiple measures):

1. College admission exams
2. Advanced coursework
3. Innovative measures
4. Course-taking behavior
5. Career preparedness assessments

The college admission exams category comprises the SAT and ACT. The advanced coursework category includes the Advanced Placement program and the International Baccalaureate Diploma Programme. Innovative measures consist of metacognitive assessments, performance assessments, and the California State Seal of Biliteracy. The course-taking behavior category includes the University of California’s a–g subject requirements, career technical education course pathways, and integrated course pathways. The career preparedness category consists of ACT’s WorkKeys, assessments from the National Occupational Competency Testing Institute, the Armed Services Vocational Aptitude Battery, and industry certifications.

Findings

The evaluation of each category of measures resulted in a rating of strong, medium, or weak on each of the ten criteria, as shown in Table 2.

Table 2. Evaluation of Measures of College and Career Preparedness

	A1: Research base	A2: Fair comparisons	A3: Stability	B1: Value to students	B2: Publicly understandable	B3: Instructional sensitivity	B4: Student performance	C1: Minimizes burden	C2: Student coverage	C3: Various pathways
S = Strong, M = Moderate, W = Weak										
College admission exams	S	M	S	M	S	M	S	S	M	M
SAT	S	M	S	M	S	M	S	S	M	M
ACT	S	M	S	M	S	M	S	S	M	M
Advanced coursework	S	M	S	S	M	S	S	M	M	M
Advanced Placement	S	M	S	S	M	S	S	M	M	M
International Baccalaureate	M	M	S	S	W	S	S	W	W	M
Innovative Measures	M	M	M	S	M	S	S	M	S	S
Metacognitive assessment	M	M	W	S	M	M	M	M	S	S
Performance assessment	M	M	M	S	M	S	S	M	S	S
California State Seal of Biliteracy	W	W	M	M	M	S	S	M	W	S
Course-taking behavior	S	M	M	S	S	S	S	S	M	M
a–g subject requirements	S	M	M	S	S	S	S	S	M	M
CTE course pathway	M	M	M	M	S	S	S	S	M	M
Integrated course pathway	S	M	M	S	M	S	S	M	W	S
Career preparedness assessments	M	M	M	M	M	S	S	S	M	M
ACT’s WorkKeys	M	M	M	M	M	M	S	S	M	M
ASVAB	W	M	S	W	W	W	S	S	M	M
NOCTI	M	M	M	M	M	S	S	M	M	M
Industry certifications	W	W	W	S	S	S	S	W	M	M

Technical Considerations

Potential technical considerations for a college and career indicator include setting the benchmark levels associated with college and career preparedness, combining multiple measures into one indicator, and aggregating student level measures to generate a summative high school-level rating.

The challenge with setting benchmark levels for tests is that they can end up being used as cut score judgments about students even though their purpose is to measure school effectiveness. Benchmark levels also tend to underrepresent the complexity of college and career preparedness.

Policymakers can use combinations of measures in several ways: 1) they can allow strengths in one area to compensate for weaknesses in others, known as a compensatory or complementary approach; 2) they can create a matrix of ratings or scores that are applied to a series of measures; or 3) they can adopt an approach where a school needs to reach a designated level on all measures, which is a conjunctive model.

In a complementary model, student performance counts only for the measure on which each student performs best. Complementary models can be compensatory in nature, which means strong scores in one area make up for weaker scores on other measures within specified ranges, or complementary in the sense that only the best performance is incorporated, regardless of how a student does on other measures. Technically, the matrix model does not combine measures. Instead, the matrix model calculates scores for each individual measure, which allows for more nuances than a single API score. Finally, the conjunctive approach requires schools to meet or exceed certain thresholds on all measures.

Additional Possible Indicators

Other measures beyond those identified by the Public Schools Accountability Act Advisory Committee could conceivably contribute information to a college and career indicator. These measures include dual/concurrent enrollment, culminating projects, coursework in languages other than English, lab science coursework, and college remediation rate. These measures all have the potential to make a distinctive contribution to understanding how well schools are preparing students for college and careers.

Recommended College and Career Indicator

The EPIC evaluation leads to the recommendation that a measure of course-taking behavior would be the single best indicator that meets the evaluative criteria used and also has the greatest probability of leading to improvements in college and career preparedness statewide.

When combined with the grades students get in courses, course-taking behavior is the best single predictor of college success. Its advantages include a well-developed research base, relative stability over time, understanding of both educators and noneducators, the ability to implement with little additional burden on schools, and the potential for all students to earn points for their school.

School scores for student course-taking behavior could potentially be weighted to take into account the nature of the students in the school. Doing so would give more points to schools whose students historically have not taken courses to prepare themselves for postsecondary education but increase the number of those types of courses they take.

A Multiple-Measure System

One conclusion reached by the EPIC researchers' evaluations of the measures considered is that all of them have potential value in certain situations, but all have limitations when applied to all students in California in a uniform fashion. This observation suggests that an indicator that incorporates multiple measures could be a more valid representation of college and career preparedness statewide than a single measure.

Several states have accountability systems that incorporate multiple measures. The challenge is to avoid excessive complexity while still including the most important measures. In a multiple-measure system, schools can receive points for student performance in more than one area, which validates a wider variety of pathways to postsecondary preparedness and a range of programs to meet their needs.

Creating a Coherent System

The API does not exist in a vacuum; quite the contrary. In fact, California schools have long attempted to meet state and federal accountability requirements that were similar to but not the same as California's own standards. With the recent introduction of an additional level of accountability at the district level in the form of Local Control Accountability Plans, educators will be challenged to manage a process that could conceivably send conflicting messages but also could be more relevant and valuable locally. A coherent system of accountability is necessary to focus educator efforts.

A state/local partnership model is one way to create more coherence. In this approach the state establishes a set of core measures that are consistently applied to all schools, and local schools then add measures that best reflect the quality of their programs and areas where they want to improve. The state measures foundational skills such as reading and mathematics and a few other key indicators, such as attendance and graduation rates. Local measures are then selected to address local programs that demonstrate school effectiveness for local student populations and address other state priorities.

Conclusion

Holding schools accountable for student performance based solely on educational outputs has proven to be challenging and nowhere near as effective as policymakers had hoped it would be. Accountability in the future will likely be more of a partnership between the state and local schools and will include more dimensions and measures than a single test in reading and mathematics. The college and career indicator that is being added to the API is a small step in that direction, but much more work remains to create an accountability system that acknowledges the full range of factors necessary to achieve sustained improvement of educational practice across all of California's diverse public high schools.

I. Introduction

By the 2015–16 academic year, California’s K–12 educational system will have new academic standards, new state standardized tests, new school funding models, and new accountability measures. This transformation began in 2010 when the state joined 44 others and the District of Columbia in adopting the Common Core State Standards in English language arts and mathematics. California replaced its previous statewide assessment system, the Standardized Testing and Reporting System, when it joined the Smarter Balanced Assessment Consortium in 2013. That same year, California introduced the Local Control Funding Formula, replacing most state categorical funding streams with base, supplemental, and concentration grants. This new system allocates considerable financial control to local school districts, which are required to submit Local Control Accountability Plans to their county office of education. Local Control Accountability Plans are intended to ensure districts direct state funds toward achieving state goals. Finally, California’s high school rating and ranking system, the Academic Performance Index (API), is currently under revision.

Since its adoption in 1999, the API has measured the quality of high schools based on students’ standardized test scores in English language arts, mathematics, history/social science, and science. In 2012, Senate Bill 1458 shifted the API for high schools from an exclusive reliance on standardized test scores to broader measures, including college and career preparedness. In accordance with Senate Bill 1458, standardized test scores can account for no more than 60% of the API by 2015–16. The remaining 40% will likely consist of high schools’ graduation rates and an indicator composed of measures of college and career preparedness. This new indicator is referred to here as the *college and career indicator*.

Project Overview

This report summarizes findings from a series of six white papers that examined (a) potential measures of college and career preparedness and (b) the technical aspects related to constructing an indicator employing multiple measures of college and career preparedness. The first section of this report describes an overview of this project’s history and background, and details the data and methods used to generate the six white papers that preceded this report. The second section describes the specific measures of college and career preparedness evaluated during this project, including their strengths, weaknesses, and trade-offs. The third section focuses on the key considerations related to constructing a multiple-measure college and career indicator, and presents a model that is informed by findings from this project and practices in other states. The fourth section presents a recommended college and career indicator. This is followed by approaches to creating a multiple-measure system. The final section discusses the role of the revised API in California’s reformed accountability system.

Project History

As a result of Senate Bill 1458, the State Superintendent of Public Instruction will consider input from six regional meetings, a webcast, an online survey of 1,768 stakeholders, and suggestions from the Public School Accountability Act Advisory Committee and Technical Design Group. In totality, these entities will provide a recommendation to the State Superintendent of Public Instruction who will then recommend to the State Board of Education the measure or measures to include in the new college and career indicator. To further support this decision-making process, the California Department of Education contracted the Educational Policy Improvement Center (EPIC) to provide research on potential measures of college and career preparedness as well as the technical aspects related to constructing a multiple-measure college and career indicator. EPIC researchers evaluated five different categories of measures with potential to gauge college and career preparedness.

The first five of the six white papers reviewed the literature and then applied a framework to evaluate the technical quality, stakeholder relevance, and system utility of a series of potential measures identified by the Public School Accountability Act Advisory Committee. EPIC researchers presented findings from these five white papers at three Public School Accountability Act Advisory Committee meetings between April and August of 2014 (see Table 3). The topic for the sixth paper (presented on August 5, 2014) was not an additional measure but a consideration of a multiple-measure system. The paper examined how states had combined measures as well as the pros and cons of multiple-measure systems. The Public School Accountability Act Advisory Committee took into consideration the results of the six white papers, this report, as well as advice from the Technical Design Group in making its recommendations for a college and career indicator that could be added to the API.

Table 3. Categories, Individual Measures, and Presentation Dates

Category of measures	Individual measures	Date presented
College admission exams	<ul style="list-style-type: none"> • SAT • ACT 	April 4, 2014
Advanced coursework	<ul style="list-style-type: none"> • Advanced Placement • International Baccalaureate 	
Innovative measures	<ul style="list-style-type: none"> • Metacognitive assessments • Performance assessments • California State Seal of Biliteracy 	June 17, 2014
Course-taking behavior	<ul style="list-style-type: none"> • a–g subject requirements • Career technical education course pathways • Integrated course pathways 	
Career preparedness assessments	<ul style="list-style-type: none"> • ACT’s WorkKeys • Armed Services Vocational Aptitude Battery • National Occupational Competency Testing Institute • Industry certification assessments 	August 5, 2014

Senate Bill 1458

In September of 2012, California Governor Jerry Brown signed into law Senate Bill 1458, which fundamentally altered the structure of the API by specifying that results of achievement tests shall constitute no more than 60 percent of the value of the API for high schools and by authorizing the State Superintendent of Public Instruction, with approval of the state board, to incorporate into the high school API valid, reliable, and stable measures of pupil preparedness for postsecondary education and career. In addition to changes discussed previously, the standardized test portion of the new API will deepen its emphasis on science and history/social science scores as the state replaces the Standardized Testing and Reporting system with the Smarter Balanced Assessment System.

The Academic Performance Index

Signed into law in 1999, the Public School Accountability Act established the Academic Performance Index (API) to monitor the academic achievement of all of California's K–12 schools and local educational agencies. The API provides a cross-sectional snapshot of school performance from one year to the next. Although the API is a measure of schools, the California Department of Education also calculates district-level APIs to satisfy requirements of the No Child Left Behind Act of 2001. All primary and secondary public schools in California receive an API. This project, however, concerns itself exclusively with high school performance. As a result, this report uses school and high school interchangeably.

Currently, each school receives an API score that ranges from 200 to 1000 points. An API score of 800 is the target for all schools. Currently, the California Department of Education calculates API scores by converting student performance on state assessments into points. The California Department of Education also disaggregates API scores for all students and numerically significant subgroups.¹

California has used API scores to compare schools one year to the next, defining improvement as the difference between schools' scores from the previous year (Base) to the subsequent year (Growth). Schools scoring less than 800 receive improvement targets from the state for the following year. The difference between Base and Growth scores determines whether schools meet those targets. The California Department of Education calculates targets as 5% of the difference between the Base and 800, with a minimum increase of five points.

API scores satisfy both state and federal accountability reporting requirements and are also used to reward and sanction schools for academic performance. Schools that meet certain API participation and growth criteria are eligible to become California Distinguished Schools or Title I Academic Achievement Awards Schools.

¹ Subgroups of students include: African American (not of Hispanic origin), American Indian or Alaska Native, Asian, English learners, Filipino, Hispanic or Latino, Pacific Islander, Socioeconomically disadvantaged, Students with disabilities, White (not of Hispanic origin), Two or More Races.

The College and Career Indicator

The college and career indicator will be a single indicator that allows multiple ways for students to contribute to their school’s API. The units of measurement for the college and career indicator are all students in the four-year graduating cohort (e.g., 12th graders on pace to graduate). However, students can contribute to the college and career indicator at any time during their high school careers. For example, students who take the SAT in 11th grade and score high enough to earn API points for their schools would have their points counted when they are 12th graders.

Measures selected for inclusion in the college and career indicator may have up to five API achievement levels. These achievement levels will not necessarily be the same for each measure included in the college and career indicator. Figure 1 presents the college and career indicator working model that was created by the Technical Design Group and the Public School Accountability Act Advisory Committee. This model is subject to change as research findings and data analyses provide greater insight into how best to determine school success for accountability purposes.

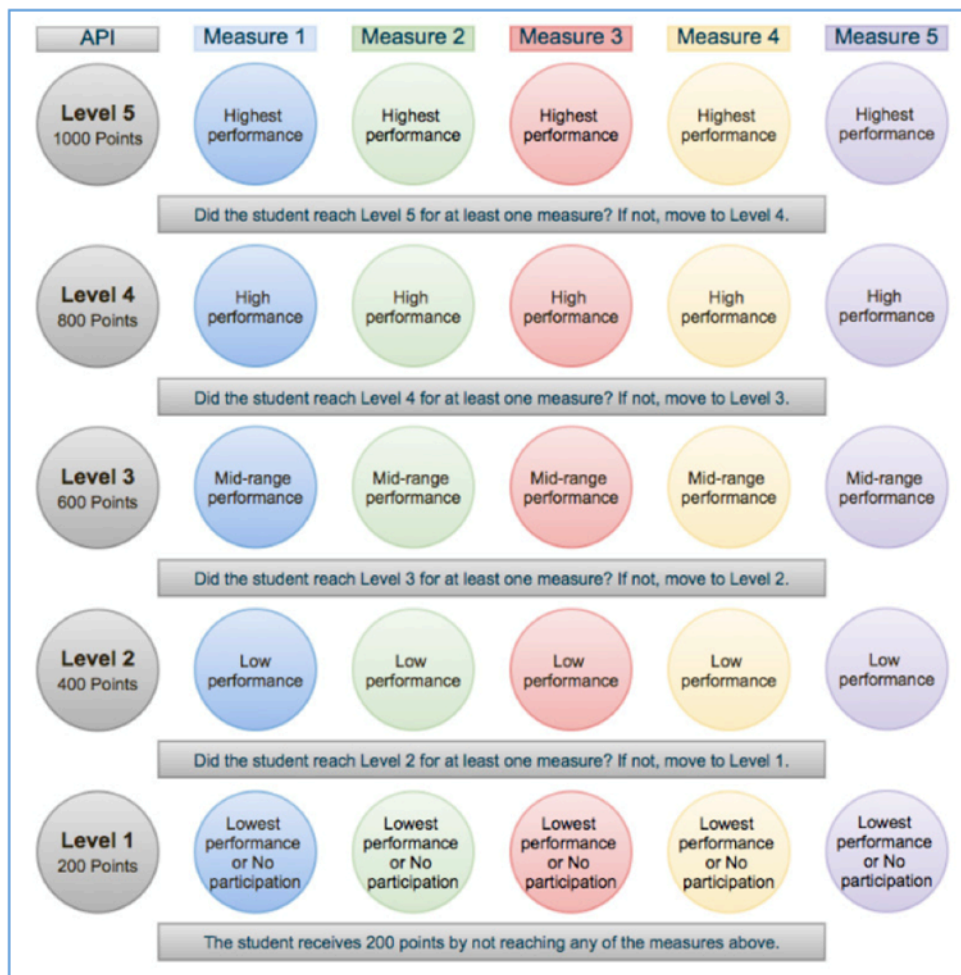


Figure 1. College and career indicator model.

Students will accumulate points toward a single measure within the college and career indicator based on the highest achievement level achieved from among all measures for which the student received a level score. For example, if a student who participated in measure 1, 2, and 3 and reached level 3 on measure 1, level 4 on measure 2, and level 5 on measure 3, only the measure 3 level would be counted, and the student would earn maximum API points for their school. Note again that this scale is for example purposes only and is expected to change.

Methodology

In collaboration with the Public School Accountability Act Advisory Committee, EPIC researchers developed a criterion-based framework that was applied to each measure that was examined for potential inclusion in the API. The criteria comprised within this framework reflect the Advisory Committee's API guiding principles and additional criteria specific to a college and career indicator. The framework is organized into three dimensions: technical quality, stakeholder relevance, and system utility. Technical quality consists of three criteria: research base, fair comparisons, and stability. Stakeholder relevance includes value to students, public understanding, instructional sensitivity, and student performance. System utility is made up of burden, student coverage, and multiple pathways. Table 4 presents the framework with explanations of each criterion.

The framework's design acknowledges that satisfaction of the above criteria is not a simple binary decision. Analyses may sometimes find criteria that are strong in one area and weak in several others, but that strength in one area offsets weaknesses in the others. To help accommodate this complexity, EPIC researchers applied the criteria both to individual measures (e.g., Advanced Placement and International Baccalaureate) as well as to the measure as a category (e.g., advanced coursework). Evaluating both individual measures and categories shed light onto how issues that exist with an individual measure may be considered differently in the context of a category.

At this stage in the process, EPIC researchers did not make recommendations but simply provided Public School Accountability Act Advisory Committee members with information regarding strengths, weaknesses, and trade-offs associated with each measure that was reviewed. This information was presented in the form of a rating for each category of measures. Each category of measures received a rating for each of the ten criteria described in Table 4 on a three-point scale: strong, moderate, or weak. EPIC researchers did not attempt to rate measures based on a summative score. Instead, the criterion-based evaluation served as a framework for a consideration of each category of measures in a more holistic fashion. The six reports include the discussions of each measure.

Table 4. Evaluative Framework by Dimension, Criterion, and Definition

Dimension	Criterion	Definition
Technical quality	Research base	<p>Has a research base that can forecast students' postsecondary success.</p> <p>Indicators with predictive validity include:</p> <ul style="list-style-type: none"> • <i>For college-bound students:</i> matriculation, persistence, course grades, grade point average, and degree completion • <i>For career-bound students:</i> extrinsic indicators (e.g., rate of employment, starting salary, or advancement) or intrinsic indicators (e.g., self-reported job satisfaction) • <i>Indicators at the high school level:</i> internships or other career exploration, certificate completion
	Fair comparisons	Allows for fair comparisons of subpopulations with attention paid to systematic bias.
	Stability	Demonstrates sufficient stability to allow trend examination over time, which requires relatively constant definitions and reasonable reliability. Measuring a dynamic construct such as college and career preparedness also requires the capacity to incorporate future measures.
Stakeholder relevance	Value for students	Is actionable and accepted by students as an indicator of their postsecondary prospects.
	Publicly understandable	Provides a clear picture of a school's status or growth in a manner understood by both educators and noneducators.
	Instructional sensitivity	Measures content, skills, and competencies that can be taught and learned in school.
	Student performance	Emphasizes student performance, not educational processes and inputs.
System utility	Minimizes burden	Indicates any additional burden to the department of education, districts, and schools in terms of time and cost to implement and collect data, as well as additional necessary resources. ²
	Student coverage	Includes as many students as possible, considering conditional and universal measures and preferring scaled or scalable measures over local or unique measures.
	Multiple pathways	Recognizes a diverse set of postsecondary pathways to determine if measure supports college-going, career-going, both, or neither.

² It should be noted that some potential measures do require more time, energy, effort, and resources, but they result in more authentic learning and greater mastery of content. Burden, as envisioned here, is meant to take into account such trade-offs and is not simply an indicator of the costs or training requirements of a measure in isolation from the potential positive effects on student preparedness for college and careers.

The Public School Accountability Act Advisory Committee provided EPIC with measures for possible inclusion in the college and career indicator. EPIC researchers conducted comprehensive literature reviews that included reviews of theoretical and empirical research, evaluation studies, and strategies adopted or considered in other states. The reviews included foundational works, as well as findings from respected nonpartisan sources such as scholarly peer-reviewed journals, articles, and books; university presses; large publishers; dissertations; government reports; and relevant websites. EPIC researchers gave priority to research conducted in the past 10–15 years. The quality of each source was assessed based on factors such as study size, relevance, and potential bias, particularly with respect to subgroups of students. Each review considered methodological issues in the study of the measure and any conflicting findings between studies as well as gaps in the literature.

During the review process, the criterion-based framework evolved in order to better reflect what was being learned about potential indicators. For example, innovative measures do not conform well to current practice by definition. They often violate the assumptions upon which traditional accountability systems operate. Therefore, the literature review for the third white paper prompted EPIC researchers to rearticulate four of the 10 evaluative criteria. For instance, *B3: Instructional sensitivity* was defined differently in order to evaluate metacognitive assessments and performance assessments than for advanced coursework or college admission exams, in part because metacognitive assessments and performance assessments transcend content areas whereas the other measures are explicitly content-specific. Another example is *C2: Student coverage* in the white paper on innovative measures, which focuses on the potential for student participation rather than current levels of participation. This recognizes that to consider innovative measures fairly it is necessary first to acknowledge that they have yet to gain universal implementation anywhere in the nation, but the literature indicates both strong possibilities for them to be good measures, and educational and policy leaders are showing interest in such measures.

The process of reviewing innovative measures in particular enabled EPIC researchers to refine the original review criteria further. For example, *A2: Fair comparisons* and *A3: Stability* were refined to include the potential for triangulating innovative measures against more traditionally stable measures until findings converge on the best means to capture metacognitive skills or to use classroom-based performance task results for higher stakes purposes.

The sixth paper, on multiple measures, illuminates strengths, weaknesses, and trade-offs of an accountability system that employs multiple measures of college and career preparedness. This paper offers criteria for designing a multiple-measure system along with examples of how multiple measures are used currently in state accountability systems. The paper reports descriptions of accountability systems in 26 states that incorporate some measure of college or career preparedness, or both. In addition, the paper includes international examples from Australia, England, Finland, Hong Kong, and Singapore.

II. College and Career Preparedness Measures

This section describes each category of measures of college and career preparedness. Each category and its individual measures is described in general terms then considered in terms of the relative strengths, weaknesses, and trade-offs for the category overall and for individual measures within the category. More complete descriptions of each category and measure can be found in the first five white papers (Conley, Beach, Thier, Lench, & Chadwick, 2014a, 2014b, 2014c, 2014d; Conley, Thier, Beach, Lench, & Chadwick, 2014a, 2014b).

College Admission Exams

As college admission exams, the SAT and ACT have more similarities than differences, particularly in terms of the correlations in scores between students who take both of the exams. The makers of each exam have designated their own benchmark scores (described later in this report) that represent college preparedness. Slightly more students now take the ACT than the SAT, but this is due largely to a number of states now requiring all students to take the ACT (ACT, 2014a; College Board, 2014a).³ However, in California, approximately 38% of the 2013 graduating class took the SAT in comparison to 17% for the ACT. Nearly 15% of the 2013 graduating class took both exams.

The SAT

The SAT was first administered in 1926 as an experimental multiple-choice exam designed to measure intelligence as general analytic aptitude. The exam represented an educational application of what were considered at the time scientific notions of human aptitude derived largely from the intelligence testing movement. Over time, the SAT moved away from aptitude to focus its measurement on “developed reasoning” skills. The SAT currently has a scale score that ranges from 600 to 2400 points, with each of three sections (mathematics, critical reading, and writing) being worth up to 800 scale points. The exam takes 3 hours and 45 minutes to complete. The College Board recently announced plans for 2016 revisions that would make the writing section optional, among several other changes.⁴

³ Maine and Delaware administer the SAT statewide to all 11th graders. Arkansas, Colorado, Hawaii, Illinois, Kentucky, Louisiana, Michigan, Montana, North Carolina, North Dakota, Tennessee, Utah, and Wyoming all administer the ACT statewide to all 11th graders. Idaho requires all 11th graders to take either the SAT or ACT.

⁴ Other revisions include removing penalties for wrong answers, replacing obscure words in the vocabulary with words students commonly encounter in college courses, making reading passages longer and more closely derived from what students read in college courses, including more informational texts in addition to literature, having students react to excerpts from “foundational documents” of Western democracy such as the Constitution and the Gettysburg Address, covering fewer topics in depth on the SATM, and removing sentence completion questions.

The ACT

ACT, Inc. developed its college admission exam in 1959 as an alternative to the SAT. The key distinction is that the ACT was designed as an achievement exam that would represent student content knowledge of the curriculum taught in the classroom. In practice, each exam now surveys teachers on their classroom practices, and the makers of each assert that their exam aligns with state content standards and the Common Core State Standards.

The ACT assesses student knowledge and skills in four mandatory sections: English, mathematics, reading, and science, and an optional writing section. Each section has a maximum scale score of 36 points. Scores are reported to colleges by individual section and as a composite score, which is the average score across the four required sections. Students have 2 hours and 55 minutes to complete the ACT and 30 minutes for the optional writing section.

Advanced Coursework

The Advanced Placement (AP) program and the International Baccalaureate (IB) Diploma Programme are distinguished by the fact that they are the only two national-level programs that offer both a curricular framework and standardized assessment in multiple subject areas. AP has 34 assessments. IB offers four programmes.⁵ To facilitate legibility, this report will use “IB” and “Diploma Programme” interchangeably. The Diploma Programme has exams in six subject areas.

Advanced Placement and International Baccalaureate

The College Board administered the first AP exams in 1956. Private school teachers from six countries created the IB Diploma Programme in Geneva, Switzerland, in 1968. Many colleges offer credit to students who take an AP exam and score 3 or higher on a 5-point scale. Some colleges use AP scores to make admission decisions or determine advanced standing of incoming students. An increasing number of colleges are also offering credit for IB exams, which are scored on a 7-point scale.

The IB Diploma Programme is a well-defined set of coursework and exams that is offered in the 11th and 12th grades and is accompanied by a series of exams. Students rarely have options about which courses or exams to take, short of leaving the program altogether. AP courses are available in nearly every California high school. The IB Diploma Programme is currently found in about 6% of California high schools.

The two programs have important differences. AP consists of separate courses, and students may choose to take as few or as many as they please. They may also take the course and not take the AP exam, which costs \$91 per exam without a fee

⁵ IB offers four programmes on a continuum that spans the (a) Primary Years Programme (typically pre-K or K through Grade 5), (b) Middle Years Programme (typically Grades 6-10), (c) Diploma Programme, and (d) IB Career-Related Certificate (both during the final two years of high school).

waiver. AP teachers submit their course syllabi to the AP Course Audit, where it is reviewed to ensure it meets the required criteria for the course. Teachers do retain considerable flexibility in terms of how they teach an AP course once they meet the course requirements verified by the Audit.

IB Diploma candidates complete exams in six groups of subjects: language and literature, language acquisition, individuals and societies, the sciences, mathematics, and the arts/electives. Additionally, students write a 4,000-word, university-level Extended Essay; take an epistemology course called Theory of Knowledge; and engage in a series of student-initiated projects and experiences called Creativity, Action, Service. To earn the diploma, students must acquire 24 out of a possible 45 points.⁶ Some schools allow students to take IB courses without pursuing the comprehensive diploma.

Innovative Measures

This category of measures includes metacognitive and performance assessments, which are not necessarily innovative in the sense that they are not entirely new or different in regular educational practice, but they are innovative in terms of including them in an accountability system. Another additional measure, the California State Seal of Biliteracy, is innovative for the reason that it is relatively new.

Metacognitive Assessments

Metacognitive assessments measure learning strategies, attitudes, and behaviors that students employ and hopefully improve upon during their learning processes. Hundreds of metacognitive assessments with varying purposes, quality, and designs are in use by educators currently for low-stakes purposes. These assessments can be formative or summative and range in format from multiple-choice tests to self-report questionnaires to situational judgment tasks to closed-ended computer-based items to video games. Metacognitive assessments can provide useful insights into how effectively students are learning, especially when these assessments complement standardized content knowledge tests (Conley & Darling-Hammond, 2013).

Performance Assessments

Performance assessments (also referred to as performance-based assessments or performance tasks) require students to construct original responses to authentic problems that are generally more complex in nature than test items. Performance assessments are different from constructed responses on standardized tests, which are generally written passages or multistep problems that are completed within the test administration period. Performance assessments are almost always administered in the classroom and may range from tasks that can be completed in a single class period to

⁶ Calculated from a maximum of seven points for each of six exams and a maximum of three points for the combined assessment of Extended Essay and Theory of Knowledge.

semester-long projects. Performance assessments are designed to elicit cognitive processing and reasoning skills critical for college and career preparedness that cannot be measured on standardized tests.

California State Seal of Biliteracy

The California State Seal of Biliteracy appears on the transcripts or diplomas of students who demonstrate proficiency in two or more languages by high school graduation. Schools or districts award the California State Seal of Biliteracy to students who:

1. Complete all English language arts requirements for graduation with an overall grade point average of 2.0 or above,
2. Pass the Grade 11 California Standards Test⁷ in English language arts at or above the “proficient” level, and
3. Demonstrate proficiency in one or more languages other than English through one of the following:
 - a) Score 3 (out of 5) or higher on an AP exam with content in a language other than English
 - b) Score 4 (out of 7) or higher on an IB exam with content in a language other than English
 - c) Successfully complete a four-year high school course of study in a language other than English with a grade point average of 3.0 or above in those courses
 - d) Pass an approved school district language examination
 - e) Score 600 or higher on a SAT II exam with content in a language other than English

Students whose first language is not English must achieve the “Early Advanced Proficiency” level on the California English Language Development Test and meet the requirements in Steps 1, 2, and 3 above.

Course-Taking Behavior

Course-taking behavior comprises the choices students make regarding the courses they take in high school. Those choices are influenced and shaped by the options available to them within their high school. For the purposes of college and career preparedness, the a–g subject requirements developed initially by the University of California (UC) system and now also used by the California State University (CSU) system align most directly with preparedness for a four-year college or university. Another option for students is a Career Technical Education (CTE) course pathway that would prepare students for postsecondary studies of a technical nature culminating either in a certificate or associate’s degree or transfer to a four-year institution to pursue

⁷ The California State Seal of Biliteracy criteria will need to be revised to reflect the shift from California Standards Test to the Smarter Balanced Assessment System.

a degree more closely associated with a technical field. Students may also choose to take a mixture of college prep and CTE courses. This integrated course pathway is pursued by nowhere near as many students as those choosing college prep or CTE, but an integrated approach has its own set of reasons to be valued and recognized as well.

The a–g Subject Requirements

The a–g subject requirements are a set of 15 college-preparatory courses in six content areas (a–f) and one elective area (g) required for admission to the 10 campuses within the UC system and the 23 campuses within the CSU system. Students must complete all 15 yearlong courses with a grade of C or higher.⁸ Requirements for UC and CSU systems are nearly identical and are slightly more rigorous than California’s high school graduation requirements (see Table 5).

Table 5. California High School Graduation Requirements, CSU Admission Requirements, and UC Admission Requirements

Subject area	CA high school graduation requirements	CSU subject requirements	UC a–g subject requirements
(a) History/social science	3 years U.S. history and geography, world history or culture and geography, and ½ government, ½ civics	2 years U.S. history or American government and a social science course	2 years world history, cultures and historical geography and either U.S. history or ½ U.S. history and ½ government
(b) English	3 years	4 years college preparatory English that includes composition and literature	4 years college preparatory English that includes literature, writing, speaking and listening
(c) Mathematics	2 years including Algebra I	3 years (4 recommended) Algebra I, geometry, Algebra II	3 years (4 recommended) Algebra I, geometry, Algebra II
(d) Laboratory science	2 years biological and physical sciences	2 years biology and physical science	2 years (3 recommended) chosen from biology, chemistry, and physics
(e) Language other than English	1 year art, foreign language, or career technical education	2 years of the same language, including American Sign Language	2 years (3 recommended) of the same language or equivalent to second level of high school instruction
(f) Visual and performing arts	1 year art, foreign language, or career technical education	1 year dance, drama/theater, music, or visual art	1 year dance, drama/theater, music, or visual art
(g) College preparatory elective	Not applicable	1 year	1 year
Physical education	2 years	Not applicable	Not applicable

⁸ Grades of D or lower can be validated by grades of C or higher in more advanced courses or by SAT, AP, or IB exam scores.

The UC Office of the President audits all high school courses offered to meet a–g requirements. High schools offering such courses must provide evidence to satisfy seven standard guidelines⁹ and other specific subject-area requirements.

Career and Technical Education Course Pathway

A career and technical education course pathway consists of three to four courses aligned to the California CTE Model Curriculum standards, which are organized across 15 industry sectors and encompass 58 career pathways. Prior to the 2006 reauthorization of the Carl D. Perkins Vocational Education Act, schools often used the terms CTE and vocational education interchangeably. However, more recently, the term CTE has come to mean a course that is both aligned with a technical field and also academically rigorous and demanding in the way it develops the foundational academic skills associated with the course’s technical area.

Integrated Course Pathway

An integrated course pathway measure is a way to acknowledge and value both college and career preparedness by awarding API points to students who complete a–g subject requirements *and* a CTE course pathway and pass a certification exam associated with the technical area. An integrated pathway may be assigned API points in a number of different ways. Examples of possible methods for assigning API points are presented in Section III.

Career Preparedness Assessments

Far fewer means exist to determine career preparedness than to measure academic preparedness for college. The three measures most commonly used to determine career preparedness are different in their design and methods. ACT’s WorkKeys and the Armed Services Vocational Aptitude Battery serve to measure general career preparedness skills. The National Occupational Competency Testing Institute offers tests of the skills within a specific occupation or industry sector.

ACT WorkKeys

In 1992, ACT introduced WorkKeys, a job readiness assessment designed to help employers hire, train, develop, and retain workers (ACT, 2014b). WorkKeys includes 11 assessments, eight of which measure foundational skills¹⁰ and three that

⁹ (a) brief course description, (b) textbooks/supplemental instructional materials, (c) course purpose, (d) course outline, (e) key assignments, (f) instructional methods and/or strategies, and (g) assessment methods and/or tools.

¹⁰ (a) Applied Mathematics, (b) Locating Information, (c) Reading for Information, (d) Applied Technology, (e) Business Writing, (f) Listening for Understanding, (g) Teamwork, and (h) Workplace Observation.

measure “soft skills.”¹¹ The foundational WorkKeys assessments result in scale scores that correspond to levels for which the highest level indicates the most difficulty and performance levels organized into four tiers: bronze, silver, gold, and platinum. The levels are used to generate the National Career Readiness Certificate. These levels correspond to the percentage of jobs in the WorkKeys database for which individuals have demonstrated the requisite skills. For example, an individual who earns the platinum National Career Readiness Certificate has the necessary foundational skills for 99% of the jobs in the WorkKeys database, which includes more than 19,000 job titles.

Armed Services Vocational Aptitude Battery

The Armed Services Vocational Aptitude Battery (ASVAB) was created in 1968 to measure the overall suitability and likelihood of success of incoming military recruits. The ASVAB is administered by the United States Military Entrance Processing Command and is used by every branch of the Armed Services for enlistee selection and classification. The Department of Defense does not endorse the use of ASVAB outside of military purposes. However, some states use the ASVAB as a proxy for career preparedness¹² and to comply with the reporting requirements of Perkins IV¹³ (Center on Education Policy, 2013). The ASVAB is a multiple-choice exam with 10 subtests covering four domains: verbal, math, science and technical, and spatial.¹⁴ The ASVAB takes approximately two-and-a-half hours to complete. Scores on word knowledge, paragraph comprehension, arithmetic reasoning, and mathematics knowledge subtests produce the Armed Forces Qualification Test (AFQT) score, which establishes minimum qualifications required to meet the enlistment standards of each branch of the armed forces. Scale scores and percentile ranks on both the ASVAB and AFQT derive from a national sample consisting of individuals aged 18 to 23.

National Occupational Competency Testing Institute

National Occupational Competency Testing Institute (NOCTI) offers job-ready, pathway, and employability assessments. Job-ready assessments span 15 industry sectors¹⁵ and measure technical skills and competencies at specific occupational levels within various industries. Pathway assessments cover 11 industry sectors¹⁶ and measure skills derived from an evaluation of the course pathway associated with a

¹¹ (a) Fit, (b) Performance, and (c) Talent.

¹² Kentucky and Missouri

¹³ Virginia

¹⁴ Science/Technical: (a) General Science, (b) Electronics Information, (c)(d) Auto and Shop Information (two separate subtests for the computer-based test), and (e) Mechanical Comprehension; Math: (f) Arithmetic Reasoning and (g) Mathematics Knowledge; Verbal: (h) Word Knowledge and (i) Paragraph Comprehension; Spatial: (j) Assembling Objects.

¹⁵ (a) Agriculture, Food, & Natural Resources; (b) Architecture & Construction; (c) Arts, A/V Technology, & Communications; (d) Business, Management, & Administration; (e) Education & Training; (f) Finance; (g) Health Science; (h) Hospitality & Tourism; (i) Human Services; (j) Information Technology; (k) Law, Public Safety, & Security; (l) Manufacturing; (m) Marketing, Sales, & Service; (n) Science, Technology, Engineering, & Mathematics; and (o) Transportation, Distribution, & Logistics.

¹⁶ All of the industry sectors available for job-ready assessment except: (a) Manufacturing; (b) Marketing, Sales, & Service; (c) Science, Technology, Engineering, & Mathematics; and (d) Transportation, Distribution, & Logistics.

specific set of industries. Employability assessments measure 21st-century skills necessary for workplace success and workplace readiness. The NOCTI assessment system includes pre- and post-tests that incorporate multiple-choice and performance assessment. State policymakers have the option to customize NOCTI assessments to fit their unique CTE standards. Students who meet a 70% benchmark on an approved multiple-choice NOCTI assessment are eligible for a College Credit Recommendation Report from the National College Credit Recommendation Service. The College Credit Recommendation Report can be presented to any of the 1,500 participating institutions for college credit consideration (NOCTI, 2014).

Industry Certifications

Assessments for industry certification vary greatly across industrial and trade sectors. In general, earning industry certifications signals completion of occupational training, an apprenticeship, or a CTE course pathway (Muller & Beatty, 2008). Employers use industry certifications to ensure applicants have acquired the skills necessary to complete the tasks and responsibilities required to perform a specific job. Fewer states use industry certifications to rate or grade schools for accountability purposes than those that use standardized and more generic career preparedness assessments such as WorkKeys, ASVAB, and NOCTI assessments. A wide variety of industry groups and trade unions—including public-private partnerships in which school districts participate—confer industry certifications, making industry certifications much more challenging to compare across states.

Comparison of Measures

Choosing measures for the college and career indicator requires an examination of strengths, weakness, and relative trade-offs both across and within each category of measures.

Across Measures

In the five papers, each category and each individual measure in a category received either a strong, medium, or weak rating for each of the 10 evaluative criteria. Table 6 reports the results from the previous papers.

As noted previously, the matrix in Table 4 should be viewed holistically and not arithmetically. The strengths of a measure may compensate for any areas of weakness. For example, the innovative measures are rated as moderate in terms of technical quality, but this is largely because not enough study has been undertaken, not because they lack technical quality. In fact, the moderate rating indicates that a reasonable amount of evidence of technical quality exists. Stronger ratings in the other two categories would be taken into consideration when evaluating innovative measures as a whole. In the case of industry certifications, although some individual measures have limitations, the category as a whole rates better, and even though most of the ratings are in the moderate range, this should not lead to the conclusion that the category as a

whole does not offer some unique potential contributions to a college and career indicator.

Several findings are not unexpected. Measures with the longest track records and the strongest backing by organizations with the resources to conduct research on their products—the SAT, ACT, and AP exam scores—demonstrated strong technical quality. These organizations have been able to improve technical quality and in fact have had to do so for their products to survive in the marketplace.

Table 6. Evaluation of Measures of College and Career Preparedness

	S = Strong, M = Moderate, W = Weak			A1: Research base	A2: Fair comparisons	A3: Stability	B1: Value to students	B2: Publicly understandable	B3: Instructional sensitivity	B4: Student performance	C1: Minimizes burden	C2: Student coverage	C3: Various pathways
College admission exams	S	M	S	M	S	M	S	M	S	S	M	M	
SAT	S	M	S	M	S	M	S	M	S	S	M	M	
ACT	S	M	S	M	S	M	S	M	S	S	M	M	
Advanced coursework	S	M	S	S	M	S	S	S	S	M	M	M	
Advanced Placement	S	M	S	S	M	S	S	S	S	M	M	M	
International Baccalaureate	M	M	S	S	W	S	S	S	S	W	W	M	
Innovative Measures	M	M	M	S	M	S	S	S	S	M	S	S	
Metacognitive assessment	M	M	W	S	M	M	M	M	M	M	S	S	
Performance assessment	M	M	M	S	M	S	S	S	S	M	S	S	
California State Seal of Biliteracy	W	W	M	M	M	S	S	S	S	M	W	S	
Course-taking behavior	S	M	M	S	S	S	S	S	S	S	M	M	
a–g subject requirements	S	M	M	S	S	S	S	S	S	S	M	M	
CTE course pathway	M	M	M	M	S	S	S	S	S	S	M	M	
Integrated course pathway	S	M	M	S	M	S	S	S	S	M	W	S	
Career preparedness assessments	M	M	M	M	M	S	S	S	S	S	M	M	
ACT's WorkKeys	M	M	M	M	M	M	M	S	S	S	M	M	
ASVAB	W	M	S	W	W	W	S	S	S	S	M	M	
NOCTI	M	M	M	M	M	S	S	S	S	M	M	M	
Industry certifications	W	W	W	S	S	S	S	S	S	W	M	M	

Other categories do not have the same backing or longevity but nevertheless offer promise. Course-taking behavior was rated highly on stakeholder relevance, which is reflective of this category's close connection to schools, teachers, and instruction. For similar reasons, course-taking behavior receives high ratings on system utility. Innovative measures—metacognitive assessments, performance assessments, and the California State Seal of Biliteracy—are yet to be fully tested and understood in terms of their value for college and career preparedness, but all received high system utility ratings for their potential to measure performance of all students. The integrated course pathway category also demonstrated the ability to be valuable for all students.

Table 6 also highlights some potential trade-offs. For instance, some tensions exist between technical quality and stakeholder relevance. Some educators and policymakers may view measures that have high technical quality as less valid or useful. While standardized college admission exams demonstrate strong technical quality, metacognitive assessments and performance assessments have the potential to provide information that is highly valid and reflective of student performance relative to the taught curriculum.

Similarly, stakeholder relevance may be in tension with system utility. For instance, measuring deeper learning requires more of schools than do external standardized tests. However, if done properly, assessment of deeper learning becomes an integral component of the teaching and learning process. When considered from that point of view, such assessment is not necessarily an increased burden on schools at all. Moreover, developing in students the adaptability they need to be prepared for a dynamically changing 21st-century economy may require more and more varied data points than standardized tests provide.

Within Measures

It is also worth noting variation within each category among technical quality, stakeholder relevance, and system utility. Due to its distinctive nature, the California State Seal of Biliteracy is discussed separately.

College admission exams. Both the SAT and ACT have demonstrated technical quality, present value to students, are relatively well understood at one level by the general public, and provide minimal burden for schools.

Technical quality has many dimensions. Predictive validity studies for college admission exams are based on correlations between student scores and grades in first-year courses (Camara & Echternacht, 2000; Kobrin, Patterson, Barbuti, Mattern, & Shaw, 2008; Morgan, 1989; Noble and Sawyer, 2002; Sanchez, 2013). This means of measurement has a series of limitations beginning with the fact that the correlation is between a subset of high school students and then those from the subset that go directly to college. Further limiting the ability to interpret the meaning of the correlation is the fact that researchers have tended to use first-year college cumulative grade point averages. This offers no control over which courses a student took and could mean that a measure of reading and mathematics skill could be correlated with courses that do not make much direct use of the content and skills tested on the exams. In essence, the

correlation could be measuring general college preparedness skills as much or more than reading and mathematics, although the research base does not entertain or explore this possibility in much depth. Controlling for demographic variables (i.e., socioeconomic status, race/ethnicity, and gender) and sample size have further attenuated correlations between college admission exams and postsecondary success. These tests do correlate at a level that explains something on the order of 15% of the variance in first-year college grades for students at four-year institutions. Other measures, such as high school grades and course-taking patterns, correlate more highly and are often used in combination with the SAT or ACT scores (Atkinson & Geiser, 2009; Berry & Sackett, 2009; Bettinger, Evans, & Pope, 2011; Geiser & Studley, 2002; Grissmer, 2000; Rothstein, 2004; Zwick, Brown, & Sklar, 2004). Unfortunately, those scores correlate highly with a student's family income (Rothstein, 2004; Sackett, Kuncel, Arneson, Cooper, & Waters, 2009; Zwick, 2012). This leads to a policy decision about how such scores should be used in interpreting the value a school is adding to a student's preparation for college.

The cost of taking the SAT is \$51 and the cost of the ACT is \$52.50 (or \$36 without writing), which currently falls on individual students.¹⁷ Some states, as noted, have mandated the test for all students and absorbed the cost. Other states do cover costs for districts wanting to administer the test to all students. Scores are sent to schools electronically and, for the most part, are now integrated seamlessly into school data systems. Both the College Board and ACT provide fee waivers to low-income students, and part of the 2016 SAT redesign plan includes increased support for low-income students. Costs would need to be taken into consideration if these tests are included in a college and career indicator.

It is also worth noting that community colleges rarely use such scores in any significant fashion, and community college enrollment is projected to grow the fastest of any segment in the California postsecondary system.

Advanced coursework. Advanced coursework measures also have research bases underlying them. These measures have the advantage of being potentially more instructionally sensitive because the end-of-course exam structure can be a measure of content, skills, and competencies taught in classrooms, assuming teachers have aligned their instruction to the exam.

AP and IB have been stable indicators over time, despite periodic revisions and significant growth in the number of students taking each exam over the past 25 years. Multiple studies show a relationship between AP exam scores and postsecondary performance (Ackerman, Kanfer, & Calderwood, 2013; Dodd, Fitzpatrick, & De Ayala, 2002; Dougherty, Mellor, & Jian, 2006; Geiser & Santelices, 2004; Hargrove, Godin, & Dodd, 2008; Morgan & Klaric, 2007; Patterson, Kobrin, & Packman, 2011; Sadler & Tai, 2007; Shaw, Marini, & Mattern, 2013). Some researchers have noted limits to the predictive validity of AP exams by subject (Klopfenstein & Thomas, 2005). It is also worth noting that less agreement exists on the value of taking an AP course and not taking the exam (Chajewski, Mattern, & Shaw, 2011; Dodd et al., 2002; Dougherty et

¹⁷ The cost of the SAT may change when the redesigned exam is released in 2016.

al., 2006; Geiser & Santelices, 2004; Hargrove, Godin, & Dodd, 2008; Klopfenstein & Thomas, 2009; Klopfenstein & Thomas, 2010; Sadler & Tai, 2007; Shaw et al., 2013; Speroni, 2011).

Researchers have also demonstrated IB's relationship to postsecondary preparedness (Caspary, 2011; Caspary & Bland, 2011; Coca et al., 2012; Halic, 2013; Shah, Dean, & Chen, 2010), including a recent study by EPIC researchers that compared college students in a college honors program who had taken IB courses in high school to students who had not been in IB (Conley, McGauhy, Davis-Molin, Farkas, & Fukuda, 2014).

One challenge of advanced coursework is that not all students will have access to these programs at their high school and, perhaps even more importantly, many students will not be prepared as a result of their middle school and initial high school courses to take advantage of advanced coursework. Some schools and even individual teachers have their own policies regarding which student can enroll in advanced coursework, which further raises equity issues regarding equal access.

AP and IB courses on a student's transcript do help them demonstrate college preparedness in a way that college admissions officers understand and take into account in a variety of ways, including granting college credits based on exam scores. Nearly every high school in California offers AP courses in at least one subject area, and 6% offer IB. Both programs are growing at a rapid rate.

As with college admission exams, students shoulder the fees for AP and IB exams. The College Board and ACT provide fee waivers to low-income students. The 2016 SAT will offer low-income test-takers free score reports. While students do not pay extra to take an AP or IB class, schools do incur costs. Creating a new AP course at a school can range from \$1,900 to \$10,000 when the costs of instructor training are taken into account. Authorized IB schools pay annual fees of more than \$10,000 in addition to five-figure start-up costs that include clear and nonnegotiable expectations for extensive teacher training. Schools have complete control over the number and type of AP courses offered. Diploma Programme schools, by contrast, must offer a minimum of seven IB courses. Both clearly focus on preparing students for postsecondary studies leading to a bachelor's degree. Recently, however, the International Organization added the IB Career Certificate, which integrates pre-university and pre-career coursework.

Metacognitive assessment and performance assessment. Metacognitive assessment and performance assessment are a distinct alternative to measures such as the ACT, the SAT, AP, and the IB. Their strength lies in their validity; they can measure content knowledge mastery and learning skills that are precursors to postsecondary success and that cannot be measured any other way. The kinds of skills these assessments measure are valued by employers, who consistently rate many nonacademic skills as being more important to success in the workplace than the type of reading and mathematics measured on standardized tests.

These types of assessments measure something different than content tests do, and in the process provide insight into aspects of learning other tests do not. They can help students become more effective learners. When properly designed, these

assessments can generate information useful for determining preparedness for both college and career postsecondary pathways. This section treats the two types of assessment as one category. However, it should be noted that performance assessment has a stronger research base and wider use in the field, while metacognitive assessments have been employed more broadly outside of education, and much of the research on their effectiveness comes from these uses.

Currently, assessments of this nature are used primarily but not exclusively for low-stakes purposes. However, they have been used in more high stakes ways in other states previously, and they are a key element of student grades in many schools that employ this type of assessment (SCALE, 2014). Kentucky, for example, used performance assessments at a state level for 20 years and achieved high rater reliabilities, and Vermont implemented teacher-scored portfolios of student writing that, after some teacher training, also achieved high reliability scores. The AP program makes use of performance assessment in several of its subject areas.

Such assessments have the potential to measure deeper learning (Conley, 2013; Darling-Hammond & Adamson, 2010; Darling-Hammond & Pecheone, 2009; Lane, 2010), and have been shown to demonstrate high levels of predictive validity in areas as diverse as student performance on the California High School Exit Exam and the clinical competence of physicians (Goldschmidt, Martinez, Niemi, & Baker, 2007; Hojay et al., 2000; Lane, 2010). One measure, the College and Work Readiness Assessment, has achieved higher predictive validity than national college admissions tests (Council for Aid to Education, 2014).

Current metacognitive assessments often rely on self-report, by students or teachers. Instruments to assess metacognitive skills have been used extensively in the private sector for decades and have a well-established track record (Ehley, 2006; Hart Research Associates, 2013). They have not been used as widely in public education on a system level. When administered without proper training and monitoring, some of these instruments can be vulnerable to social-desirability bias in responses (Huws, Reddy, & Talcott, 2009). This tendency can be mitigated in part by triangulating results against other measures to detect cases of overt faking in a manner similar to that used on standardized achievement tests to identify possible instances of cheating.

Seven California school districts that serve more than one million students, known as the California Office to Reform Education or “CORE” districts, are now designing an accountability system to meet federal requirements that will include measures of student metacognitive skills. These districts could conceivably be a source of longitudinal data that could help demonstrate the reliability of metacognitive assessments, particularly when the results are triangulated with other student performance results, such as grades and test scores.

These types of assessments have several inherent challenges. Creating the prompts or tasks requires a different process than item development, one that employs both evidence-centered design and criterion-based decision-making to ensure tasks will elicit the desired behaviors and that they can be scored against a standardized guide in order to maximize scorer consistency and reliability. The tasks are easier to remember and thus have to be replaced or rotated out more frequently, which adds to the expense

of development. Scorer reliability is always raised as an issue, and the process to maximize reliability begins as noted with task design and alignment with clear scoring guides and continues to training that ensures scorers are operating from the same points of reference regarding performance levels (Conley, 2013; Darling-Hammond & Adamson, 2010). Significant attention is being given to automated scoring systems, although at this point, no artificial intelligence scoring method is sufficiently sophisticated to deal with anything more complex than short essays.

Results from complex performance assessments are taken into consideration in the admissions process at many highly selective institutions that have the resources to look at collections of evidence on student preparedness. Few, if any, less-selective institutions currently have the ability to take into consideration performance tasks completed in high school (Conley, 2013). However, colleges could conceivably use the results from metacognitive assessments as additional information, not to make an admission decision, but to help the admitted student make a more successful transition to college (Conley, Beach, et al., 2014d). Postsecondary institutions such as Boston College, DePaul University, Oregon State University, and Tufts University use information gleaned from metacognitive assessments for admissions purposes. Some employers use metacognitive components of ACT's WorkKeys to hire, train, develop, and retain workers (ACT, 2014b).

Combining or triangulating information from metacognitive and performance assessments with information from content tests and grade point averages could lead to a more complete picture of student preparedness. Student self-reports could be compared to teacher reports of student characteristics such as persistence and goal focus as well as to standardized content test scores to detect inconsistencies and anomalous patterns (Conley, Beach, et al., 2014d).

The concept of performance assessment is well understood by anyone who has taken a driving test, undergone CPR certification, auditioned for a school play, or competed in a tryout for a sports team. However, these results have not been used to rate school quality. This leads to a general unease or distrust of these measures because the public and policymakers have come to believe that tests of reading and mathematics are the only valid and legitimate accountability measures. This fact presents a challenge for the use of performance assessments for accountability purposes.

Unlike current standardized tests, metacognitive and performance assessments provide educators with actionable, immediate feedback that facilitates greater classroom-level responsiveness to learning needs. This helps counteract educator perceptions that what is measured does not reflect their classrooms.

California State Seal of Biliteracy. The California State Seal of Biliteracy is a new and promising measure of college and career preparedness that as yet does not have a longitudinal record to determine its precise value and effects. Students earning the California State Seal of Biliteracy may have access to career opportunities not available to those who know only one language, especially in California where nearly 44% of residents over the age of five speak a language other than English at home (Ryan, 2013). It is worth noting that only some of the pathways to earning the California

State Seal of Biliteracy meet the criteria set forth in this paper for fairness and stability as a measure of postsecondary success due to characteristics of some of the measures included in the Seal, such as approved school district language examinations, the SAT II, and coursework grade point averages. As a relatively new program, public understanding of the California State Seal of Biliteracy is naturally low. Part of the success of the California State Seal of Biliteracy will be dependent on the ability of local districts to secure the necessary instructors for languages other than English.

Course-taking behavior. The course-taking behavior category is rated strong in six of the 10 evaluative criteria, more than any other category of measures. The grades generated by a–g course requirements have some issues in terms of their technical quality, as do grades for CTE pathways, but the proposed measure is of course-taking behavior, not of the grades earned in those courses.

Course-taking behavior shows the strongest statistical relationship with postsecondary outcomes (Adelman, 1999). Advanced course-taking in high school increases entry into and performance in college (Adelman, 1999; Iatarola, Conger, & Long, 2011), particularly for mathematics (Adelman, 2006; Finkelstein & Fong, 2008; Finkelstein, Fong, Tiffany-Morales, Shields, & Huang, 2012; Goodman, 2012; Levine, & Zimmerman, 1995; Rose & Betts, 2004; Trusty & Niles, 2003), sciences (Schwartz, Sadler, Sonnert, & Tai, 2009), and the arts (Catterall, 2009). Integrated course pathways of the type developed and promoted by the Linked Learning Alliance lead to improved graduation rates, eligibility for university admission, and lifetime earnings (Center for Advanced Research and Technology, 2011). Not all career courses necessarily show the same relationship to postsecondary preparedness. Career academies, which have a relatively long track record, have produced mixed results. Researchers have found both significant and insignificant differences in the postsecondary preparedness (college enrollment, remediation rates, and a–g completion rates) of career academy and nonacademy students (Dayton, Hester, & Stern, 2011; Kemple, 2004; Kemple & Willner, 2008; Leikes et al., 2007; Maxwell & Rubin, 1997).

Course quality and consistency is a big issue when considering course-taking as a measure of college and career preparedness. Presently, entry-level college and job-training program courses with the same or similar course titles across different institutions show variations in the prerequisite knowledge, skills, and abilities expected of incoming students (WestEd and EPIC, 2013; EPIC, 2014). The UC Office of the President audits all a–g courses offered in California. The result is a degree of consistency in challenge level and a level of assurance that approved courses address key areas in their subject area. At the moment, no equivalent quality control method exists for CTE pathway courses. An audit system would likely improve course quality and be used to improve alignment to the Common Core State Standards and to the CTE Model Curriculum standards. Creating an audit process for CTE courses could strengthen the confidence in the rigor of these courses while still allowing for diversity of offerings statewide. Such a requirement would lead to considerable course development or redesign at the local level to ensure alignment with audit standards. While such a process would have costs associated with it, the result would be to place

CTE courses on the same level as a–g requirements in terms of their alignment to a single set of external standards while still allowing for local variation.

Both a–g and CTE course pathways provide tangible value to students. The a–g subject requirements guide and structure the high school program of study for students by ensuring they are satisfying course-taking admission requirements applicable to all public institutions of higher education in California. Completing CTE course pathways prepares students for postsecondary certificate programs and can, in some cases, lead directly to industry certifications (Foster & Pritz, 2006). They may be designed to be the initial two years of a “2+2” program in which a student spends two years in high school and then two years in college studying a career or occupational area in order to receive an associate-level certificate in the area.

One key advantage to incorporating a course-taking behavior measure into the API is the absence of student fees or other additional requirements of students. All schools already offer courses in these areas, and the UC Office of the President already collects data on a–g courses offered. Data collection systems for CTE courses would need to be developed. Integrated pathways would create their own set of data collection needs and challenges.

Career preparedness assessment. In general, career preparedness assessments have more technical challenges than academic content knowledge or aptitude-based measures due to their relatively limited use in educational settings until recently. Career preparedness assessments show considerable variation in their value to students, understandability, instructional sensitivity, and system-level burdens.

Technical quality varies widely for career preparedness assessments. Studies have established a relationship between WorkKeys scores and career preparedness (ACT, 2008a, 2008b, 2008c, 2014c, 2014d; Hendrick, 2006; Swaney, Allen, Casillas, Hanson, & Robbins, 2012) and ASVAB (Campbell & Knapp, 2001; Welsh, Kucinkas, & Curran, 1990; Wolfe, Larson, & Alderton, 2006). The vendor conducted these studies (Hendrick, 2006), which does not necessarily invalidate them but is worth noting. Also worth noting is that the studies have not examined college success as an outcome criterion (Bowles, 2004; Lindon, 2010). Another of the major measures, ASVAB, has not been demonstrated to have a relationship to career success in civilian occupations. NOCTI has strong processes to measure internal consistency, evaluate the reliability of test items, and detect and eliminate bias (NOCTI, 2012), but the research on effects is limited.

The lack of extensive peer-reviewed findings should not be surprising, given that interest in using these measures broadly as college or career preparedness measures is relatively recent. More research is needed to determine comparability across schools, longitudinal patterns across schools, effects on student subgroups, and the potential for social desirability bias (Huws, Reddy, & Talcott, 2009). The impact on schools of using these assessments should be relatively low because all are developed and administered externally.

Given their relative newness as sources of information to be incorporated into an accountability system, these assessments may be more appropriate as contributors to a multiple measures approach rather than as independent measures.

III. Constructing a College and Career Indicator

This section discusses the issues involved in constructing a college and career indicator. These include technical aspects, equity issues, postsecondary pathways addressed, and possibilities for the indicator to evolve over time. The goal is to ensure that the classifications of schools are as valid and accurate as possible and that students and educators have incentives to act in ways consistent with state goals and their own best interests (Conley, 2012).

Technical Considerations

Technical considerations are those aspects of the system that establish the mechanics of it. They include setting the benchmark levels associated with college and career preparedness, combining measures when more than one measure is included in the calculation of the college and career indicator, aggregating student-level indicators to generate a summative school-level rating, and considering the ways measures can be used in a combination of low- and high-stakes ways.

Setting Benchmarks

Testing organizations such as the College Board, ACT, and the International Baccalaureate Organization have already set benchmarks that they believe represent college- and career-prepared performance. Each organization uses a slightly different method based on the relationship between scores on assessments and subsequent performance in college in the case of college-preparedness measures or preparedness for occupations in the case of career preparedness measures.

Both the College Board and ACT caution against using benchmarks to make college admission determinations about individual students. Instead, they recommend that exam scores can be used to identify the overall level of college preparedness of students in a district, state, or nationally. The College Board and International Baccalaureate Organization recommend that colleges award credit to students who score above the benchmarks listed for AP and IB tests as listed in Table 7. States use completion of an industry certification, either through NOCTI or other organizations, as the measure of career preparedness. Metacognitive assessments, performance assessments, and the State Seal of Biliteracy have not been incorporated into any state accountability system and therefore have no established or consistent benchmark levels. States measuring course-taking behavior for college and career preparedness (most notably in Georgia, Maryland, New Mexico, New York, and North Carolina) require completing pathways of courses consisting of three to four courses.

The challenge with setting benchmark levels is that they can become cut scores, with judgments made about students who are above or below the cut line, even though the larger purpose may be school-level accountability. Additionally, benchmark levels overlook the continuous nature of student performance and the complexity of college and career preparedness in general. While benchmark levels may be necessary for policy purposes, they should not be confused with a comprehensive system that captures preparedness more on a continuum. The rest of this section considers a broader range of measures that could be used in combination with or instead of benchmark levels.

Table 7. College and Career Preparedness Benchmarks by Measure

Measure	Vendor-set benchmarks	Highest possible score	Examples of state-determined benchmarks
SAT	<ul style="list-style-type: none"> • 500 per section, or • 1550 total 	<ul style="list-style-type: none"> • 800 per section, or • 2400 total 	<ul style="list-style-type: none"> • States typically use vendor-set benchmarks; however, Florida sets college preparedness at 440 for each SAT section or 18 on ACT reading and 19 on ACT mathematics
ACT	<ul style="list-style-type: none"> • 18 on English • 22 on mathematics • 22 on reading • 23 on science 	<ul style="list-style-type: none"> • 36 on each section 	
Advanced Placement	<ul style="list-style-type: none"> • 3 out of 5 	<ul style="list-style-type: none"> • 5 out of 5 	<ul style="list-style-type: none"> • States typically use vendor-set benchmarks
International Baccalaureate	<ul style="list-style-type: none"> • 4 out of 7 	<ul style="list-style-type: none"> • 7 out of 7 	
ACT's WorkKeys	<ul style="list-style-type: none"> • Silver tier 	<ul style="list-style-type: none"> • Platinum tier 	<ul style="list-style-type: none"> • Kentucky: Silver level (third highest of four levels), which translates to preparedness for 75% of 19,000 jobs in the ACT database
ASVAB	N/A	<ul style="list-style-type: none"> • 99th percentile 	<ul style="list-style-type: none"> • Kentucky: 55th percentile • Missouri: 30th percentile (with increasing API point levels for higher percentiles)

Complementary (or Compensatory), Matrix, and Conjunctive Methods

The complementary model combines measures, which allows different aspects of preparedness to be valued. A complementary method can be designed in ways that make a group of measures interchangeable; a student who is successful on one measure does not need to be as successful on all the others. In practice, this means that strong performance in one area can compensate for weaker performance in other areas (Conley, 2012). This approach allows schools to offer a wider range of programs that are responsive to student needs and still get credit from an accountability

perspective for being “good” schools. Complementary or compensatory models derive from the notion that college and career preparedness encompass a range of skills and competencies that no single measure can capture adequately (Conley, Thier, et al., 2014b).

In a complementary model, meeting or exceeding the threshold of a single measure indicates sufficiency, even if the thresholds of other measures are not met. Therefore, student performance counts only for the measure on which they perform best. As student performances are aggregated to the school level, the college and career indicator operates in a compensatory fashion: higher performance on one or some measure(s) offsets lower performances on other measures. For instance, if a college and career indicator has three measures, one school could perform well on one measure (e.g., advanced coursework) and achieve a high API, while another school could perform equally well by emphasizing another measure, such as student course-taking of quality CTE courses. Doing well on CTE course participation might offset lower participation in advanced coursework or lower scores on a third measure, such as SAT/ACT scores.

Clearly, this approach works only if all the measures included in the combination are equally important and can be measured clearly and accurately. While schools in some districts might seek the “easiest” route to a high API score, most districts would want to link their strategy to their Local Control Accountability Plans, which are developed with significant community input and are subject to annual review to determine progress and achievements. The state could also act as a safeguard by questioning combinations of measures that do not seem to challenge the student population sufficiently.

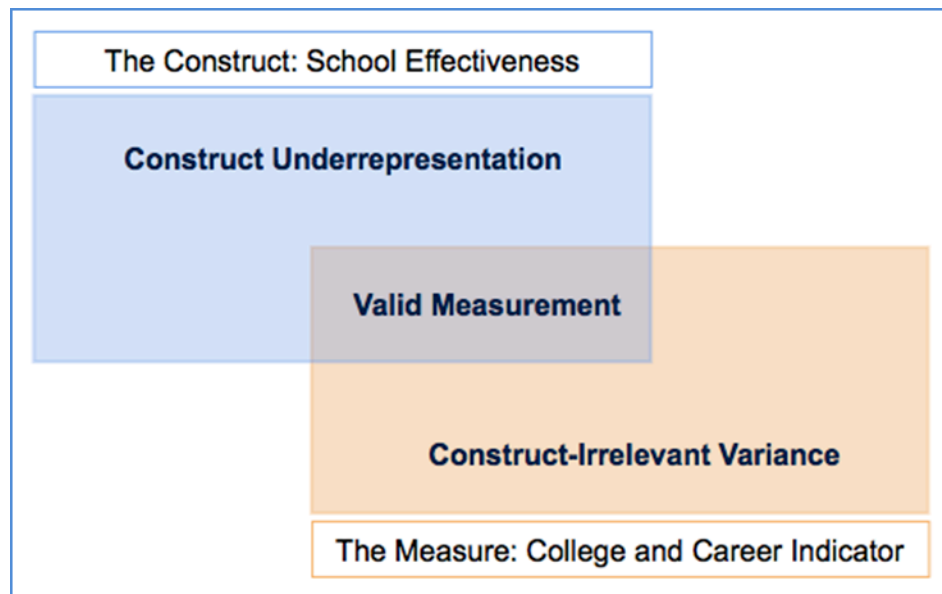
Another way to think about the use of multiple scores is a matrix model. Technically, this approach does not combine measures. Instead, it uses levels (e.g., high, mid-range, and low) to demonstrate school performance relative to each measure. A matrix method requires a more complex calculation than a single API measure. An important advantage of a matrix is that it allows for more nuances than a single number and can be used to target improvement areas more specifically.

Finally, a conjunctive approach to multiple measures would require schools to meet or exceed certain thresholds on all measures. For example, the state could set a target API score of 800 for each college and career preparedness measure. The key issue here is that all of the measures truly need to be mastered at the cut score level for a school to be meeting the criterion of producing students who are prepared for college and careers. A conjunctive model has the disadvantage of standardizing the definition of a quality school, an act that might force schools to create programs that do not suit their local contexts. One other problem is the more criteria schools must meet, the fewer school that will be able to do so. This has the effect either of having few schools meet accountability standards or of having to lower the benchmark level or cut score on some measures to ensure enough schools can achieve them.

Aggregating Student-Level Indicators to Measure School Effectiveness

A potential problem of aggregating student-level indicators is the interaction between construct-irrelevant variance and construct underrepresentation (Hamilton, Stecher, & Klein, 2002). The API currently aggregates student-level indicators to measure school effectiveness. This can be problematic when the factors associated with a student succeeding do not capture the range of factors necessary for a school to be effective. In such cases, some schools that score well may not be adding a great deal of value to their students' education due to the nature of their student population, while others that do not score as well may be adding significant value given the characteristics of their students.

The relationship between the construct being measured and the assessment measuring that construct is shown in Figure 2.



Source: Revised from Hogan, T.P (2007)

Figure 2. Construct-irrelevant variance and construct underrepresentation.

Additionally, construct underrepresentation occurs when an indicator fails to capture the entire domain of interest. Using the example of differences in student populations, a college and career indicator with measures relevant only to the college-going pathway will provide an incomplete picture of career preparedness. Similarly, an indicator that does not measure metacognitive skill attainment will not sufficiently capture information critical to understanding how well students may perform who are highly motivated and goal driven but still have some content-area deficiencies as indicated by test scores. Construct-irrelevant variance occurs when an indicator includes content not pertinent to measuring the construct of interest.

Equity Considerations

An accountability system needs to provide information on student learning in a fashion that is not based solely or primarily on aptitude or privileged opportunity. These dimensions are not particularly actionable by schools, nor do they represent a school's value-added to a student's educational development. Creating an equitable accountability system requires careful attention to measures that are fair and on which all students can conceivably perform well if given sufficient opportunity and quality instruction.

Biases of Race, Ethnicity, and Socioeconomic Status

Teacher and resource inequalities pervade high-minority and low-income schools (Carroll, Krop, Arkes, Morrison, & Flanagan, 2005; Oakes & Sanders, 2004), and these inequalities raise fairness concerns about student performance on all of the measures discussed in this white paper series. For example, parents' highest degree attained and socioeconomic status (Buchmann, Condrón, & Roscigno, 2010; Byun & Park, 2012; Card & Rothstein, 2007; Rothstein, 2004; Sackett et al., 2009; Zwick, 2012; Zwick & Himelfarb, 2011) and race (ACT, 2013; Dixon-Román, Everson, & McArdle, 2013) correlate highly with student scores on college admission examinations. Similarly, students from minority or economically disadvantaged backgrounds are less likely to pass AP exams even after statistically controlling for prior achievement and educational expectations (The Broad Foundations, 2013).

High correlations between SAT score and parent income level are particularly ironic given that Harvard University first adopted the exam in 1933 as a method to identify and select students who did not come from high-income families and did not attend prestigious boarding schools (Lemann, 1999). Although AP and IB were designed initially for schools with higher concentrations of students from higher-income families, both have since evolved to emphasize much greater inclusion. Now, nearly 14,000 U.S. high schools offer AP. Participation among high school graduates has increased dramatically, from 19% in 2003 to 33% in 2013. AP exam-taking among students from low-income families increased fourfold in the last decade (College Board, 2014b). IB still serves an international clientele outside of the U.S. that skews toward higher socioeconomic status, but the schools participating in IB are becoming more diverse. Worldwide, more than half of IB schools are public, including 88.9% in the U.S., and more schools are adopting IB with low- and middle-income student populations.

Although not a panacea, nonstandardized assessments have the potential to measure the performance of some subgroups more equitably. For example, well-designed performance assessments can improve accessibility for English learners and students with disabilities when compared to multiple-choice assessments (Darling-Hammond & Adamson, 2010). Similarly, the gap between white students, English-only students, and traditionally disadvantaged students was larger on the standardized Stanford Achievement Test than on the language arts performance task of the California High School Exit Examination (Goldschmidt et al., 2007).

The caution here is that extra attention to the needs of students from groups historically denied full access to educational opportunity will be necessary if measures that have been shown to be strongly related to aptitude or socioeconomic status are included in the college and career indicator.

Access and Opportunity to Learn

Opportunity-to-learn plays a central role in judging the equity of an accountability system that hopes to improve educational outcomes. It is a common-sense notion that students will not be able to perform well in areas where they have not received adequate instruction, and yet this is the case in many schools whose students are then judged to be inadequate.

McDonnell (1995) identified three assumptions about opportunity-to-learn as a policy instrument: educational standards represent a social contract between schools and communities, standards should focus on the factors that most directly affect student learning, and indicators have to be able to be developed to measure schools' ability to deliver the standards. Therefore, measuring student performance without accounting for inputs and processes will not convey a full picture of school quality. To that end, college admission examinations and advanced coursework may produce blind spots if used as isolated measures of college and career preparedness because they may be measures of opportunity to learn as much as of student capabilities. Other measures more directly gauge the opportunities students have had. Course-taking behavior, in particular, is a direct indicator of opportunity to learn, particularly if the quality of the courses has been determined via an external audit process.

Beyond basic access, high schools need to present equal opportunities to students in terms of rigorous coursework and the resources and conditions necessary for effective instruction (Adelman, 2006; Carroll et al., 2005; Oakes & Sanders, 2004). Economically disadvantaged students have less access to language other than English courses than their advantaged peers due to lack of qualified instructors, feeder middle-school programs, and resources for those courses (Sung, Padilla, and Silva, 2006). Furthermore, an EPIC evaluation revealed that CTE pathway completion in California is higher among middle-income students, who also enjoy the lowest student-to-teacher ratios in CTE courses. Schools that have invested heavily in AP or IB programs may in the process provide fewer resources and opportunities to career-oriented students.

Pathway Choices Between College and Career

It is important to bear in mind that schools have historically sorted students into those who will go onto college and those who will go to work. The historical core of the American high school is in college preparation. Vocational education was added in the 1910s and 1920s and for a while achieved a level of parity with the college preparatory mission of most high schools. However, since the end of World War II, vocational education has gradually come to be viewed much more as a place for students who were not capable of or interested in pursuing academic studies rather than an equal educational opportunity. While recent efforts to revitalize and redesign traditional

vocational education programs so that they metamorphose into career-oriented education are gaining momentum, many high schools still bear the cultural remnants of a stigmatization of vocational programs. Careful attention will need to be given to the cultivation and validation of career preparation programs and pathways in the college and career indicator in order to help schools move away from the prejudice that preparation for a four-year college education should be the primary goal for most students.

Prior to the reauthorization of the Carl D. Perkins Vocational Education Act in 2006, most schools tracked students into college or vocational pursuits. Many local education authorities cast some students as winners and others as losers on this basis. Typically, educators, policymakers, researchers, and other constituents have reinforced this dichotomy by showing greater interest in college preparedness than career preparedness. Though many schools have dismantled infrastructure and abandoned practices that segment students based on their employment goals, vestiges remain. It is common for schools' best teachers to serve students in advanced programs (Klopfenstein & Thomas, 2010).

Simply treating college and career preparedness as "the same" has been encouraged in some quarters. However, college and career preparation are not identical (Conley & McGaughy, 2012). College preparedness depends on a broader range of academic skills across multiple areas of English and mathematics. Career preparedness also requires a strong foundation in these areas, but the specific language and number skills students must master are much more varied in nature and dependent on the context of the career area. This makes it more difficult to use one measure of reading and mathematics as a universal indicator of college and career preparedness (Conley, McGaughy, Brown, van der Valk, & Young, 2009; EPIC, 2014; WestEd & EPIC, 2013).

Choosing the measures to include in the college and career indicator requires sensitivity to the continuum of college and career preparedness. Figure 3 displays where each category of measures evaluated in these papers falls along the college/career preparedness continuum, with the exception of course-taking behavior, which is dependent on the decisions about which courses are taken.

SAT and ACT scores demonstrate a statistical relationship with first-year college grades, but little evidence exists on the degree to which scores on these exams relate to career preparedness (Conley, Beach, et al., 2014b). The inverse is true for career preparedness assessments, and no independent evidence supports claims that they also predict college preparedness (Conley, Beach, et al., 2014a). This remains the case even though each assesses a core of academic and cognitive skills that should generalize across both settings.

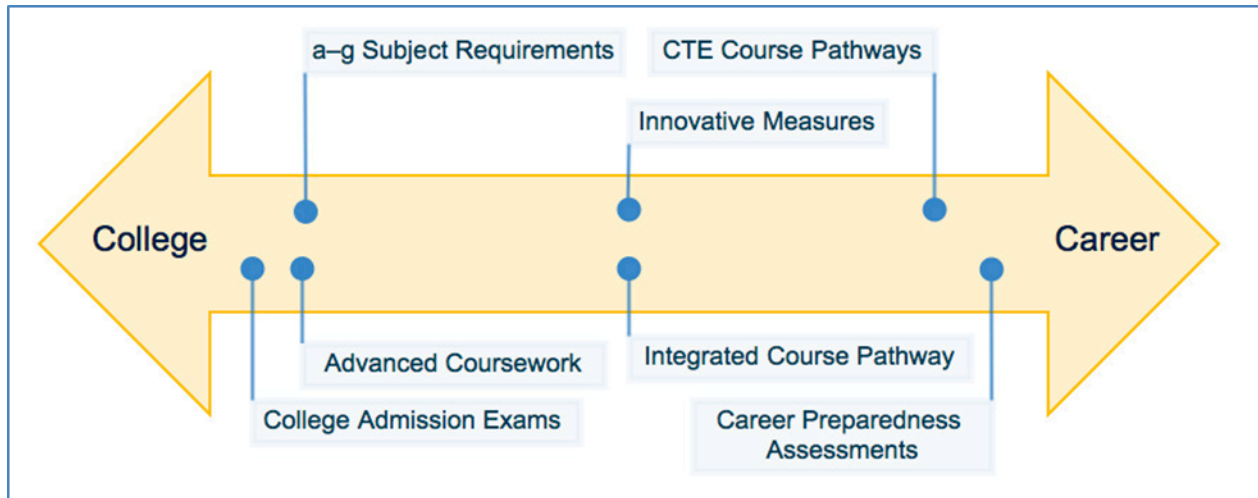


Figure 3. Continuum of college and career preparedness measures.

Including advanced coursework in the college and career indicator creates similar challenges. Though studies have found students taking advanced coursework to be more likely to attend college (Chajewski et al., 2011; Coca et al., 2012; Speroni, 2011), schools often restrict enrollment in these courses, which makes it more difficult to interpret the findings about their relationship to college preparedness. The relationship between AP or IB and career success is also not well understood, in part because these programs do not necessarily explicitly develop important workplace skills such as time management, following directions, remembering instructions, communicating with coworkers and customers, and making good decisions (Ng, Eby, Sorensen, & Feldman, 2005), although many students in these programs may already have these skills or may improve them while in the program. Teachers interviewed for a mixed-methods study of AP and dual enrollment reported the need for students to learn skills to study and manage time (Haley, 2013). A small-scale study at one university did demonstrate that IB fosters academic *and* metacognitive skills that enhance college preparedness (Conley et al., 2014).

The New Mexico example: Emphasize college preparation. New Mexico's system emphasizes college preparedness at the expense of career preparedness. Table 8 contains a list of all the measures schools can use to earn points. Eight of the indicators measure college preparedness, one measures career preparedness, and two measure elements of both college and career preparedness.

Of the two dual measures, College Board's ACCUPLACER measures college preparedness much more directly than career preparedness. ACCUPLACER's four sections assess skills in reading, writing, math, and computer skills. Only the computer skills section aligns with career preparedness, but does so broadly. Whether a dual/concurrent enrollment program offers opportunities for college or career preparation depends upon the courses the high school chooses to offer and that students choose to take.

Table 8. New Mexico’s College and Career Measures by Potential to Measure Preparedness for College, Career, or Both

Indicator	College, Career, or Both
ACT	College
ACT’s Plan	College
ACT’s Compass	College
Advanced Placement exam	College
CTE course pathway completion	Career
College Board’s ACCUPLACER	Both
Dual/concurrent enrollment	Both
International Baccalaureate exam	College
National Merit Scholarship Qualifying Test	College
PSAT	College
SAT	College

The Georgia and Kentucky approach: Integrated pathway. An integrated pathway, of which Georgia and Kentucky are examples, allows students to earn API points by completing a–g subject requirements *and* by completing a Career and Technical Education pathway. Georgia’s Post High School Readiness metric includes an indicator for the percentages of graduates completing a Career Technical and Agriculture Education Pathway, an advanced academic pathway, a fine arts pathway, or a world language pathway within their programs of study.

Students in Kentucky have the opportunity to earn the status of college-prepared, career-prepared, or both. To be deemed college prepared, students must meet a benchmark score on the ACT test, ACT’s Compass (college placement test), or the Kentucky Online Testing Program (KYOTE). Career-prepared students must meet a benchmark on either WorkKeys or ASVAB *in addition to* meeting the benchmark on the Kentucky Occupational Skills Standards Assessment (KOSSA) or earning an industry certificate. Students can earn college- and career-prepared status by meeting a benchmark on the ACT test, ACT’s Compass, or KYOTE *and* meeting the benchmark on KOSSA or earning an industry certification.

Additional Possible Contributors to the College and Career Indicator

Other measures beyond those explored in the six papers prepared by EPIC could conceivably contribute information on college and career preparedness. Several are discussed in this section.

Dual/Concurrent Enrollment

Dual and concurrent enrollment programs have been mentioned several times already in this paper as potential contributors to a multiple-measure system.¹⁸ Dual enrollment has shown positive relationships with college aspirations (Smith, 2007), preparedness (Kim, 2008), grade point average (Spurling & Gabiner, 2002), enrollment (Karp, Calcagno, Hughes, Jeong, & Bailey 2007), persistence (Karp et. al., 2007), and graduation (Adelman, 2006; Swanson, 2008). Dual enrollment programs provide direct value to students who earn college credit and indirectly benefit students by easing the transition from high school to college (Bailey, Hughes, & Karp, 2002).

Concerns stem from the variation in quality across dual enrollment programs (Karp, Bailey, Hughes, & Fermin, 2004). This derives from the fact that courses are taught by high school instructors in some cases or that the courses may be designed and implemented without sufficient oversight by postsecondary faculty.¹⁹ It can be challenging for high school teachers to establish college-level expectations for their students, who are used to being treated like high school students throughout the rest of their school day.

Dual-enrollment programs could conceivably be included in a course-taking behavior measure as an indicator of challenging coursework or as an outcome measure based on the grades students receive and even how well they fare in college subsequently.

Culminating Projects

Culminating projects, also known as graduation projects or exit projects, typically require students to complete a complex set of requirements over an extended period. For example, for Rhode Island students must complete at least two district-designed performance-based assessments, including graduation portfolios, exhibitions, and comprehensive course assessments (Darling-Hammond, Wilhoit, & Pittinger, 2014). Students in the 40 high schools that make up the New York Performance Standards Consortium substitute a graduation portfolio with performance tasks across three to four content areas for the New York Regents examinations (Conley & Darling-Hammond, 2013). Similar to performance tasks, culminating projects can be scored reliably using a common rubric and include elements that can be adapted to college and career standards (Conley & Darling-Hammond, 2013).

Languages Other Than English Coursework

Course-taking in languages other than English is a necessary step to increase biliteracy, a skill that can enhance employment opportunities (Bureau of Labor Statistics, 2014) and can also better prepare students for postsecondary language

¹⁸ Concurrent enrollment programs allow students to simultaneously earn college and high school credit by taking college courses.

¹⁹ Considerable controversy exists regarding these points, and the evidence about who is qualified to teach a college course in high school continues to evolve.

studies. The flagship universities in 36 of 50 states require a minimum of two years of courses in languages other than English for admission, and more selective public schools have more stringent requirements (Conley, Beach, et al., 2014d).

Biliteracy has been shown to strengthen brain functioning (Adesope, Lavin, Thompson, & Ungerleider, 2010; Rodriguez, Carrasquillo, & Lee, 2014; Soveri, Laine, Hamalainen, & Hugdahl, 2011) and is associated with higher performance on achievement tests (Armstrong & Rogers, 1997; Dumas, 1999). Biliteracy is increasingly important in a global economy, creating and enhancing career opportunities not available to those who know only one language. This competency is particularly useful in California, where the percentage of people age 5 and older who speak a language other than English at home has increased from 11.0% in 1980 to 20.6% in 2010 (Ryan, 2013).

Lab Science Coursework

Current a–g requirements call for two years of biological and physical sciences, which support preparedness for many college majors and many careers. Science coursework can increase student understanding of how to use evidence, an important skill in the Common Core State Standards and an element that will be emphasized more on the revised SAT that will be given beginning in 2016. (College Board, 2014c).

College Remediation Rate

College remediation rate is being used increasingly by states to judge how well high schools have prepared students for college. Remediation delays enrollment in credit-bearing courses, which increases cost and decreases college success (Bailey, Jeong, & Cho, 2010). Studies find college remediation rate to be an indicator of the preparedness for postsecondary education (Scott-Clayton & Rodriguez, 2012; Mikulecky & Christie, 2014). Hawaii and Missouri measure remediation rates for accountability purposes.

One limitation of remediation rates is that they are based on placement test scores that measure content that is not always required for the student's major or field of technical study. In other cases, the tests cover material that test-takers may have learned several years prior and with which they are no longer as familiar. Because few students actually study for a placement exam, the results may not accurately reflect what students know (or could relearn in short order) that is directly related to their proposed field of study.

Another potential problem is that the college and career indicator is not supposed to cover areas already tested by California exams, and tests such as ACCUPLACER and Compass appear to do just that. Alignment studies would be necessary to determine the degree to which placement tests overlap with the content tested by the Smarter Balanced Assessment Consortium and the California High School Exit Examination.

IV. Recommended College and Career Indicator

All of the various potential measures have their strengths and limitations. In other words, no one measure is ideal in all respects. EPIC's ten criteria explore the extent to which each measure under consideration:

- has a *research base* demonstrating a relationship with postsecondary success
- allows for *fair comparisons*
- has *stability*
- has *value for students*
- is *publicly understandable*
- has *instructional sensitivity*
- emphasizes *student performance*, not educational processes
- *minimizes burden*
- provides as much *student coverage* as possible
- recognizes *various postsecondary pathways*

A course-taking behavior measure has a number of advantages including the following:

- a well-developed research base for college-prep courses and the potential to develop a research base for career-related courses
- the ability to examine opportunities and performance by subgroup
- relative stability over time in most schools
- the potential to communicate clearly the status of the school and any efforts it makes to increase student participation in high-quality courses associated with preparedness for college and careers
- a clear and primary focus on content, skills, and competencies that are taught in schools
- a strong emphasis on student performance, particularly if additional credit is given for students who receive a specified grade in the courses, one associated with subsequent postsecondary success
- the potential to minimize the burden on educational agencies and organizations to collect data, although some new data collection systems will need to be developed
- the ability to be fully inclusive of students at all schools, if schools make opportunities to take such classes available to all students

- validation of multiple postsecondary pathways by means of a range of high quality course offerings meeting both a–g requirements and developing CTE skills necessary for participation in postsecondary certificate programs

If the goal is to select a single category of measures consistent with the language of the legislation and with the criteria EPIC researchers developed to screen measures, then the best option is some measure of *course-taking behavior*. While all of the measures examined in the six papers meet some of the legal requirements and review criteria, a course-taking measure does the best job of addressing all of them.

A course-taking measure can include and recognize several types of courses. Courses certified by the UC Office of the President to meet a–g subject requirements address college preparedness, while CTE courses certified by some yet-to-be-determined form of audit process address career preparedness. AP and IB courses contribute to college preparedness, as do dual and concurrent enrollment courses, which, as noted previously, can count for either college or career preparedness depending on their content. The net effect is to create an incentive for high schools to increase their offerings of high-quality, challenging courses that prepare students for a range of postsecondary pathways. It is worth noting that course-taking behavior and grades are the strongest single predictive measure of postsecondary success and that work by schools on improving the quality of their courses leads to the most foundational and sustainable improvements of all the potential measures examined in this paper.

The score a school received on the college and career indicator could potentially be augmented in two ways. First, student contributions to the school college and career indicator score could be weighted in the same fashion that students are weighted for the Local Control Funding Formula. Doing so would create incentives for schools to devote more of the LCFF resources to ensuring that these students take high-quality courses. Second, an additional value could be awarded for all students who met or exceeded a specified grade in qualifying courses. A grade of B or better, for example, in any of these courses might add points to that student’s contribution to the school college and career indicator score in addition to the base points for taking the course and any weighting derived from LCFF categories. Figure 4 presents a potential API formula based on course-taking.

$$\begin{array}{l} \text{course base points +} \\ \text{student LCFF weighting points +} \\ \text{student grade weighting points =} \\ \text{college and career indicator} \end{array}$$

Figure 4. Potential API formula for college and career indicator based on course-taking.

Adding these two levels of weightings would create strong incentives for schools to direct resources and attention to the needs of students who could benefit from more challenging courses. In the process, schools would also have incentives to upgrade the skills of their teaching force in order to offer more of these types of courses. A measure of course-taking would not be as subject to gaming if the quality of courses had to be

certified, as is already the case with the a–g courses, AP courses (via the AP Course Audit), and IB courses through the International Baccalaureate Organization’s quality review processes and requirements.

The primary issue to be addressed for course-taking behavior to be a valid measure of college and career preparedness is a mechanism to ensure course quality. This does require some state engagement, although the process could be independent of the state for the most part. As noted, the UC Office of the President already has a course review and approval system of courses seeking to meet a–g subject requirements as does the College Board for its AP courses, and IB has a quality control process that could be adapted to this purpose (and most IB courses would already meet a–g requirements). This leaves only CTE courses as needing a review process. It is feasible to roll CTE courses into either the a–g or AP review processes, if the sponsors would consent, thereby avoiding the need to create an entirely new review system. Other mechanisms may exist, such as a review by Linked Learning. However, it does seem feasible to have all courses that contribute to the college and career indicator be subject to a quality review of some sort.

Including student grades in college and career preparatory courses is somewhat more problematic, given the wide variation in teacher grading practices. Here, again, the external review process is a potential tool to address this variation. The material submitted for review would need to include sufficient information on grading practices to allow reviewers to ascertain if the grading criteria were sufficiently rigorous and valid to justify inclusion of student grades as potential bonus points in college and career indicator calculation.

The challenge the review process would encounter is the quality of high school syllabi in general. Most currently lack sufficient information to make valid determinations regarding their quality and challenge level. The necessary process of improving syllabi so that they meet review criteria would lead to general course improvement. In this way, the entire process would be more than a simple compliance exercise and would be a kick start for course redesign and refocusing in many instances. Sample and exemplary syllabi could be identified or created to help guide teachers without controlling course content or mandating specific instructional practices. Explicit criterion-based scoring guides that would be used to review courses would be made broadly available, further enhancing course improvement and the approval rate for courses. This process is an example of the state setting quality criteria but the field then administering those criteria in ways that lead to the best quality curriculum and instruction at the local level.

V. Considering a Multiple-Measure System

One conclusion reached by the EPIC researchers’ evaluations of the measures considered is that all of them have potential use in certain situations, but that all have limitations when applied to all students in California in a uniform fashion. This observation suggests that an indicator that incorporated multiple measures could be a more valid representation of college and career preparedness statewide than a single measure.

Although a single measure of college and career preparedness is consistent with the legislation, it would be limited in many ways. Other options exist beyond the recommended measure, course-taking behavior. Other approaches could be constructed around multiple measures. A multiple measure approach offers many advantages but is more challenging in several respects to implement. This section presents examples of potential multiple measures that could conceivably be included in a college and career indicator that would meet the legislative requirements. However, a multiple-measure approach could also go beyond a college and career indicator component for API to serve as a framework for a new type of accountability system, one that captures state and local priorities. This idea is explored further in Section VI.

Combining high-quality elements of the current accountability systems in Georgia, Missouri, and Kentucky (see Conley, Thier, et al., 2014b) informs one approach to redesigning California's API. Georgia's system incorporates the nation's most comprehensive approach to capturing both college and career preparedness. Georgia includes eight separate indicators with a total of 24 measures of postsecondary preparedness, which generates useful information at the state level about preparedness in a wide range of categories. However, its effects on individual schools are more diffuse for the same reason. Schools have a hard time figuring out how to improve their overall rating because of the interaction effects among the various indicators. Will efforts to improve in one area lead to a lower score in another area? Can schools actually organize programs to monitor and improve performance in 24 categories and, if not, how much improvement in how many areas can be managed and will prove to be sufficient?

Figure 5 is an example of multiple measures that aggregate into a college and career indicator, but with far fewer measures than the Georgia example. Nothing in this example is meant to suggest a final model. However, Figure 5 does offer one way to operationalize the college and career indicator component of the API. Other options will be discussed subsequently.

In this example, the standard is set at Level 3 for all measures. Score levels for college admission exams are derived from reported means and standard deviations for the two-section SAT and composite ACT (National Center for Education Statistics, 2009; 2012). Score levels for exams that assess advanced coursework in high school draw upon standards colleges typically set for these measures (AP score ≥ 3 and IB score ≥ 4). Industry Certifications, NOCTI, and an integrated course-taking pathway have no published benchmarks for college and career preparedness, so the levels shown here are for illustrative purposes only.

The model employs a complementary approach at the student level, meaning student performance on any or all of the measures can contribute points to the school college and career indicator, and a compensatory approach at the school level because schools can emphasize a range of strengths by combining multiple measures in a purposive and intentional fashion. A multitiered, multidimensional model offers greater possibilities for achieving the accuracy, consistency, and reliability necessary for constituents to value the results, for schools to feel they can improve their performance over time, and for policymakers to understand better what is happening in schools

statewide (Brookhart, 2009; Gong & Hill, 2001; Marion & Gong, 2003). This approach captures elements of performance (scores on tests) and of participation (number of challenging courses taken). The compensatory component of this approach takes a step down the road of recognizing and validating local variation in emphasis on college and career preparedness, and also creates incentives for schools to pursue both.

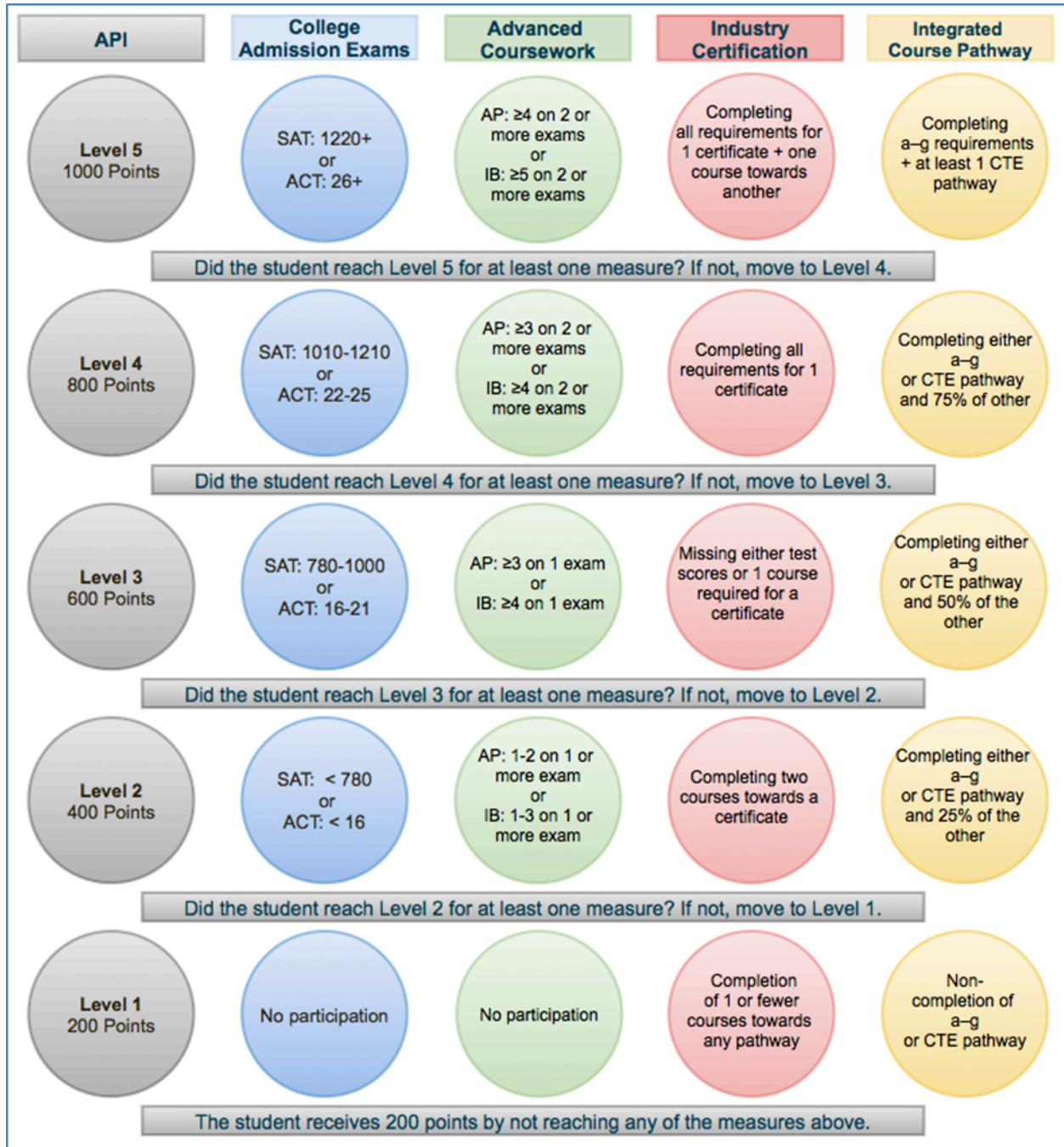


Figure 5. College and career indicator example.

Recognition Approaches

Two states provide examples of two different approaches to recognizing schools that perform well in their accountability system.

Oklahoma employs a bonus structure (see Conley, Thier, et al., 2014b) that could be incorporated to create incentives for schools to improve performance on one or more measures. For example, schools could receive 20 bonus API bonus points for each measure on which they exceeded a specified improvement rate. For example, if a school had a base API score of 740 points and then increased by a specified percent in students taking a dual-enrollment course and graduates receiving the California Seal of Biliteracy, that school's API score would receive bonus points from a table designating the number of bonus points in relation to the percent increase on these measures.

Texas uses a system of performance categories to recognize schools that rate at the highest levels in its accountability system (see Conley, Thier, et al., 2014b). Schools scoring within the top quartiles for each measure are recognized as Gold. Those in the next quartile are designated Silver. Recognizing schools in this fashion provides for additional recognition of quality beyond the API's single number. Schools could also be recognized for amount of improvement and for relative value added to student preparedness for college and careers.

VI. Creating a Coherent Accountability System

The API does not exist in a vacuum; quite the contrary. In fact, California schools have long attempted to meet state and federal accountability requirements that were similar to but not the same as California's own standards. With the introduction now of an additional level of accountability at the district level in the form of Local Control Accountability Plans (LCAPs), educators will be challenged to manage a process that could conceivably send conflicting messages but also could be more relevant and valuable locally. Without a measure of coherence, it is almost certain that accountability in California will become unwieldy and unmanageable in complexity and result in educators retreating to a compliance-based follow-the-rules mentality. This type of thinking rarely, if ever, leads to significant improvements in schooling. The net effect will be to defeat what should be one of the primary purposes of accountability, namely, to improve educational practices in ways that result in enhanced student learning outcomes.

Interacting Accountability Components

In addition to the revised API, California's K–12 system of accountability is marked by the introduction of several new components, including LCAPs and the Smarter Balanced Assessment Consortium (SBAC) assessments. These are overlaid on existing, well-established measures such as high school graduation requirements and graduation rates. These varying accountability elements interact in ways that may or may not result in improved education for California schools. Any accountability

system that incorporates multiple measures needs to consider the interaction among measures, their relative importance to educators, and their likely effects on practice, both intended and unintended.

This section briefly outlines several key accountability demands and measures to which schools must respond, then considers a model that suggests the distinct state and local responsibilities for school accountability. The intent here is not to lay out the design of a comprehensive, coherent model per se but to demonstrate the type of relationships that may be necessary for the state to achieve the goal of accountability that informs state policy decisions, drives improvement at the local level, and is owned locally.

Local Control Funding Formula, Local Control Accountability Plans, and the California Collaborative for Excellence in Education

Funding in California has for the past 50+ years been characterized by a proliferation of categorical programs that direct funds to specific purposes or programs. The Local Control Funding Formula (LCFF) replaces this categorical funding system with a model that shifts most spending authority from the state to the local level. Each school district receives a base grant. Supplemental and concentration grants are allotted to schools based on their need, defined by the percentage of students in poverty, English-language learners, and foster children. Schools then are expected to use their spending authority to allocate funds to priorities and programs that improve student performance, especially for high-need students, who constitute 63% of California K–12 student population. Each district designs an LCAP that specifies how the district will address eight state priority areas. The LCAP is the vehicle by which the local district translates state goals into local responses.

The LCAP is developed with input from parents, educators, and the community. The annual goals span a three-year period to address the eight state priorities in three categories: conditions of learning, pupil outcomes, and engagement.²⁰ The LCFF and LCAP are still in the definition and specification stages. The State Board of Education will adopt evaluation rubrics that govern how the State Superintendent of Public Instruction will intervene when districts fail to improve or achieve their annual goals. In addition, the rubrics are to help LEAs and charter schools to evaluate their strengths and weakness, to identify districts and schools that need assistance, and for the Superintendent to identify districts in need of intervention.

The state has created the California Collaborative for Educational Excellence (CCEE), an agency with the mission of providing high-quality technical assistance to districts in need of improvement. The CCEE, still under development, is intended to serve as a centralized entity that can help districts improve, not to punish them, when districts are consistently not meeting state priorities.

²⁰ The eight priority areas are: (1) Basic, (2) Implementation of Common Core State Standards, (3) Parental Involvement, (4) Pupil Achievement, (5) Pupil Engagement, (6) School Climate, (7) Course Access, and (8) Other Pupil Outcomes.

Multiple Levels of Accountability and the Challenge of Compliance Versus Goal-Oriented Mindsets

Districts and schools will be subject to federal requirements specified in the Elementary and Secondary Education Act, state requirements as represented in part by the API, and local accountability standards manifested in the LCAPs. The existence of three levels of accountability almost ensures either some level of redundancy or of differing expectations across levels. The state expects pupil achievement as reported in LCAPs to be gauged by performance on standardized tests, score on the API, share of pupils who are prepared for college and career, share of English learners who become English proficient, English learner reclassification rate, share of pupils who pass AP exams with 3 or higher, and share of pupils determined to be prepared for college by the Early Assessment Program.

Federal requirements will be met largely by scores on the SBAC exams, which will contribute information that can be used to determine performance for some, but not all, of the state priorities. The SBAC assessment replaces California's Standardized Testing and Reporting system (STAR). SBAC assesses content knowledge attainment in mathematics and English language arts as well as cognitive processing skills related to completing performance tasks. The SBAC assessments include multiple choice and performance task questions and have descriptors for four achievement levels. Level three is designated as college-prepared. The career-prepared level will likely be a state-by-state decision. These cut scores will be used to judge schools for the standardized test portion of the API.

The LCAPs will potentially incorporate a wider and more diverse set of measures that could include course-taking patterns, attendance, graduation rates, measures, proportion of students completing a college application, and assessments in other subject areas. Or, districts could choose simply to adopt federal and state requirements and not go beyond them to meet as many of the eight state priority areas as possible, although it seems likely it would still be necessary for districts to adopt some additional measures and metrics.

The key question is whether this potential thicket of measures will overwhelm districts and cause them simply to do what they have always done and let the cards fall where they may, or whether these three accountability systems can be made to be complementary in ways that concentrate the energy and effort of educators in a laser-like focus on areas of most importance to students in greatest need. It will be very easy for districts to adopt a compliance mindset rather than engaging in the more challenging process of making organizational decisions that reflect a goal-oriented perspective on improvement.

If districts adopt the lowest common denominator approach to accountability, the results will likely be similar to those seen in most places across the country for the past 15 years. The evidence suggests that some system-level improvements have been made in reading and mathematics at the elementary level. However, far less evidence can be found to demonstrate overall sustained, multiyear improvement by most schools. This is particularly true of high schools, which are in many respects the most in need of

dramatic improvement. At least part of the explanation is that educators may perceive their potential to make real changes and improvements as highly limited. By attending only to outcomes in a few narrow areas, educators may be neglecting improvements in the organizational and learning processes that will lead to changes in educational outcomes. A compliance mindset precludes thinking about the systems issues that must be addressed to achieve the broader goal of well-educated students prepared for more challenging futures.

A State/Local Model

The options presented to this point have been built around a set of common state-level indicators that also address federal requirements, for the most part. Some options allow combining indicators in various ways; others report results in ways that go beyond a single score. But in all cases, the pool of accountability indicators is consistent statewide. In part, this is due to the language of the law that governs the creation of the college and career indicator. This section entertains an option not expressly entertained by the law, but one that perhaps holds promise to strike a balance between state and local needs. That option is a state/local model that incorporates a common set of core measures at the state level and allows for local variation in accountability measures in addition to the common state measures.

The basic principle of a state/local model is that the state has a set of core measures that are consistently applied to all schools and that local schools in addition have the ability to add measures that best reflect the quality of their programs and the areas in which they wish to demonstrate improvement. The state measures continue to focus on foundational skills such as reading and mathematics and on a few key organizational health indicators, such as attendance and graduation rates. The goal of these indicators is to establish that all schools are meeting the basic needs of all students and that students are present and able to take advantage of the school's programs.

The local measures address state priorities, as the LCAPs are designed to do currently. The LCAP process is scaled down to the school level so that each school has a set of improvement goals aligned with state priorities that go beyond federal requirements. These measures are compiled and scaled to allow them to be aggregated and combined with the state measures into a more comprehensive picture of school performance. Although the natural tendency here is to generate one score and one overall rating, the distinguishing difference of this approach is that it allows for and encourages more of a profile approach to reporting school performance.

As a practical matter, a multiple-measure approach to accountability might work something like this. Rather than focusing solely on the performance level of a school, the state indicators would also track the trend line of a school on core indicators of a foundational education. These would include reading and math scores based on examinations that took into account the deeper dimensions of mastery and application of these topics and other organizational health indicators such as graduation and attendance rates. The results would be reported both as a performance score

(approaches, meets, exceeds) and as a trend score (improving, static, declining) in the form of a rolling three-year average. In this way, schools could be recognized for improvement as much as for current performance. Doing so would enhance the sense of self-efficacy among school staffs, particularly at lower-performing schools, and would challenge higher-performing schools to continue to improve.

The second piece would be local measures. The district LCAP model would be extended to the school level, and staff and community members at each school would identify the state priority areas for which they were holding the school accountable to improve. The state could set certain of the eight priorities as mandatory for all schools (college and career preparedness would be an example) while allowing some choice among other priorities, based on the school's circumstances. Eight areas is probably more than any school can address at once anyway, so providing some choice would likely not have a significant effect on the actual amount of improvement occurring at each school.

Having identified key areas, educators at each school would then be responsible to specify the educational outcomes they were committed to achieving and the measures they would use to determine achievement. For example, in the area of college and career preparedness, some schools could target increases in AP course participation while others emphasize high-quality CTE/Linked Learning courses for all students. Some might adopt an IB program for all, while others might develop close relationships with the local community college and emphasize expanded dual and concurrent enrollment opportunities. Others might choose to have all students enroll in a–g courses.

The current measure of accumulating points on a per-student basis could still be employed with this complementary method. Schools would calculate their points based on the method that yielded the most points for each individual student and, in the example of college and career preparedness, could receive multiple points for the student who engaged in more than one of the areas for which the school was calculating a metric. Once again, as in the case of the state-level indicators, trends would be reported so that schools would have both a performance and trend score for each priority area.

The example in Table 9 paints a more complex picture of a school. This school is meeting reading performance targets, but its trend line is flat. Math scores are below the target but are improving. Students are attending at a level that meets standard, and their attendance is improving, but they are not yet graduating at an acceptable rate nor are they improving the graduation rate over time. They are, however, successful in getting more students than the standard to apply to college, and that rate is remaining constant.

This profile suggests the school is taking steps to improve college and career preparedness by getting more students to apply to college, supporting community involvement around college and careers, improving math performance, maintaining strong attendance, and delivering positive results on reaching performance. At the same time, some challenges can be identified. These include math performance, graduation rate, and student aspirations.

Table 9. Example of a School-Level Accountability Measure for State-Level Priority Areas and Selected Local Priorities

Sample School: Grade 11	Performance Score	Trend Score
State-level Indicator 1. Reading	Meets	Static
State-level Indicator 2. Mathematics	Approaches	Improving
State-level Indicator 3. Attendance	Meets	Improving
State-level Indicator 4. Graduation Rate	Approaches	Declining
State-level Indicator 5. Application Rate	Exceeds	Static
LCAP Indicator 1. College/Career Preparedness	Approaches	Improving
LCAP Indicator 2. Community Involvement	Meets	Improving
LCAP Indicator 3. Student Interest Explorations to Increase Aspirations	Approaches	Static

This information, in profile form, can be easily communicated to staff and community in ways that help focus the school on key improvement areas. This type of goal-oriented improvement is more powerful than simply seeking to raise a test score through any means available. A profile approach takes into account more of the variables and factors that affect student performance and begins to view the school as a learning system rather than one or two test scores.²¹

VII. Conclusion

Holding schools accountable for student performance based solely on educational outputs has proven to be challenging and nowhere near as effective as policymakers had hoped it would be. Accountability in the future will likely be more of a partnership between the state and local schools and will include more dimensions and measures than a single test in reading and mathematics. The college and career indicator that is being added to the API is a small step in that direction, but much more work will remain to be done to create an accountability system that leads to sustained improvement of educational practice and not just a postmortem perspective on current practice.

Linda Darling-Hammond, Gene Wilhoit, and Linda Pittenger (2014) lay out a new approach to accountability for learning in which all parties—students, educators, and policymakers—share the accountability burden. It is based on measures of what the authors describe as “meaningful learning,” combined with measures of the professional capacity of staff to deliver meaningful learning to all students and indicators of resource adequacy sufficient to hold educators and students accountable for achieving state goals. This notion goes well beyond a multimeasure API, but at its heart it is based on

²¹ For a description of profiles as applied to students more broadly, see “New Conceptions of College and Career Ready: A Profile Approach to Admission,” by D. T. Conley, 2014, *The Journal of College Admission* (223), pp. 12–23.

the notion that multiple measures are key to ensuring that the type of learning that is being encouraged is what students need in order to be prepared for college and career. Resource allocation includes allocating adequate resources to schools, but also entails the responsibility of schools to offer high-quality curriculum and instructional materials to all students and to ensure that all teaching staff are properly prepared to meet student needs and continuously add new skills.

Including a college and career indicator in the API is a positive first step in the process of expanding notions of accountability beyond scores on reading and math tests. It is, however, only a first step. As the deliberations of the Public School Accountability Act Advisory Committee illustrate, it is not a simple task to identify one indicator from among perhaps a dozen possibilities, each of which hold the potential to offer insights into college and career preparedness, but not necessarily the same insights for all schools.

When searching for accountability elements, the desire for simplicity needs to be balanced against the complex and dynamic nature of schools as systems and the tremendous variation among California's schools in terms of the make-up of their student population and the values and priorities of the local communities. While all students need to meet key baseline performance measures in which the state has a strong and clear vested interest, success for most students will be defined by a wider range of factors, many of which will be determined by the decisions local educators make regarding the programs and opportunities they choose to offer, their efforts to continuously improve their curricular and instructional programs, and the support services they provide to their students and community. This broader, more expansive notion of an effective school will require a larger toolbox of measures and the ability to have some of those measures selected locally if accountability is ever to be owned fully by educators and result in the type of schools policymakers and the public want and expect.

References

- ACT. (2008a). *WorkKeys: Locating Information technical manual*. Iowa City, IA: Author.
- ACT. (2008b). *WorkKeys: Reading for Information technical manual*. Iowa City, IA: Author.
- ACT. (2008c). *WorkKeys: Applied Mathematics technical manual*. Iowa City, IA: Author.
- ACT. (2013). *The condition of college & career readiness 2013: African American students*. Retrieved from <http://www.act.org/newsroom/data/2013/states/pdf/AfricanAmerican.pdf>
- ACT. (2014a). *The condition of college & career readiness 2014: National*. Retrieved from <http://www.act.org/research/policymakers/cccr14/pdf/CCCR14-NationalReadinessRpt.pdf>
- ACT. (2014b). ACT WorkKeys: Overview. Retrieved from <http://www.act.org/products/workforce-act-workkeys>
- ACT. (2014c). Boosting health care employees' skills and promotion opportunities. Retrieved from <http://www.act.org/workforce/case/owensboro.html>
- ACT. (2014d). Northrop Grumman Ship Systems develops a ready-to-train workforce. Retrieved from <http://www.act.org/workforce/case/northrop.html>
- Ackerman, P. L., Kanfer, R., & Calderwood, C. (2013). High school Advanced Placement and student performance in college: STEM majors, non-STEM majors, and gender differences. *Teachers College Record*, 115(10), 1–43.
- Adelman, C. (1999). *Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment*. Washington, DC: U.S. Department of Education.
- Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college*. Washington, DC: U.S. Department of Education.
- Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, 80(2), 207–245.
- Armstrong, P. W., & Rogers, J. D. (1997). Basic skills revisited: The effects of foreign language instruction on reading, math, and language arts. *Learning Languages*, 2(3), 20–31.
- Atkinson, R. C., & Geiser, S. (2009). Reflections on a century of college admissions tests. *Educational Researcher*, 38(9), 665–676.
- Bailey, T., Jeong, D. W., & Cho, S. W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review*, 29(2), 255–270.

- Bailey, T. R., Hughes, K. L., & Karp, M. M. (2002). *What role can dual enrollment programs play in easing the transition between high school and postsecondary education?* New York, NY: Teachers College, Columbia University.
- Berry, C. M., & Sackett, P. R. (2009). Individual differences in course choice result in underestimation of the validity of college admissions systems. *Psychological Science, 20*(7), 822–830.
- Bettinger, E. P., Evans, B. J., & Pope, D. G. (2011). *Improving college performance and retention the easy way: Unpacking the ACT exam* (NBER Working Paper No. 17119). Cambridge, MA: National Bureau of Economic Research.
- Bowles, F. E. (2004). *WorkKeys assessments and their validity as academic success predictors* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3157119)
- The Broad Foundations. (2013). *The road to equity: Expanding AP access and success for African-American students*. Los Angeles, CA: Author.
- Brookhart, S. M. (2009). The many meanings of "multiple measures." *Educational Leadership, 67*(3), 6–12.
- Buchmann, C., Condrón, D. J., & Roscigno, V. J. (2010). Shadow education, American style: Test preparation, the SAT and college enrollment. *Social Forces, 89*(2), 435–461.
- Bureau of Labor Statistics. (2014). Occupational outlook handbook: Interpreters and translators. Retrieved from <http://www.bls.gov/ooh/media-and-communication/interpreters-and-translators.htm>
- Byun, S. Y., & Park, H. (2012). The academic success of East Asian American youth: The role of shadow education. *Sociology of Education, 85*(1), 219–239.
- Camara, W. J., & Echternacht, G. (2000). *The SAT [R] I and high school grades: Utility in predicting success in college* (Research Notes RN-10). New York, NY: The College Board.
- Campbell, J. P., & Knapp, D. J. (Eds.). (2001). *Exploring the limits in personnel selection and classification*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Card, D., & Rothstein, J. (2007). Racial segregation and the black–white test score gap. *Journal of Public Economics 91*(11), 2158–2184.
- Carroll, S. J., Krop, C., Arkes, J., Morrison, P. A., & Flanagan, A. (2005). *California's K-12 public schools: How are they doing?* Santa Monica, CA: RAND Corporation.
- Casparly, K. (2011). *Research brief: Postsecondary enrollment patterns of IB certificate and diploma candidates from U.S. high schools*. Menlo Park, CA: SRI International.
- Casparly, K., & Bland, J. (2011). *Research brief: First college courses taken by Florida IB students*. Menlo Park, CA: SRI International.

- Catterall, J. S. (2009). *Doing well and doing good by doing art*. Los Angeles, CA: I-Group Books.
- Center for Advanced Research and Technology. (2011). *A model for success: CART's Linked Learning program increases college enrollment*. Clovis, CA: Author.
- Center on Education Policy. (2013). *Assessment profiles*. Retrieved from http://www.cep-dc.org/cfcontent_file.cfm?Attachment=AssessmentsProfiles_CareerReadiness_10.29.13.pdf
- Chajewski, M., Mattern, K., & Shaw, E. J. (2011). Examining the role of Advanced Placement® exam participation in 4-year college enrollment. *Educational Measurement: Issues and Practice* 30(4), 16–27.
- Coca, V., Johnson, D., Kelley-Kemple, T., Roderick, M., Moeller, E., Williams, N., & Moragne, K. (2012). *Working to my potential: The postsecondary experiences of CPS students in the International Baccalaureate Diploma Programme*. Chicago, IL: The University of Chicago Consortium on Chicago School Research.
- College Board. (2014a). *2014 college-bound seniors: Total group profile report: Total group*. Retrieved from <https://secure-media.collegeboard.org/digitalServices/pdf/sat/TotalGroup-2014.pdf>
- College Board. (2014b). *10 years of Advanced Placement exam data show significant gains in access and success; Areas for improvement*. Retrieved from <https://www.collegeboard.org/releases/2014/class-2013-advanced-placement-results-announced>
- College Board. (2014c). The College Board announces bold plans to expand access to opportunity; Redesign of the SAT. Retrieved from <https://www.collegeboard.org/releases/2014/expand-opportunity-redesign-sat>
- Conley, D. T. (2012). Does common mean “the same”? Implementing new state standards and assessments equitably and fairly. *The State Education Standard*, 13(1), 23–28.
- Conley, D. T. (2013). *Getting ready for college, careers, and the Common Core: What every educator needs to know*. New York, NY: John Wiley & Sons.
- Conley, D. T., Beach, P., Thier, M., Lench, S. C., & Chadwick, K. L. (2014a). Measures for a college and career indicator: College admission exams. Eugene, OR: Educational Policy Improvement Center.
- Conley, D. T., Beach, P., Thier, M., Lench, S. C., & Chadwick, K. L. (2014b). Measures for a college and career indicator: Innovative measures. Eugene, OR: Educational Policy Improvement Center.
- Conley, D. T., Beach, P., Thier, M., Lench, S. C., & Chadwick, K. L. (2014c). Measures for a college and career indicator: Course-taking behavior. Eugene, OR: Educational Policy Improvement Center.

- Conley, D. T., Beach, P., Thier, M., Lench, S. C., & Chadwick, K. L. (2014d). Measures for a college and career indicator: Career preparedness assessments. Eugene, OR: Educational Policy Improvement Center.
- Conley, D. T., & Darling-Hammond, L. (2013). *Creating systems of assessment for deeper learning*. Stanford, CA: Stanford Center for Opportunity Policy in Education.
- Conley, D. T., & McGaughy, C. (2012). College and career readiness: Same or different? *Educational Leadership*, 69(7), 28–34.
- Conley, D. T., McGaughy, C., Brown, D., van der Valk, A., & Young, B. (2009). *Validation study III: Alignment of Texas College and Career Readiness Standards with courses in two career pathways*. Eugene, OR: Educational Policy Improvement Center.
- Conley, D. T., McGaughy, C., Davis-Molin, W., Farkas, R., & Fukuda, E. (2014). *International Baccalaureate Diploma Programme: Examining college readiness*. Eugene, OR: Educational Policy Improvement Center.
- Conley, D. T., Thier, M., Beach, P., Lench, S. C., & Chadwick, K. L. (2014a). Measures for a college and career indicator: Advanced coursework. Eugene, OR: Educational Policy Improvement Center.
- Conley, D. T., Thier, M., Beach, P., Lench, S. C., & Chadwick, K. L. (2014b). Measures for a college and career indicator: Multiple measures. Eugene, OR: Educational Policy Improvement Center.
- Conley, D. T. (2014g). New conceptions of college and career ready: A profile approach to admission. *The Journal of College Admission* (223), 12–23.
- Carroll, S. J., Krop, C., Arkes, J., Morrison, P. A., & Flanagan, A. (2005). *California's K-12 public schools: How are they doing?* Santa Monica, CA: RAND Corporation.
- Council for Aid to Education. (2014). "CWRA+ overview." Retrieved September 5, 2014, from <http://cae.org/participating-institutions/cwra-overview/>
- Darling-Hammond, L., & Adamson, F. (2010). Beyond basic skills: *The role of performance assessment in achieving 21st century standards of learning*. Stanford University, CA: Stanford Center for Opportunity Policy in Education.
- Darling-Hammond, L., & Pecheone, R. (2009). Reframing accountability: Using performance assessments to focus learning on higher-order skills. In L. M. Pinkus (Ed.), *Meaningful measurement: The role of assessments in improving high school education in the twenty-first century* (pp. 25–53). Washington, DC: Alliance for Excellent Education.
- Darling-Hammond, L., Wilhoit, G., & Pittenger, L. (2014). Accountability for college and career readiness: Developing a new paradigm. *Education Policy Analysis Archives*, 22(86).
- Dayton, C., Hester, C. H., & Stern, D. (2011). *Profile of the California Partnership Academies, 2009-10*. Berkeley, CA: University of California.

- Dixon-Román, E. J., Everson, H. T., & McArdle, J. J. (2013). Race, poverty and SAT scores: Modeling the influences of family income on black and white high school students' SAT performance. *Teachers College Record*, 115(4).
- Dodd, B. G., Fitzpatrick, S., & De Ayala, R. J. (2002). *An investigation of the validity of AP grades of 3 and a comparison of AP and non-AP graduate groups* (Report No. 2002–9). New York, NY: College Entrance Examination Board.
- Dougherty, C., Mellor, L., & Jian, S. (2006). *The relationship between advanced placement and college graduation* (2005 AP Study Series, Report 1). Austin, TX: The National Center for Accountability.
- Dumas, L. S. (1999). Learning a second language: Exposing your child to a new world of words boosts her brainpower, vocabulary, and self-esteem. *Child*, 72(74), 76–77.
- Educational Policy Improvement Center. (2014). *National Assessment of Educational Progress grade 12 preparedness research project: College course content analysis*. Eugene, OR: Author.
- Ehley, L. (2006). *Digital portfolios: A study of undergraduate student and faculty use and perceptions of Alverno College's diagnostic digital portfolio* (Doctoral dissertation). Cardinal Stritch University, Milwaukee, WI.
- Finkelstein, N., Fong, A., Tiffany-Morales, J., Shields, P., & Huang, M. (2012). *College bound in middle school and high school? How math course sequences matter*. Sacramento, CA: The Center for the Future of Teaching and Learning at WestEd.
- Finkelstein, N. D., & Fong, A. B. (2008). *Course-taking patterns and preparation for postsecondary education in California's public university systems among minority youth*. (REL 2008-No. 035). Sacramento, CA: Regional Educational Laboratory West.
- Foster, J. C., & Pritz, S. G. (2006). The certification advantage. *Techniques*, 81(1), 14–20.
- Foster, J. C., & Pritz, S. G. (2006). The certification advantage. *Techniques*, 81(1), 14–20.
- Geiser, S., & Santelices, V. (2004). The role of advanced placement and honors courses in college admissions. In P. Gandara, G. Orfield, & C. Horn (Eds.), *Expanding opportunity in higher education: Leveraging promise* (pp. 75–113). Albany: State University of New York.
- Geiser, S., & Studley, W. R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, 8(1), 1–26.
- Goldschmidt, P., Martinez, J. F., Niemi, D., & Baker, E. L. (2007). Relationships among measures as empirical evidence of validity: Incorporating multiple indicators of achievement and school context. *Educational Assessment*, 12(3–4), 239–266.
- Gong, B., & Hill, R. (2001). *Some considerations of multiple measures in assessment and school accountability* [PowerPoint slides]. Retrieved from http://www.nciea.org/publications/MultiMeasures_Gong01.pdf

- Goodman, J. (2012). *The labor of division: Returns to compulsory math coursework* (RWP12-032). Cambridge, MA: Harvard University John F. Kennedy School of Government.
- Grissmer, D. W. (2000). The continuing use and misuse of SAT scores. *Psychology, Public Policy, and Law*, 6(1), 223–232.
- Haley, K. A. (2013). *A comparative study of high school Advanced Placement and dual enrollment programs using a mixed methods analysis* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3599429)
- Halic, O. (2013). *Postsecondary educational attainment of IB Diploma Programme candidates from US high schools*. Geneva, Switzerland: International Baccalaureate Organization.
- Hamilton, L. S., Stecher, B. M., & Klein, S. P. (Eds.). (2002). *Making sense of test-based accountability in education*. Pittsburgh, PA: Rand Corporation.
- Hargrove, L., Godin, D., & Dodd, B. (2008). *College outcomes comparisons by AP and non-AP high school experiences*. New York, NY: The College Board.
- Hart Research Associates. (2013). *It takes more than a major: Employer priorities for college learning and student success*. Washington, DC: Author.
- Hendrick, R. Z. (2006). *Evaluating WorkKeys profiling as a pre-employment assessment tool to increase employee retention* (Unpublished doctoral dissertation). Old Dominion University, Norfolk, VA.
- Hogan, T. P. (2007). *Psychological testing: A practical introduction* (2nd ed.). New York, NY: Wiley.
- Hojay, M., Erdmann, J. B., Veloski, J. J., Nasca, T. J., Callahan, C. A., Julian, E., & Peck, J. (2000). A validity study of the writing sample section of the Medical College Admission Test. *Academic Medicine*, 75(10), S25–S27.
- Huws, N., Reddy, P. A., & Talcott, J. B. (2009). The effects of faking on non-cognitive predictors of academic performance in University students. *Learning and Individual Differences*, 19(4), 476–480.
- Iatarola, P., Conger, D., & Long, M. C. (2011). Determinants of high schools' advanced course offerings. *Educational Evaluation and Policy Analysis*, 33(3), 340–359.
- Karp, M. J. M., Bailey, T. R., Hughes, K. L., & Fermin, B. J. (2004). *State dual enrollment policies: Addressing access and quality*. Washington, DC: U.S. Department of Education, Office of Vocational and Adult Education.
- Karp, M. M., Calcagno, J. C., Hughes, K. L., Jeong, D. W., & Bailey, T. R. (2007). The postsecondary achievement of participants in dual enrollment: An analysis of student outcomes in two states. New York, NY: Community College Research Center, Columbia University.
- Kemple, J. J. (2004). *Career academies: Impacts on labor market outcomes and educational attainment*. New York, NY: MDRC.

- Kemple, J. J., & Willner, C. J. (2008). *Career academies: Long-term impacts on labor market outcomes, educational attainment, and transitions to adulthood*. New York, NY: MDRC.
- Kim, J. (2008). *The impact of dual and articulated credit on college readiness and total credit hours in four selected community colleges: Excerpts from a doctoral dissertation literature review* (Doctoral dissertation). Champaign, IL: Office of Community College Research and Leadership, University of Illinois.
- Klopfenstein, K., & Thomas, M. K. (2005, January). The advanced placement performance advantage: Fact or fiction. In M. Ferreyra (Chair), *Women's Acquisition of Human Capital*. Symposium conducted at the annual meeting of the Allied Social Sciences Association, Philadelphia, PA.
- Klopfenstein, K., & Thomas, M. K. (2009). The link between advanced placement experience and early college success. *Southern Economic Journal*, 75(3), 873–891.
- Klopfenstein, K., & Thomas, M. K. (2010). Evaluating the policies of states and colleges. In P. M. Sadler, G. Sonnert, R. H. Tai, & K. Klopfenstein (Eds.), *AP: A critical examination of the Advanced Placement program* (pp. 167–188). Cambridge, MA: Harvard Education Press.
- Kobrin, J. L., Patterson, B. F., Barbuti, S. M., Mattern, K. D., & Shaw, E. J. (2008). *Validity of the SAT for predicting first-year college grade point average* (Research Report No. 2008-5). New York, NY: The College Board.
- Lane, S. (2010). *Performance assessment: The state of the art*. Stanford, CA: Stanford Center for Assessment, Learning and Equity.
- Lekes, N., Bragg, D., Loeb, J. W., Oleksiw, C. A., Marszalek, J., Brooks-LaRaviere, M.,...Hood, L. K. (2007). *Career and technical education pathway programs, academic performance, and the transition to college and career*. St. Paul, MN: University of Minnesota National Research Center for Career and Technical Education.
- Lemann, N. (1999). *The big test: The secret history of the American meritocracy*. New York, NY: Farrar, Straus, Giroux.
- Levine, P. B., & Zimmerman, D. J. (1995). The benefit of additional high-school math and science classes for young men and women. *Journal of Business & Economic Statistics*, 13(2), 137–149.
- Lindon, J. (2010). *Utilizing WorkKeys as a measure of community and technical college student success* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3412661)
- Marion, S., & Gong, B. (2003). *Evaluating the validity of state accountability systems* [PowerPoint slides]. Retrieved from http://www.nciea.org/publication_PDFs/CCSSO_Validity_BGSM03.pdf

- Maxwell, N. L., & Rubin, V. (1997). *The relative impact of a career academy on post-secondary work and education skills in urban, public high schools*. Berkeley, CA: University of California National Center for Research in Vocational Education.
- McDonnell, L. M. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis*, 17(3), 305–322.
- Mikulecky, M., & Christie, K. (2014). *Rating states, grading schools: What parents and experts say states should consider to make school accountability systems meaningful*. Retrieved from <http://www.ecs.org/docs/rating-states,grading-schools.pdf>
- Morgan, R. (1989). *Analysis of the predictive validity of the SAT and high school grades from 1976 to 1985* (Report No. 89-7). New York, NY: College Entrance Examination Board.
- Morgan, R., & Klaric, J. (2007). AP students in college: An analysis of five-year academic careers (Report No. 2007-4). New York, NY: The College Board.
- Muller, R. D., & Beatty, A. (2008). *Work readiness certification and industry credentials: What do state high school policy makers need to know?* Retrieved from <http://www.achieve.org/files/WorkReadinessCertificationandIndustryCredentials.pdf>
- National Center for Education Statistics. (2009). ACT score averages and standard deviations, by sex and race/ethnicity, and percentage of ACT test takers, by selected composite score ranges and planned fields of study: Selected years, 1995 through 2010. Retrieved from http://nces.ed.gov/programs/digest/d10/tables/dt10_155.asp
- National Center for Education Statistics. (2012). Fast Facts: SAT Scores. Retrieved from <http://nces.ed.gov/fastfacts/display.asp?id=171>
- National Occupational Competency Testing Institute. (2012). *Technical manual: Accounting basic*. Big Rapids, MI: Author.
- National Occupational Competency Testing Institute. (2014). National College Credit Recommendation Service (NCCRS). Retrieved from <http://www.nocti.org/nccrs.cfm>
- Ng, T. W., Eby, L. T., Sorensen, K. L., & Feldman, D. C. (2005). Predictors of objective and subjective career success: A meta-analysis. *Personnel Psychology*, 58(2), 367–408.
- Noble, J., & Sawyer, R. (2002). *Predicting different levels of academic success in college using high school GPA and ACT composite score* (ACT Research Report Series 2002-4). Iowa City, IA: ACT.
- Oakes, J., & Saunders, M. (2004). Education's most basic tools: Access to textbooks and instructional materials in California's public schools. *The Teachers College Record*, 106(10), 1967–1988.

- Patterson, B. F., Kobrin, J. L., & Packman, S. (2011). *Advanced Placement exam-taking and performance: Relationships with first-year subject area college grades*. New York, NY: The College Board.
- Rodriguez, D., Carrasquillo, A., & Lee, K. S. (2014). *The bilingual advantage: Promoting academic development, biliteracy, and native language in the classroom*. New York, NY: Teachers College Press.
- Rose, H., & Betts, J. R. (2004). The effect of high school courses on earnings. *Review of Economics and Statistics*, 86(2), 497–513.
- Rothstein, J. M. (2004). College performance predictions and the SAT. *Journal of Econometrics*, 121(1), 297–317.
- Ryan, C. (2013). *Language use in the United States: 2011 (ACS-22)*. Retrieved from <http://www.census.gov/prod/2013pubs/acs-22.pdf>
- Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., & Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance? *Psychological Bulletin*, 135(1), 1–22.
- Sadler, P. M., & Tai, R. H. (2007). Advanced Placement exam scores as a predictor of performance in introductory college biology, chemistry and physics courses. *Science Educator* 16(2), 1–19.
- Sanchez, E. D. (2013). *Differential effects of using ACT college readiness assessment scores and high school GPA to predict first-year college GPA among racial/ethnic, gender, and income groups* (ACT Research Report Series 2013-4). Iowa City, IA: ACT.
- Schwartz, M. S., Sadler, P. M., Sonnert, G., & Tai, R. H. (2009). Depth versus breadth: How content coverage in high school science courses relates to later success in college science coursework. *Science Education*, 93(5), 798–826.
- Scott-Clayton, J., & Rodriguez, O. (2012). *Development, discouragement, or diversion? New evidence on the effects of college remediation* (No. 18328). Cambridge, MA: National Bureau of Economic Research.
- Shah, S., Dean, M., & Chen, Y. C. (2010). *Academic performance of IB students entering the University of California System from 2000-2002*. Geneva, Switzerland: International Baccalaureate Organization.
- Shaw, E. J., Marini, J. P., & Mattern, K. D. (2013). Exploring the utility of Advanced Placement participation and performance in college admission decisions. *Educational and Psychological Measurement*, 73(2), 229–253.
- Smith, D. (2007). Why expand dual-credit programs? *Community College Journal of Research and Practice*, 31(5), 371–387.
- Soveri, A., Laine, M., Hamalainen, H., & Hugdahl, K. (2011). Bilingual advantage in attentional control: Evidence from the forced-attention dichotic listening paradigm. *Bilingualism: Language and Cognition*, 14(3), 371–378.

- Speroni, C. (2011). *Determinants of students' success: The role of Advanced Placement and dual enrollment programs*. New York, NY: National Center for Postsecondary Research.
- Spurling, S., & Gabriner, R. (2002). *The effect of concurrent enrollment programs upon student success at City College of San Francisco: Findings*. Retrieved from http://www.ccsf.edu/Offices/Research_Planning/pdf/eceps02.pdf
- Stanford Center for Assessment Learning and Equity. (2014). Envision Schools college success portfolio. Retrieved September 5, 2014, from <https://scale.stanford.edu/content/envision-schools-college-success-portfolio>
- Sung, H., Padilla, A. M., & Silva, D. M. (2006). Foreign language education, academic performance, and socioeconomic status: A study of California schools. *Foreign Language Annals*, 39(1), 115–130.
- Swaney, K. B., Allen, J., Casillas, A., Hanson, M. A., & Robbins, S. B. (2012). Interests, work values, and occupations predicting work outcomes with the WorkKeys Fit assessment. *Journal of Career Assessment*, 20(4), 359–374.
- Swanson, J. L. (2008). *An analysis of the impact of high school dual enrollment course participation on post-secondary academic success, persistence and degree completion* (Doctoral dissertation). Retrieved from <http://nacep.org/wp-content/uploads/2010/02/Dissertation-2008-Joni-L.-Swanson.pdf>
- Trusty, J., & Niles, S. G. (2003). High-school math courses and completion of the bachelor's degree. *Professional School Counseling*, 7(2), 99–107.
- Welsh, J. R., Kucinkas, S. K., & Curran, L. T. (1990). *Armed services vocational battery (ASVAB): Integrative review of validity studies* (Technical Report No. 90-22). Brooks Air Force Base, TX: Air Force Systems Command.
- WestEd & EPIC. (2013). *National Assessment of Educational Progress grade 12 preparedness research project job training programs curriculum study: Final report*. San Francisco, CA, and Eugene, OR: Authors.
- Wolfe, J. H., Larson, G. E., & Alderton, D. L. (2006). *Incremental validity of new computerized aptitude tests for predicting training performance in nine Navy technical schools* (No. NPRST-TR-06-3). Millington, TN: Navy Personnel Research Studies and Technology.
- Zwick, R. (2012). The role of admissions test scores, socioeconomic status, and high school grades in predicting college achievement. *Pensamiento Educativo*, 49(2), 23–30.
- Zwick, R., Brown, T., & Sklar, J. C. (2004). *California and the SAT: A reanalysis of University of California admissions data*. Berkeley: Center for Studies in Higher Education, University of California.
- Zwick, R., & Himelfarb, I. (2011). The effect of high school socioeconomic status on the predictive validity of SAT scores and high school grade-point average. *Journal of Educational Measurement*, 48(2), 101–121.