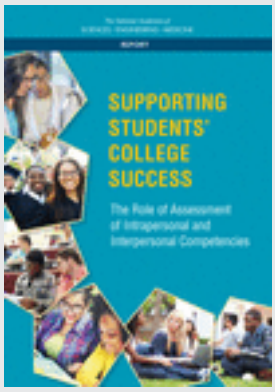

Supporting Students' College Success: The Role of Assessment of Intrapersonal and Interpersonal Competencies

DETAILS

184 pages | 6 x 9 | PAPERBACK
ISBN 978-0-309-45605-0 | DOI: 10.17226/24697

CONTRIBUTORS

Joan Herman and Margaret Hilton, Editors; Committee on Assessing Intrapersonal and Interpersonal Competencies; Board on Testing and Assessment; Division of Behavioral and Social Sciences and Education; National Academies of Sciences, Engineering, and Medicine

PREPUBLICATION COPY, UNCORRECTED PROOFS

Supporting Students' College Success: The Role of Assessment of Intrapersonal and Interpersonal Competencies

Committee on Assessing Intrapersonal and Interpersonal Competencies

Joan Herman and Margaret Hilton, Editors

Board on Testing and Assessment

Division of Behavioral and Social Sciences and Education

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

This activity was supported by Contract /Grant No. 1460028 between the National Academy of Sciences and the National Science Foundation. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any organization or agency that provided support for the project.

International Standard Book Number-13: 978-0-309-XXXX-X
International Standard Book Number-10: 0-309-XXXXX-X
Digital Object Identifier: https://doi.org/10.17226/24697
Library of Congress Control Number

Additional copies of this report are available for sale from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu.

Printed in the United States of America

Suggested citation: National Academies of Sciences, Engineering, and Medicine. (2017). *Supporting Students' College Success: The Role of Assessment of Intrapersonal and Interpersonal Competencies.* Washington, DC: The National Academies Press. https://doi.org/10.17226/24697.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia K. McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. C. D. Mote, Jr., is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at **www.national-academies.org.**

PREPUBLICATION COPY, UNCORRECTED PROOFS

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Reports document the evidence-based consensus of an authoring committee of experts. Reports typically include findings, conclusions, and recommendations based on information gathered by the committee and committee deliberations. Reports are peer reviewed and are approved by the National Academies of Sciences, Engineering, and Medicine.

Proceedings chronicle the presentations and discussions at a workshop, symposium, or other convening event. The statements and opinions contained in proceedings are those of the participants and are not necessarily endorsed by other participants, the planning committee, or the National Academies of Sciences, Engineering, and Medicine.

For information about other products and activities of the National Academies, please visit nationalacademies.org/whatwedo.

PREPUBLICATION COPY, UNCORRECTED PROOFS

COMMITTEE ON ASSESSING INTRAPERSONAL AND INTERPERSONAL COMPETENCIES

JOAN HERMAN *(Chair),* National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles
DAVID BILLS, College of Education, University of Iowa, Iowa City
CORBIN CAMPBELL, Teachers College, Columbia University
TABBYE CHAVOUS, Rackham Graduate School, University of Michigan, Ann Arbor
GREG DUNCAN, School of Education, University of California, Irvine
SYLVIA HURTADO, Graduate School of Education and Information Sciences, University of California, Los Angeles
PATRICK KYLLONEN, Center for Academic and Workforce Readiness, Educational Testing Service, Princeton, New Jersey
DAN MCADAMS, Department of Psychology, Northwestern University, Evanston, Illinois
FREDERICK OSWALD, Department of Psychology, Rice University, Houston, Texas
JONATHAN PLUCKER, School of Education and Center for Talented Youth, Johns Hopkins University, Baltimore, Maryland
K. ANN RENNINGER, Department of Educational Studies, Swarthmore College, Swarthmore, Pennsylvania
BRIAN STECHER, Pardee RAND Graduate School, RAND Corporation, Santa Monica, California

MARGARET HILTON, *Study Director*
PATRICIA MORISON, *Acting Board Director, Board on Testing and Assessment*
HEIDI SCHWEINGRUBER, *Director, Board on Science Education*
JUDITH KOENIG, *Senior Program Officer*
KELLY ARRINGTON, *Senior Program Assistant*

BOARD ON TESTING AND ASSESSMENT

Acknowledgments

This report was made possible by the contributions of the National Academies of Sciences, Engineering, and Medicine (the National Academies); the study committee; and many other experts. First, the committee thanks the National Science Foundation (NSF) for sponsoring the study and the William and Flora Hewlett Foundation for providing additional dissemination funding. Particular thanks go to Susan Singer (former director, NSF Division of Undergraduate Education) and Marc Chun (program officer, the William and Flora Hewlett Foundation).

During the course of its work, the committee benefited from presentations by and discussions with several experts who participated in its three fact-finding meetings. At the first meeting, the committee discussed its charge with Susan Singer, representing NSF. In addition, Stephen M. Fiore, University of Central Florida, and David Yeager, University of Texas-Austin presented recent research on assessment of intra- and interpersonal competencies.

The committee's third meeting included a workshop designed to explore the changing context of higher education; the growing importance of social skills in the workplace; and the relationships among conscientiousness, motivation, and academic achievement. The committee found the presentations and discussions at this meeting enlightening and thanks all who participated, including Nicholas Bowman (University of Iowa), David Deming (Harvard University), Alicia Dowd (Pennsylvania State University), Carol Geary Schneider (Association of American Colleges and Universities), Judith Harackiewicz (University of Wisconsin), Joshua Jackson (Washington University at St. Louis), and Alex (Sandy) Pentland (Massachusetts Institute of Technology).

At the committee's fourth meeting, Nicholas Bowman and David Deming presented findings from research commissioned by the committee. The committee thanks them for their work and insights. Following the fourth meeting, Sabrina Solanki (University of California, Irvine), quickly assembled and synthesized literature on interventions for developing intrarpersonal competencies. This synthesis proved invaluable to the committee's deliberations.

The committee gratefully acknowledges the efforts of the staff of the Board on Testing and Assessment (BOTA) and the Board on Science Education (BOSE). The committee thanks Patricia Morison, current acting director of BOTA, for her support and guidance at key stages of this project; Judith Koenig, BOTA senior program officer, for sharing her assessment expertise; and Heidi Schweingruber, director of BOSE, for her advice throughout the course of the study. Special thanks are due to Kelly Arrington, senior project assistant, for her exceptional organizational skills and close attention to detail in handling all the administrative details associated with the committee's in-person and virtual meetings and the workshop, and providing critical support in preparing the manuscript for this report. We also thank Vanessa Lazar, research assistant, for her adept research skills, and Rebecca Morgan, National Academies senior research librarian, for conducting several literature searches.

The committee also thanks members of the Office of Reports and Communication of the Division of Behavioral and Social Sciences and Education for their work on this report. Thanks go to Rona Briere for her sage editorial advice on this manuscript, Kirsten Sampson-Snyder for

her work in coordinating the review process, and Yvonne Wise for shepherding the manuscript through production.

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the Report Review Committee of the National Academies. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

The committee thanks the following individuals for their review of this report: Roger Benjamin, President, Council for Aid to Education; Jacquelynne S. Eccles, School of Education, University of California, Irvine; Susan T. Fiske, Psychology and Public Affairs, Princeton University; Catherine Lord, Center for Autism and the Developing Brain, Weill Cornell Medicine; Matthew J. Mayhew, College Impact laboratory, College of Education and Human Ecology, The Ohio State University; James W. Pellegrino, Learning Sciences Research Institute, University of Illinois at Chicago; Ann Marie Ryan, Department of Psychology, Michigan State University; Paul R. Sackett, Department of Psychology, University of Minnesota; Xueli Wang, Educational Leadership and Policy Analysis, University of Wisconsin–Madison.

Although the reviewers listed above provided many constructive comments and suggestions, they were not asked to endorse the report's conclusions or recommendations, nor did they see the final draft of the report before its release. The review of this report was overseen by Susan J. Curry (National Academy of Medicine), University of Iowa, and Milton D. Hakel, Bowling Green State University (emeritus). Appointed by the National Academies, they were responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

Finally, we wish to thank the committee members for their extraordinary efforts. Their broad range of expertise related to educational measurement, assessment, and accountability policy in higher education; educational psychology; persistence in higher education; and economics was critical to the ability to address the study charge successfully. The committee members freely contributed their time to accomplish the myriad tasks associated with assembling information and preparing this report. They actively assisted in all stages of the study, including planning the meetings and the workshop, as well as writing and rewriting multiple versions of this report. They gave generously of their time to ensure that the final product would accurately represent the committee's consensus findings, conclusions, and recommendations. These efforts reflected the committee members' commitment to improving student success in higher education.

Joan Herman, *Chair*
Margaret Hilton, *Study Director*
Committee on Assessing Intrapersonal and Interpersonal Competencies

Contents

Summary

The importance of higher education has never been clearer. Educational attainment—the number of years a person spends in school—strongly predicts adult earnings, as well as health and civic engagement (National Research Council, 2012a). Yet relative to other developed nations, educational attainment in the United States is lagging, with young Americans who heretofore led the world in completing postsecondary degrees now falling behind their global peers (OECD, 2013b). As part of a broader national college completion agenda aimed at increasing college graduation rates, higher education researchers and policy makers are exploring the role of intrapersonal and interpersonal competencies in supporting student success. These sets of competencies represent two of three domains of competence outlined in a previous study (National Research Council, 2012a):

- *Intrapersonal competencies* involve self-management and the ability to regulate one's behavior and emotion to reach goals.
- *Interpersonal competencies* involve expressing information to others as well as interpreting others' messages and responding appropriately.
- *Cognitive competencies* involve thinking, reasoning, and related skills.

That study found that these competencies across all three domains were closely intertwined with—and supported the acquisition of—core knowledge and skills in science, mathematics, and English language arts. Because it also found more available research to guide the teaching and assessment of cognitive competencies relative to the other two domains, the committee that carried out that study recommended additional research on the assessment of intrapersonal and interpersonal competencies that traditionally have not been targeted as educational goals (National Research Council, 2012a).

To address this recommendation and the national imperative to increase college completion, the National Science Foundation (NSF) commissioned the National Academies of Sciences, Engineering, and Medicine (National Academies) to convene an ad hoc committee to conduct a study of the assessment of intrapersonal and interpersonal competencies in higher education, with the following charge.

> …examine how to assess interpersonal and intrapersonal competencies (e.g., teamwork, communication skills, academic mindset, and grit) of undergraduate students for different purposes. This examination will include identifying a range of competencies that may be related to postsecondary persistence and success, and that evidence indicates can be enhanced through intervention. The committee will author a report that establishes priorities for the development and use of

> assessments related to the identified intrapersonal and interpersonal competencies that influence higher education success, especially in STEM [science, technology, engineering, and mathematics].

In fulfilling this charge, the committee was tasked with undertaking three principal analytical tasks:

> Task 1: Review the relevant research to more clearly define interpersonal and intrapersonal competencies, to examine whether and to what extent a range of these competencies may be related to each other and to persistence and success in undergraduate education (especially in STEM) and to examine the extent to which these competencies can be enhanced through intervention.
> Task 2: Examine available assessments of the interpersonal and intrapersonal competencies or competency clusters that are most strongly related to undergraduate persistence.
> Task 3: Establish priorities for development and use of these assessments for different purposes.

FRAMING THE STUDY

When interpreting this charge, the committee defined "postsecondary persistence and success" as "persistence and success in undergraduate education," as is clearly stated in Task 1. Thus defined, persistence and success are reflected in such measures as retention from one semester or one year to the following semester or year, retention and/or success in STEM, grade point average (GPA), and graduation. This definition followed on a prior, related report, which found that educational attainment was strongly predictive of labor market success, a finding that was upheld even with rigorous research approaches designed to approximate random assignment (National Research Council, 2012b). The committee interpreted "especially in STEM" in Task 1 to mean that it should give special attention to research on the role of intra- and interpersonal competencies in supporting persistence and completion within STEM majors. However, the committee found insufficient evidence to reach any conclusions specifically about these majors.

The committee adopted a broad definition of "competency" to include a range of attitudes, behaviors, beliefs, and dispositions that reside within the individual student and that may also be influenced by college environments and contexts. These competencies incorporate both intrapersonal and interpersonal dimensions, as well as cognitive dimensions. In light of the growing diversity of the undergraduate student population and its charge to focus on persistence and success especially in STEM, the committee viewed diversity and inclusion as central themes of its work. The committee gave special attention to research on student groups that have historically experienced lower persistence and success than other groups, both in postsecondary education generally and in STEM specifically. These include three racial/ethnic minority groups (black, Hispanic, and American Indian), along with first-generation college students, students from low-income families, and, in certain STEM disciplines, women. In this report, the committee refers to these diverse student groups as "underrepresented groups" and to black, Hispanic, and American Indian student groups as "underrepresented minorities."

When considering the meaning of Task 2, the committee defined "examine available assessments" primarily as a charge to examine available assessment *methods* that are currently

being used or could potentially be used to assess the competencies identified in this report. The committee also reviewed prominent existing assessment instruments. Finally, while focusing primarily on competencies related to persistence and success in undergraduate education, the committee recognized that higher education leaders have identified certain intrapersonal and interpersonal competencies (e.g., teamwork, ethics, and intercultural sensitivity) as desired learning *outcomes* for graduates. In response to this development, the committee briefly explored research on competencies identified as important learning outcomes.

IDENTIFYING COMPETENCIES FOR COLLEGE SUCCESS

To address its charge to identify competencies that are related to persistence and success in undergraduate education and can be enhanced through intervention, the committee conducted an extensive search of the literature. It found that the research base is limited, especially in the area of relationships between interpersonal competencies and college success. There were also gaps related to other important topics.

Conclusion: Only limited research has been conducted to date on the potential relationships between various intra- and interpersonal competencies and students' college success. There are major gaps in the research evidence:

- ***Little research is available on the possible relationship between interpersonal competencies and students' college success.***
- ***The available research has been conducted almost entirely in 4-year institutions; very little experimental evidence is available on the possible relationship between intra- and interpersonal competencies and students' success in community colleges.***
- ***There is a paucity of evidence on the possible relationships between intra- and interpersonal competencies and the success of students intending to major in STEM fields.***

Recommendation 1: Federal agencies and foundations should invest in research exploring the possible relationships between various intra- and interpersonal competencies and students' college success. To address gaps in the research base, these investments should include support for research examining

- **how interpersonal competencies may be related to student success in 4-year colleges;**
- **how intra- and interpersonal competencies may be related to student success in community colleges; and**
- **how intra- and interpersonal competencies may be related to student success in 2- and 4-year STEM programs and majors.**

Promising Competencies

Based on its review of the limited available research, the committee identified promising competencies that appear to be related to college success. Correlational research suggests that,

among the competencies reviewed by the committee, the most robust predictor of college success is dispositional conscientiousness—the tendency to be self-controlled, responsible to others, hardworking, persevering, rule-abiding, and achievement oriented. Conscientiousness is closely related to other constructs such as self-control, self-discipline, persistence, and grit. Indeed, scores on measures of conscientiousness are nearly as predictive of college success as are measures of general cognitive ability. Yet evidence supports conscientiousness as a deeply ingrained dispositional trait that is difficult to change, at least in the short term. Nevertheless, a few interventions have targeted specific behaviors associated with conscientiousness, to date yielding significant but small effects on college success.

> ***Conclusion: Beyond cognitive factors, correlational research has shown that individual differences in intrapersonal competencies predict college success and completion. These competencies include the broad personality trait of conscientiousness. Although an individual's relative standing on conscientiousness tends to be highly stable over time, some interventions have successfully targeted task management and other specific manifestations of this trait.***

Moving beyond dispositions and traits, the committee reviewed the available research on the relationships between other intra- and interpersonal competencies and college success. Through this process, the committee identified eight intrapersonal competencies. (Reflecting a lack of research evidence, this list includes no purely interpersonal competencies.) These eight competencies have been studied using a range of methods, and the committee identified them based on correlational and experimental research. However, the committee judged the strength of the evidence related to each competency based exclusively on research that has developed and tested them through interventions using random assignment.

> ***Conclusion: The limited intervention studies conducted to date have generated promising evidence that the competencies of sense of belonging, growth mindset, and utility goals and values are related to college success and are malleable in response to interventions. Available intervention studies provide more modest evidence that five other competencies are similarly related to college success and malleable, yielding a total of eight identified competencies:***
>
> - ***Behaviors related to conscientiousness—behaviors related to self-control, hard work, persistence, and achievement orientation;***
> - ***Sense of belonging—a student's sense that he or she belongs at a college, fits in well, and is socially integrated;***
> - ***Academic self-efficacy—a student's belief that he or she can succeed in academic tasks;***
> - ***Growth mindset—a student's belief that his or her own intelligence is not a fixed entity, but a malleable quality that can grow and improve;***
> - ***Utility goals and values—personal goals and values that a student perceives to be directly linked to the achievement of a future, desired end;***

- ***Intrinsic goals and interest—personal goals that a student experiences as rewarding in and of themselves, linked to strong interest;***
- ***Prosocial goals and values—the desire to promote the well-being or development of other people or of domains that transcend the self; and***
- ***Positive future self—a positive image or personal narrative constructed by a student to represent what kind of person he or she will be in the future.***

Interventions that often required very little time and money to implement have helped students develop these eight competencies. Some of these interventions have been particularly effective for underrepresented student groups that are most at risk for academic failure.

Conclusion: Notably, evidence shows that low-cost interventions aimed at developing sense of belonging, growth mindset, and utility goals and values have sometimes generated the largest benefits for underrepresented student groups that are most at risk for academic failure. Although encouraging, this evidence is limited and recent, and further research is needed to replicate and extend it.

Recommendation 2: Federal agencies and foundations should invest in intervention research using random assignment and research employing a range of other methods to understand better the competencies identified in this report, their relationship to college success, and the mechanisms through which they operate to improve college success. Research focused on supporting the college success of underrepresented student groups should be a priority.

Recommendation 3: Colleges and universities should support the intervention research proposed in Recommendation 2 by facilitating the implementation and evaluation of random-assignment interventions, thereby gaining valuable information about their students and building the knowledge base on effective interventions needed to increase student retention and success.

Sensitivity to Context and Subgroup Effects

The committee's review of the research indicates that issues of race, ethnicity, gender, social class, and culture need to be carefully considered when educators, administrators, researchers, and policy makers think about competencies and their contribution to college success. Certain competencies, and the problems they might help address or solve, may be more salient or useful for certain groups of students than others. Underrepresented minority students, for example, may bring to college such competencies as a strong racial or cultural identity that may help them navigate academic or social environments. More research also is needed to examine how particular educational and cultural contexts currently influence the development of motivation, intrinsic goals and interest, and other intra- and interpersonal competencies among

underrepresented groups (minority students, first-generation college students, students from low-income families, and women).

Conclusion: Certain competencies develop and function differently for different groups and within different cultural and educational contexts. For example, although a strong sense of belonging in college is an important factor for success among underrepresented student groups, members of these groups may find it difficult to develop this competency if they experience campus environments that are discriminatory, negative, or unwelcoming.

Recommendation 4: To help reduce disparities in college success among student groups, institutions of higher education should evaluate and improve their social and learning environments to support the development of the eight identified competencies, especially among underrepresented student groups.

ASSESSMENT METHODS FOR THE IDENTIFIED COMPETENCIES

The committee reviewed the nature and quality of existing competency assessments, focusing particularly on the eight identified competencies, together with research and professional standards related to the overall process of developing; validating; implementing; and interpreting, evaluating, and using the results of assessments. The test development practices used to create assessments of cognitive knowledge and skills that meet these professional standards are equally applicable to intra- and interpersonal competency assessments.

Current Assessments

The committee examined the assessments used in the intervention studies targeting the eight identified competencies and commissioned a literature search on measurement of these competencies. Drawing on both sources, the committee also identified and closely analyzed a small sample of established assessment instruments targeting one or more of the eight competencies. Overall, the review revealed that self-report methods, with their known limitations, predominated in the assessments of the eight competencies. Analysis of the quality of the assessments used in the intervention studies revealed spotty attention to reliability and almost no reported evidence of validity or fairness. However, more evidence of assessment quality was found for some established assessment instruments used in higher education research, particularly those efforts that have received funding for assessment research and development. These instruments provide evidence on reliability and validity but lack evidence on fairness. Assessments developed by professional testing companies provide even more evidence of quality, including fairness data; however, these assessments target a wider range of competencies, only partially addressing some of the eight competencies.

Conclusion: Most current assessments of the eight identified competencies are uneven in quality, providing only limited evidence to date that they meet professional standards of reliability, validity, and fairness.

Assessments for High-Stakes Purposes

Developers of all types of assessments, whether they aim to measure cognitive, intrapersonal, or interpersonal competencies, must exercise particular care when an assessment will serve a high-stakes purpose. Assessments are considered high-stakes when their results carry serious consequences for individuals or institutions.

Conclusion: The development and validation of assessments of intra- and interpersonal competencies for high-stakes purposes is a rigorous, time-consuming, and expensive process that depends critically on expertise in assessment and psychometrics. Validity, reliability, and fairness are essential considerations in evaluating assessment quality.

Recommendation 5: When developing and validating intra- and interpersonal competency assessments to be used for high-stakes purposes, stakeholders in higher education (e.g., faculty, administrators, student services offices) should comply with professional standards, legal guidelines, and best practices to enable appropriate interpretations of the assessment results for particular uses.

Recommendation 6: Institutions of higher education should not make high-stakes decisions based solely on current assessments of the eight identified competencies, given the relatively limited research to date demonstrating their validity for predicting college success.

Assessments for Low-Stakes Purposes

Researchers and practitioners in higher education also use assessments for low-stakes purposes, such as to evaluate the quality of interventions, policies, and instructional practices or simply to monitor student change over time. When used for these low-stakes purposes, assessments need not meet the high evidentiary requirements of individual high-stakes student assessments, such as college admissions tests. Professional testing standards clearly state that the amount and type of evidence needed to support a test's validity may vary depending on the use or interpretation of the test scores. At the same time, even when assessments are not used for high-stakes purposes, they need to be sensitive to the competencies they are intended to measure.

Conclusion: Even low-stakes uses of intra- and interpersonal competency assessments require attention to validity, reliability, and fairness, although they need not meet the high evidentiary requirements of high-stakes assessments.

Recommendation 7: Those who develop, select, or use intra- and interpersonal competency assessments should pay heed to, and collect evidence of, validity, reliability, and fairness as appropriate for the intended high-stakes or low-stakes uses.

Definition of Constructs Being Assessed

After reviewing both general principles for assessment development and use and recent research on measurement of the eight identified competencies, the committee concluded that defining each competency clearly and comprehensively is a critical first step in developing high-quality assessments. Clear definitions are especially important in light of the wide variety of terms used for these competencies. For example, conscientiousness, grit, and persistence are closely related constructs, despite being named differently. In fact, assessments of all of these constructs may contain very similar item content. Conversely, assessments bearing the same name may in fact comprise items measuring different competencies.

Conclusion: High-quality assessment begins with a clear definition of the competency to be measured, and identifies how the assessment will be used and what kinds of inferences it will support.

Competency definitions guide assessment development and selection by making it possible to evaluate how well the assessment represents the competency it is intended to measure, thereby supporting appropriate inferences about the construct for particular uses. High-quality assessments avoid construct underrepresentation, represent the breadth and depth of the competency, and minimize any distortions caused by competency-irrelevant influences.

Innovative Methods and Technologies for Assessment

Self-report measures, such as those frequently used to assess the eight identified competencies, have several limitations. First, individuals responding to both high- and low-stakes assessments may be motivated to present themselves in a favorable light. In addition, people often express themselves on a response scale in habitual or characteristic ways, such as tending to mark the extremes (e.g., "strongly agree" or "strongly disagree") or to agree or respond positively regardless of the question. In addition, respondents' tendency to compare themselves with those around them can compromise the use of the responses to measure growth or to compare groups of individuals because such comparisons depend on an absolute rather than a relative standard. Because self-report measures are widely used, these limitations affect a broad swath of current intra- and interpersonal competency assessments.

Recent research has identified various methods that can mitigate these limitations. For example, the use of forced-choice and ranking methods for collecting self-evaluations avoids response-style bias by circumventing traditional rating scales altogether. The use of anchoring vignettes also addresses response-style bias by having raters make use of detailed objective anchors, and may potentially deal with reference group effects as well. Other nontraditional measures include situational judgment tests, as well as games or simulations, which avoid many of the documented limitations of self-ratings. Further research is needed to develop, extend, and refine these and other promising new approaches.

Conclusion: Most existing assessments of the eight identified competencies, as well as many existing assessments of other intra- and interpersonal competencies, use self-report measures, which have well-documented limitations. These limitations may constrain or preclude certain uses of the

results. Innovative approaches for assessing intra- and interpersonal competencies can address these limitations.

Recommendation 8: Federal agencies and foundations should support additional research, development, and validation of new intra- and interpersonal competency assessments that address the shortcomings of existing measures.

Fairness in Assessment

The 2014 *Standards for Educational and Psychological Testing* make clear that fairness to all individuals for whom an assessment is intended should be a driving concern throughout the development, validation, and use of all types of assessments. Assessment development should minimize construct-irrelevant characteristics that would interfere with the ability of some individuals or subgroups to show their standing on a competency or lead to individual or subgroup differences in the meaning of test scores. Whenever differences in subgroup scores are observed, follow-up research may be needed to examine the reasons, the potential sources of bias, and the comparability of score interpretations across individuals and subgroups in light of the intended uses of the assessment results. The committee applied these fairness principles in its review of current assessments of the eight identified competencies:

Conclusion: Despite the ever-increasing diversity of undergraduate student populations, attention to fairness for diverse populations is often inadequate in the development, validation, and use of current assessments of the eight identified competencies.

Because these fairness principles apply broadly to all types of assessments, the committee recommends:

Recommendation 9: Researchers and practitioners in higher education should consider evidence on fairness during the development, selection, and validation of intra- and interpersonal competency assessments.

Consideration of Contextual Factors

Self-, peer, or instructor ratings of an intrapersonal competency such as conscientiousness or an interpersonal competency such as teamwork may vary depending on local norms (e.g., reference group effects). In addition, contextual variables may mediate or moderate the relationships between intra- and interpersonal competencies and educational outcomes. For example, an intervention intended to develop sense of belonging may be effective only for underrepresented student groups.

Conclusion: Appropriate interpretation of the results of intra- and interpersonal competency assessments requires consideration of contextual factors such as student background, college climate, and department or discipline.

Recommendation 10: Higher education researchers and assessment experts should incorporate data on context (e.g., culture, climate, discipline) into their analyses and interpretations of the results of intra- and interpersonal competency assessments.

Implementing this recommendation will require that higher education researchers use appropriate statistical analyses that incorporate data on context when examining assessment results. Such analyses include use of multilevel statistical models, measurement invariance analyses, application of differential item functioning, and mediator and moderator analyses. These analyses can enhance understanding of the complex interactions and processes entailed in students' individual competencies and of features of higher education contexts that contribute to students' persistence and success. Multiple measures also can be used to minimize the possibility that inferences about a student's intra- and interpersonal competencies are due to a particular measurement approach.

ASSESSMENT USES AND STAKEHOLDERS

Addressing its charge to prioritize the use of assessments of intra- and interpersonal competencies, the committee reviewed research on how institutions of higher education are using assessments of cognitive, intrapersonal, and interpersonal competencies. This review revealed four major uses of such assessments in higher education:

- selection and placement of individual students;
- formative improvement of local educational processes, practices, and programs;
- research and evaluation supporting knowledge generation; and
- accountability.

Assessments for these four purposes are carried out by a variety of stakeholders, including families, K-12 schools, faculty members, college administrators, accreditors, and state and federal policy makers. To understand how these stakeholders presently and potentially could use data resulting from these assessments, the committee reviewed relevant higher education literature and reports on current practice.

Assessment Processes Supporting Student Success

Different higher education stakeholders may have different needs for assessments of intra- and interpersonal competencies, depending on the immediacy of those needs, the purposes to be served by the data, and stakeholders' assessment-related knowledge and skills. Variations in assessment uses necessitate different assessment instruments, different levels of evidence and aggregation, and different kinds of buy-in for the assessment process and its uses. It is important to consider these contextual aspects of the assessment process when implementing an intra- or interpersonal competency assessment in practice.

Conclusion: Assessments of intra- and interpersonal competencies in higher education are most valuable for supporting student success when their

selection, design, analysis, and interpretation are guided by stakeholder information needs, intended uses, and users

Recommendation 11: Leaders in higher education should select, design, analyze, and interpret data from assessments of intra- and interpersonal competencies based on stakeholder information needs, intended uses, and users.

The research literature contains convincing evidence that institutions of higher education can benefit from using assessments for both institutional improvement and accountability purposes, and these uses can ultimately be mutually reinforcing. However, assessment processes that emphasize improvement tend to garner more institutional support, including faculty buy-in, relative to those emphasizing accountability. Indeed, some administrators are concerned that external accountability mandates may focus institutional conversations about assessment on bottom-line compliance rather than institutional improvement, especially given limited assessment resources. College stakeholders also tend to be more receptive to assessment processes when they are internally derived, sensitive to specific institutional and disciplinary contexts, and driven by a belief that the assessment process can serve the goal of improving student learning outcomes. Therefore, institutional improvement requires planning for needed resources and putting systems in place to support moving assessments from data collection to improvement processes.

Conclusion: Assessments are more likely to be implemented and used by stakeholders to improve student success when they are motivated by internal institutional improvement purposes than when they are motivated by accountability purposes.

In one example, assessment results were used to catalyze multiple stakeholders' collaborative work toward the shared goal of improving students' leadership abilities. In another example, university advisers and student affairs staff used assessment data individually with students to tailor support services, while central administrators found the data useful to support strategic initiatives aimed at retaining diverse and underprepared students through graduation. Overall, research has highlighted the need for multiple stakeholders at various levels (students, staff, faculty members, administrators) to work together in an assessment process if they wish to effect pervasive change on a college campus.

Conclusion: Assessments are more likely to contribute to student retention and completion if efforts to use their results involve stakeholders at multiple levels of the organization (e.g., student support services, faculty, diversity officers, administrators) as opposed to involving individual stakeholders acting alone.

Support for Stakeholders' Assessment Capacity

Administrators and faculty in institutions of higher education may not have specialized training or expertise in educational assessment with regard to instrument design and selection, test administration, data analysis, or the best uses of assessment data. Yet while some

stakeholders on campus, such as institutional researchers and assessment experts, can help with educating the broader campus community about assessment, they may not be familiar enough with the specific issues involved in intra- and interpersonal competency assessment. Therefore, training targeted at specific stakeholders may be necessary for the full value of these assessments to be realized. In addition, although data on intra- and interpersonal competencies can potentially add substantial value to efforts to enhance the success of underrepresented student groups, faculty may not be familiar with this particular use of the data.

Conclusion: Some stakeholders in higher education will require support and training to develop the knowledge and skills needed to select, use, and interpret data from assessments to improve student success in higher education. Such training can also help stakeholders understand how these assessments can contribute to the success of underrepresented student groups in particular and how to engage stakeholders who are resistant to assessment in general.

Research Needs

As noted above, it is important when implementing an assessment in a college or university to consider the contextual aspects of the assessment as a process. Research has yielded preliminary evidence of the importance of the eight identified competencies to success in college, and case studies of the use of cognitive assessment data by institutions of higher education for purposes of institutional and instructional improvement also are widely available. By contrast, evidence on how data from assessments of the eight identified competencies, or other intra- or interpersonal competencies, can be used for these purposes is relatively sparse. As additional assessments of these competencies take place on college campuses, they may yield a more robust understanding of how such assessments can lead to improvement within specific institutional, disciplinary, and student contexts.

Conclusion: Limited evidence is available from an organizational science perspective on how stakeholders in higher education can use data on intra- and interpersonal competencies for improvement and evaluation purposes.

Recommendation 12: To broaden understanding of how assessments of intra- and interpersonal competencies can lead to greater student retention and success, institutions of higher education should study and report on their use of these assessments for improvement purposes (e.g., enhancing student support services, developing underrepresented students' sense of belonging, improving courses, identifying effective programs).

INTRA- AND INTERPERSONAL COMPETENCIES AS COLLEGE OUTCOMES

The intra- and interpersonal competencies of ethics, lifelong learning/career orientation, intercultural/diversity competence, civic engagement/citizenship, communication, and teamwork have been identified as valued outcomes of college education. Although it might seem intuitive that these competencies would predict academic success, there is little evidence to date that these desired outcomes for graduates are actually improving in college itself and also contribute to

persistence, GPA, and graduation. There simply are too many large gaps in the research literature and in the available data to say with any certainty that these competencies do or do not matter for students' college success.

Conclusion: To date, only limited research has been conducted on the intra- and interpersonal competencies that have been identified as important learning outcomes for college graduates. Therefore, little is known about whether and under what conditions these competencies are related to persistence and success in college.

Recommendation 13: Federal agencies and foundations should invest in research examining whether, and under what conditions, the intra- and interpersonal competencies identified as outcomes for college graduates may also be related to students' persistence and success in college.

When considering the research needs, the committee identified three issues. First, the state of measurement of most of these competencies is still markedly underdeveloped. This issue echoes the committee's conclusion that most current assessments of the eight competencies showing some evidence of a relationship with students' college success are uneven in quality and reinforces the need for further assessment research and development, as called for in Recommendation 8. Second, much theoretical and conceptual work remains to be done before statistical analysis is undertaken to explore potential areas of overlap between competencies identified as college outcomes and predictors of college persistence. Finally, the committee endorses a broad research agenda incorporating multiple methods to better understand the role of these competencies in students' college success.

1

Introduction

Educational attainment—the number of years a person spends in school—strongly predicts adult earnings, and also predicts health and civic engagement. Relative to those with less education, highly educated adults tend to be more productive and capable of adapting to technological change, fueling the U.S. economy's productivity in a competitive global environment (National Research Council, 2012b). Yet despite the importance of educational attainment for individual and societal success, young Americans, who heretofore have led the world in completing postsecondary degrees, are now falling behind their global peers (OECD, 2013a). The college completion agenda, discussed further below, has emerged as a national priority, and with it the need to resolve persistent disparities in graduation rates by gender, race, socioeconomic status, and parental education (Mann and DiPrete, 2013; National Academies of Sciences, Engineering, and Medicine, 2016a; Xie et al., 2015).

STUDY CHARGE

This college completion imperative provided the impetus for this study. The National Science Foundation (NSF) requested that the National Academies of Sciences, Engineering, and Medicine (National Academies) form a committee to examine whether intra- and interpersonal competencies that have been identified as significant for success in K-12 education, work, and life may also contribute to persistence and success in higher education. The committee was asked to explore how institutions of higher education can marshal the assessment and development of these competencies to increase college completion, focusing in particular on assessment as a key lever for supporting, stimulating, and informing change (Campbell, 2015). The charge to the committee is presented in Box 1-1.

BOX 1-1
Charge to the Committee

NSF charged the National Academies of Sciences, Engineering, and Medicine to

> …examine how to assess interpersonal and intrapersonal competencies (e.g., teamwork, communication skills, academic mindset, and grit) of undergraduate students for different purposes. This examination will include identifying a range of competencies that may be related to postsecondary persistence and success, and that evidence indicates can be enhanced through intervention. The committee will author a report that establishes

priorities for the development and use of assessments related to the identified intrapersonal and interpersonal competencies that influence higher education success, especially in STEM.

The committee will undertake three principal analytical tasks:

Task 1: Review the relevant research to more clearly define interpersonal and intrapersonal competencies, to examine whether and to what extent a range of these competencies may be related to each other and to persistence and success in undergraduate education (especially in STEM) and to examine the extent to which these competencies can be enhanced through intervention.
Task 2: Examine available assessments of the interpersonal and intrapersonal competencies or competency clusters that are most strongly related to undergraduate persistence.
Task 3: Establish priorities for development and use of these assessments for different purposes.

[End Box 1-1]

As a first step in addressing its charge, the committee reached agreement on its interpretation of the charge and its principal tasks. With regard to Task 1, the committee defined "postsecondary persistence and success" as "persistence and success in undergraduate education," as is clearly stated in Task 1. The committee focused on success during college, reflected in such indicators as retention from one semester or one year to the next, grade point average (GPA), and completion of a degree or certificate. In focusing solely on persistence and success during college, the committee followed on a prior, related report, which found that educational attainment was strongly predictive of labor market success, a finding that was upheld even with rigorous research approaches designed to approximate random assignment (National Research Council, 2012b). The committee thus focused on identifying those competencies for which there was empirical research investigating their relationships with indicators of progress toward and completion of undergraduate education. The committee interpreted "especially in STEM" in Task 1 to mean that it should give special attention to research on the role of intra- and interpersonal competencies in supporting persistence and completion within STEM majors. As discussed further below, however, the committee found insufficient evidence to reach any conclusions specifically about these majors. The committee adopted a broad definition of "competency" encompassing malleable attitudes, behaviors, beliefs, and dispositions that reside within the individual student and may also be influenced by college environments or contexts.

The committee was committed to maintaining a primary focus on persistence and success in undergraduate education. At the same time, it recognized that leaders in higher education are beginning to target intrapersonal competencies such as ethics and lifelong learning and interpersonal competencies such as teamwork and intercultural sensitivity as desired learning *outcomes* that are expected to be valuable for work and in life after college graduation. To address this current interest, the committee briefly considered selected intra- and interpersonal competencies that have been identified by higher education leaders as important learning outcomes. The committee examined whether these competencies targeted as desired learning

outcomes might also be related to persistence and success in undergraduate education and how they can be assessed (see Chapter 5).

With respect to Task 2, the committee interpreted "examine available assessments" as a charge to examine and evaluate assessment methods that are currently being used to measure the competencies identified in Chapter 2. The committee viewed this task as a charge to provide guidance to leaders and researchers in higher education by outlining important assessment principles, examining the strengths and weaknesses of currently used assessment methods, and recommending approaches for strengthening future assessments.

STUDY CONTEXT

Recent data illustrate the clear value of and growing demand for higher education. The earnings premium for college graduates relative to those with only a high school degree, which rose continuously from the 1970s to 2000, remains at a near-record high (Autor, 2014; Matsudaira, 2015). Reflecting this growing demand, the fraction of recent high school graduates who immediately entered postsecondary institutions grew from only 45 percent in 1960 to 60 percent in 1990 and 68 percent in 2014 (Kena et al., 2016). Employer demand, as reflected in growing wages and low unemployment, is especially strong for 2- and 4-year graduates with majors and specific skills in STEM (Carnevale et al., 2011; Rothwell, 2013). Although there is debate about projected future demand for STEM graduates (e.g., Lowell and Salzman, 2007), wages in both STEM and STEM-related occupations (e.g., health care, management in STEM-intensive industries) have been rising in recent years, suggesting that demand for STEM graduates is currently strong. In 2013, as the nation emerged from recession and the national unemployment rate stood at 8.1 percent, the unemployment rate among scientists and engineers was less than half as high, at 3.8 percent (National Science Board, 2016).

The Completion Challenge

Many college entrants drop out before completing their degrees. On average, among all full-time students who first entered 4-year institutions in 2008, only 60 percent had completed a degree at that institution 6 years later. Among full-time students who first entered 2-year institutions in 2011, only 28 percent had completed a degree or certificate within 150 percent of the expected time[1] (Kena et al., 2016). For students aspiring to major in STEM, completion rates within one of the STEM fields are even lower: Eagan and colleagues (2014) found that only 40 percent of first-time full-time STEM aspirants entering 4-year institutions in fall 2004 had completed a STEM degree 6 years later. The authors report wide variation across STEM fields of study in the probability that students completed the originally intended major, switched to another STEM field, switched to a non-STEM field, or dropped out of college entirely. Among students who intended to major in the life sciences, for example, 37 percent completed that major, another 6 percent completed a bachelor's degree in a different STEM field, and nearly one-third (31 percent) switched to a non-STEM field but still completed a bachelor's degree within 6 years. Overall, 74 percent of original life sciences aspirants had completed a 4-year degree in some field of study after 6 years.

[1]Both this 28 percent completion rate and the 60 percent rate for 4-year students exclude transfers. Using a separate database, Shapiro and colleagues (2015) estimate that among all students who entered 2- or 4-year institutions in 2007, 53 percent had graduated from the initial institution or another 6 years later.

With these low rates of graduation, the nation's educational attainment rate has slipped relative to that in other countries. By 2012, the United States, once the global leader in young adults' attainment of postsecondary degrees, ranked 14th among the 37 OECD and Group of 20 (G20) nations in the fraction of 25- to 34-year-olds who had graduated, placing it above average but far behind the top-ranked nations. The rate of increase in in educational attainment in the United States is lower than that across all OECD and G20 countries. Between 2000 and 2010, for example, postsecondary attainment in the United States grew an average of 1.3 percentage points a year, compared with 3.7 percentage points for OECD countries overall (OECD, 2013a).

Failure to enter or complete college has adverse effects on individuals and the national economy. DiPrete and Buchmann (2013) identify three categories of young people: those who do not go beyond high school, those who enter college but do not finish, and those who complete college. Those who enter 2- or 4-year institutions but fail to complete their degrees are in some ways worse off than those with only a high school diploma because they are likely to have accumulated student loan debt and lack marketable credentials. At the same time, they a have incurred the opportunity costs of lost labor force experience (including both foregone wages and on-the-job learning and training opportunities).

Calls for Accountability and Improvement

The loss of U.S. competitiveness in rates of completion of higher education relative to other advanced economies and the costs of noncompletion, together with rapidly rising college costs and student debt, have led policy makers to call for greater accountability and improvement in higher education. In a February 2009 address to a joint session of Congress, then-President Obama proposed the goal that "by 2020, the U.S. will once again have the highest proportion of college graduates in the world" (Obama, 2009). Many institutions of higher education, state governments, foundations, businesses, and federal policy makers are now working together on this college completion agenda (Hughes, 2013).

Retaining and graduating students in STEM majors is considered an essential component of the broader completion agenda, focused on building and diversifying the STEM workforce (President's Council of Advisors on Science and Technology, 2012). Yet colleges and universities continue to face challenges in supporting students—especially women, minorities, those from low-income families, and first-generation students—in the pursuit of STEM majors. Studies suggest that many students who have been successful in high school STEM enter college intending to major in a STEM discipline, but later switch to other fields (e.g., Ohland et al., 2008; Seymour and Hewitt, 1997). Switching fields may lower future earnings, as individuals with bachelor's degrees in science, engineering, and related fields (e.g., health care, management of scientists and engineers) enjoy an earnings premium over the course of their careers compared with those with degrees in other fields (National Science Board, 2016).

Relative to previous generations, today's undergraduates are more likely to be from minority groups, to be single parents, and to earn credits from multiple public and private institutions (National Academies of Sciences, Engineering, and Medicine, 2016a). This shift compounds the challenge and the need to close existing achievement gaps. Between 1990 and 2012, the share of students from low-income families increased by nearly 18 percent, while the share of non-Hispanic white students fell from 77 percent to 57 percent (National Academies of Sciences, Engineering, and Medicine, 2016a). And women make up the majority of college students today—about 57 percent. Disparities exist among these different groups in where they

enroll and how frequently they graduate: minorities and students from low-income families are more likely than more advantaged students to enroll in community colleges or less selective 4-year institutions, where completion rates are much lower than in highly selective 4-year institutions (U.S. Department of Education, 2015). As noted earlier, on average about 60 percent of 4-year students entering in 2008 had completed degrees 6 years later. However, completion rates were lower for blacks (40.9 percent), Hispanics (53.5 percent), and American Indians (39.7 percent) than for whites (63.2 percent) and Asian Americans (70.6 percent) (U.S. Department of Education, 2015).

Focusing specifically on STEM, Eagan and colleagues (2014) found that 40 percent of all students entering 4-year institutions in 2004 with the intent of pursuing STEM majors had completed such degrees after 6 years. Completion rates in STEM majors after 6 years were lower for black (21.8 percent), Hispanic (29 percent), and American Indian (24.9 percent) students than for white (43 percent) and Asian American (52 percent) students.

The Range of Competencies for Life and Work

Research demonstrates the value for students of developing a full complement of competencies to support their success in education, life, and work. As part of the broader national effort to increase completion rates, researchers and policy makers in higher education are exploring the role of intra- and interpersonal competencies in supporting student success. As noted earlier, these sets of competencies represent two of three domains of competence outlined in a previous National Academies report (National Research Council, 2012b):

- *Intrapersonal competencies* involve self-management and the ability to regulate one's behavior and emotions to reach goals.
- *Interpersonal competencies* involve expressing information to others as well as interpreting others' messages and responding appropriately.
- *Cognitive competencies* involve thinking, reasoning, and related skills.

That study found that these competencies are closely intertwined with—and support the acquisition of—core knowledge and skills in science, mathematics, and English language arts. Because much more research is available to guide teaching and assessment in the cognitive domain relative to the other two domains, the report calls for greater attention to teaching and assessing the intra- and interpersonal competencies that traditionally have not been targeted as educational goals (National Research Council, 2012b).

The value of intra- and interpersonal competencies in the workplace is beginning to be documented in research. In separate recent studies, economists David Deming (2015) and Catherine Weinberger (2014) have demonstrated that jobs requiring high levels of both social and cognitive competencies are growing more rapidly and also tend to be rewarded with higher wages relative to other jobs. Their research augments earlier studies revealing that jobs requiring higher levels of cognitive, intrapersonal, and interpersonal competencies grew rapidly over the past half-century at the same time that computers eliminated jobs involving more routine tasks (Autor et al., 2003; Levy and Murnane, 2013). Although emerging evidence suggests that technology is beginning to displace workers at increasingly higher levels of the skill distribution (Brown et al., 2010), multiple employer surveys over the past decade (e.g., American Management Association, 2010; Casner-Lotto and Barrington, 2006; Hart Research Associates,

2010, 2015) also suggest a growing need for more labor market entrants with such competencies as communication, teamwork, ethics, and intercultural sensitivity.

To meet this growing labor market demand, educational policy makers increasingly have identified students' attainment of intra- and interpersonal competencies as key goals of both K-12 and higher education, and are working to assess them. The Association of American Colleges and Universities (AAC&U) has identified teamwork, problem solving, and lifelong learning as essential learning outcomes for all college graduates and has also developed rubrics for assessing these competencies.[2] The requirement of the Accreditation Board for Engineering and Technology (ABET) that undergraduate engineering programs develop students' ability to function on multidisciplinary teams and to engage in lifelong learning (Accreditation Board for Engineering and Technology, 2015) has spurred faculty members, researchers, and administrators to develop assessments aligned with this requirement. At the K-12 level, the Partnership for 21st Century Skills (an association of K-12 leaders) targets critical thinking, communication, collaboration, and creativity as key learning and innovation skills[3] and is working with state and local school systems and researchers to assess these skills.

In the push for accountability, policy makers, parents, and students are demanding that colleges and universities provide data demonstrating the quality of their degree programs (Campbell, 2015; Matsudaira, 2015). Although often focusing on graduation rates and graduates' earnings, policy makers also are asking institutions to define college outcomes for students more clearly and to provide reliable indicators of students' progress toward and attainment of these outcomes (Campbell, 2015). As noted above, the expected outcomes increasingly include intra- and interpersonal competencies, and research is under way to develop assessments of these competencies (e.g., Torney-Purta et al., 2015).

This report focuses primarily on competencies that 2- and 4-year institutions and faculty members could develop and assess to improve their students' college persistence and completion. This focus reflects the strong evidence that increased educational attainment is related to higher earnings, greater health, and civic engagement (National Research Council, 2012b), together with suggestive evidence that intra- and interpersonal competencies can help support educational achievement and attainment. At the K-12 level, well-designed social and emotional learning programs have been shown to improve students' academic achievement, as well as their social skills and behavior (Durlak and Weissberg, 2011; Durlak et al., 2010). Conversely, children with persistently high levels of antisocial behavior across elementary school are less likely to graduate high school and are far less likely to attend college than children who never have these problems (Duncan and Magnuson, 2011). At the undergraduate level, research has shown that such competencies as conscientiousness and self-regulated learning are positively correlated with grades and persistence (National Research Council, 2012b). And in a recent review of the literature on STEM persistence, Xie and colleagues (2015) found that, although both socioeconomic status and social-psychological factors predicted general educational attainment, the latter factors were more important correlates of participation and achievement in STEM than in other fields.

These emerging research findings on the potential links between intra- and interpersonal competencies and educational success have spurred further efforts to assess these competencies. Some faculty members are measuring these competencies in their research on student learning for the purpose of improving instruction. In a composite study of introductory geoscience

[2]See http://www.aacu.org/leap/vision.cfm [July 2016].
[3]See http://www.p21.org/our-work/p21-framework [July 2016].

classes, for example, student performance was significantly and positively correlated with scores on a measure of self-efficacy, an intrapersonal competency (McConnell and van der Hoeven Kraft, 2011; McConnell et al., 2010). In fact, students with low initial content knowledge but high self-efficacy earned the same grades as students with higher initial content knowledge but low self-efficacy. In another example, test publishers have created new assessments of intrapersonal competencies such as determination, goal striving, and self-management, and some colleges are using these assessments to identify incoming students who may need additional supports or may bring previously unrecognized strengths to their studies (Fain, 2015).

In sum, efforts are under way to assess intra- and interpersonal competencies in undergraduate education. However, these efforts are diffuse, seeking to assess a wide range of competencies that are often poorly defined—or called by different names while potentially representing the same underlying constructs—and they often rely on methods that fail to meet professional testing standards of reliability and validity. This report identifies competencies that show some evidence of specifically supporting persistence and success in undergraduate education and reviews methods and standards for their assessment.

STUDY APPROACH

The committee's approach to addressing its charge entailed reviewing the relevant research; considering competencies, diversity, and contexts; examining available assessments of intra- and interpersonal competencies; and establishing priorities for the development and use of assessments.

Identifying Competencies Related to College Success

To address Task 1, the committee conducted an extensive search of the literature examining the relationships between various skills, competencies, attitudes, and abilities and persistence and success in undergraduate education. Within the available time and resources for the study, the committee focused primarily on studies addressing key indicators of undergraduate persistence (e.g., GPA, persistence over semesters, completion of a degree or certificate) and gave less attention to the broader literature related to undergraduate learning. Individual committee members identified relevant materials (e.g., book chapters, articles, reports), and the committee also invited several outside experts to share their research on various constructs and their relationships to student success (see Appendix D). At the committee's request, the National Academies library staff conducted four literature searches seeking research on links between intra- and interpersonal competencies and college persistence (see Appendix A), using terms for various skills, abilities, and competencies that had been identified and organized in the prior report referenced above (National Research Council, 2012b). Although four of these searches included the term "interpersonal competencies" and related terms, such as "teamwork," "oral communication," "leadership," "interpersonal skills," and "intercultural competence," they yielded no rigorous research providing evidence that any of these competencies is related to persistence and success in undergraduate education.

To identify competencies that are not only related to success during college but also malleable in response to interventions, the committee assembled and reviewed experimental evidence from interventions (see Appendix B). In addition, the committee commissioned two original data analyses on the possible relationship between various competencies and success in

undergraduate education. First, as it continued to search for any evidence that one or more interpersonal competencies might be related to college persistence, the committee asked economist David Deming (professor, Harvard University) to analyze data from the National Longitudinal Survey of Youth (NLSY) on the relationship between social skills and college graduation. Second, educational researcher Nicholas Bowman (associate professor, University of Iowa) was asked to explore whether intra- and interpersonal competencies that have been targeted as desired learning outcomes for college graduates might also be predictors of postsecondary persistence and success. The committee requested that he analyze data from the Wabash Study of Liberal Arts Education on the relationships between several competencies and college success, breaking the data down for different student groups (males and females, minorities, and first-generation students; see Chapter 5).

The committee used the results of these literature searches and commissioned analyses in developing a list of competencies for which there is suggestive evidence of a relationship to persistence and success in undergraduate education and malleability in response to interventions (see Chapter 2). In so doing, the committee took into account the quality and quantity of the research evidence obtained from each source, privileging experimental evidence from interventions while also including studies addressing the key outcomes of undergraduate persistence and success using correlational and other methods. Reflecting the lack of research evidence, the list of competencies detailed in Chapter 2 includes no purely interpersonal competencies, and the committee points to the need for further research on competencies in this domain.

Recognizing that 2-year colleges play a significant role in undergraduate education at large and specifically in STEM fields (National Academies of Sciences, Engineering, and Medicine, 2016a), the committee sought out studies focused on community colleges. It was able to locate only one study of an intervention related to conscientiousness that included community college students (Liu et al., 2012 and one study of the correlation between mathematics self-efficacy and mathematics achievement in a sample of community college students (Nordstrom, 2012). These studies are discussed in Chapter 2, and the committee identified the need for additional research examining how intra- and interpersonal competencies may be related to college persistence and success for this important population. Similarly, to address its charge to identify a range of competencies related to persistence and success "especially in STEM," the committee reviewed the literature seeking studies focused specifically on persistence in STEM. However, the available evidence was not sufficient to conclude that any competency is related to persistence and success specifically in STEM undergraduate education, and the committee highlights the need for further research in this area as well.

Considering Competencies, Diversity, and Contexts

Considering the growing diversity of the undergraduate population and the disparities in graduation rates across different student groups, both generally among all fields of study and specifically in STEM fields, the committee identified the theme of inclusion and diversity as central to responding to its charge. As a first step toward addressing this theme, the committee identified the student groups to be the focus for inclusion and diversity. To this end, the committee built on a prior National Science Foundation (2013) analysis identifying "underrepresented groups" as those constituting smaller percentages of recipients of science and engineering degrees and employed scientists and engineers relative to their percentage

representation in the general population. These underrepresented groups include women in certain STEM fields, persons with disabilities, and three racial/ethnic groups—blacks, Hispanics, and American Indians. Although women now constitute the majority of the undergraduate population (Kena et al., 2016), their proportionate representation in some STEM fields has not increased substantially since the 1980s, (Mann and DiPrete, 2013). In the life and social sciences, for example, women have earned the majority of degrees since the 1980s, but they remain significantly underrepresented among degree recipients in engineering, the physical sciences, mathematics, and computer science (Xie and Killewald, 2012; Xie et al., 2015). In computer science, the percentage of women earning degrees declined from 28 percent in 2000 to 17.9 percent in 2009 (National Science Board, 2012). Because of these disparities in graduation rates, women remain significantly underrepresented in the engineering and computer-related occupations that make up more than 80 percent of STEM employment (Landivar, 2013).

Expanding the NSF definition, the committee noted that the above three racial/ethnic groups—blacks, Hispanics, and American Indians—have experienced lower graduation rates generally as well as in STEM (Eagan et al., 2014; National Center for Education Statistics, 2016). It also noted that students from low-income families and those who are the first in their families to attend college also have experienced lower graduation rates than other student groups (Cahalan et al., 2016; DeAngelo et al., 2011). Therefore, the committee uses the term "underrepresented groups" throughout this report to encompass all of these student groups that have historically experienced lower graduation rates than other student groups. In addition, the committee uses the term "underrepresented minorities" to refer to the above three racial/ethnic student groups.

To explore the theme of inclusion and diversity, the committee drew on committee members' own research, as well as a literature search focused on this theme. The committee also invited educational researcher Alicia Dowd (professor of education, Pennsylvania State University) to present her work addressing assessment, accountability, and institutional change. She discussed a project in Colorado that engaged teams of STEM faculty at three institutions in action research. The teams analyzed quantitative data on equity gaps in student progress toward degrees and qualitative data (e.g., observations, interviews) on teaching practices and departmental and institutional practices. These inquiries catalyzed changes in practices and policy, and one institution saw marked improvements in mathematics outcomes for African American and Hispanic students, along with an increase in success rates in college-level math among all students who began in developmental mathematics classes (Dowd, 2015).

This literature indicates that the success of diverse student groups is affected by systems of influence that include not only the competencies possessed by individual students but also students' interactions with others in different learning and living environments. Sato and colleagues (2015) observe that "noncognitive factors" (i.e., intra- and interpersonal competencies) are shaped by sociocultural contexts that affect how students experience and respond to situations, including learning activities and assessment prompts or tasks. For example, researchers have found that the discrimination, microaggression, and "chilly" climate in some STEM classrooms and laboratories negatively affect students' beliefs about their ability to engage seriously in STEM and undermine their feelings of competence (see Henderson et al., 2011). A sense of belonging to social and academic environments is important, and it may be especially predictive of persistence and achievement for underrepresented minority students in STEM (e.g., Darling et al., 2008). Students' perceptions of STEM contexts can influence their

abilities to identify and make use of institutional supports to continue pursuing STEM (e.g., Chang et al., 2014; Hurtado and Carter, 1997).

Similarly, Dowd and colleagues (2011) argue that efforts to measure and develop students' sense of belonging should include the recognition that college environments are not culturally neutral. They suggest that leaders and researchers in higher education should "measure a broader range of factors...in ensuring a sense of belonging, membership, and validation among students from all racial-ethnic groups" (p. 18).

For example, student identity is a multifaceted construct, encompassing attitudes, values, beliefs, and connections to social and cultural groups. Identity influences self-evaluation, contextual appraisal, and academic motivation, and it is associated with postsecondary persistence and success. Research has shown that group and individual identity experiences influence participation and persistence in STEM fields specifically (e.g., Borum and Walker, 2012; Carlone and Johnson, 2007; Kyoung Ro and Loya, 2015; Malcom and Malcom, 2011; Ong et al., 2011).

More generally, the identity-context congruence perspective of Byrd and Chavous 2011, 2012) provides a frame for considering how students' experience of college context norms may influence their engagement and persistence. In samples of African American secondary and postsecondary students, the authors found that perceiving positive racial climates (defined by inclusion and equity) was related to students' stronger sense of belonging to and connectedness with their academic settings; this relationship was particularly pronounced for students with a strong, positive connection to their racial identity. Furthermore, a sense of belonging and connectedness was positively related to students' motivation (as measured by engagement in and enjoyment of learning in classes).

In another example, Harackiewicz and colleagues (2015) found that among students taking introductory biology, underrepresented minority groups and first-generation students generally performed more poorly than other student groups. However, although first-generation underrepresented minority groups tended to have the weakest biology backgrounds and lowest incoming GPAs, they also were the most highly motivated to do well and to give back to their families and communities. The authors tested a brief intervention highlighting the relevance and value of the biology course and found that it helped students from all groups find utility value in the course content, which in turn tended to improve course performance. The intervention's effect size and the improvement in course grades were statistically and practically significant only for the first-generation underrepresented minority groups. The authors hypothesize that the preexisting desire of this group of students to give back to their families and communities may have increased their receptiveness to the utility value intervention, which in turn improved their academic performance (see Chapter 2 for further discussion of utility value).

Based on such findings, the committee gave some consideration to the literature examining, for different specific student groups, how the development of various competencies is influenced by different contexts at different points in the college trajectory. The committee's exploration of this emerging area of research informed the list of promising competencies discussed in Chapter 2. At the same time, this research suggested that systems of interrelationships between competencies and contexts, taken together, may be more likely to support persistence and postsecondary success than any single, isolated intervention targeting the development of a specific competency (Sato et al., 2015; see also Bailey et al., 2015). Nevertheless, to maintain focus on its charge, which emphasizes the role of individual competencies in supporting student success, the committee did not conduct a comprehensive

review of the literature on systems of competencies in context. Further research is needed in this area.

Examining Available Assessments

Addressing its charge to "examine how to assess intrapersonal and interpersonal competencies," the committee reviewed relevant current research, development, and practice. The committee searched the literature, deliberated on committee members' own work in this area, and reviewed relevant National Academies reports (e.g., National Research Council, 2012b, 2015a, 2015b) to identify the strengths and weaknesses of various assessment methods. At the committee's request, the National Academies library conducted two literature searches focused on assessment of intra- and interpersonal competencies (see Appendix A). To gain additional insights on assessment, the committee invited engineering education researcher Ashley Ater Kranov (affiliated assistant professor, Washington State University) to provide an overview of a new assessment of engineering professional skills that includes intra- and interpersonal competencies. The committee also invited Alex Pentland (professor, Massachusetts Institute of Technology) to present his social sciences computational research on team processes and outcomes, using unobtrusive badges to track individual behaviors and interactions (Pentland, 2015).

The committee examined primarily methods and strategies that could be used to define more clearly and assess the eight intrapersonal competencies identified in Chapter 2. However, recognizing that faculty and policy makers in higher education are keenly interested in assessing other competencies they view as valuable outcomes for college graduates, the committee briefly considered how to define and assess a selected number of these competencies as well (see Chapter 5).

Establishing Priorities for the Development and Use of Assessments

To establish priorities for the development and use of assessments, the committee deliberated on all the various strands of research discussed above, and also reviewed research and practice on assessment for improvement and accountability in undergraduate education (e.g., Campbell, 2015). The committee invited a presentation by Carol Geary Schneider (2015a) (president emerita, AAU&C) to discuss AAC&U's development of rubrics for assessing various learning outcomes, including intra- and interpersonal competencies.

ORGANIZATION OF THE REPORT

Following this introduction, Chapter 2 reviews the research evidence on the relationship between various dispositions, behaviors, attitudes, and beliefs and college persistence and success, identifying a small number of promising competencies and calling for further research. Chapter 3 examines the methods and instruments currently used to assess intra- and interpersonal competencies, particularly those identified in Chapter 2; the quality of these methods and instruments in light of foundational principles of assessment; and approaches for strengthening assessment of these competencies. Chapter 4 details different purposes for assessing competencies in higher education and the various higher education stakeholders who can use assessments for these purposes. It offers conclusions and recommends further research on how

higher education organizations can use competency assessments appropriately to enhance student success. Chapter 5 identifies a set of intra- and interpersonal competencies that have been targeted by higher education leaders as desired outcomes for college graduates, examines whether these competencies are related to college success, and considers their assessment. Chapter 6 summarizes the committee's recommendations. Appendix A describes the literature searches commissioned by the committee, while Appendix B, available online at [insert website], summarizes the experimental evidence from evaluations of interventions designed to develop the eight competencies discussed in Chapter 2. Appendix C describes the assessments used in the evaluations of interventions, Appendix D presents biosketches of the committee members, and Appendix E provides the agenda for the committee's December 2015 workshop.

2

Competencies for College Success

This chapter addresses the committee's charge to identify a range of intra- and interpersonal competencies that may be related to persistence and success in undergraduate education and that evidence indicates can be enhanced through intervention. It opens with an overview of intra- and interpersonal competencies in general; the committee's perspective on these competencies within home, community, and academic contexts; and the committee's approach to addressing this part of its charge. The second section presents a developmental framework of intra- and interpersonal competencies that may be relevant to college experiences: (1) broad dispositions, (2) beliefs, (3) specific motivations, and (4) future identity. The following four sections examine each of these concepts and identify eight specific competencies within them, which appear, based on the limited evidence available, to be related to persistence and success in undergraduate education:

- behaviors related to conscientiousness,
- sense of belonging,
- academic self-efficacy,
- growth mindset,
- utility goals and values,
- intrinsic goals and interest,
- prosocial goals and values, and
- positive future self.

As noted in Chapter 1, the committee defines "competencies" broadly to include malleable attitudes, behaviors, beliefs, and dispositions. Although such competencies may be influenced by college environments or contexts, as they exist and are perceived, they ultimately reside within the individual student. As discussed further below, for example, controlled experiments have shown that interventions can develop students' sense of belonging, that this competency persists over time, and that it is related to higher grade point average (GPA) relative to control groups of students in the same college context that did not receive the intervention. For each of these competencies, the committee reviews the evidence for a relationship to college persistence and success and its malleability in response to interventions. The chapter ends with conclusions and recommendations.

OVERVIEW

What student competencies (skills, abilities, beliefs, and attitudes) produce success in college? It is well documented that individual differences in general cognitive ability, as well as in many specific intellectual skills, such as critical thinking, writing ability, and computational expertise, are associated with individual differences in academic success. Simply put, students with a stronger and wider range of cognitive skills tend to do better in college. But other factors beyond cognitive abilities also appear to be predictive.

The most frequently used standardized measures of cognitive indicators, such as the SAT and ACT, account for a portion of the variance in college success and degree completion, but other factors, such as high school grade point average (GPA), also account for some of this variance (Bowen et al., 2011; Massey et al., 2003; Vars and Bowen, 1998). There is evidence as well that traditional cognitive assessments may differentially predict performance across demographic groups. Across 110 U.S. colleges and institutions (151,316 students), the SAT was found to underpredict GPA for women relative to men (i.e., women tend to perform better than predicted by the SAT) and to overpredict GPA for minorities relative to the white majority (i.e., minorities tend to perform less well than predicted by the SAT) (Mattern et al., 2008).

Subgroup performance differences on tests and in college have been attributed to a wide range of societal and historical inequalities and messages (Massey et al., 2003; Steele, 1997). Likewise, at the level of the individual student, it is almost certain that a wide range of factors extending well beyond measured cognitive skills play important roles in determining the extent to which a student will succeed in college. Although many of these factors reside in the student's environment (e.g., social class disparities, family resources, differential experiences of various kinds), others—including intra- and interpersonal competencies—reside more clearly within individual students and therefore are more within their control. This chapter introduces the intra- and interpersonal competencies identified by the committee as those that appear to be most important for college success, based on its review of the research evidence.

The committee reviewed research on intra- and interpersonal qualities that speak to the social and emotional lives of human beings and to human motivation. This research examines intrapersonal competencies such as self-control and motivation to achieve and interpersonal competencies such as teamwork, social communication, and community involvement. In some studies, the distinction between "intra" and "inter" appears to be blurred because the competencies were found to be interrelated or to operate both within the individual and in interpersonal relationships. For example, teamwork is viewed as an interpersonal skill promoting positive involvement with other people, teams, and communities. However, effective teamwork also involves intrapersonal skills such as self-regulation and decision making. Overall, the committee found little rigorous research examining how interpersonal competencies may be related to undergraduate persistence and success.

The committee's conceptualization considers the critical interplay between the characteristics of students and their academic, social, and cultural contexts. For instance, contextual factors such as family socialization and resources, campus climate and culture, quality of instruction, and other conditions of the student's environment may influence—either independently or jointly—the development of competencies that affect how students respond to academic and social challenges. Students from diverse backgrounds may experience these contextual factors, and therefore college, in very different ways. For example, underrepresented

groups[1] may find certain campus climates and disciplinary cultures and practices to be uncongenial or even aversive, yet find other campus climates supportive. For students to be successful in college, their unique experiences in these contexts may require that they draw on special strengths they have developed over the course of their lives, such as their capacity for adaptation and resilience.

With regard to success in science, technology, engineering, and mathematics (STEM), the joint consideration of competencies and the college context is especially relevant. For instance, a recent National Academies of Sciences, Engineering, and Medicine (2016a) report characterizes today's overall culture of STEM and STEM college education as tending to emphasize inherent, or "natural," cognitive ability as necessary for success. But the report notes that STEM skills can be improved, and that many studies indicate that student attributes and beliefs are in fact malleable and can be improved through intervention. Furthermore, the report highlights how racial and gender stereotypes around ability are salient in STEM contexts, and can manifest themselves in how women and students of color are treated and responded to. When prejudice, discrimination, and perceptions thereof are not prevented at the group and individual levels, even the most successful high school students can be discouraged from pursuing a college major in STEM, with the ultimate effect of pushing many highly competent women and minority college students out of these fields (e.g., Ohland et al., 2008; Seymour and Hewitt, 1997).

Thus, this chapter identifies competencies that appear to be causally related to engagement and success in college and completion to degree, above and beyond the demonstrated effects of cognitive skills. The focus is on competencies that help explain persistence in college across students from diverse backgrounds. As noted, the committee sought to identify competencies that research suggests are malleable, and therefore can be improved in college through programs or interventions designed to enhance student development and thereby boost academic engagement, academic success, and college completion rates.

As discussed in Chapter 1, to identify such competencies, the committee searched for literature across a range of fields, – including personality, development, educational, and social psychology– examining the relationship between intra- and/or interpersonal competencies and students' college persistence and success. . The committee also invited presentations by outside experts and commissioned National Academies library staff to conduct four searches of literature databases (see Appendix A). These literature searches used various terms for intra- and interpersonal competencies that had been identified and organized in the prior report referenced earlier (National Research Council, 2012b). For example, one early search included the terms "interpersonal competence," "teamwork," "communication," "intrapersonal skills," "motivation," "self-regulation," "metacognition," and "college persistence." Although all four searches included various terms for interpersonal competencies, they yielded no rigorous research suggesting that these competencies are related to persistence and success in undergraduate education. Therefore, the committee began to focus on intrapersonal competencies showing evidence of a relationship to undergraduate success. Next, to identify competencies that are malleable in response to interventions and find strong, causal evidence of a relationship with college success, the committee conducted a search of the intervention research.

In this process, the committee considered the potential role of students' mental health or illness in their college success. The committee commissioned National Academies library staff to

[1]As discussed in Chapter 1, the committee's definition of "underrepresented groups" encompasses low-income and first-generation college students; women in certain STEM disciplines; and black, Hispanic, and American Indian student populations.

conduct a search of the literature using terms related to mental health derived from the above report (National Research Council, 2012b; see Appendix A). Some studies found that students who enjoy high levels of mental health are more likely than those suffering from depression, chronic anxiety, and related mental illnesses to make friends at school, join campus organizations, and perform up to their full academic potential (Megivern et al., 2003; Salzer, 2012). In addition, the committee considered research on how the personality trait of neuroticism[2] may influence students' college success. As discussed in the following section, however, traits such as neuroticism are not considered malleable in response to the kinds of interventions that colleges can implement, and there is weak evidence that neuroticism is related to college persistence. Whereas a few studies suggest that neuroticism has a modest negative association with class grades (Chamorro-Premuzik and Furnham, 2003; Rigdell and Lounsbury, 2004), others show no such association (McAbee and Oswald, 2013).

After deliberating on this research, the committee determined that the broad domain of mental health lay outside its charge. In the committee's view, improving mental health is a very different goal from enhancing college success and completion. Moreover, the kinds of interventions that are required for addressing problems in mental health include therapy, counseling, and clinical work. In contrast, interventions designed to develop competencies aimed at college persistence are more academic by nature, sometimes implemented directly within courses.

Through this process, the committee arrived at the list of eight intrapersonal competencies presented above. The identification of these competencies was based on both correlational and experimental research. The committee, however, made judgments about the strength of the evidence related to each competency based exclusively on experimental research, for two reasons. First, as noted in Chapter 1, the committee was charged to identify competencies that are malleable in response to interventions. Because experiments attempt to isolate an intervention as the cause for changes in competencies and outcomes (e.g., through random assignment and the use of control groups), the experimental evidence from intervention research can provide strong evidence on a competency's malleability. Second, experimental manipulation can be stronger than nonexperimental approaches in reducing biases from omitted variables.[3] The committee's search for relevant intervention studies is described in Box 2-1; the effects on academic outcomes found in these studies are summarized in Table 2-1. Each study is described in more detail in the Intervention table, found on the report home page. Go to: www.nap.edu and search for "Supporting Students College Success." On the home page, scroll down and to the right to locate RESOURCES AT A GLANCE and click on "Intervention Table."

[2]As discussed further below, neuroticism is one of the "Big Five" (Goldberg, 1993; McCrae and Costa, 2008) personality traits, or broad dispositions.

[3]The committee acknowledges that both published experimental and correlational studies can suffer from a host of problems, including post hoc "fishing expeditions" that involve multiple measures, subgrouping categories, and so on (Gelman and Loken, 2013; Simmons et al., 2011). As one countermeasure, researchers encourage the preregistration of one's study design, hypotheses, and analysis plan (e.g., on AsPredicted at http://aspredicted.org [January 2017]; and through the Open Science Framework at http://osf.io [January 2017]). Exploration is still encouraged, but preregistration helps distinguish the researchers' a priori plan from their post hoc analyses. Also, the committee acknowledges that even under experimental control, there are many potential reasons for a treatment's effectiveness that can be explored further in future research (e.g., further experiments that isolate aspects of the intervention, use of different higher education institutions and samples).

Copyright © National Academy of Sciences. All rights reserved.

BOX 2-1
Search of the Intervention Literature

To gather evidence on relevant interventions, the committee searched the social-psychological intervention literature in several ways. The search started with all of the references cited in chapters in *Motivational Interventions* (Karabenick and Urdan, 2014), a recent edited volume summarizing work on the kinds of interventions the committee sought to describe. The committee also conducted searches in Google Scholar and searched websites of researchers known to be conducting social-psychological interventions. Whenever relevant intervention studies were found, the committee reviewed their reference lists for additional studies. All told, the search uncovered 49 articles describing 61 studies that met the inclusion criteria. These criteria required that the intervention (1) sought to manipulate one of the competencies identified in this chapter, (2) included clearly defined treatment and control and comparison groups comprising college students or individuals who were about to matriculate in a college, (3) was based on at least 10 subjects per group (most samples were much larger), (4) incurred less than 50 percent attrition between its start and the time at which the outcome was measured, and (5) employed random assignment. The search produced the following numbers of studies, grouped by competency: behaviors related to conscientiousness (7), sense of belonging (10), academic self-efficacy (2), growth mindset (17), utility goals and values (15), intrinsic goals and interest (3), prosocial goals and values (2), and positive future self (5). Among these 61 studies, 29 were conducted after 2010.

Regarding the impact of the interventions, the committee coded information on the size of the impact reported for all of the academic and competency outcomes included in each study. Some 47 studies among those coded report results for academic achievement outcomes; an equal number report impacts on the committee's eight targeted competencies, although not all studies provide enough data for the committee to calculate the size of the impact, with some merely reporting whether findings were statistically significant. For academic outcomes, course grades and GPA are reported most frequently. Impacts on college retention are reported in 5 studies. Regarding long-term impact, very few studies report impacts more than 1 year after the end of the intervention.

SOURCE: Prepared by the committee.

[End of BOX 2-1]

TABLE 2-1 Intra- and Interpersonal Competencies Associated with College Persistence and Success

		Strength and Nature of Intervention Evidence (random-assignment studies)	
Competency	**Definition**	**Strength**	**Nature of Evidence on Statistically Significant ($p \leq .05$) Effects**
Behaviors related to conscientious-ness	Behaviors related to self-control, responsibility, hard work, persistence, and achievement orientation	Modest	A student coaching intervention implemented at both public and private universities increased the 12-month college persistence rate by 5.3 percent. A goal-setting intervention improved students' college semester grade point average (GPA) (effect size = .65 standard deviation [SD]).
Sense of belonging	A student's sense that he or she belongs, fits in well, or is socially integrated at college	Promising	Six of seven studies show effects on college GPA (effect size range = .25-1.10 SD). Many impacts are limited to sample subgroups, such as first-generation college students, underrepresented minority groups, and women.
Academic self-efficacy	A student's belief that he or she can succeed in academic tasks	Modest	One of one study shows moderate to large effect sizes for achievement-related outcomes (effect size range = .42-.90 SD).
Growth mindset	A student's belief that his or her own intelligence (or any other important personal attribute) is not a fixed entity but a malleable quality that can grow and improve	Promising	Six of eight studies show effects on college GPA (effect size range = .38-.86 SD). Five of six studies show effects on final course grade (effect size range = .39-.74 SD). One of one study shows effects on course test performance (effect size = .96 SD). Many impacts are limited to sample subgroups, such as

		Strength and Nature of Intervention Evidence (random-assignment studies)	
Competency	**Definition**	**Strength**	**Nature of Evidence on Statistically Significant ($p \leq .05$) Effects**
			underrepresented minorities and students at risk for academic failure.
Utility goals and values	Personal goals and values that a student perceives as directly linked to the achievement of a desired end in the future	Promising	Five of six studies show effects on final course grade (effect size range = .06-.55 SD). Two of two studies show effects on college GPA (effect size range = .33-.52 SD). Many impacts are limited to sample subgroups, such as first-generation college students, underrepresented minorities, and women.
Intrinsic goals and interest	Personal goals and values that a student experiences as rewarding or meaningful in and of themselves, linked to strong interest	Modest	Two of two studies show effects on academic outcomes (written test of comprehension) (effect size range = .39-1.25 SD).
Prosocial goals and values	Personal goals and values aimed at helping others, furthering goals/values of a group or society as a whole, or promoting a prosocial religious or political agenda or some other endeavor that transcends self-interest	Modest	Two of two studies show positive effects on deeper learning behavior; however, neither study reports achievement impacts.

		Strength and Nature of Intervention Evidence (random-assignment studies)	
Competency	**Definition**	**Strength**	**Nature of Evidence on Statistically Significant ($p \leq .05$) Effects**
Positive future self	A positive image, picture, imagined trajectory, or personal narrative that a student constructs to represent what kind of person he or she will be in the future	Modest	One of one study shows positive effects on exam performance (0.56 SD).

SOURCE: Created by the committee.

A DEVELOPMENTAL FRAMEWORK

Every college student's experience is unique. Gender, race/ethnicity, age, and social class—and their unique configurations—shape students' college experiences in profound and innumerable ways. Students who attend elite private colleges tend to have different characteristics, enjoy different opportunities, and face different challenges relative to those who attend 4-year state institutions or community colleges, or those who take online courses, or those enrolled in technical schools aimed at training students for a particular niche in the economy. Graduation rates are higher at highly selective institutions and private nonprofit institutions than at less selective public colleges and universities (Kena et al., 2016). A 35-year-old minority commuter student who simultaneously holds down a full-time job would typically perceive the nature and value of her education very differently from an 18-year-old Caucasian freshman whose parents just drove him 500 miles to campus and helped him move into his dorm room. Likewise, two students with the same levels of achievement and aspirations in STEM fields in high school may be differentially challenged by a college environment if one is a minority who is reminded that few members of his race/ethnicity enter or succeed as STEM majors, while the other is a Caucasian student who typically experiences an environment of inclusion and high expectations (National Academies of Sciences, Engineering, and Medicine, 2016a). Thus, no single developmental course fully explains how all students move through their college years.

At the same time that these wide differences in situational contexts and individual characteristics influence college success, many college students face certain challenges in common, and in roughly the same order. Even before they take their first class, for example, students bring with them a set of competencies that reflect, in part, prior experiences in education and in life. These competencies speak to a basic question that many students may ask themselves even before entering college: *What are my strengths*? Such strengths may include *broad dispositional resources*, such as the ability to sustain academic effort even in the face of boredom or the abilities to recruit positive emotion and avoid or redirect negative emotion under conditions of stress. These basic dispositions may continue to develop through the college years, and the student may continue to draw on dispositional strengths in meeting the many challenges, both academic and personal, that arise in college.

Upon entering college, one of the first challenges a student tends to face concerns feelings or questions about fit or comfort in the new college environment. Whether the new student is trying to adjust to dormitory life or commuting from home to take courses in the evening, a central question that arises very early on is this: *Do I belong here*? As the student begins to attend classes and attempts to do the required academic work, another question is likely to follow: *Can I succeed*? Over the course of college—from the initial formative years until graduation—students will develop and refine *beliefs*, explicit or implicit, about the extent to which they belong in their college and the extent to which they are likely to be successful. These beliefs themselves constitute competencies in their own right and may also inform the development and expression of other competencies instrumental to college success.

As students continue to take classes, develop new relationships, participate in extracurricular activities, and become involved in college activities, their *specific motivations* regarding college may become clearer to them, as well as to others. Most students eventually ask more pointed questions about their college experience, such as: *What are my goals*? *What most interests me? What do I value*? Of course, some students have goals and values regarding college

long before they take their first college class. High school counselors, teachers, parents, and peers may ask prospective college students to explain why they want to go to college in the first place or what they plan to get out of college in the long run. For instance, some students have developed interest in and established academic goals for a STEM discipline. For these students, the departmental and classroom climates, interactions with faculty, the perceived relevance of the curriculum, and methods of instruction may be the most important factors in their commitment to completing a major (Carlone and Johnson, 2007; Chang et al., 2011; Seymour and Hewitt, 1997). Other entering students, however, may not initially self-identify as scientists (Chang et al., 2011) and may therefore tend to undervalue STEM course content and the particular instructional methods used to present it (e.g., Darling et al., 2008).

College success may hinge not only on goals and values but also on the student's identity, or understanding of self, in the context of the college environment. Over the course of college, the following questions may become increasingly salient: *Who am I*? *Whom do I want to become*? Developing a positive *identity* regarding the current and future self and linking that identity in meaningful ways to one's experiences in college is a daunting challenge for any college student. Indeed, the ability to do so may itself be considered an important intrapersonal competency, one that may be relevant at any point in a college student's career. It is likely to be especially important toward the end of college, as students who anticipate graduation look ahead to their postcollege future. It also may be the case that students who develop positive identities tied to their early college experiences are more likely to persist and graduate compared with students who are unable to do so.

BROAD DISPOSITIONS: *WHAT ARE MY STRENGTHS?*

Before prospective college students ever set foot on campus or enroll in their first class, they bring with them a unique life history and profile of general competencies. The intra- and interpersonal competencies that travel with the student into college, and evolve over time, include broad dispositional tendencies that pertain to emotional well-being and mental health, social and instrumental effectiveness, and attitudes and orientations regarding the self and the world. As an initial step in developing a list of competencies that appear to be related to persistence and success in undergraduate education, the committee drew on 50 years of research in personality psychology. This research has resulted in identification of the "Big Five" framework (Goldberg, 1993; McCrae and Costa, 2008) as a useful and generally comprehensive accounting of broad dispositions relevant to intra- and interpersonal competencies.

These five broad dispositions are typically labeled as conscientiousness, neuroticism (the opposite of which is emotional stability), extraversion, openness to experience, and agreeableness. Each of the five describes a set of intra- and interpersonal strengths or weaknesses students may bring with them upon entrance to college. Whereas all five dispositions can, in principle, have bearing on student success in college, the empirical evidence suggests that conscientiousness is the one trait among the five that is correlated most consistently with college persistence and success. Indeed, the correlational literature suggests that *conscientiousness* is the strongest predictor of college persistence and success among all intra- and interpersonal competencies.

Conscientiousness and Related Behaviors

Conscientiousness refers to a spectrum of dispositional tendencies that describe individual differences in the propensity to be self-controlled, responsible to others, hardworking, orderly, and rule-abiding (Roberts et al., 2014). The broad rubric of conscientiousness encompasses a large number of overlapping dispositional constructs that have, over the decades, been variously described in different theories of personality with such labels as self-control, ego control, effort control, delay of gratification, self-discipline, future orientation, constraint, planfulness, socialization, and grit. Individuals who score high on conscientiousness are typically described (by themselves and others) as especially well-organized, hardworking, dependable, disciplined, dutiful, and persevering. By contrast, those scoring low on the trait are seen as disorganized, impulsive, lazy, undependable, undisciplined, irresponsible, and lacking in self-control. At the temperamental heart of conscientiousness lies a capacity to set aside impulses and other short-term distractions to pursue long-term ends. In a useful conceptualization, Hill and Jackson (2016) describe conscientiousness as a general tendency for individuals to invest time and energy with a focus on long-term, socially valued payouts.

Within the Big Five taxonomy of dispositional traits, individual differences in conscientiousness are the strongest and most consistent predictor of success in various life domains, including marital stability, health, and occupational attainment (Roberts et al., 2007). High levels of conscientiousness are prospectively associated with success in nearly every type of occupational field, but especially in jobs and careers that demand high levels of autonomy (Barrick and Mount, 1993).

The correlational research indicates that conscientiousness predicts academic success long before college. For example, a number of studies have found that high scores on self-reported assessments of conscientiousness are associated with better high school grades (e.g., Freudenthaler et al., 2008; Poropat, 2009). In examining the possible reasons for the link between conscientiousness and GPA in high school, researchers have pointed to the role of academic effort: teenagers with greater conscientiousness tend to work harder in school. One recent study of nearly 5,000 German students in grades 5-8 showed that conscientiousness was a strong predictor of academic effort, especially when students found a school subject to be uninteresting (Trautwein et al., 2015). The results of this study suggest that in the face of boredom, many students may curtail their effort, but those high in conscientiousness may continue to work hard, which pays off in higher grades.

At the postsecondary level, empirical studies (based on correlational research) consistently demonstrate that high levels of conscientiousness predict better academic performance and higher levels of college completion (e.g., Chamorro-Premuzic and Furnham, 2003; Trapmann et al., 2007), even in research controlling for general cognitive ability (e.g., Okun and Finch, 1998; Poropat, 2009). Examining multiple measures of the Big Five dispositions, McAbee and Oswald (2013) found that measures of conscientiousness showed consistently strong predictive correlations with college GPA, whereas the other four dispositions (neuroticism, extraversion, openness to experience, and agreeableness) did not. In their authoritative review of psychological correlates of academic performance, Richardson and colleagues (2012) report that high effort regulation and low procrastination were statistically significantly associated with college GPA. Effort regulation is virtually synonymous with the self-control and perseverance themes of conscientiousness. Procrastination would appear to track

the low end of conscientiousness, as depicted in facets of conscientiousness that tap into punctuality, orderliness, and achievement striving (De Young et al., 2007; McCrae and Costa, 2008).

In an effort to tease apart the relationship between conscientiousness and academic success in college, Corker and colleagues (2012) examined the strategies, goals, and academic behaviors of nearly 350 college students over two semesters. They found that the link between high conscientiousness and high grades was explained by students with high conscientiousness engaging in such concrete behaviors as completing homework assignments on time, studying hard for tests, and persevering even when the course material was boring. Highly conscientious students put more time into their studies, and they completed projects in a timely manner. They also managed to suppress distractions that might divert them from completing their assignments, even when the assignments were tedious and difficult.

Although conscientiousness generally is a powerful predictor of academic achievement, some evidence suggests that the strength of this correlation varies across different student populations. In their study of Austrian adolescents, for example, Freudenthaler and colleagues (2008) found that the correlation between conscientiousness and GPA was statistically significant for girls but not for boys, whereas school-related intrinsic motivation, school anxiety, and performance-avoidance goals explained additional variance in GPA only for boys. As noted above, researchers (e.g., Corker et al., 2012) suggest that conscientiousness influences high grades through effort. However, DeLuca and Rosenbaum (2001) analyzed data from the High School and Beyond study and found that the rewards to effort varied with socioeconomic status. Although high school effort had a statistically significant relationship with later educational attainment net of socioeconomic status, the attainment benefits of effort varied by socioeconomic status. De Luca and Rosenbaum (2001) suggest that, even if students of low socioeconomic status strive very hard, their educational outcomes may not be improved or rewarded as much as those of other students, so they may have less incentive for school effort.

Lundberg (2013) conducted a study that identified substantial differences across types of families in the personality traits that predicted successful completion of college, particularly for men. Using data from the National Longitudinal Study of Adolescent Health to model the associations between personality traits and college graduation for different groups, she found that these relationships varied by gender, socioeconomic status (based on mother's level of education), and race. Among men, conscientiousness was statistically significantly related to college completion for those from families of high socioeconomic status but not for those from less advantaged families. Among women, however, conscientiousness was statistically significantly related to college completion regardless of socioeconomic status. For both men and women, openness to experience was significantly correlated with college completion for those from families of low socioeconomic status, but not for those from advantaged families. Estimating the model separately for black and white subsamples, Lundberg (2013) found that, regardless of socioeconomic status, the positive effect of conscientiousness on educational attainment was weaker for black men and women, and the marginal effect of openness to experience was stronger. As noted earlier, the committee believes further research is needed to understand how family and academic contexts influence the development of intra- and interpersonal competencies and their relationships with academic achievement.

Most personality psychologists see all of the Big Five traits, including conscientiousness, as deeply ingrained and relatively stable tendencies that are difficult to alter. As with all dispositional personality traits, individual differences in conscientiousness reveal, in part, genetic

variability among people. As measured by self-report tests and ratings of observers, individual differences in conscientiousness are at least 50 percent heritable (Bouchard, 2004; Jang et al., 1996; Power and Pluess, 2015). This means that at least half of the variation in conscientiousness scores within a population is attributed to genetic differences among people. Moreover, individual differences in conscientiousness tend to be relatively stable, and their stability increases over the adult life course (Roberts and DelVecchio, 2000). Put simply, the most conscientious person in a high school graduating class will probably score somewhere near the top of a distribution of conscientiousness scores at the time of his or her 10-year reunion.

Nevertheless, longitudinal studies also demonstrate that people tend to show a gradual increase in conscientiousness from late adolescence through midlife (Roberts et al., 2006). Underlying this tendency, different people change in different ways. A growing body of life-span personality studies demonstrates that major changes in social roles (e.g., getting married, becoming employed, having a child) are associated with changes in conscientiousness, although one study found that the changes in conscientiousness may have preceded these events (Specht et al., 2014). Nonetheless, major changes in social roles also constitute extremely intensive "interventions"—much more intensive than what might take place in a college setting.

Magidson and colleagues (2014) suggest that if an intervention aims to target conscientiousness directly, it should focus on the discrete behavioral and motivational factors that inform the broad trait of conscientiousness, ideally providing "detailed structure, a focus on values, guided action, goal-setting, immediate feedback on progress and challenges, clear accountability, and an opportunity for remediation" (p. 1445). The logic behind this approach suggests that conscientiousness is supported by, and articulated through, a collection of more specific intra- and interpersonal constructs that entail particular behaviors, goals, values, identities, and the like. These constructs and processes are perhaps not technically part of conscientiousness itself, but may work to support it and/or work with it to produce positive life outcomes related to academic success.

Moving beyond correlational research designs, the committee identified seven random-assignment evaluations of interventions that, as suggested by Magidson and colleagues (2014), target discrete motivations and behaviors related to conscientiousness. Most of these interventions consisted of setting, monitoring, and (in some cases) coaching various study goals and work habits. Four of the seven interventions assessed impacts on academic outcomes. The largest of these studies is that of Bettinger and Baker (2014), who evaluated InsideTrack, a for-profit student coaching service that is now the largest provider of student coaching in the country. Most of the more than 13,000 participating students were nontraditional college students. The exact nature of the intervention activities is proprietary information, but one of the study authors reports that InsideTrack coaches reached out to assigned students (via phone, text, email, and social media). The goals of the meetings varied according to student need, but the coaches generally assessed potential risk factors, worked to develop both cognitive and noncognitive skills (e.g., time management, self-advocacy) and guided students in seeking and leveraging campus supports. Coaches usually worked with their students over the course of two semesters.

Bettinger and Baker (2014) were forced to rely on data that the company supplied, although extensive checks showed broad-based comparability between treatment and control groups. They found that this goal-setting and support intervention increased the fraction of students persisting in the university 6 months after the end of the intervention by 5.1 percentage points, the fraction persisting 12 months after the intervention by 5.3 percentage points, and the

fraction persisting 24 months after the intervention by 3.3 percentage points. Rates of college graduation were 4.0 percentage points higher for the treatment relative to the control group. All of these differences were statistically significant.

Morisano and colleagues (2010) report that their web-based goal-setting intervention increased retention rates significantly but do not report the sizes of those effects. Liu and colleagues (2012) sought to motivate students with information that their performance on a test battery might be reported to employers, and found impacts of around .40 standard deviation on a proficiency test, scores on an essay, and a self-reported motivational assessment. Duckworth and colleagues (2016) found (in Study 3) that psychology students randomly assigned to a condition that set study goals and attempted to remove temptations that might interfere with meeting those goals had greater success at meeting the goals 1 week after the intervention relative to controls, although they did not test for achievement or attainment impacts.

Taken together, the evidence that conscientiousness-based interventions improve college persistence is modest (see Table 2-1). Most random-assignment studies failed to measure academic outcomes, and the one that did (Bettinger and Baker, 2014) has proprietary restrictions on revealing the exact nature of the intervention. The Bettinger and Baker (2014) intervention showed positive effects on the 12-month college persistence rate of 5.3 percent, and the goal-setting intervention (Morisano et al., 2010) showed significant improvement in students' college semester GPA (effect size = .65 standard deviation [SD]).

BELIEFS: *DO I BELONG HERE? CAN I SUCCEED?*

The transition to college presents an exciting opportunity for many new students, but it can also present daunting personal and interpersonal challenges. Many new students, just out of high school, are leaving home to live on their own for the first time. Some are leaving their families to live on a distant campus they may have seen only once or twice before. Many are breaking away from established friendship networks while establishing new ones. Others may be trying to juggle part-time jobs or existing family commitments to take on a demanding new role. Those who are the first in their family to attend college may have few identifiable role models in their immediate college environment to help them navigate the college experience. At the same time, first-generation students may also feel the pressure of high expectations, as if the hopes and dreams of the family are now riding on their efforts and success. Students from homogeneous communities may be experiencing racial and socioeconomic diversity for the first time in dorms and classrooms. Academically and socially, many new college students may feel as if they are now strangers in a strange land.

Indeed, models of voluntary student departure and persistence suggest that the transition from high school to college is a critically important period, as the early development of goals/intentions and institutional commitment is critical to engagement and eventual completion (DeAngelo et al., 2011; Nora and Crisp, 2009; Tinto, 1994). It is not surprising, therefore, that researchers and institutions have devoted significant attention to the transition experience and to interventions that can help new students navigate college.

The question *Do I belong here*? may arise in a multitude of contexts for students beginning their college experience. New students want to know, Can I make new friends in this new place? Are there groups on campus that will accept me? Do I feel comfortable here? Do I feel that I fit in? After placement tests and by midterm of the classroom experience, another question is certain to arise for many students: *Can I succeed here*? Students wonder, Can I meet

expectations and perform well in these classes? If the academic strategies I have used in the past do not seem to work here, what should I do?

These questions are very personal, yet the ways in which a student answers them depend on many factors that go well beyond the student's own dispositional strengths. *Institutional* factors include the general ethos of the college environment itself—its demographic makeup and peer climate, the level of resources devoted to student life and instruction, the opportunities made available for meaningful involvement, the quality of instruction, the selectivity of the institution, and the financial aid awarded (Bowen et al., 2011; Oseguera and Rhee, 2009; Singell, 2004; Titus, 2004). Experiences in high school and financial concerns also are likely to have a broad impact on both initial college choice and retention (Cabrera et al., 1993; Paulsen and St. John, 2002).

Relevant as well are many *individual* factors that go well beyond the broad dispositional trait of conscientiousness. An important internal factor may be the particular beliefs about college that a student develops early in his or her career. Indeed, what a student believes or perceives about the environment may be nearly as important as the objective realities of the environment itself. For example, perceptions of the campus climate (e.g., hostility, competition) affect the successful transition to college for students from many different racial/ethnic groups and majors (STEM and non-STEM) in terms of managing academic adjustment and integration in college (Cabrera et al., 1999; Hurtado and Carter, 1997; Hurtado et al., 2007; Locks et al., 2008). Similarly, what the student believes about his or her own abilities to succeed may be nearly as important as the strength of those abilities.

When it comes to beliefs about college, then, the research literature suggests that at least three intrapersonal competencies may be involved in determining how well students do in college: *a sense of belonging*, *academic self-efficacy*, and a *growth mindset*.

Sense of Belonging

A student's sense of belonging is an intrapersonal competency that is influenced by interpersonal/social relationships as well as academic/professional concerns. It exists as a complex and evolving set of beliefs in the mind of the student regarding the extent to which he or she fits into important niches or features of the college environment.

Motivational theory posits that individuals are most likely to engage and perform positively in settings in which they feel a sense of connectedness or relatedness to others (Deci and Ryan, 1991). At the heart of human nature lies a basic need to be part of a group, also referred to as a psychological sense of integration (Tinto, 1994) or sense of belonging (Hurtado and Carter, 1997; Strayhorn, 2012). The transition to college marks a critical period in a person's life when the sense of belonging is likely to be strongly activated. As a complex social ecology, college presents a panoply of groups with which students may affiliate, including formal organizations with rules for membership (e.g., sororities and fraternities), informal friendship networks, religious and political student organizations, and groups organized around academic concentrations (e.g., major and professional career clubs) and other personal interests (e.g., hobbies and sports).

Research shows that students with a higher sense of belonging in STEM are "more likely to report having friends in the major, socializing with peers and faculty in the field, and feeling like their 'friends would miss [them]' if they left" college (Strayhorn, 2012, p. 68). In the minds of some students, moreover, the college itself may resemble an extended group, as participation

in these social niches enables them to feel that they are an integral part of the college writ large (Hurtado and Carter, 1997; Hurtado et al., 2007). A sense of belonging also is associated with whether students feel that the "climate" for learning is welcoming or unwelcoming in their first year of college (e.g., instructional practices that promote cooperation and inclusiveness rather than competition and stereotyping).

The research literature using correlational methods suggests that feeling a sense of belonging, bonding, or solidarity with others in college tends to be positively associated with college retention (Robbins et al., 2009). Although most students' intention to persist in college tends to decline over the first year, a random-assignment intervention designed to cultivate a higher sense of belonging among students was found to stem this decline (Hausmann et al., 2009). The intervention, which entailed providing community-oriented letters from university administrators and gifts with university logos versus general letters from faculty and generic gifts, was found to have similar effects for both white and black students, on average.

Sense of belonging is positively associated with the ability to manage academic adjustment, with grades, with self-rated change in the ability to conduct research, and with perceptions of the relevance of coursework in the first year of college (Hurtado et al., 2007). Although performance in the first year of college is a strong determinant of degree completion within 6 years, where one goes to college (level of selectivity) and how students feel about belonging in their college during this transition year are associated with 6-year degree attainment across all racial/ethnic groups (Hurtado and Ruiz Alvarado, 2016). Pittman and Richmond (2008) found that higher scores on sense of belonging predicted higher grades, academic competence, and positive self-worth. On the other hand, Richardson and colleagues (2012) found little evidence in their meta-analysis that such variables as social integration and social support were associated with college GPA. Their review, however, focused not on specific measures of sense of belonging but on general social integration measures that often include a variety of social activities, nor was it specific to the first year of college.

Although belonging and connectedness to an educational context are important for all students, they may be especially important to college achievement and persistence for student groups that are underrepresented in higher education (Strayhorn, 2012). Some studies have shown that underrepresented minority groups report a lower sense of belonging relative to their white counterparts (Johnson et al., 2007), in part because of perceptions and experiences of discrimination (Hurtado and Carter, 1997). The former groups are especially likely to experience challenges to their identity (in the form of negative climate, discrimination, and perceived microaggression) that can undermine their sense of belonging and connectedness to the academic context (National Academies of Sciences, Engineering, and Medicine, 2016a). Students from these backgrounds tend to be particularly underrepresented in some STEM contexts (e.g., being the only Latina in a science class), which can make it even more challenging to develop or maintain a strong sense of belonging.

The committee identified 10 intervention studies addressing students' uncertainty about "belonging" in college settings (see Appendix B). Of these 10 studies, 7 report academic outcomes, including GPA in 5 of the studies, and 8 report competency outcomes. In a laboratory intervention, Walton and Cohen (2007) attempted to mitigate doubts about social belonging in college. Participants in the treatment condition received survey information from upperclassmen indicating that most students worry about belonging during their first year, but that these concerns lessen over time. Control students received information unrelated to the issue of belonging. Relative to their control group counterparts, black students (but not others) in the

treatment group experienced a substantial increase in their GPAs. The intervention also improved students' achievement behavior, again for black participants only. Black treatment group students reported studying an average of almost 1.5 hours more per day compared with black control group students, and also sent more email queries to their professors. The black treatment group students also felt a greater sense of academic fit, both immediately after the intervention and 7 days later. The treatment, however, actually had adverse effects for white students, whose sense of academic fit had decreased immediately after the intervention and 7 days later.

In a subsequent study, Walton and Cohen (2011) implemented a similar social belonging intervention, in this case lasting 1 hour in a laboratory setting, involving students in their second semester of their first college year. As in the Walton and Cohen (2007) study, black students in the treatment group experienced a significantly higher increase in their GPA relative to black students in the control group, with the GPA effect persisting through their senior year. GPA trajectories of white students in the treatment condition were not significantly different from those of their control group counterparts. Overall, the intervention closed the minority achievement gap by 52 percent.

Walton and colleagues (2015) report on two interventions targeted at women in engineering. The first entailed communicating to 228 female and male students in their first college semester that uncertainty about belonging in engineering is common during the first year of college, but it is temporary. The second, a values-affirmation intervention, encouraged students to incorporate personally important self-identities and values in their daily lives as an explicit strategy for managing stress and threat. This intervention was designed to empower the female students to maintain their own well-being. Students' transcript data for their first year and survey data from questionnaires administered immediately after the intervention both showed that the first intervention was generally effective for women in male-dominated but not in gender-diverse engineering majors. Both interventions raised GPAs meaningfully (by about 11 points on a 100-point GPA scale) among women in male-dominated majors relative to the control condition, but no statistically significant differences emerged for women in the gender-diverse majors. Walton and colleagues (2015) also report positive competency outcomes. The two interventions combined generally improved women's felt experience in engineering—as measured by a sense of belonging, self-efficacy, and enjoyment—relative to the control condition (a moderately large effect size equal to a standardized mean difference of 0.67). This outcome persisted to the second semester. The interventions also improved women's confidence in their prospects for succeeding in engineering.

An intervention study conducted by Stephens and colleagues (2014) targeting first-year, first-generation college students provides further support for the efficacy of interventions targeting sense of belonging. The intervention consisted of students in both treatment and control conditions attending a panel discussion. Both sets of panelists highlighted how they had adjusted to and succeeded in college. The key difference between the treatment and control panels was whether the panelists' stories highlighted how their social class backgrounds mattered for their college experience. First-generation students' end-of-year GPAs were substantially higher for the treatment than for the control group, and not statistically significantly different from the GPAs of non-first-generation students. A number of psychosocial outcomes also were measured. Treatment participants (regardless of generational status) scored significantly higher than controls on the following measures: psychological well-being, social fit, academic identification,

and maintaining relationships. They did not differ from control participants on psychological distress, social identity threat, and social support.

A study by Folger and colleagues (2004) represents the last one in this category to include GPA among its achievement outcomes. In this case, three different GPA outcomes were measured: first-semester, second-semester, and cumulative GPA. The study participants were academically at-risk first-generation college students who participated in the Freshmen Empowerment Program (FEP). Unlike the other interventions in this category, the FEP was intensive, not brief. As part of this program, treatment students met for 6 weeks to discuss a range of topics (e.g., academics, college resources, adjustment to college). At each time point, the mean GPA of the intervention group was statistically significantly higher than that of the control group, but effect sizes are not reported.

A belonging intervention study conducted by Walton and colleagues (2012) examined the relationship between social connections and achievement motivation. In this study, students were randomly assigned to a skill-promotive context condition or to a relational context condition. In both conditions, students read a fictional report written by a recent graduate of the math department. The reports were identical; they varied only in their characterization of the math department. In the skill-promotive condition, the department was portrayed as providing students with opportunities to develop their personal ability and interest in math, whereas in the relational context condition, the report portrayed opportunities for positive, collaborative social interactions. Students in the latter condition reported greater motivation for and sense of social connectedness with math.

Taken together, the evidence from interventions targeting sense of belonging is promising. Among the eight intrapersonal competencies identified in this report, interventions targeting this competency showed the most consistent impacts on college performance. Six of seven studies testing the relationships between these interventions and GPA showed significant positive effects (effect size range = .25-1.10 SD), although their effects were often selective—significant for some groups but not others. Importantly, the selective impacts tended to be stronger for underrepresented student groups—women in male-dominated engineering majors in one study, first-generation students in another, and black students in two others.

Research using correlational methods has shown that a strong sense of belonging early in college is associated with engagement in specific college activities and with success in college, and a growing body of intervention research demonstrates that it is possible to foster such a sense of belonging. More research is clearly needed on how to develop a sense of belonging among college students and how this competency promotes student success. It is especially important to disaggregate group data on this construct in order to examine differences and similarities across gender and other demographic groups. Sense of belonging may pose an especially vexing problem for underrepresented student groups (e.g., first-generation college students, underrepresented minority groups), who may feel disenfranchised or on the periphery in a given college context. Programs and policies aimed at promoting sense of belonging need to focus on the specific needs of and challenges faced by particular student groups, identified by particular demographic profiles and life history experiences.

Academic Self-Efficacy

For nearly four decades, psychologists have conceived of self-efficacy as a person's belief that he or she can successfully carry out "courses of action required to deal with

prospective [future] situations containing many ambiguous, unpredictable, and often stressful elements" (Bandura and Schunk, 1981, p. 587). According to proponents of this construct, self-efficacy is specific to particular tasks or realms of life. Therefore, a person might have a different level of self-efficacy in, say, the realm of sports compared with the realms of work, friendships, health and diet, and so on.

Academic self-efficacy, then, is a student's belief that he or she can carry out actions that will lead to success in school. For college students, academic self-efficacy is a subjective belief in one's own competence to meet the academic demands that arise in college. The belief is surely influenced by prior experiences of success and failure in high school. Not surprisingly, one's belief in one's competence is a function of one's actual competence in a given area. Nonetheless, the subjective belief itself is viewed as having strong motivational power, above and beyond objective competence itself. Indeed, the subjective belief can boost (or hinder) objective competence.

Self-efficacy beliefs are construed as empowering the individual to become a more effective agent in a given goal domain. Psychologists typically assume that self-efficacy beliefs derive from past events in which the person experienced mastery or success in the given domain. In expectancy-value theories of motivation (e.g., Eccles, 2009; Eccles et al., 1983; see also Wigfield and Cambria, 2010), expectancy refers to an individual's subjective expectation of success for a given task—not unlike self-efficacy. According to expectancy-value theories, people are motivated to pursue goals or ends (1) wherein they expect success and (2) to which they assign importance or value.

A considerable body of research has shown that high levels of academic self-efficacy are positively associated with college students' academic performance and college retention (e.g., Chemers et al., 2001; Matten and Shaw, 2010). It is important to note that this research provides no causal evidence. In a meta-analysis of psychosocial and study skills in college, for example, Robbins and colleagues (2004) found that academic self-efficacy was the strongest predictor of GPA. In their meta-analysis of a broad set of psychological factors implicated in college success, Richardson and colleagues (2012) identified performance self-efficacy as a significant predictor of college GPA, even after controlling for cognitive ability as measured by high school GPA and SAT or ACT scores. Lent and colleagues (2005) found that academic self-efficacy enhanced motivation and performance for minority students pursuing engineering careers. Similarly, Nordstrom (2012) found that mathematics self-efficacy predicted achievement in basic-skills math courses for Hispanic students in a community college.

Developing and maintaining strong academic self-efficacy can be uniquely challenging for underrepresented student groups. A robust body of scholarship in the area of stereotype threat (e.g., Steele, 1997) has demonstrated how stigmatizing experiences based on race/ethnicity, gender, or social class serve as "identity threats" that may undermine students' academic task performance. These threat experiences are represented by academic settings that make stereotypes salient or by students' encounters of stereotype-based treatment in their academic settings. Such experiences are consistently documented among racial/ethnic minority students and women in STEM (National Academies of Sciences, Engineering, and Medicine, 2016a). Repeated threat experiences over time are said to lead to students' disidentifying with, or disconnecting their personal identity from, the academic domain in order to protect their self-concept from the deleterious effects of threat experiences. Ironically, the experience of threat is most likely for those students who identify most strongly with the academic domain. These processes have been demonstrated in STEM specifically. Chang and colleagues (2011), for

example, report that students who possess a high science identity (i.e., a strong connection to and self-efficacy in science) but encounter stereotype-based experiences in college are more likely than students who do not have such experiences to change majors in the first year of college.

Research on improving self-efficacy is often specific to a particular knowledge domain or context. Fitch and colleagues (2012), for example, designed a "solution focused, goal-setting, group work" intervention for college students that resulted in statistically significantly higher scores on measures of academic self-efficacy for the experimental relative to the control group.

Two intervention studies that met the committee's criteria targeted academic self-efficacy, although only one measured impacts on academic outcomes. Luzzo and colleagues (1999) assessed the academic impacts of an intervention that consisted of a short video presentation from two graduates of the university who described their early indecision over choice of major. The two were undeclared majors early in college, but after successful encounters with STEM courses had gone on to graduate and begin successful STEM-related careers. Intervention subjects (94 in all) were college students who at the time of the intervention had not yet decided what career to pursue and possessed at least a moderate degree of math ability. The intervention generated quite large positive impacts on STEM-related academic self-efficacy, as well as large impacts (standardized mean difference of .90) on subsequent enrollment in STEM courses and moderate impacts (standardized mean difference of .40) on opting to major in a STEM field.

Overall, evidence from interventions designed to boost academic self-efficacy is modest. A single study showed significant, moderate to large effect sizes for achievement-related outcomes (effect size range = .42-.90 SD). Although correlational research identifies this competency as a powerful predictor of college success, the modest evidence from interventions highlights the need for further research to determine whether, and to what extent, simply believing that one can do well will help one do well. Future studies need to examine how self-efficacy develops and how it can be promoted in different student groups. It may be that specific kinds of efforts are needed to boost and/or help maintain high academic self-efficacy among underrepresented minority groups, women, students from low-income families, and first-generation students in STEM fields. Such research would attend to cultural or structural features of the academic context that may represent threats with the potential to negatively impact academic self-concept and self-efficacy among these underrepresented groups.

Growth Mindset

Whereas academic self-efficacy refers to a student's beliefs regarding his or her ability to be successful in a particular academic pursuit, growth mindset refers the extent to which a person understands his or her own ability to be fixed or malleable (Dweck and Molden, 2005). The target ability most often studied is general cognitive ability (intelligence). An *entity-based mindset* applies when a person understands his or her own intelligence to be a fixed entity: The person may believe that he or she has high ("I am smart") or low ("I am dumb") intelligence, or somewhere in between, but whatever the level is, the person assumes that the level is more or less unchangeable. A fixed-entity mindset on intelligence becomes especially problematic when the person encounters challenges that seem to go beyond his or her self-ascribed level of intelligence: "Because I cannot accomplish this task, I must be dumb; and because my intelligence is fixed, I cannot make myself any smarter than I am to accomplish this task." In the face of difficult cognitive challenges, then, a person with an entity-based mindset regarding

intelligence may give up. By contrast, a *growth mindset* assumes that intelligence is malleable and, therefore, improvable through effort. When faced with difficult cognitive problems, the person with a growth mindset regarding his or her intelligence may show more resilience and may be better able to learn from mistakes.

Believing that one's intelligence can grow through experience thus may constitute an intrapersonal competency that has implications for both identity and motivation. From the standpoint of identity, the growth mindset indicates that "I am the kind of person who can learn from mistakes and improve over time; I can get smarter and accomplish more through hard work." This understanding of the self may motivate the person to work harder, which may ultimately result in deeper learning and greater success.

A recent meta-analysis shows that growth mindset tends to be associated with greater self-control and with positive expectations in the pursuit of goals (Burnette et al., 2013; see also Robins and Pals, 2002). A growth mindset also tends to be positively associated with academic achievement, although the effect sizes here are relatively modest (Blackwell et al., 2007).

Most social psychologists who study mindset assume that an individual's mindset is readily modified, an assumption that has led to many interventions in this area. Among all the intervention studies meeting the committee's criteria (see Appendix B), the largest number (17) focused on mindset. Specifically, these studies examined the way in which students responded to failure events, and were meant to increase the likelihood that students would attribute failure to controllable causes, such as effort, and decrease the likelihood that they would attribute it to uncontrollable causes, such as ability. Within this group of mindset interventions, 16 assessed impacts on academic outcomes, although not all of them report impacts in a readily interpretable form. Aronson and colleagues (2002) targeted stereotype threat among African American students with a pen-pal condition that encouraged students to view intelligence as malleable rather than fixed. The intervention boosted beliefs about the malleability of intelligence several months after it ended, as well as enjoyment of education and identification of academic achievement. A number of these impacts were statistically significant for both African American and white students. Relative to controls, later GPAs were boosted for African American students only.

A number of studies consisted of an attributional retraining (AR) intervention. Prior to completing an achievement task, students viewed a video depicting two students discussing the ways in which poor performance can improve. The video dialogue focused on controllable attributions; after performing poorly on an exam, for example, students discussed how putting more effort into studying helped improve their grades. The video generally was followed by an activity that reinforced the importance and functionality of effort attributions.

In their intervention study, Hamm and colleagues (2014) were unable to detect an effect of their AR treatment on cumulative first-year GPA. Similarly, Haynes and colleagues (2008) were unable to find a direct relationship between AR and cumulative GPA; however, they detected an indirect relationship such that AR predicted mastery motivation, which in turn predicted GPA. Perry and colleagues (2010) found a significant treatment effect, with AR participants having higher GPAs compared with their no-AR counterparts (a difference of .26 grade points). In this case, AR participants also performed better on a postintervention class exam and earned significantly higher course grades relative to the no-AR group (effect sizes were in the moderate range for low- and average-performing students compared with controls). On the other hand, Hall and colleagues (2006) were unable to detect a statistically significant treatment effect for the final course grade. Finally, in a study of AR training, Ruthig and

colleagues (2004) found a marginally significant treatment effect on cumulative GPA and a statistically significant impact on voluntary course withdrawal (AR-trained students dropped fewer credit hours relative to those in the control condition). This finding is somewhat consistent with the finding of the Hamm and colleagues (2014) intervention study that that among students characterized as failure acceptors, those in the AR treatment group were 61 percent less likely to drop a class relative to their peers in the no-AR condition.

In their intervention study, Eskreis-Winkler and colleagues (2016) sought to improve math performance among college students. The intervention was similar to prior AR interventions in that participants were taught that talent and effort both contribute to success. In this study, however, the intervention was designed to motivate participants to engage in deliberate practice (an especially effortful activity) and improve their achievement. Moderate treatment effects were found for end-of-year GPAs.

Yeager and colleagues (2016) report on an intervention study that took place after students had been admitted to college but before they had arrived on campus. The participants were primarily African American and first-generation college students, and their achievement outcomes were measured during their college years. In this study (Study 1 in Yeager et al., 2016) the authors tested two interventions—one focused on social belonging and the other on growth mindset. Both interventions— designed to teach students a lay theory for the transition to college. In essence, they provided students with a framework that helped them understand that challenges are common in the transition to college and that students often question their academic success and sense of belonging. They also taught students how to face those challenges effectively. The main outcome measured was full-time enrollment during students' first year in college. While the growth mindset intervention showed no treatment effects, the social belonging intervention increased full-time enrollment over the next academic year among the African American and first-generation treatment group participants by 34 percent relative to the control group.

Taken together, the evidence on the effectiveness of growth mindset/AR interventions for college performance and persistence is promising. Six of seven studies that examined GPA as an outcome showed significant effects (effect size range = .38-.86 SD); two of four studies showed significant effects on final course grade (effect size range = .57-.78 SD); and four of four studies showed significant effects on course test performance (effect size range = .73-1.56 SD). Many impacts were limited to sample subgroups, such as underrepresented minorities and students at risk for academic failure. A growth mindset appears to influence motivation in a positive way, encouraging students to work harder even in the face of failure.

SPECIFIC MOTIVATIONS: *WHAT ARE MY GOALS? WHAT MOST INTERESTS ME? WHAT DO I VALUE?*

By the time they enter college, most students have some sense of why they are there. Some are attending college because doing otherwise is nearly unthinkable. Their parents, and perhaps even their grandparents, attended college and have always expected that they will follow suit. Other students come from families in which they are the first to attend college, and they may pursue college for an education and a degree that will allow them to attain employment and social and economic mobility not available to others in their families and communities. Many students hope that college will prepare them for gainful employment down the road. Some may even value education for its own sake, expressing a love for learning that may motivate a lifelong

quest to expand and refine their minds. Still others have already developed a specific interest in a field of study. They know what their major is going to be, and they may even have a plan for what job or postgraduate studies they will pursue later on.

For many students, however, such goals, values, and areas of interest are fluid early on, and these students have only a vague understanding of their goals and values when they begin college. Moreover, those whose motivational agendas are more solidified at first may later change their aspirations dramatically. Many develop major and career interests they never knew they had. Many also confront intellectual and motivational challenges in the discipline they expected to pursue. Over time, students change majors; switch schools; and try out many different plans, projects, and programs. They are looking to identify their goals, values, and areas of interest and then to fulfill them. Questions concerning motivation thus may become more salient as students move through their college years: *What are my goals*? *What most interests me? What do I value*?

The research literature in developmental and social psychology and in higher education suggests that the specific ways in which students approach, work with, and seek to realize their goals constitute critical skills for success in college. Beyond broad dispositional strengths and particular beliefs about college, students' motivations also contribute to their intrapersonal competencies.

Utility Goals and Values

According to expectancy-value theories of human motivation (Eccles, 2009; Eccles et al., 1983; see also Wigfield and Cambria, 2010), students will work hard to succeed when they (1) *expect* to succeed in (2) a *valued* goal domain. Whereas academic self-efficacy speaks directly to a student's expectations of success, the extent to which the student values the particular goal area within which success may be expected is an important factor to consider as well. After all, a person may expect to succeed in a given goal domain but may not value the goal, and therefore may not pursue it with the vigor or commitment that is necessary for success. Such a student may say to herself, "I could succeed in Task A if I wanted to, but I don't value Task A, and therefore I see no point in striving to achieve Task A." Expectancy-value theories propose the common-sense idea that people do their best when they are striving to achieve goals they value. Of course, extrinsic values can be more immediate (e.g., valuing a goal because it will produce a good grade) or more distant (e.g., valuing graduation because it will lead to paid and sustained employment).

Immediate and distant values are clearly related, and understanding the utility of a goal is essential to appreciating value in all its forms. A student's perception of the utility of coursework, for example, can be an important extrinsic payoff. For instance, achieving a high grade in a chemistry class may enhance the prospects for getting into medical school. The student may not value chemistry per se, but excelling in the coursework is a means to an end if she clearly values the goal of becoming a physician in the future. Similarly, a student may work hard to achieve a good grade in a boring introductory survey course because passing that course opens a gateway to other courses that are expected to be much more interesting. Another student may join a campus organization focused on addressing a particular societal problem. The student may not be especially interested in that problem itself, at least initially, yet may know that accumulating such experiences may pay off later on when he applies to law school. This same student may develop a real intrinsic interest in the societal problem, and may even change his

career aspirations as a result. But in each of these hypothetical cases, the student is initially motivated to pursue a goal not because it is valued for its own sake but because it may facilitate attainment of the student's long-term aims or long-range anticipated rewards.

The committee located 15 intervention studies that targeted utility goals or values. Each assessed at least one academic outcome, and only 5 assessed at least one competency outcome.

Harackiewicz and colleagues (2014) conducted a values affirmation intervention intended to close the achievement gap for first-generation college students studying STEM. Targeting students taking an introductory biology class, this intervention was delivered twice during the semester through a brief writing exercise in which students wrote about values most important to them. Students in the experimental group wrote about why two or three values selected from a list were important to them. By contrast, the control group students wrote about why values least important to them might be important to someone else. The effect of this type of intervention on GPAs was significant for first-generation students but not for "continuing-generation" students. Further, the intervention promoted continued enrollment in the biology course sequence; first-generation students in the values affirmation group were more likely to enroll in the next biology course compared with continuing-generation students, a difference in enrollment of roughly 10 percent (75 versus 86%). They also were more likely to enroll in the next course compared with their counterparts in the control group—a difference in enrollment of roughly 20 percent (66 versus 86%).

In a subsequent study, Harackiewicz and colleagues (2015) implemented both a values affirmation intervention and a utility value intervention in an introductory college biology course. In both interventions, students wrote brief essays at three different times during the course. Those in the values affirmation treatment group wrote essays similar to those in the Harackiewicz et al. (2014) study, whereas those in the utility value treatment group were asked to select a concept or question relevant to the course and answer it, drawing on course material and explaining why this information was directly relevant or useful to their own lives. Although there were no significant effects for the values affirmation intervention, the utility value intervention significantly improved student performance relative to the control group. The impacts were largest in the case of 32 underrepresented minority students who were also first-generation students assigned to the treatment condition; their grades rose more than half a letter grade on average compared with their 32 counterparts in the control condition (slightly over 2.5 versus slightly under 2.1 on the GPA scale).

Brady and colleagues (2016) conducted a values affirmation intervention study focused on both achievement and competency outcomes, but here the sample consisted of 183 first- and second-year Latino and white college students. As with the Harackiewicz et al. (2015) intervention, the students completed a brief writing exercise in which those in the affirmation condition wrote about their most important value, while those in the control group wrote about a low-ranked value (ranked 9th out of 11) and why this value might be important to someone else. This study also was similar to that of Harackiewicz et al. (2014) in that the intervention's effect size was positive and moderate—but only for Latino students. For Latino participants, GPAs collected 2 years after the end of the intervention were significantly higher for students in the affirmation condition than for their counterparts in the control group—on par with those of the white students in the control group. It appeared that the initial affirmation intervention, during the sensitive first year of college, helped Latino students see subsequent adversities in a more optimistic light and presumably helped them overcome those adversities, which further increased

their confidence. Among white students, on the other hand, GPAs for those in the treatment group were significantly lower than GPAs for those in the control group.

The affirmation intervention also had a large positive impact (a standardized mean difference of 0.94) for Latino students on "adaptive adequacy"—a measure of self-integrity, self-esteem, and hope. On the other hand, the difference in adaptive adequacy between white treatment and control group participants was quite small and not statistically significant. A similar pattern emerged for academic belonging: Latino students in the affirmation condition felt a stronger sense of belonging in school compared with control group participants, whereas the difference between whites in the two conditions was not statistically significant.

Miyake and colleagues (2010) conducted a brief (15-minute) values affirmation writing intervention targeting students in an introductory physics course. Student performance was measured in two ways: with an overall exam score in the course (an average score of the percent correct for four exams) and with a Force and Motion Conceptual Examination (FMCE). As in other values affirmation studies, this intervention proved to be effective in reducing the achievement gap, but in this case only for women. Final course grades shifted from the C to the B range for women in the affirmation condition but not for women in the control condition, and there was no such final grade improvement for men. A similar pattern emerged for the FMCE score.

Martens and colleagues (2006) evaluated two values affirmation interventions designed to alleviate the performance gap between men and women due to stereotype threat. In both studies, women who had been exposed to stereotype threat were randomly assigned to a self-affirmation group or a no-affirmation control group. In both studies, students in the treatment condition participated in a standard affirmation exercise in which they wrote about their most important value. Women in the treatment condition performed statistically significantly better on a difficult math test (in Study 1) and a spatial rotation test (in Study 2) relative to the control group in each study. The authors do not present effect sizes for the two studies.

Turning from academic outcomes to competencies, a small number of studies demonstrate how these types of brief writing interventions impact the target competency. In a laboratory study, for example, Hulleman and colleagues (2010) addressed the targeted competency directly, as measured by situational interest in a new method for solving two-digit multiplication problems that was taught in the intervention. Before using this new technique to solve a series of problems, students participated in a writing activity that resembled the Harackiewicz et al. (2015) utility value intervention. Those in the relevance condition wrote about how the math activity could relate to their life or to the lives of college students in general. Control participants wrote about a topic unrelated to the math activity. Competency measures were assessed following treatment. Compared with those in the control condition, students in the relevance condition became more interested in the new math technique, and they indicated that they were more inclined to use the technique in the future.

Similar interventions were implemented and studied by Durik and colleagues (2015) and Schechter and colleagues (2010). Instead of participating in a relevant writing task, however, students were simply told about the usefulness of the new technique for their performance in future classes, preparation for graduate school admissions tests, and future careers. Treatment outcomes for situational interest differed depending on the type of student. In the Durik et al. (2015) study, for example, treatment students high in their initial perceived ability in math were more situationally interested in the math technique relative to the control group. Treatment students low in their initial perceived ability in math were less situationally interested in the math

technique, relative to the control group. A similar pattern emerged when students were grouped by initial interest in math; however, treatment effects were not statistically significant. In the Schechter et al. (2010) intervention study, the intervention appeared to be particularly effective for East Asian learners with initially low math interest. This group showed more interest in the new math technique relative to Westerners.

Taken together, the evidence from interventions targeting utility goals and values is promising. Five of six studies show effects on final course grade (effect size range = .06-.55 SD) and two of two studies show effects on college GPA (effect size range = .33-.52 SD). Although many of these impacts are limited to sample subgroups such as first-generation college students, underrepresented minority groups, and women, these underrepresented student groups are most at risk of leaving STEM or leaving college altogether. The ability to recognize the utility or value in a given college activity (e.g., classwork) can potentially motivate a student to work harder on that activity even if the activity itself is not especially interesting, and even if the student is not dispositionally inclined to work hard in the first place (that is, the student is not especially high in trait conscientiousness). However, further research is needed to explore this possibility and demonstrate more conclusively that the development of utility goals and values increases college persistence and success.

Intrinsic Goals and Interest

Another way in which a goal may be infused with value is if pursuing it is viewed to be intrinsically enjoyable or worthwhile and is valued for its own sake. The pursuit of intrinsically motivated academic goals involves the development of content knowledge (deeper learning, or sense making) and the development of interest in the discipline (see Renninger and Hidi, 2016).

A substantial body of research on the concept of intrinsic motivation suggests that pursuing goals perceived as rewarding or valuable in and of themselves is associated with many positive outcomes in life, such as higher levels of well-being and greater success (Deci and Ryan, 1991). Findings from neuroscience have established the association between interest, or seeking behaviors, and the reward circuitry of the brain (e.g., Gruber et al., 2014; Panksepp, 1998). This research provides evidence that once interest is developed, it becomes its own reward. Accordingly, those who do not have an interest are likely to need support to find learning rewarding, and once they receive such support (e.g., through interactions aimed at identifying meaning in content to be learned), their interest can be sustained and developed. Once interest begins to develop, moreover, it may promote the development of academic self-efficacy (e.g., Bong et al., 2015) and self-regulation (e.g., Sansone et al., 2015).

The development of interest (and its accompanying value) can be supported at any age and in any disciplinary context (see reviews in Hidi and Renninger, 2006; Renninger and Hidi, 2016). For the development of a new interest in college—say, mathematics—support from other people, such as mentors and instructors, often is required, as well as the design of engaging text, software, exhibits, and so on. In early phases of interest development, students may need to discover the utility or relevance of the content for their own lives.

Research on college students shows that infusing goals with intrinsic value is associated with more successful pursuit of those goals. Turner and colleagues (2009), for example, showed that both intrinsic motivation and academic self-efficacy predicted academic performance. Some studies have found mastery goals (typically seen as intrinsically rewarding) to be positively correlated with sustained interest in college subjects, as measured by course taking and choice of

major (Harackiewicz et al., 2000, 2002). Harackiewicz and colleagues (2002) also found that early interest and final grade, measured in an introductory psychology course, were positively correlated with the likelihood of majoring in psychology. In an examination of interventions designed to broaden participation of underrepresented minority students in STEM disciplines, a group of investigators found that mastery goals (associated with intrinsic motivation) were correlated with higher cumulative GPAs (Hernandez et al., 2013).

Three intervention studies addressing intrinsic goals and interest met the committee's inclusion criteria. One of these—Hamm et al. (2014), discussed above in the section on growth mindset—was designed to investigate the effects of a traditional attribution retraining intervention on at-risk students who were low in perceived control of their studies and either low or high in preoccupation with failure. The first-year students who participated in the study were classified, based on questionnaire responses, into four different performance orientations: (1) failure-acceptors, who were low on both preoccupation with failure and perceived control; (2) failure-ruminators, who were high on preoccupation with failure and low on perceived control; (3) achievement-oriented students, who were low on preoccupation with failure and high on perceived control; and (4) over-strivers, who were high on both preoccupation with failure and perceived control. Intrinsic motivation was measured as one of three competency outcomes. Students in the treatment condition reported higher motivation relative to their control group peers, but this was the case only for the subset of students characterized as failure-acceptors and failure-ruminators.

An intervention study by Vansteenkiste and colleagues (2004b Study 1) provides possible insights into the process through which intrinsic motivation is developed and influences academic performance. In this study, 200 Belgian college students studying to become preschool teachers were randomly assigned to intrinsic and extrinsic motivation conditions before participating in a target activity. The intervention manipulation was embedded within the activity instructions for each condition. Following the intervention, the intrinsic motivation participants' self-reported superficial processing scores (measuring the degree to which students had engaged with the material in a superficial way) were on average lower than those of the extrinsic motivation participants. Further, the intrinsic motivation participants' deep processing scores (measuring the degree to which students engaged with the material deeply by questioning its underlying meaning and relating it to other concepts) were on average statistically significantly higher than those of the extrinsic motivation participants. This intervention study also measured achievement outcomes. The performance of intrinsically motivated students, measured by a written test of comprehension and group discussion, was significantly higher than that of their extrinsic group peers; additionally, they visited the library more often following treatment to learn more about the activity topic.

In summary, the evidence from intervention studies designed to foster intrinsic motivation and interest is modest. Two of two studies show significant effects on academic outcomes (measured through a written test of comprehension) by framing academic exercises in terms of intrinsic (versus extrinsic) goals (effect size range =.39-1.25 SD).

Further research is needed to examine whether development of these competencies is causally related to college persistence and success. Interest is a powerful motivator (Renninger and Hidi, 2011, 2016), and future interventions could potentially trigger and support the development of intrinsically motivated goals infused with interest in at least two ways. First, certain interventions might lead students to identify connections between new and important academic ideas and opportunities on the one hand and what they already know on the other,

encouraging them to develop their understanding further. Second, these interventions might encourage students to continue to pursue intrinsic goals and to develop and deepen their understanding of already existing interests. Research is needed to test such future interventions, as is more basic research on the processes that underlie the development of intrinsic motivation and interest.

Prosocial Goals and Values

Many parents, educators, and citizens believe that college should encourage students to formulate goals and values that are aimed at promoting the common good. One of the rationales traditionally invoked to justify a liberal arts education is the idea that a well-educated and well-rounded student will be prepared to make a positive contribution to society. In a related vein, recent years have witnessed a proliferation of civic engagement initiatives, service-learning opportunities, and other programs designed to stimulate and organize students' prosocial motivations (see Chapter 5 for further discussion). These efforts provide new ways for students to become engaged in rich and meaningful college life. Cultivating prosocial goals and values also may promote college success, as indexed by such traditional metrics as grades and completion to degree.

In this light, prosocial goals and values are focused explicitly on promoting the well-being or development of other people, or of domains that transcend the self. Included under this broad rubric are student goals and values aimed, for example, at helping others, benefiting society, promoting a prosocial religious or political agenda, and contributing something positive to the next generation. A closely related construct may be what Wolniak and colleagues (2012) call "socially responsible leadership"—the extent to which "students possess values such as equity, social justice, self-knowledge, citizenship, and commitment towards social change" (p. 814).

Research in life-span developmental psychology suggests that prosocial goals tend to increase in salience from young adulthood through late midlife (Freund and Riediger, 2006). Evidence also suggests that valuing prosocial outcomes tends to be associated with positive psychological health and well-being in adulthood. Many college teachers and counselors contend that nurturing prosocial goals and values, including values linked to citizenship and cultural awareness, promotes a civically engaged and enlightened identity among college students.

To date, research suggesting that the cultivation of prosocial goals and values predicts college success is sparse. Still, Wolniak and colleagues (2012), analyzing data from the Wabash National Study of Liberal Arts Education, found that socially responsible leadership predicted persistence in college. Likewise, Yeager and colleagues (2014) found that holding a self-transcendent purpose for learning was associated with higher levels of academic self-regulation among adolescents and college students.

The committee identified two intervention studies focused on prosocial goals and values that met its inclusion criteria. Both are reported by Yeager and colleagues (2014). In the authors' Study 3, students assigned to a transcendent purpose condition and a control group completed an online exam review activity, which consisted of answering 100 multiple-choice questions. Students in the transcendent purpose condition reviewed self-transcendent purpose materials prior to the review. Online software tracked the amount of time students spent on each problem; this measure of time spent served as the main dependent variable indicating students' intention to learn from the material. The results show that a transcendent purpose intervention can promote

deeper learning behavior: students in the transcendent purpose condition spent twice as much time on each question compared with the control group, a difference that represents a moderate effect size.

Yeager and colleagues' (2014) Study 4 included 429 students enrolled in a college-level psychology course; the intervention procedure was identical to that of Study 3. The goal of this study was threefold. First, it tested whether a purpose condition fostered deeper learning behavior in the face of desirable alternatives. As in Study 3, students reviewed a series of math problems; however, they were told they could quit any time and entertain themselves online. Second, the study examined whether a purpose intervention benefited students during the most uninteresting portion of a task. Indeed, students in the purpose condition completed 26 percent fewer problems during the most tedious portion of the diligence task (i.e., a set of math problems) than during the less tedious portion, whereas this differential was 44 percent for control participants. Lastly, the study examined whether self-oriented beliefs alone promoted academic self-regulation. To address this question, the study included another intervention group, one that emphasized self-oriented benefits (i.e., intrinsic motives for learning) as opposed to purpose beliefs. Self-regulation on the diligence task did not improve among the self-oriented intervention group compared with the control group as measured by the number of math problems solved.

In summary, these two interventions provide only modest evidence that the development of prosocial/transcendent goals and values is related to college persistence and success. Although both studies show significant, positive effects on deeper learning behavior, neither reports achievement impacts. Fostering such goals during the college years could potentially produce long-term benefits in the realms of personal happiness, fulfillment, and citizenship, as well as promote meaningful interpersonal relationships and group affiliations. However, whether developing these kinds of goals and values directly promotes success in college, as measured by grades and graduation rates, remains an open question.

Research is needed to examine in much greater detail how students' prosocial goals and values develop in college, how they relate to intrinsic interest and more utilitarian goals, and the extent to which pursuing a prosocial or self-transcendent agenda in college may influence college persistence and academic performance. Moreover, the relationships and interactions that may be revealed in such research need to be understood in terms of students' unique profiles of gender, race/ethnicity, class, and other characteristics. First-generation and underrepresented minority students, for example, may find strong motivation in a desire to give back to their communities of origin. If colleges could capitalize on this desire by explicitly connecting coursework and college activities to these kinds of prosocial goals and values, underrepresented student groups might achieve higher levels of college success.

IDENTITY: *WHO AM I? WHOM DO I WANT TO BECOME?*

Many theories in developmental psychology contend that the college years are prime time for the development of identity (Erikson, 1963; McAdams, 2015). In its broadest sense, identity refers to a person's overall conception of self and the self's relationship with the world. The central questions in identity development are *Who am I*? and *Whom do I want to become?* Identity questions thus are directed both backward and forward in time. The person aims to link the remembered past with an anticipated future, forging links among conceptions of past, present, and future selves. Identities typically refer to "the kind of person I see myself as being or

becoming" or "the kind of life I hope to live." Among other things, identity functions to provide a person's life with a sense of unity, purpose, and coherence.

Do certain forms of identity promote success in college? Over the span of their college experience, students may begin to develop images, pictures, imagined trajectories, and narratives about the kind of person they hope to become in the future. Many social psychologists argue that these imagined future selves have motivational power (Markus and Nurius, 1986; Smith and Oyserman, 2015). Future selves may help organize and give meaning to a person's goals, strivings, and values. They may impel the person to work harder to succeed in the field of study to which the future self may be connected. If a student articulates a clear image or story in her mind about becoming an engineer or a scientist, for example, she may find greater meaning and interest in her engineering or science classes and may work harder and perform at a higher level in those classes. She also may work harder to do well in college more generally and to complete it successfully so she can take the next step toward actualizing her imagined future self. Indeed, a student's identification as a scientist, or with the broader scientific community, has been cited as a significant predictor of persistence in STEM fields (Graham et al., 2014).

Positive Future Self

Developing a positive future self may increase the utility value of certain college activities and imbue those activities with clearer purpose and meaning. Many educators believe that linking academic work to a student's developing image or story of whom he or she hopes to become in the future presents a powerful opportunity to promote psychosocial development and college success (Hanson, 2014; Markus and Nurius, 1986). With this in mind, Landau and colleagues (2014) designed and tested several interventions intended to encourage students to think of their academic life as a "journey." They argue that the identity metaphor of life as a journey enabled the students to think more clearly and productively about how their current activities might lead to the realization of positive future selves.

The committee identified five intervention studies meeting its criteria that address the competency of positive future self. Each of these studies reports outcomes related to the target competency; only two, however, report academic outcomes. Three of the five interventions are found in the Landau and colleagues (2014) article described above. Their intervention addressed identity exploration by helping students make connections between their current identity and their possible academic identity. In all three studies, students in the treatment condition were primed to frame their academic career as a journey toward becoming an academically accomplished college graduate. Students in the comparison groups were primed to frame their academic careers differently (for example, to see each college year as a separate container) and/or to imagine different future identities. In Study 3, following the treatment, students in the journey-framed group indicated that they planned to dedicate 23 percent more time to coursework relative to the container-framed group. Further, in this study, the journey-framed group earned significantly higher scores than those in the container-framed group on a final exam administered 1 week after the intervention.

Landau and colleagues (2014) also report competency outcomes (in Studies 1 and 2). In Study 1, journey-framed participants reported being more academically engaged and having more academic intentions relative to students in the comparison groups. In Study 2, journey-framed participants showed significantly higher academic engagement than students in the

comparison groups, as measured by effort on an academic task (solvable math problems). However, the authors do not report the specific sizes of these effects.

In another intervention study meeting the committee's inclusion criteria, Harrison and colleagues (2006) showed how academic conditions can impede identity development. In this study, students in both treatment and control conditions were assigned to complete an academic test assessing math and verbal ability. Prior to the exam, treatment participants were exposed to stereotype threat; in particular, the exam instructions stated that the exam would be used in research to understand why lower-income students generally perform worse than their higher-income counterparts. Lower-income students in the treatment condition performed significantly worse than the control group on both the math and verbal portions of the exam. Relative to the control group, lower-income students in the treatment condition also indicated that they identified less with English and math subjects.

Taken together, the evidence from intervention studies on the competency of positive future self is modest. A single study showed a significant increase in student engagement with academic identity and subsequent significant improvement in exam performance (0.56 SD). Although the intervention results suggest that promoting positive future self may be able to improve students' academic engagement and at least their short-run test performance, long-term effects on college persistence are not known.

Identity achievement may represent both a *predictor* of success in college and a valued *outcome* of undergraduate education. The process of developing a positive identity may draw on a range of competencies enumerated above, from deeply ingrained dispositional traits such as conscientiousness, to particular beliefs about college, to the specific motivational agenda—the goals, values, and interests—that a student formulates and begins to pursue over the course of college. Further research is needed to explore this potential.

Identity also is shaped by the exigencies of the college environment and by broad cultural forces that establish life expectations that may be contingent on gender, race, and class categories. In a sense, identity development is the culmination of a student's educational experiences in college. In the best-case scenarios, the student has learned something important about who he or she was, is, and may someday become, and has embarked on a directed journey to become the positive future self now envisioned. Further research is needed to explore how to support the development of a positive identity during the college years.

CONCLUSIONS AND RECOMMENDATIONS

This chapter has reviewed the available research, using a developmental framework, to articulate the competencies and that may be implicated in students' college success.

Research Needs

Over the course of an extensive search for research evidence of possible relationships between intra- and interpersonal competencies and persistence and success in college, the committee found that the research base is limited, especially in the area of interpersonal competencies and college success. There were also particular gaps related to other important topics.

Conclusion: Only limited research has been conducted to date on the potential relationships between various intra- and interpersonal competencies and students' college success. There are major gaps in the research evidence:

- ***Little research is available on the possible relationship between interpersonal competencies and students' college success.***
- ***The available research has been conducted almost entirely in 4-year institutions; very little experimental evidence is available on the possible relationship between intra- and interpersonal competencies and students' success in community colleges.***
- ***There is a paucity of evidence on the possible relationships between intra- and interpersonal competencies and the success of students intending to major in STEM fields.***

Recommendation 1: Federal agencies and foundations should invest in research exploring the possible relationships between various intra- and interpersonal competencies and students' college success. To address gaps in the research base, these investments should include support for research examining

- **how interpersonal competencies may be related to student success in 4-year colleges;**
- **how intra- and interpersonal competencies may be related to student success in community colleges; and**
- **how intra- and interpersonal competencies may be related to students' success in 2- and 4-year STEM programs and majors.**

Promising Competencies

Based on its review of the limited available research, the committee identified promising competencies that appear to be related to college success. Correlational research suggests that, among the competencies reviewed by the committee, the most robust predictor of college success is dispositional conscientiousness—the tendency to be self-controlled, responsible to others, hardworking, persevering, rule-abiding, and achievement oriented. Conscientiousness is closely related to other constructs such as self-control, self-discipline, persistence, and grit. Indeed, scores on measures of conscientiousness are nearly as predictive of college success as are measures of general cognitive ability. Yet evidence supports conscientiousness as a deeply ingrained dispositional trait that is difficult to change, at least in the short term. Nevertheless, a few interventions have targeted specific behaviors associated with conscientiousness, to date yielding only a little evidence of effects on college success.

Conclusion: Beyond cognitive factors, correlational research has shown that individual differences in intrapersonal competencies predict college success and completion. These competencies include the broad personality trait of conscientiousness. Although an individual's relative standing on conscientiousness tends to be highly stable over time, some interventions have

successfully targeted task management and other specific manifestations of this trait.

Moving beyond dispositions and traits, the committee reviewed the available research on the relationships between other intra- and interpersonal competencies and college success. Through this process, the committee identified eight intrapersonal competencies. (Reflecting a lack of research evidence, this list includes no purely interpersonal competencies.) These eight competencies have been studied using a range of methods, and the committee identified them based on correlational and experimental research. However, the committee judged the strength of the evidence related to each competency based exclusively on research that has developed and tested them through interventions using random assignment.

Conclusion: The limited intervention studies conducted to date have generated promising evidence that the competencies of sense of belonging, growth mindset, and utility goals and values are related to college success and are malleable in response to interventions. Available intervention studies provide more modest evidence that five other competencies are similarly related to college success and malleable, yielding a total of eight identified competencies:

- ***Behaviors related to conscientiousness—behaviors related to self-control, hard work, persistence, and achievement orientation;***
- ***Sense of belonging—a student's sense that he or she belongs at a college, fits in well, and is socially integrated;***
- ***Academic self-efficacy—a student's belief that he or she can succeed in academic tasks;***
- ***Growth mindset—a student's belief that his or her own intelligence is not a fixed entity, but a malleable quality that can grow and improve;***
- ***Utility goals and values—personal goals and values that a student perceives to be directly linked to the achievement of a future, desired end;***
- ***Intrinsic goals and interest—personal goals that a student experiences as rewarding in and of themselves, linked to strong interest;***
- ***Prosocial goals and values—the desire to promote the well-being or development of other people or of domains that transcend the self; and***
- ***Positive future self—a positive image or personal narrative constructed by a student to represent what kind of person he or she will be in the future.***

Interventions that often required very little time and money to implement have helped students develop these eight competencies. Some of these interventions have been particularly effective for underrepresented student groups that are most at risk for academic failure.

Conclusion: Notably, evidence shows that low-cost interventions aimed at developing sense of belonging, growth mindset, and utility goals and values have sometimes generated the largest benefits for underrepresented student groups that are most at risk for academic failure. Although encouraging, this

evidence is limited and recent, and further research is needed to replicate and extend it.

Recommendation 2: Federal agencies and foundations should invest in intervention research and research employing a range of other methods to understand better the competencies identified in this report, their relationship to college success, and the mechanisms through which they operate to improve college success. Research focused on supporting the college success of underrepresented student groups should be a priority.

Recommendation 3: Colleges and universities should support the intervention research proposed in Recommendation 2 by facilitating the implementation and evaluation of random-assignment interventions, thereby gaining valuable information about their students and building the knowledge base on effective interventions needed to increase student retention and success.

Success in boosting the eight identified competencies will also require more research on their development and their role in undergraduate success generally, especially in the STEM disciplines. The findings in this regard to date are promising, but very limited. Research is needed on the interrelationships of these competencies with each other and with student success, both generally and for different student groups. For example, further research on sense of belonging could potentially find that it plays a fundamental role in supporting the development of academic self-efficacy, utility goals and values, and other competencies. Developing a robust body of research on the identified competencies will in turn require that funding agencies, journals, professional societies, and promotion and tenure committees all value and even prioritize replication studies. Studies should be conducted with adequate statistical power, and if they find a lack of statistical significance (e.g., null results), those results will also be important contributors to the base of scientific knowledge about these competencies. This commitment to rigorous research will allow institutions to know when an investment in developing these competencies is warranted.

Sensitivity to Context and Subgroup Effects

The committee's review of the research indicates that issues of race, ethnicity, gender, social class, and culture need to be carefully considered when educators, administrators, researchers, and policy makers think about competencies and their contribution to college success. Certain competencies, and the problems they might help address or solve, may be more salient or useful for certain groups of students than others. Underrepresented minority students, for example, may bring to college such competencies as a strong racial or cultural identity that may help them navigate potentially unwelcoming academic or social environments. More research also is needed to examine how particular educational and cultural contexts currently influence the development of motivation, intrinsic goals and interest, and other intra- and interpersonal competencies among underrepresented groups (minority students, first-generation college students, students from low-income families, and women in certain STEM disciplines). Such research would inform efforts to improve educational and cultural contexts to ensure that

they do not hinder, but rather support, the positive development of promising competencies among these groups. The need for such research is especially salient in the STEM fields, in support of the aim of enhancing diversity and promoting opportunities in these fields.

Conclusion: Certain competencies develop and function differently for different groups and within different cultural and educational contexts. For example, although a strong sense of belonging in college is an important factor for success among underrepresented student groups, members of these groups may find it difficult to develop this competency if they experience campus environments that are discriminatory, negative, or unwelcoming.

Recommendation 4: To help reduce disparities in college success among student groups, institutions of higher education should evaluate and improve their social and learning environments to support the development of the eight identified competencies, especially among underrepresented student groups.

3

Assessment Methods for College Competencies

The previous chapter identifies eight competencies that the available research suggests are related to undergraduate persistence and success: (1) behaviors related to conscientiousness; (2) sense of belonging; (3) academic self-efficacy; (4) growth mindset; (5) utility goals and values; (6) intrinsic goals and interest; (7) prosocial goals and values; and (8) positive future self. This chapter turns to the measurement of students' status relative to these competencies, a necessary precursor to research and practice that can enable both deeper understanding of the relationship between these competencies and student success and evidence-based decisions and programmatic actions that capitalize on these competencies to promote students' persistence and success. But measurement is a complex science. Just because an assessment claims to or is intended to measure a given competency does not ensure that it does so fairly and with appropriate levels of precision. The quality of assessment matters.

This chapter focuses on the nature and quality of existing assessments of the identified competencies; summarizes well-established principles of assessment development and validation that higher education stakeholders should keep in mind as they develop, select, and/or use assessment to support student success; and considers available options and future directions for improving assessment practices. Importantly, the chapter draws on long-standing tenets that have been honed largely in the context of measuring cognitive competencies. Although the formats used in typical assessments of this study's eight competencies (e.g., self-report surveys) may differ from the formats used to assess students' cognitive skills, the underlying principles of measurement remain the same. Further, although this chapter focuses on the measurement of individuals' competency, the committee fully recognizes that the identified competencies influence and are influenced by the college environment and other cultural and contextual variables. The climate on a college campus, for example, can greatly affect a student's sense of belonging, and it may be the college context that needs to be measured and improved rather than the student's competency.

The chapter is divided into four major sections. The first reports on the nature of the methods currently being used to assess intra- and interpersonal competencies, focusing especially on methods used to assess the eight identified competencies. The second section presents key principles of assessment quality, introducing the concepts of validity, reliability, and fairness. The third section applies these principles to evaluate the quality of current assessments of the eight competencies. The fourth section lays out a pathway to better measurement through a professional approach to developing and validating assessments for serious use and through current and future innovations in measurement and analysis. The chapter ends with conclusions and recommendations.

CURRENT ASSESSMENT METHODS

The committee examined the status of current assessments based on its review of the literature on intra- and interpersonal competencies, focusing primarily on the eight identified competencies. In its examination, the committee drew predominantly on three sources of evidence: (1) analysis of the assessment instruments used in the intervention studies that the committee used to judge the strength of evidence supporting each competency, (2) general review of the results of a literature search on assessments of the identified competencies (see Appendix A), and (3) close analysis of a small sample of established assessment instruments targeting one or more of the eight competencies. To provide a sense of the current landscape of these assessments, the following sections describe the types of assessment formats that could potentially be used to measure intra- and interpersonal competencies and note the extent to which each format is used in existing assessment instruments. The final subsection provides a summary of the formats used in current assessments of the eight competencies.

Self-Ratings (Selected Response)

A typical self-rating presents a trait term, or a statement, and requires the respondent to indicate the extent to which he or she agrees with the statement on a Likert scale (e.g., strongly agree, agree, neutral, disagree, strongly disagree) or to indicate the frequency of engaging in the thought or behavior described in the statement (e.g., never or rarely, sometimes, often, always, or almost always). For example, the Programme for International Student Assessment (PISA) 2000 measured "sense of belonging" using a four-point agreement scale for the statements, "School is a place where…I feel like an outsider (or left out of things)," "…I feel awkward or out of place," "…I feel lonely," "…I feel like I belong," and "…other students seem to like me" (Willms, 2003, p. 64). Note that of these five statements, the latter two would be positively keyed and the others negatively keyed.

The most common approach for scoring a set of self-ratings is to average the scale values (e.g., 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree) across items. This approach is commonly used because (1) the average is interpretable on the same 1-5 scale, and (2) low scores are then the result of an actual low standing on the construct, not missing responses. However, appropriate interpretation of the results of this approach rests on three assumptions: no item effects (i.e., every item has the same mean response); no item x person effects (i.e., a person who responds more highly than others on one item will respond more highly than others on all other items); and an equal interval scale (i.e., the difference in the underlying trait is the same at every score interval). All of these assumptions can be tested, and violations can be addressed with more sophisticated scoring approaches (e.g., item response theory-based scoring methods; latent class analyses).

Self-ratings are by far the most common means used for measuring intra- and interpersonal competencies in general (Duckworth and Yeager, 2015; Robins et al., 2007) and the eight identified competencies in particular. Among the 87 assessment instruments used across the intervention studies discussed in Chapter 2, 74 (85 percent) used self-ratings with selected response formats (mostly Likert scales; see Appendix C). In addition, the available meta-analyses of the correlations between these competencies and college outcomes reviewed in Chapter 2 are based almost exclusively on data from assessments using self-rating formats

(Noftle and Robins, 2007; Poropat, 2009; Richardson et al., 2012). Finally, the committee's search for and review of measures of the eight competencies similarly revealed reliance on self-report scales. Among these, both ACT and the Educational Testing Service (ETS) have developed self-report Likert surveys of admitted students' intra- and interpersonal competencies as an aid to identifying students' likelihood of success, providing students with direct feedback and support services, and placing them in appropriate course levels (e.g., developmental versus college-level mathematics; see further discussion in Chapter 4). Both tests measure some of the eight competencies, along with others. For example, the ACT Engage College subscales of Academic Discipline, General Determination, Goal Striving, and Study Skills generally align with behaviors related to conscientiousness, while the subscales of Academic Self-Confidence and Social Connection relate to academic self-efficacy and sense of belonging, respectively. Similarly, ETS's SuccessNavigator Tools and Strategies subscale includes items on organization, a behavior related to conscientiousness, while the Self-Management subscale includes items used to measure academic self-efficacy.

Despite their prevalence in assessments of the eight competencies, self-rating scales have several well-known limitations, which are discussed further later in this chapter. The first is social desirability: respondents may distort their responses to avoid embarrassment and project a favorable image (Zerbe and Paulhus, 1987). In this regard, studies of biodata have suggested that asking respondents to provide a rationale for their ratings, even when such content is not analyzed, may result in more accurate ratings (Schmitt and Kunce, 2002; Schmitt et al., 2003).

A second issue is the interpretation of item rating scales. Individual respondents may vary on how they interpret or use the rating scale, and as a result, ratings will not have the same meaning across individuals. For example, response style refers to the systematic tendency to respond a particular way regardless of the construct being measured. Common response styles are acquiescence (or "yea saying," the tendency to respond positively) and extreme response style (the tendency to choose the extremes of the response scale) (Stricker, 1963). These tendencies can distort relationships, and research continues to focus on determining whether corrections for them might improve the quality of resulting data (Falk and Cai, 2016; He and van de Vijver, 2015).

Biographical and Personal Essays and Statements

Personal statements providing biographical information (biodata) and admissions essays that commonly have been required in college admissions and scholarship award contexts for some time (Willingham and Breland, 1982; Willingham et al., 1987) provide information about intrapersonal competencies such as those identified in Chapter 2. A study conducted by ETS found that college administrators and faculty members reported using personal statements to draw inferences about students' intra- and interpersonal competencies (Kyllonen, 2008; Walpole et al., 2002), including some of those identified in Chapter 2 (e.g., perseverance, a behavior related to conscientiousness). Such use of personal statements persists despite the shortage of evidence for their predictive validity, especially after controlling for grades and test scores (Murphy et al., 2009), the potential validity threat posed by the possibility that the essay was prepared by someone other than the applicant (Willingham and Breland, 1982), and the apparent lack of standard treatments for scoring or evaluating these statements.

A rather different use of essays for assessment is found in one of the intervention studies cited in Chapter 2. Harackiewicz and colleagues' (2015) utility value intervention required

students enrolled in an introductory biology course to write short (one- to two-page) essays about the personal value of course material, such as how an animal physiology course might inform their personal workout and exercise program. Three such writing assignments, staggered over the semester, were integrated into the course. Research assistants coded the essays on a 0-4 scale based on how specific and personal the utility value connection was to the individual, providing a measure of implementation fidelity and strength of utility value, with 0 indicating no utility and 4 indicating a strong connection to the individual, reflecting deep appreciation of the material. In another example, involving less extended responses, Walton and colleagues (2012) asked students participating in a sense of belonging intervention to generate two reasons for their success and/or failure in math. Two raters independently coded whether each reason indicated (1) social-relational factors, or sense of belonging with others in math; (2) nonrelational social factors (e.g., interest in math relative to other students); (3) academic self-efficacy in math; or (4) unspecified factors. The authors calculated a valence score for each category by subtracting the number of negative from the number of positive reasons generated. This valence score and scores from a separate self-report measure of the perceived warmth and fairness of the math department were standardized and averaged to form a composite measure of social connectedness to math.

Others' Ratings, Including Letters of Recommendation

Ratings by others generally take the same form as self-ratings and personal statements, except they are provided by knowledgeable others. That is, others are asked to respond to similar selected-response surveys, often involving Likert scales, for purposes of reporting on an individual's traits, behaviors, attitudes, skills, or dispositions, or to produce letters of recommendation that provide biodata and comment on the individual's competencies or qualifications and personal history. Faculty members in fact report using information from letters of recommendation to draw inferences about a range of competencies of student applicants (Walpole et al., 2002). Similarly, although not apparent in the studies reviewed for this report, Likert-type scales are common in others' ratings of individuals.

Current assessments of the eight competencies identified in Chapter 2 do not rely on others' ratings, although ETS formerly offered the Personal Potential Index (PPI), which used this approach to measure several competencies, including planning and organization—behaviors related to conscientiousness. The PPI (Kyllonen, 2008) was a standardized letter of recommendation used for collecting information on graduate school applicants' cognitive, intrapersonal, and interpersonal competencies as determined from recommenders' ratings on 24 statements in six dimensions (knowledge and creativity, communication skills, teamwork, resilience, planning and organization, and ethics and integrity), along with an overall evaluation. Preliminary research showed that the instrument predicted graduate school cumulative grade point average (GPA), above and beyond that based on undergraduate GPA and Graduate Record Examination (GRE) scores (Klieger et al., 2015), and showed smaller subgroup differences than these other measures.

In general, ratings by others have been found to be substantially more predictive of academic achievement and job performance outcomes relative to self-ratings (Connelly and Ones, 2010), and others' ratings add to the predictions derived from self-ratings (whereas the reverse is not true). A meta-analysis of studies of letters of recommendation (Kuncel et al., 2014), for example, found that they predicted various higher education outcomes, such as GPA

and degree attainment, and for degree attainment, some of this prediction provided unique information beyond that derived from other quantitative aspects of students' academic records. Letters typically are not standardized, however, and a key question is whether standardized letters might be fairer for applicants; be more consistent and less fatiguing for recommenders; and provide better prediction of academic outcomes, including intra- and interpersonal competencies.

Interviews

Interviews are common in higher education admissions, particularly for graduate school, medical school, and other professional schools. They likely are often intended to determine some of the eight identified intrapersonal competencies, such as growth mindset, even if implicitly and imperfectly, and they suffer from the same limitations that characterize other self-report measures.

Interviews vary widely in their structure, content, interpretation or scoring, and use. They fall into three broad types: standardized, behavioral, and informal. Standardized interviews ask each respondent the same questions, such as "Tell me about why you want to pursue x" or "Tell me about your qualifications to pursue y." Behavioral interviews ask respondents such questions as "Tell me about a time when you had to give up a planned event to meet a deadline." And informal interviews give the interviewer free reign to pursue various questions, such as those conditional on prior responses. Interviews also vary as to whether they are administered face-to-face or online (Kell et al., in press). Although there has been little work on interviews in higher education, a meta-analysis of studies of employment interviews (McDaniel et al., 1994) found that standardized interviews generally had higher correlations with employment outcomes relative to informal interviews.

Performance Assessments and Behavioral Measures

Performance assessments encompass a wide variety of methods, but all involve an individual's creating or constructing a response as opposed to answering a multiple-choice question, responding to an interview question, or being rated by a peer or teacher. Students are asked to engage in a given task, and their behavioral response to the task is then tracked and/or evaluated.

Performance assessment has become common as a way to assess students' deeper learning and ability to apply their knowledge to solve real-world problems—for example, devising solutions for a given social problem, creating a business plan for a new product, mounting an advertising campaign to convince the members of the public to change some aspect of their behavior, or engineering a new approach to developing solar cars. Team projects can be a context for assessing students' interpersonal competencies (see Chapter 5), and students' involvement in complex individual performance tasks can serve as context for assessing their intrapersonal competencies. In the K-12 arena, for example, it is becoming increasingly common to ask students to assess their effort and efficacy in completing their work.

Similarly, students' responses to challenging problems provided indicators of behaviors related to conscientiousness in a number of the intervention studies discussed in Chapter 2. For example, Walton and colleagues (2015) monitored the time students spent on an insoluble math puzzle as a measure of motivation and self-regulation, while Yeager and colleagues (2014)

measured time spent on completing math problems versus consuming online media to assess these same conscientiousness-related behaviors.

While performance assessment requires that students respond to a given task, behavioral measures similarly monitor student responses or behaviors but need not be tied to a specific project or task. Instead, the assessment monitors specific behavioral indicators in given contexts and/or over specific periods of time. For example, Vansteenkiste and colleagues (2004a) tracked the number of times students visited the library and/or the recycling center to learn more about recycling as a measure of their intrinsic goals and interest.

Technology-based projects can serve as a means of easily and unobtrusively monitoring students' behaviors, including self-regulation and other behaviors related to conscientiousness, as in the examples cited above. At the same time, however, performance assessment can be time-consuming and costly in terms of the time required for both task completion and scoring, which often involves human scoring. Thus performance assessment currently can be difficult to scale up for larger studies. Further, measures often are based on responses to a single task, which raises questions about the generalizability of the scores, since research in K-12 settings has demonstrated substantial variation in performance across different tasks and topics. Students' behaviors related to conscientiousness (e.g., self-regulation) when approaching different problem sets, for example, may well depend on the topics of the problems.

Situational Judgment Tests (SJTs)

SJTs differ from self- and others' ratings in that they provide a hypothetical situation and ask the respondent to select the most appropriate response to that situation from a set of possibilities or to rate the appropriateness of each possibility. To date, assessment instruments designed to measure the eight competencies identified in Chapter 2 have not used this format. However, in a study conducted by Oswald and colleagues (2004), college students were administered an SJT containing hypothetical situations explicitly designed to measure 12 competencies, including perseverance, a behavior related to conscientiousness (the others were knowledge, learning, artistic, multicultural, leadership, interpersonal, citizenship, health, career, adaptability, and ethics). The following is an example used to measure the competency of perseverance:

> You realize about the fourth week of the term that you have too much coursework and other activities to get them all done—at least within the amount of time you are currently doing homework. What would you do?
>
> a. Drop the non-academic activities.
> b. Use your time management skills to figure out a new study plan, putting the most important coursework at the top of the list.
> c. Analyze how much time you are spending on the homework and consider getting help if the work seems to be taking longer than it should.
> d. Evaluate whether you can reduce the number of credits you are taking, or reduce the number of activities you are involved in.
> e. Put homework and schoolwork first—it's why you're at school.

SJTs are most commonly administered in paper-pencil format, with written situations and responses, but video-based versions also have been developed. Meta-analyses suggest that SJTs provide information not available from self-report personality measures or from cognitive ability tests (McDaniel et al., 2001), but are not evident in any of the studies reviewed for this report.

Summary of Current Assessment Methods in Higher Education

The committee's review of assessments used in higher education revealed evidence that institutions use a number of different types of assessment formats. Biographical and personal essays and statements; others' ratings, including letters of recommendation; and interviews are common in college admissions. These assessment methods tend to address implicitly some of the eight identified competencies, including behaviors related to conscientiousness, intrinsic goals and interest, growth mindset, and academic self-efficacy. However, these methods are not evident in the research reviewed in Chapter 2. Rather, self-report surveys, particularly those using Likert-type scales, are ubiquitous in the correlational and intervention studies reviewed. A limited number of examples of performance and behavioral measures were found in research or evaluation studies.

The committee's analysis of the assessment instruments used in the intervention studies reviewed (which is discussed further below) revealed these same patterns. The committee analyzed all assessments of the eight competencies found in the 61 intervention studies meeting its criteria for inclusion in the literature search conducted for this study (see Box 2-1 in Chapter 2). Overall, of 87 total instruments, 74 (85 percent) used self-report scales, and of these, most were Likert scales. Nevertheless, at least one assessment format other than self-report scales was used in the intervention studies that assessed five of the eight competencies (see Appendix C). These included six performance and behavioral measures of behaviors related to conscientiousness, one behavioral measure of intrinsic goals and interest, and four performance and behavioral measures of sense of belonging. For example, as one measure of female engineering students' sense of belonging in the field, Walton and colleagues (2015) used a behavioral measure, asking students to list the initials, gender, and major of their five closest friends. The authors used the number of male friends in engineering as one indicator of sense of belonging in the field.

Further, it is noteworthy that the great majority of assessments found in the intervention research were investigator-developed, although the development process often relied on previously published assessment instruments. For example, in a study of growth mindset, Brady and colleagues (2016) evaluated a values affirmation intervention among a sample of 183 Latino and white students. To assess the outcomes, the authors created a new assessment of adaptive adequacy by combining three preexisting self-report measures that loaded on a single factor (α = .86): (1) self-integrity, measured with seven items rated on a six-point Likert scale (adapted from Sherman et al., 2009; α = .87); (2) self-esteem, measured with the ten-item Rosenberg self-esteem scale and rated on a six-point Likert scale (Rosenberg, 1965; α = .93); and (3) hope, measured with a eight-item adult hope scale (Snyder et al., 1991; α = .82). In another example, Hausmann and colleagues (2009) used a three-item sense of belonging subscale created earlier by Bollen and Hoyle (1990), with responses on a five-point Likert scale.

In a few cases, different investigators used all or parts of the same, previously published instrument to assess the same competency. Fitch and colleagues (2012) used five scales from the 44-item version of the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich et al.,

1991)[1] to measure behaviors related to conscientiousness, while Haynes and colleagues (2008) used 8 items from the same questionnaire also to measure behaviors related to conscientiousness. And to measure intrinsic goals and interest, Hamm and colleagues (2014) used the Manitoba Motivation and Academic Achievement database, which includes data for two decades of separate cohorts of introductory psychology students (1992-2012). Data for the 2001-2002 cohort include responses to the Intrinsic Motivation Scale (Hall et al., 2007), which in turn was adapted from the MSLQ (Pintrich et al., 1993). In another example, Walton and Cohen (2007, Study 2) used an investigator-developed, 17-item self-report assessment of sense of belonging, which they refer to as a "social fit scale." In a related study of an intervention to increase sense of belonging, Walton and colleagues (2011) used these 17 items, along with others, in a daily online/email survey. These examples are exceptions, however. Overall, the assessments used to measure each of the eight identified competencies differed across different investigators.

KEY PRINCIPLES OF ASSESSMENT QUALITY

As essential background for the evaluation of the quality of assessments currently being used, this section describes three foundational concepts that the measurement community uses to judge assessment quality.

Overview of Key Measures of Assessment Quality

The committee's perspective on assessment quality is shaped largely by the *Standards for Educational and Psychological Testing (Standards)*, sponsored jointly by the American Educational Research Association (AERA), American Psychological Association (APA), and National Council of Measurement in Education (NCME) (2014). The *Standards* define a trinity of principles—validity, reliability/precision, and fairness—as the foundation for sound measurement. *Validity* refers to the nature and weight of the evidence that supports the intended interpretation and use of an assessment—that is, evidence demonstrating the extent to which the assessment actually measures what it is intended to measure, and does so in a manner that serves its intended purpose(s), such as making accurate construct-relevant distinctions among the individuals or groups being assessed. *Reliability* reflects the consistency, precision, and replicability of scores from a measure. *Fairness* in measurement actually is a validity issue: it refers to the validity of score inferences for all individuals and groups in the intended population for the test. A fair measure does not disadvantage some individuals because of characteristics that are unrelated to the construct being measured. Because reliability is a necessary but not sufficient precursor to both validity and fairness, the discussion below starts with it.

Reliability

Reliability refers to the degree to which "test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable and consistent for an individual test taker" (American Educational Research Association et al., 2014, p. 223). Reliability can be estimated statistically given the availability of

[1]Fitch and colleagues (2012) mention that another version of the MSLQ (Pintrich et al., 1993) includes 81 items in 15 subscales.

appropriate data. Such estimates can take many forms, but in general provide a basic level of confidence that test scores may offer useful results.

When one student obtains a higher assessment score than another on academic self-efficacy, for instance, does this really mean that he or she has higher levels of this competency? Students' assessment scores reflect both their status on a given construct and error caused by transient errors of measurement and other conditions (deemed random noise or measurement error variance), which can cause students' observed score or rank ordering on an assessment to differ from their true score or rank ordering on the underlying competency. When reliability is higher, certain types of random errors are reduced, and observed scores on the assessment become more aligned with students' actual standing on the intended construct (assuming that the assessment measures the intended construct).

Reliability is therefore one of several necessary conditions for scores on assessments of any of the eight identified competencies to be appropriate reflections of the actual competencies, and it is a prerequisite for obtaining solid validity evidence for an assessment. High levels of estimated reliability help ensure that decisions based on assessment scores will lead to consistent, justifiable decisions in many contexts, such as making admissions decisions, awarding scholarships, adapting instruction to a student's performance or ability level, conducting outcomes assessment, and evaluating proficiency with respect to standards.

Reliability is estimated in multiple ways, depending on the statistical framework used and on those factors on which the user seeks to generalize the assessment (e.g., over time, across different forms). Three major statistical frameworks are used for estimating the reliability of an assessment: classical test theory (Lord, 1955), generalizability theory (Cronbach et al., 1972), and item response theory (de Ayala, 2009). All three are based on the idea that an assessment score and the item responses contributing to it contain both variance related to the construct of interest and variance reflecting some form (or forms) of measurement error. Regardless of the statistical framework used for estimating the reliability of an assessment, however, it is generally the case, all other things being equal, that the larger the sample of behavior (across observations, items, or raters), the higher is the reliability of the measurement.

Validity

Although validity traditionally referred to the degree to which an assessment measures what it purports to measure (McDonald, 1999), more recent treatments of the concept also highlight the importance of the specific intended use(s) and interpretations(s) of test scores in evaluating validity. This shift is reflected in the *Standards*, which define validity as "the degree to which accumulated evidence and theory support a specific interpretation of test scores for a given use of a test" (American Educational Research Association et al., 2014, p. 225). Validity thus does not refer to a test, but to particular uses and interpretations of it. The *Standards* lay out five sources of evidence for evaluating the validity of an assessment for a particular purpose.

Evidence Based on Test Content

Evidence based on test content refers to the degree to which the content of the assessment aligns with the construct-relevant behaviors, attitudes, activities, etc. that the assessment is designed to measure (Achieve, 2010). Typically, such alignment studies involve expert panels conducting detailed reviews of test content, which provide evidence and judgments about how

well the test represents its intended construct(s), including the adequacy of construct relevance (representation) or evidence of construct underrepresentation. The *Standards* point to construct underrepresentation—defined as " the extent to which a test fails to capture important aspects of the construct domain that the test is designed to measure, resulting in test scores that do not fully represent that construct" (American Educational Research Association et al., 2014, p. 217)—as a major concern in evaluating validity. If a construct is inadequately represented by an assessment, any conclusions drawn from it are limited and may need additional verification. Expert content reviews also typically address the possibility of construct-irrelevant aspects of an assessment that may influence individual scores—for example, issues of construct-item characteristics such as items with unnecessarily complex language or reference to privileged experiences that could bias the assessment of a particular student's standing on the construct (see also the discussion of fairness below).

Content-related evidence is stronger and easier to collect when there is clear definition of and agreement on the construct to be measured. However, in the domain of intrapersonal competencies, there often is a problem with overlapping definitions and fluctuating terminology. Consider that conscientiousness, grit, self-regulation, persistence, pluck, and stick-to-itiveness are closely related if not identical competencies, despite being named differently (Jackson and Roberts, 2015). In fact, assessment instruments used to measure of all of these competencies might contain similar or identical item content. Credé and colleagues (2016) recently found that scores on assessments of grit were highly correlated with scores on assessments of persistence, which in turn is a facet of the broader trait of conscientiousness. Conversely, assessments may use the same name when in fact measuring different competencies. For example, the concept of grit encompasses both perseverance of effort and consistency of interest over time (Duckworth et al., 2007), and different assessments of grit may measure different dimensions or both (e.g., Bowman et al., 2015; Duckworth and Quinn, 2009). Such errors have been termed "jingle-jangle fallacies" (Kelley, 1927; Pedhazur and Schmelkin, 1991; Roeser et al., 2006). That is, one cannot necessarily assume that two assessments with the same competency label actually measure the same construct (jingle fallacy) or that two assessments bearing different competency labels actually reflect different competencies (jangle fallacy).

Evidence Based on Response Processes

Evidence based on response processes ideally demonstrates that an individual's response draws primarily on the specific competency being assessed, such that all other reasons for a response are essentially random and minimized. Interviewing respondents for what they are thinking while responding to a test item or survey question has become standard practice prior to pilot or field testing of an instrument. These interviews go by such names as cognitive interviews, cognitive labs (Ruiz-Primo, 2015), protocol analysis (Ericsson and Simon, 1993), talk-aloud protocols, and verbal reports (Leighton et al., 2009). Typically, cognitive interviews involve in-depth, semistructured interviews with a small number of respondents similar to those that will be targeted in the assessment. They are commonly conducted while an individual is responding to the assessment, or sometimes retrospectively, after the examinee has completed a potential assessment item.

Cognitive interviews also are used to establish that respondents understand the questions in the manner intended. Such interviews can be valuable because the respondent may have a different interpretation of a word or phrase in a question from what is intended by the survey

developers or may interpret a response category differently from how others might interpret it. For example, the respondent may choose the "neither agree nor disagree" category because of a neutral stance or because he or she is unsure. Through cognitive interviews, interpretational differences can be determined in terms of either variation at the student level or thematic differences among relevant subgroups (e.g., gender, race/ethnicity, grade level).

Evidence Based on Internal Structure

Evidence based on internal structure concerns the relationships among items in an assessment. Item factor analyses, cluster analyses, and reliability analyses are commonly conducted to examine the internal structure of items on an assessment. If an assessment is designed to measure one particular construct, such as sense of belonging, items in the assessment should be positively correlated with each other. From a factor analysis perspective, items should empirically support unidimensionality, meaning that only one factor (in the present case, the competency) is responsible for the interitem correlations. Furthermore, because items almost always differ in the strength of their wording, they tend to elicit different levels of response, so that the mean levels of a student's response will differ across items measuring a competency, such as sense of belonging (e.g., "I strongly believe that I belong at my university." versus "I feel comfortable on campus.").

Evidence Based on Relations to Other Variables

Evidence based on relations to other variables involves the empirical association between scores on the measure of the intended competencies with scores on other, validated measures. There are two general sources of evidence in this category: convergent and discriminant evidence and test-criterion relationships.

Convergent and discriminant evidence examines the extent to which scores from an assessment of a given construct are more strongly related to that of another measure of the same or closely related construct (convergent) than to a measure of a different construct (divergent). For example, because sense of belonging and growth mindset are conceptually distinct constructs, it would make sense that different assessments of sense of belonging should correlate more highly with one another than with assessments of growth mindset, and vice versa. Similarly, one can look for convergent and discriminant relationships with other variables in relation to how specific conditions are expected to relate to an individual's standing with regard to a construct. For example, an experiment designed to increase one's sense of belonging (see Chapter 2) can be evaluated for actually doing so, and it also can be evaluated for not showing similar increases in other constructs, such as achievement motivation or emotional stability. If a treatment increases levels of these rival constructs (perhaps even more than sense of belonging itself), the treatment may appropriately be understood as a broader one than initially envisioned with effects other than those intended.

Test-criterion relationships When people say they "validated a test," they often mean to imply that they correlated the scores from the test with a criterion or predicted outcome of the intended construct (e.g., the score on the SAT with first-year GPA or graduation rates). A typical or more traditional validation study might involve first administering an intra- or interpersonal

competency assessment (e.g., of sense of belonging or growth mindset), and then correlating scores from the assessment with such outcome measures as graduation, GPA, absenteeism, or time to degree, or any number of academic outcomes valued by higher education leaders. Evidence for interpretation of test scores as indicating readiness for college success comes from such correlations between the test scores and such indicators of college success.

Evidence of Consequential Validity

Evidence of consequential validity involves the consequences that follow on the use of a test and decisions based on it (see Kane, 2006; Messick, 1995), intended and unintended. Given that tests often signal the skills that institutions or employers find important, changing what is tested is a strategy for encouraging curricular change (Fredericksen, 1984). Likewise, some colleges and universities already are signaling the importance of students' intra- and interpersonal competencies by explicitly assessing and developing them (see Appendix B). If assessment of these competencies in higher education were to grow, it could promote greater emphasis on teaching intra- and interpersonal competencies in high school.

In evaluating the use of an assessment, it also is important to consider whether it may have unintended negative consequences. In considering whether to augment its current admissions criteria with measures of intra- or interpersonal competencies, for example, an institution would want to evaluate whether such a change might have adverse consequences for some subgroups of students. Such evaluation would include not only the aforementioned expert review of assessment items, but also consideration of broader influences on recruitment, admissions, student development, and student success.

When assessments produce results that have serious consequences for individuals, they often are known as high-stakes tests, defined in the *Standards* as "a test used to provide results that have important, direct consequences for individuals, programs, or institutions involved in the testing" (American Educational Research Association et al., 2014, p. 219). By contrast, low-stakes tests carry little or no consequences for those who are assessed. The stakes associated with an assessment can engender behavior that in turn affects validity.

Some Validity Threats: Faking, Cheating, and Motivation

Validity can be compromised much more quickly than it can be established. For assessment in general, security is a major issue, and this is particularly true for high-stakes tests,

An entire industry is devoted to security topics, as are annual conferences (e.g., the 2015 Conference on Test Security), and most of the major test publishers have offices or departments dedicated to the matter. At this point, because assessments are not being used in a high-stakes manner, issues of test security have not yet become prominent. Nevertheless, as results from such assessments begin being used in decision making, affecting individuals and institutions, security measures may soon be needed in securing items, in storing and handling test results, and in exploring the possibilities for security breaches, as is done with cognitive test materials. The *Standards* include extensive discussions and standards pertaining to these issues.

A more immediate validity issue concerns faking responses on intra- and intrapersonal competency assessments, particularly if they include Likert-scale response items, which is the case for 85 percent of the assessments used in the interventions reviewed. Respondents can be tempted to use the extremes of the scale—for example, to endorse all positive statements as

"most like me" and all negative statements as "least like me." Indeed, students and applicants are often motivated to appear diligent, enthusiastic, and appreciative to those for whom they are completing an assessment (faculty members, potential employers, even researchers). High-stakes settings, such as college admissions, also provide incentives for test takers to present themselves in the most desirable light. This is not necessarily a problem: self-presenting effectively may (1) imply higher intra- and intrapersonal competencies in the first place, (2) predict the outcomes that intra- and interpersonal competencies are supposed to predict, and (3) be related to the sort of self-presentation that is required to result in the outcomes of interest (e.g., landing a job by interviewing more successfully than others).

Several approaches can be used to reduce the effects of faking, cheating, or other phenomena that distort and threaten the validity of assessment scores. One approach is to provide pretest warnings against faking (Converse et al., 2008; Dwight and Donovan, 2003; McFarland, 2003); another is to provide such warnings during the assessment itself, tailoring them to students who appear to be responding in a way that implies faking or cheating (see Pace and Borman, 2006, who illustrate other types of warnings); and a third is to require respondents to elaborate and provide follow-up justification for their test responses (Schmitt and Kunce, 2002; Schmitt et al., 2003). Note, however, that an elaboration requirement could impose an unnecessary burden and cognitive load, which could be responsible for lowering test scores in addition to serving the intended purpose of keeping the respondent honest.

As a final note, although a large literature is focused on social desirability (i.e., the tendency to think of and present oneself in a favorable light), measures of social desirability typically have not been proven useful for adjusting scores on an assessment to control for this tendency statistically (Sackett, 2012). Situational judgment tests and performance tests, discussed earlier, and forced-choice formats, discussed below, are intended to reduce the possibility of faking on intra- and interpersonal competency assessments and thereby increase the validity of scores. Regardless of the method employed, however, the higher the stakes associated with an assessment, the greater is the need for strong evidence of validity, as it would be unfair to make important decisions about an individual based on an assessment score that lacked credibility.

For low-stakes tests, defined by the *Standards* as those tests yielding data that have relatively minor consequences for individuals, programs, or institutions involved in the testing, the primary threat to validity is lack of motivation. In cognitive tests, differences in motivation among examinees can lead to substantial differences in scores (Liu et al., 2012).

In summary, the stakes associated with assessment results influence the validity issues that require investigation. The committee agrees with the *Standards* (American Educational Research Association et al., 2014, p. 22), which specify that

> the amount and character of evidence required to support a provisional judgment of validity often vary between areas and also within an area as research on a topic advances. For example, prevailing standards of evidence may vary with the stakes involved in the use or interpretation of test scores. Higher stakes may entail higher standards of evidence.

Fairness

In the assessment context, fairness is a broad concept that generally refers to the degree to which a test measures the same construct and scores have the same meaning for different individuals, or more commonly, for different subgroups of the population for which the test is intended. In other words, fairness is a question of the validity of scores for all intended individuals and subgroups.

Typical subgroup characteristics are sex (male, female), race/ethnicity (e.g., white, Hispanic, black, American Indian), culture, first language (e.g., English versus another language), language/culture group more broadly, socioeconomic status, and immigrant status. One might also consider college-specific subgroup categories, such as first- versus continuing-generation students, domestic versus international students, or on-campus students versus commuters.

The *Standards* lay out an overarching standard for fairness:

> All steps in the testing process, including test design, validation, development, administration and scoring procedures, should be design in such a manger as to minimize construct irrelevant variance and to promote valid score interpretations for the intended uses for all examinees in the intended population. (American Educational Research Association et al., 2014, p.63)

This means that fairness needs to be designed in and investigated at multiple points of the test development, validation, and use cycle, using multiple methods. A first issue is being clear on for whom the test is intended and taking into consideration the diversity of the examinee pool in designing a test. For example, if some in the intended population are not fully English proficient or have disabilities, care must be taken to design an assessment that is accessible for those students—for example, using principles of universal design, avoiding unnecessarily complex language (if language is not related to the construct of interest), and developing accommodations for those who otherwise would be unfairly disadvantaged in demonstrating their competency (universal design and accommodations are discussed further below). During the test development process, test items and forms usually undergo fairness reviews by committees comprising members of the pertinent subgroups as one step in ensuring fairness.

In addition, fairness is evaluated statistically following pilot or field testing and/or based on ongoing large-scale operational testing. These analyses help determine whether there is empirical support that the construct is being measured in the same manner across subgroups at both the item and test levels (Dorans and Holland, 1992). Items and tests that appear to function differently across subgroups are said to exhibit *differential item functioning* and *differential test functioning*, respectively. When tests function similarly across subgroups, they are said to show *measurement equivalence* or *measurement invariance*. Generally speaking, testing for measurement invariance entails administering successive tests to determine whether subgroups are equivalent in terms of (1) the number of factors underlying the assessment, then (2) the extent to which each item reflects the construct of interest, and then (3) whether the underlying subgroup means on the construct are unbiased and can therefore be used to make subgroup comparisons. The literature provides the technical details of these tests in both the item response

theory and confirmatory factor analysis frameworks (Raju et al., 2002; Vandenberg and Lance, 2000).

The possibility of differential validity also is a major concern in evaluating fairness. Evidence of differential validity means that correlations between an assessment and an outcome may differ statistically between subgroups (e.g., males versus females, black versus white students), indicating as well that the meaning of the scores differs across groups (Young and Kobrin, 2001). A distinct but related concept is differential prediction (Cleary, 1966), where regression lines show statistically different slopes between subgroups when an outcome is being predicted, implying different predicted values for each subgroup. To address fairness, the measurement of intra- and interpersonal competencies needs to consider both differential validity and differential prediction when test-criterion relationships are being analyzed and explored.

Fairness in the assessment of intra- and interpersonal competencies also becomes especially important in multicultural contexts because of issues of the comparability of the constructs across cultures and language groups and thus the comparability of measurements of such constructs. Intra- and interpersonal competency assessments will raise unique challenges related to cultural judgments of fairness, given that different cultures may express and value these competencies differently (Sato et al., 2015). The appropriateness of construct definitions for different cultural groups and the comparability of the constructs and of the measures across cultural groups are central concerns in deriving valid score-based inferences from an assessment (Ercikan and Oliveri, 2016).

THE QUALITY OF CURRENT ASSESSMENTS OF THE IDENTIFIED COMPETENCIES

This section considers evidence on the reliability, validity, and fairness of existing assessments of the eight intrapersonal competencies identified in Chapter 2. The discussion is based on an examination of the assessments used in the intervention studies cited in Chapter 2 and on close analysis of two specific, commercially developed assessment instruments. One of the latter instruments was selected because it measures several of the eight competencies and includes subscales that were used in several of the intervention studies reviewed; the other was selected because it measures sense of belonging, a competency for which the committee found promising evidence of a relationship to college success.

Assessment Quality in Intervention Studies

As noted earlier, the committee closely examined the intervention studies cited in Chapter 2 to identify what assessments were used in each study and what evidence of their reliability, validity, and fairness was provided in the study reports (see Appendix C). This review indicated that overall, the investigators paid little attention to the quality of the assessments used. As shown in Table 3-1, of the 46 studies that assess at least one of the eight competencies, fewer than half provide evidence of the reliability of the assessments used. Studies reporting on reliability almost uniformly report coefficient alpha, a measure of internal consistency. Internal consistency ranges from .60 to .93, indicating a modest to high level of reliability. Only one study includes any evidence of validity: Cohen and Garcia (2005) cite evidence of convergent validity in the strong correlation of their Racial Identification scale with the Race Centrality subscale of the Multidimensional Inventory of Black Identity (Sellers et al., 1997) (r(34) = .79,

p<.01) (Cohen and Garcia, 2005, pp. 571-572). None of the 46 studies explicitly report evidence of fairness.

TABLE 3-1 Evidence of Measurement Quality in Intervention Studies

Competency	No. of Studies Assessing Competency	Studies with Reliability Evidence	Studies with Validity Evidence	Studies with Fairness Evidence
Behaviors related to Conscientiousness	15	5 (range 0.67-0.98)	0	0
Academic Self-Efficacy	2	2 (range 0.76-0.95)	0	0
Growth Mindset	12	2 (range 0.63-0.88)	0	0
Intrinsic Goals and Interest	2	1 (.72)	0	0
Positive Future Self	3	3 (range 0.60-0.92)	0	0
Prosocial Goals and Values	0	0	0	0
Sense of Belonging	8	6 (range 0.63-0.93)	1	0
Utility Goals and Values	4	3 (range 0.78-0.93	0	0

Assessment Quality in Established Instruments

To date, relatively few assessment instruments measuring one or more of the eight competencies have undergone enough research, development, and testing to yield durable evidence of reliability, validity, and fairness. That said, there are a few notable exceptions. The focus here is on the MSLQ (Pintrich et al., 1993), which measures several of the eight competencies (along with a few other more cognitive competencies), along with the Sense of Belonging scale (Bollen and Hoyle, 1990). Brief mention also is made of the extensive validation research that has been conducted on ACT's Engage and ETS's SuccessNavigator, two assessments for admitted students that address a broad range of intra- and interpersonal competencies, including a few of those identified in Chapter 2.

Motivated Strategies for Learning Questionnaire

The MSLQ is a self-report survey used widely in higher education research. The assessment underwent a 10-year development, refinement, and validation process, with funding

from the U.S. Department of Education's research division, then known as the Office of Educational Research and Improvement (Pintrich et al., 1991). It was initially developed by Duncan, Pintrich and McKeachie as a product of their theoretical model of college student motivation and self-regulated learning (Duncan and McKeachie, 2005).

The full 81-item instrument is composed of two sections, each focused on student responses to any particular classroom course. The first section, consisting of six related scales, focuses on motivation. Five of the six tap constructs related to the competencies identified in Chapter 2 (the sixth is test anxiety). These five scales reflect students' course goals (including intrinsic goals and interest), perceived control of learning (i.e., growth mindset), value beliefs (i.e., utility value), and sense of self-efficacy for learning and performance (i.e., academic self-efficacy). The second section, which assesses learning strategies, contains nine scales addressing students' cognitive and metacognitive strategies and their management of learning resources. Of these, five appear congruent with behaviors related to conscientiousness: organization, metacognitive self-regulation, time/study environmental management, and effort regulation. For each item, students are presented with a construct-related statement and are asked to rate themselves on a seven-point scale (from 1 = not true at all of me to 7 = very true of me). Coefficient alphas for the various scales relevant to this report range from .62 to .93, with the majority in the range of .68 to .80, thus revealing moderate to good reliability; the task value and self-efficacy scales show alphas at or above .90.

Validity for the MSLQ can be attributed in part to the content derived from its strong theoretical base, which is situated in a social-cognitive view of motivation and learning strategies. Complementing this, both confirmatory factor analysis and structural modeling in a large sample of college students (N = 380) provided empirical support for the constructs within the motivation and learning strategies sections (Pintrich et al., 1993). The statistical fit of these models to the data was reasonable according to a range of goodness-of-fit indices, including the chi-squared test, goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), and root mean square residual (RMR). The relationship between the 15 scales and final course grades showed modest evidence of predictive validity, with self-efficacy for learning and performance showing the highest validity (r = .41). Overall, given the patterns of convergent and discriminant validity found among the scales and in their correlations with final course grades, the authors propose that the MSLQ scales are "valid measures of the motivational and cognitive constructs" (p. 811). Validation studies on their own provide no evidence related to fairness in operational settings toward relevant subgroups, such as gender and race/ethnicity, with the latter warranting further exploration.

Sense of Belonging Scale

The three-item Sense of Belonging scale is based in Bollen and Hoyle's (1990) work on measuring perceived cohesion and has been part of the Diverse Learning Environment Survey, funded by the Ford Foundation and conducted by the Cooperative Institutional Research Program (CIRP) of the Higher Education Research Institute (HERI) at the University of California, Los Angeles. The authors based the scale on a theoretical definition of the construct "Perceived cohesion encompasses an individual's sense of belonging to a particular group and his or her feelings of morale associated with membership in the group" (p. 482). Three of the scale's six items measure students' sense of belonging to a community or institution, and the remaining three relate to their feelings of morale. The three sense of belonging items are (1) I

feel a sense of belonging to_____; (2) I feel that I am a member of the _____ community; and (3) I see myself as part of the _____ community. Students respond to these statements on a 10-point Likert scale (from 0 = strongly disagree to 10 = strongly agree).

Bollen and Hoyle (1990) tested the validity of their scale in a study involving two samples of respondents, the first a randomly selected group of 102 undergraduates from a small, northeastern college reputed to have strong school spirit, and the second a random sample of 110 residents of a midsized northeastern city. The study tested the hypothesis that the college students would show greater cohesion than the city residents. Analyses of both samples found that a two-factor model, reflecting, respectively, items on sense of belonging and feelings of morale, was a good fit to the data based on a number of model fit indices (chi square test, GFI, AGFI). However, the unrestricted model revealed that the two factors were empirically indistinguishable. Nonetheless, the authors argue that although the two may be empirically indistinct, they remain theoretically useful as an overall construct of cohesion. The analysis revealed the measures were reliable with a high degree of structural invariance across the two samples (high and equal factor loadings). Further, the latent group means for the college group on these two constructs underlying cohesion were both higher than those of the city resident group—as the authors hypothesize, supporting the validity of interpreting scores as students' standing on the construct of group cohesion. The sole example of evidence related to fairness was an investigation of potential bias for individuals from the middle versus the working class. None was found.

As noted, the Sense of Belonging scale currently is part of HERI's CIRP, which conducted additional expert and practitioner review and psychometric analysis of the scale as part of the pilot and field testing of the Diverse Learning Environment Survey (Hurtado and Guillermo-Wann, 2013). The three-item scale again showed high internal consistency reliability ($\alpha = .93$); in such a short scale, this means the three items were highly correlated (people who responded at a high or low level on a given belongingness item generally did the same for the others).

Engage and SuccessNavigator

ACT and ETS, the major publishers of college admissions tests, have each developed assessments of college readiness—Engage and SuccessNavigator, respectively—that measure a few of the eight competencies identified by the committee, along with a range of other competencies. Both instruments are designed for use with students already admitted into college to identify proactively those who may require additional support and to assist colleges in identifying the developmental interventions that will increase the likelihood of students' persistence and academic success. Both instruments have undergone extensive research and development (Le et al., 2005; Robbins et al., 2004, 2006; Markle et al., 2013a; Rikoon et al., 2014; Wiley et al., 2010). Similar to the above examples, work included thorough grounding in available theory; pilot and field testing to establish reliability; first- and second-order factor analyses to provide empirical support for the instruments' conceptual design; and analysis of the relationships between the assessment scores and various criteria, including course grades, retention, and graduation (e.g., Moore et al., 2015). Further supporting the validity of score use, studies have examined the relationship between the use of the scores for placement and subsequent student success with coursework (Rikoon et al., 2014). Fairness also has been

explicitly examined. For example, SuccessNavigator was subjected to tests for both measurement (reliability) and structural (validity/prediction) invariance by gender and race/ethnicity (Markle et al., 2013a).

TOWARD BETTER MEASUREMENT

The committee's analysis of the quality of existing assessments of the eight identified competencies indicates room for improvement. Such improvement starts, the committee believes, with a professional approach to assessment development, considers new measurement options that may mitigate existing shortcomings, and includes the use of multiple measures and multiple levels of analysis.

Assessment Development

Validity and fairness are driving concerns throughout a rigorous test development and validation process. As the *Standards* note:

> …all steps in the testing process, including design, validation, development, administration, and scoring procedures, should be designed in such a manner as to minimize construct-irrelevant variance and to promote valid score interpretations for the intended uses for all examinees in the intended population (American Educational Research Association et al., 2014, p. 63).

Borrowing from the *Standards,* Downing and Haladyna's (2006) 12-step framework, and Mislevy and colleagues' (2003) evidence-centered design, the subsections below outline a systematic process of test specification, item development and review, administration, and validation.[2] Although the *Standards* are a primary source for the discussion, there are several other significant, but perhaps more specialized, sets of assessment standards also worthy of consideration.[3]

Test Specifications: Overall Plan for Assessment Development

Because the quality of an assessment is evaluated largely by how well it measures intended constructs and serves the intended purposes, it is important to begin the assessment development process with a clear articulation of the constructs to be assessed and the purposes, or set of purposes, to be served. Test specifications provide an overall plan for assessment development that begins with a description of these purposes.

The specified *content definition* elaborates the meaning of the construct to be measured. Dweck (2006), for example, argues that growth mindset is the belief that intelligence is not fixed,

[2]While it is beyond the scope of this chapter to provide a detailed discussion of the full range of issues involved in assessment development, the interested reader is referred to more extensive treatments elsewhere (Downing and Haladyna, 2006; Lane et al., 2015; Schmeiser and Welch, 2006).

[3]These include the Society for Industrial and Organizational Psychology's (2003) *Principles for the Validation and Use of Personnel Selection Procedures, 4th edition*; the International Test Commission's (2014) *Guidelines on Quality Control in Scoring, Test Analysis, and Reporting of Test Scores*; and the Educational Testing Service's (2014) *ETS Standards for Quality and Fairness*.

but is malleable; it changes with learning and experience, and it can be taught (see Chapter 2). Thus a content definition of the growth mindset construct would include these elements, and a sampling plan would articulate how an assessment would be based on sampling those elements (what percentage of items would address one's beliefs in a fixed versus malleable view of intelligence, one's beliefs about the efficacy of changing intelligence, and so on).

The test purpose(s) and score interpretations also are specified in the assessment development process. Doing so is important because different purposes and score interpretations have implications for the assessment's design and psychometric characteristics. For example, an assessment intended to determine and interpret a student's level of growth mindset for the purpose of informing decisions about admission or placement will need to differentiate students at various specified levels in a psychometrically reliable manner. By contrast, an assessment that will be used to evaluate how well an intervention influences the development of the construct needs to measure growth mindset over time, perhaps at the group level.

In addition, test specifications lay out a detailed plan for who will be tested and how the test will work. The *Standards* (American Educational Research Association et al., 2014, p. 85) present a comprehensive and detailed plan for what should be included:

- the purpose(s) of the test,
- the definition of the construct or domain measured,
- the intended examinee population,
- interpretations for intended uses,
- the content of the test,
- the proposed test length,
- the item formats,
- the desired psychometric properties of the test items and the test, and
- the intended sequencing of items and sections.

The *Standards* also suggest that the process of documenting the validity of the interpretation of test scores starts with the rationale for the test specifications. The specifications should be subject to external review of their quality by qualified and diverse experts (American Educational Research Association et al., 2014, p. 87, Standard 4.6) who can provide objective judgments. Particularly important is an evaluation of the content definition and the extent to which it represents the intended intra- or interpersonal construct(s). This definition not only will guide test development but also will serve as a critical touch point in evaluating existing assessments for use. Critically, the *Standards* point to construct underrepresentation as a major issue in assessment validity. If a construct is inadequately represented by an assessment, any conclusions drawn from the assessment results are limited and may need additional verification.

Close examination of test specifications when selecting assessment instruments also can help stakeholders in higher education guard against the "jingle-jangle" fallacies (Kelley, 1927; Pedhazur and Schmelkin, 1991; Roeser et al., 2006) mentioned earlier. Examining test specifications can help guard against such erroneous assumptions.

Item Development and Review

According to the *Standards*, assessment items are developed, based on the test specifications, by teams of individuals trained and qualified in the process, which typically

would mean that they are familiar with the construct, with the item types used to measure the construct, and with how to interpret psychometric findings from a pilot or field test.

Universal design and accommodations in item development Attention to individuals with disabilities and those for whom English is not the first language, as well as issues surrounding accommodations for those individuals, is one way in which assessment development in the United States has changed most dramatically over the past 15 years, (Thurlow et al., 2006). These concerns have become particularly important since passage of the 2008 Americans with Disabilities Amendments Act.[4] A central issue concerns test development procedures that avoid biased scores due to construct-irrelevant variance related to a student's disability or English language proficiency, thereby supporting comparability of scores and comparability of inferences about the intended construct across all intended test takers.

Universal design—a concept that originated in architecture to make buildings accessible to all—has now become standard practice in assessment development (Johnstone et al., 2008) and increasingly is required by state and federal education agencies (Laitusis, 2007). In the present context, universal design means designing items that will be accessible to as wide a range of the intended examinee population as possible—for example, by eliminating unnecessarily complex language (when such language is construct-irrelevant) so as not to bias results for non-English-fluent students or those with reading disabilities. Accommodations for students with disabilities include such things as screen and text readers (software that converts text from a screen into speech and allows use of a keyboard rather than a mouse) and tactile graphics (figures raised from the text, often with captions in braille) (Educational Testing Service, 2010; Laitusis, 2007). Test administrators also may allow for alternative response modes (e.g., speaking, pointing), additional testing time, and unproctored testing (Beaty et al., 2011). Because intra- and interpersonal competencies often are assessed through written self-report instruments, and language ability is not the target of such measures, the use of simplified English expression to reduce the language demands of a test may be particularly important for ensuring accessibility to higher education for students who are not fully fluent in English.[5]

Item review Once assessment items and any accommodations have been developed, they are subject to content and fairness reviews, typically conducted by committees of substantive and psychometric experts, along with experts knowledgeable about the populations to be tested. These experts evaluate the extent to which the items match their construct targets, reflect the test specifications, and are accurate in content. Items also are reviewed for fairness to ensure that they do not contain construct-irrelevant characteristics that could impede some students' ability to understand or respond appropriately, which would distort " the meaning of the scores and thereby decrease the validity of the proposed interpretation" (American Educational Research Association et al., 2014, p. 217).

[4]Public Law 110-325, http://www.access-board.gov/about/laws/ada-amendments.htm [July 2016].

[5]The chapter in the *Standards* on fairness (Chapter III) provides additional detail on several key principles in the use of appropriate test accommodations and modifications for English language learners and students with disabilities and in the reporting of scores from such assessments.

Test Design, Assembly, and Field Testing

Test design and assembly refers to the process of compiling the collection of items and tasks that will form the actual test (also called a "test form") so that they conform to the test specifications. Test assembly often involves selecting items to satisfy psychometric requirements, such as test form reliability, as well as content requirements. There are increasingly sophisticated approaches for conducting test design and assembly, including evidence-centered design (Mislevy and Riconscente, 2006), automatic item generation (Gierl and Haladyna, 2013; Irvine and Kyllonen, 2002), optimal test design (van der Linden, 2005), and assessment engineering (Luecht, 2013).

Alternative test forms are those that measure the same construct, are built to the same set of test specifications, and are to be administered under similar conditions. Alternative forms are useful because the security of item content can be compromised through item exposure, and alternative forms can help minimize exposure of the content over time for any particular item. Alternative forms may also be useful in the context of a program evaluation in which an assessment is conducted both prior to and following the treatment (e.g., pretest, posttest, delayed posttest). In other contexts, items may partially overlap across alternative forms, such as in adaptive testing, where items are tailored to each examinee's responses, or when one seeks to measure a construct at the group level, but each student receives only a limited number of items.

Test Administration

Test administration encompasses, for proctoring, the qualifications of those involved in the test administration, security, timing, translations, and issues associated with accommodations for test takers with special needs. All of these issues are considered to ensure that an assessment measures the construct it is intended to measure and to minimize the effects of cheating, adverse testing conditions, and other factors that might otherwise induce construct-irrelevant bias or variance on test scores. Although these issues are routinely considered critical for cognitive assessments, they are just as important in assessing intra- and interpersonal competencies. Conditions under which assessments are administered, including timing and instructions, also should be standardized to ensure fairness and comparability of scores across sessions.

Validation of Score Inferences

Evidence supporting the validity of score interpretation and use is collected throughout the test development and administration process, including through special validation studies. Content and bias reviews of test specifications, items, and forms can provide content-related evidence about the extent to which an assessment measures the intended competency and the range and depth of its construct representation, and ensure that test items are free of extraneous attributes that otherwise could constrain some students' ability to demonstrate their competence. Pilot testing of test items often includes think-aloud protocols that can elicit evidence both that students understand the assessment questions and/or expected responses and that response processes actually reflect the intended competency.

Field testing and/or operational test administration then follows, which typically generates evidence of reliability; internal structure-related validity evidence; and, assuming

adequate subgroup representation, various differential item functioning (DIF) analyses related to fairness. Special studies then need to be conducted to collect additional validity evidence related to hypothesized relationships between test scores and other variables. These studies include convergent-divergent studies, analyses of test-criterion relationships and/or convergent-divergent evidence, and studies exploring the possibility of differential prediction and/or criterion relationships for diverse student groups (see prior sections of this chapter on validity, reliability, and fairness).

New Testing Techniques Supporting Improvement

Rigorous test development processes can help improve the quality of existing measures of the eight competencies, as can new approaches to ameliorating some of the shortcomings of self-report measures, the most common type of measures currently used to assess the eight identified competencies.

Anchoring Vignettes

Anchoring vignettes are a technique developed to mitigate response style bias in Likert measures among individuals or groups, particularly cross-culturally (King et al., 2004). The technique requires respondents to rate one or more anchoring vignettes, which are brief descriptions of a hypothetical person or situation. Respondents then rate themselves or their own situation on the same rating scale they used to rate the anchoring vignettes. Comparison of the self-ratings with the anchored ratings is then used to create an adjusted score; both parametric and nonparametric scoring methods are employed (King and Wand, 2007). Data from the Programme for International Student Assessment (PISA) 2012 indicate that this approach has improved the comparability of intra- and interpersonal competency scores for individuals and groups (Kyllonen and Bertling, 2013).

To illustrate the concept of anchoring vignettes, consider a vignette pertaining to the intrapersonal competency of sense of belonging. An example item reflecting sense of belonging can be found in the PISA 2012 survey (OECD, 2013b): "I feel like I belong at school." Students report their response on the four-point Likert scale "strongly agree," "agree," "disagree," and "strongly disagree." A corresponding anchoring vignette for this item might be something like the following: "After a class lecture, Rodrigo will discuss the class with his peers comfortably and without a sense of competition. He also shows a sense of humor interacting with his peers on his intramural volleyball team. Indicate how much you agree that Hidalgo believes he belongs at school: 'strongly agree,' 'agree,' 'disagree,' or 'strongly disagree.'" Based on the self-rating and the vignette rating, an anchoring vignette-adjusted score would then be computed. This score would be related to the difference between the two ratings, reflecting the degree to which respondents rated themselves higher or lower than they rated the hypothetical Rodrigo.

Anchoring vignettes such as this operate under two assumptions: the *vignette equivalence assumption*, that all respondents interpret the vignette in the same way (Bago d'Uva et al., 2011); and the *response consistency assumption*, that respondents use the same scale when evaluating themselves and the person in the vignette (Kapteyn et al., 2011).

Forced Choice

The forced-choice, or ipsative, method requires respondents to select from among two or more alternatives, such as two statements. In a strict forced-choice format, a respondent might be asked to indicate which statement is "most like me" from the statements "I enjoy working with others" and "I set high personal standards." Choosing one statement necessarily means not choosing the other. This can be difficult for the respondent, which is often the point—that a choice must be made despite the difficulty, and a better assessment should result. Another common approach is a more relaxed form of forced choice that entails providing four statements from which the respondent is asked to select the one "most like me," as well as the one "least like me." The four statements might include, for example, "I enjoy working with others," "I set high personal standards," "I manage to relax easily," and "I am careful about detail" (Brown and Maydeu-Olivares, 2011).

The advantage of the forced-choice format is that respondents can be made to choose between pairs of statements that have equal levels of social desirability (i.e., equally desirable or equally undesirable), which can reduce bias due to social desirability. Although the ipsative method has been used for decades (Edwards, 1957), a traditional assumption was that the information gathered was relevant only for understanding within-person choices or relative preferences, not for understanding differences among people (Cattell, 1944). More recently, however, quasi-ipsative test designs and psychometric approaches to forced-choice measures have been developed. There now is evidence that scores on assessments using both approaches (quasi-ipsative and forced choice), as well as ranking approaches, provide similar evidence for a respondent's trait standing. Results of an experimental laboratory study suggest that Likert scale and forced-choice methods provided similar information about respondents with respect to the traits measured, although assessment scores in both formats were affected by instructed faking conditions (Heggestad et al., 2006). A meta-analysis (Salgado and Táuriz, 2014), however, indicates that across studies, forced-choice methods tended to provide stronger predictions of educational outcomes (GPA) and workforce outcomes (performance, training) relative to rating-scale methods.

New Methods Enabled by Technology

Advances in technology will continue to push the options for assessment. For example, automatic scoring methods based on natural language processing (NLP) techniques, that are under development, may be able to alleviate the scoring burden of performance tasks, essays, and other constructed-response items and thus enable larger-scale implementation. Beigman Klebanov and colleagues (2016) demonstrated that NLP techniques could be used to assess the utility value students perceive in biology topics. Specifically, the authors used these techniques successfully to evaluate the degree to which essays were in compliance with the following instructions:

> Write an essay addressing this question and discuss the relevance of the concept or issue to your own life. Be sure to include some concrete information that was covered in this unit, **explaining *why* this specific information is relevant to**

> **your life or useful for you.** Be sure to explain *how* the information applies to you personally and give examples (Harackiewicz et al., 2015).

The authors showed that essay writing features such as the use of appropriate general and genre-topic vocabulary, affect and social vocabulary, and argumentative and narrative elements were useful in measuring the degree to which student essay writers complied with instructions to reflect on and express the value of the biology topics in their personal and social lives.

In another example from the intervention literature, Yeager and colleagues (2014, Study 4) used a technology-assisted "diligence task" to measure academic self-regulation, a behavior related to conscientiousness. Designed to mirror students' real-world behavior when trying to complete homework in the face of digital distractions, the task gave participants the choice of completing boring single-digit subtraction problems or consuming media (either brief, viral videos or playing the video game Tetris). The software unobtrusively tracked the number of math problems completed successfully to assess academic self-regulation.

To assess female engineering students' sense of belonging in the field, Walton and colleagues (2015) used a modified version of the Implicit Attitude Test (Yoshida et al., 2012), a computerized assessment designed to address the challenge that individuals may provide a socially desirable response rather than a true response when asked to self-report on attitudes that may be viewed socially as biased or prejudiced. Specifically, the test measured the reaction time when students were asked to associate the concept "most undergraduates at your university like" (versus "most undergraduates at your university don't like") with the concept "female engineers." Participants were asked to categorize a series of words and images as quickly and as accurately as possible using keys on the left and right sides of the keyboard to indicate the category to which each word or image belonged. Students holding negative associations with most people's evaluation of female engineers were expected to find the task more difficult, and respond more slowly, when "most people like" and "female engineers" were presented together than when "most people don't like" and "female engineers" were presented together. Higher scores represented more positive implicit attitudes toward female engineers, suggesting a greater sense of belonging.

A more novel, recent example of the promise of technology is Pentland's (2008) work using unobtrusive badges that automatically track how frequently individuals speak to each other, turn toward each other, mirror each others' gestures, and so on to assess authentic communication and teamwork. The author notes that humans express subtle interaction patterns that can be interpreted as honest signals in the form of timing, energy, and the variability of expressions. These honest signals provide clues regarding one's degree of influence on other people, the degree to which one unconsciously mimics others, how one expresses interest and excitement in the form of activity and demonstrativeness, and consistency in speech and movement that convey focus or signal openness to influence from others. Likewise, building on the idea that certain cues (e.g., facial expressions, gestures, vocal prosodity) can provide information about a person's intra- and interpersonal competencies, recent work in computer science is concerned with automating the evaluation or scoring of interviews based on multimodal emotion detection methods (Chen et al., 2014). Applications of such technologies are beginning to emerge, and as the technologies become more available and less expensive, they are likely to influence assessment.

There are other rapidly developing areas of assessment based on advances in technology and the concomitant advances in psychometrics that support new assessment types. These

include game-based assessments (Mislevy et al., 2014) and collaborative problem-solving tasks (von Davier et al., 2017).

Use of Multiple Measures

It is axiomatic that any single measure can provide only one perspective on any given construct. Consider the construct of conscientiousness, which can be measured in a variety of ways, for a variety of purposes. Conscientiousness might be measured with a self-report rating scale (e.g., "I continue working on tasks until they are finished. Select one: strongly disagree, disagree, agree, strongly agree.") or a peer or teacher rating (e.g., "How often does X turn in assignments on time? Select one: rarely or never, seldom, sometimes, often, always, or almost always."). Conscientiousness also might be measured by college administrative records (e.g., available records on class attendance); with a situational judgment test (e.g., "You have a test the next day and don't feel fully prepared. You are very tired and you are not thinking clearly. What do you do?"); or in a behavioral interview (e.g., "Tell me about a time when you had to persist on a task despite many barriers in your way."). Or it might be measured indirectly with a performance task, such as working on an impossible puzzle (Walton et al., 2012); choosing a difficult over an easy anagram task (Gerhards and Gravert, 2015) or a difficult over an easy addition task (Alan et al., 2015); or working quickly on a tedious task, such as a picture-number lookup task (Segal, 2008).

Because each of these methods can have its benefits and drawbacks, using more than one method can increase the precision of measurement and the strength of inferences that can be drawn. A few of the intervention studies in Chapter 2 used multiple assessment methods to gather additional information about the target competency (see Appendix C). For example, Walton and colleagues (2012) assessed motivation, a behavior related to conscientiousness, using both a self-report instrument and the amount of time spent on an insoluble math puzzle. In another example, Vansteenkiste and colleagues (2004a, Study 1) assessed intrinsic interest in the intervention topic (recycling) using both a self-report, selected-response instrument and a behavioral measure of visits to the library and/or the recycling center. Because students' library visits were automatically recorded with a card swipe, data on this measure of intrinsic interest were readily available.

Furthermore, even within a specific method, one often can use a wide range of equally legitimate content when developing an assessment. In addition, time and other practicalities always limit the number of items that can be used for an assessment, and in general, as noted earlier, the longer the assessment, the more reliable it is likely to be. By extension, it is a truism of assessment that longer tests are associated with (but do not guarantee) higher levels of validity, a point that the *Standards* note as particularly critical when high-stakes decisions depend on assessment results.

Ultimately, the usefulness of any given assessment depends on the fundamental characteristics of validity, reliability, and fairness discussed above. When put to their best use, assessments are designed to provide data with which to answer one or more specific questions (e.g., What level of conscientiousness does the student have? Is that level enough to help him/her persist through challenges at college? Does the new intervention being used to increase students' conscientiousness actually work?). The use of multiple measures of the targeted competency is likely to yield more valid answers to such questions.

Recognizing the Multilevel Context

It is important to emphasize that, beyond the factors discussed thus far in this chapter, educators and researchers need to understand and be sensitive to the factors that define the context in which assessment occurs if they are seeking to measure, understand, and intervene on key intra- and interpersonal competencies that ultimately influence student success. This systems perspective is broader than the charge to this committee and the scope of this report, and practically speaking, any research study necessarily will focus on a small portion of such a system. Nonetheless, it must be recognized that the context of intra- and interpersonal competency assessment is wide-ranging and encompasses numerous individual, group, and institutional entities operating and interacting simultaneously (e.g., diverse students and peer groups, instructors with varying roles and experience, classrooms with the potential to create and facilitate opportunities to exhibit and develop intra- and interpersonal competencies, institutions and departments that help establish both mission and culture).

Today, available statistical methods and computational power can be used to analyze assessments and context at the same time. Without providing a comprehensive review of those methods, it is worth noting that *multilevel models* **(**Gelman and Hill, 2007) provide both a conceptual and statistical basis for determining whether levels in a hierarchy are related. Referring to Figure 3-1, suppose one wanted to know whether new disciplinary accreditation standards calling for assessment of particular intra- and interpersonal competencies influenced a community college president, who then implemented departmental policies that affected faculty members, who in turn interacted with students who were assessed on their intra- and interpersonal competencies before and after the policy was formally enacted. In addition to answering this sort of question, multilevel models can incorporate longitudinal data (e.g., whether a construct measured among instructors at time A affected another construct among students at time B); measurement error variance (e.g., modeling relationships while accounting for the fact that psychological measures are never perfectly reliable); and different error structures (e.g., cyclical trends experienced year to year in a department or institution, or autocorrelation between neighboring events in any span of time).

The overall point is that the choice of an assessment strategy needs to be sensitive to the assessment context across all levels, such as those just mentioned and depicted in Figure 3-1. In addition, this context will have a significant influence on what range of best practices should be considered in developing, administering, scoring, and interpreting an assessment, regardless of whether it is to be used for high- or low-stakes purposes. Researchers and assessment experts in higher education are encouraged to incorporate data on context (e.g., culture, climate, department) into their analyses and interpretations of intra- and interpersonal competency assessments.

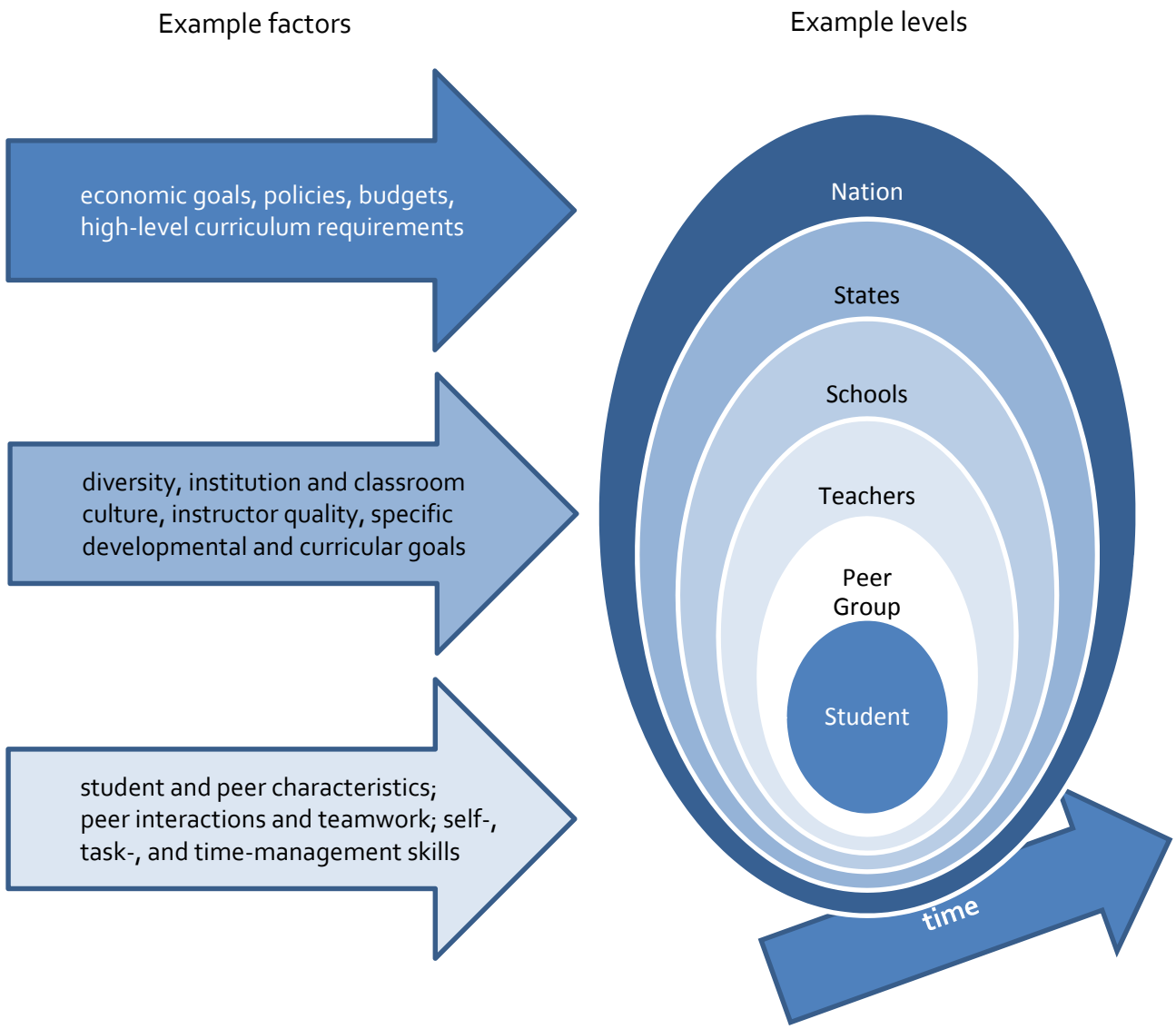

FIGURE 3-1 A multilevel perspective on intra- and interpersonal competencies.
SOURCE: Created by the committee.

CONCLUSIONS AND RECOMMENDATIONS

The committee reviewed the nature and quality of existing assessments of the eight competencies identified in Chapter 2, together with research and professional standards related to the overall process of developing; validating; implementing; and interpreting, evaluating, and using the results of assessments of intra- and interpersonal competencies. The test development practices used to create assessments of cognitive knowledge and skills that meet professional standards are equally applicable to intra- and interpersonal competency assessments.

Current Assessments

The committee examined the assessments used in the intervention studies targeting the eight competencies identified above and commissioned a literature search on measurement of these competencies. Drawing on both sources, the committee also identified and closely analyzed a small sample of established assessment instruments targeting one or more of the eight competencies. Overall, the review revealed that self-report methods, with their known limitations, predominated in the assessments of the eight competencies. Analysis of the quality of the assessments used in the intervention studies revealed spotty attention to reliability and almost no reported evidence of validity or fairness. However, more evidence of assessment quality was found for some established assessment instruments used in higher education research, particularly those efforts that have received funding for assessment research and development. These instruments provide evidence on reliability and validity but lack evidence on fairness. Assessments developed by professional testing companies provide even more evidence of quality, including fairness data; however, these assessments target a wider range of competencies, only partially addressing some of the eight competencies.

Conclusion: Most current assessments of the eight identified competencies are uneven in quality, providing only limited evidence to date that they meet professional standards of reliability, validity, and fairness.

Assessments for High-Stakes Purposes

Developers of all types of assessments, whether they aim to measure cognitive, intrapersonal, or interpersonal competencies, must exercise particular care when an assessment will serve a high-stakes purposes. Assessments are considered high-stakes when their results carry serious consequences for individuals or institutions.

Conclusion: The development and validation of assessments of intra- and interpersonal competencies for high-stakes purposes is a rigorous, time-consuming, and expensive process that depends critically on expertise in assessment and psychometrics. Validity, reliability, and fairness are essential considerations in evaluating assessment quality.

Recommendation 5: When developing and validating intra- and interpersonal competency assessments to be used for high-stakes purposes, stakeholders in higher education (e.g., faculty, administrators, student

services offices) should comply with professional standards, legal guidelines, and best practices to enable appropriate interpretations of the assessment results for particular uses.

Recommendation 6: Institutions of higher education should not make high-stakes decisions based solely on current assessments of the eight identified competencies, given the relatively limited research to date demonstrating their validity for predicting college success.

Assessments for Low-Stakes Purposes

Researchers and practitioners in higher education also use assessments for low-stakes purposes, such as to evaluate the quality of interventions, policies, and instructional practices or simply to monitor student change over time. When used for these low-stakes purposes, assessments need not meet the high evidentiary requirements of individual high-stakes student assessments, such as college admissions tests. Professional testing standards clearly state that the amount and type of evidence needed to support a test's validity may vary depending on the use or interpretation of the test scores. At the same time, even when assessments are not used for high-stakes purposes, they need to be sensitive to the competencies they are intended to measure.

Conclusion: Even low-stakes uses of intra- and interpersonal competency assessments require attention to validity, reliability, and fairness, although they need not meet the high evidentiary requirements of high-stakes assessments.

Recommendation 7: Those who develop, select, or use intra- and interpersonal competency assessments should pay heed to, and collect evidence of, validity, reliability, and fairness as appropriate for the intended high-stakes or low-stakes uses.

Definition of Constructs Being Assessed

After reviewing both general principles for assessment development and use and recent research on measurement of the eight identified competencies, the committee concluded that defining each competency clearly and comprehensively is a critical first step in developing high-quality assessments. Clear definitions are especially important in light of the wide variety of terms used for these competencies. For example, conscientiousness, grit, and persistence are closely related constructs, despite being named differently. In fact, the content of items used to assess all of these constructs may be quite similar. Conversely, assessments bearing the same name may in fact measure different constructs.

Conclusion: High-quality assessment begins with a clear definition of the competency to be measured, and identifies how the assessment will be used and what kinds of inferences it will support.

Construct definitions guide assessment development and selection by making it possible to evaluate how well the assessment represents the construct it is intended to measure, thereby

supporting appropriate inferences about the construct for particular uses. High-quality assessments avoid construct underrepresentation, represent the breadth and depth of the construct, and minimize any distortions caused by construct-irrelevant influences.

Innovative Methods and Technologies for Assessment

Self-report measures, such as those frequently used to assess the eight identified competencies, have several limitations. First, individuals responding to both high- and low-stakes assessments may be motivated to present themselves in a favorable light. In addition, people often express themselves on a response scale in habitual or characteristic ways, such as tending to mark the extremes (e.g., "strongly agree" or "strongly disagree") or to agree or respond positively regardless of the question. Respondents also tend to compare themselves with those around them, such as their close peers. This tendency can compromise the use of the responses to measure growth or to compare groups of individuals because such comparisons depend on an absolute rather than a relative standard. Because self-report measures are widely used, these limitations affect a broad swath of current intra- and interpersonal competency assessments.

Recent research has identified various methods that can mitigate these limitations. Ratings by others avoid some of the problems of self-reports and have been found to yield more reliable and predictive data in many contexts. The use of forced-choice and ranking methods for collecting self-evaluations avoids response-style bias by circumventing traditional rating scales altogether. The use of anchoring vignettes also addresses response-style bias by having raters make use of detailed objective anchors, and may potentially deal with reference group effects as well. Other nontraditional measures include situational judgment tests as well as games or simulations, which avoid many of the documented limitations of self-ratings. Further research is needed to develop, extend, and refine these promising new approaches.

Conclusion: Most existing assessments of the eight identified competencies, as well as many existing assessments of other intra- and interpersonal competencies, use self-report measures, which have well-documented limitations. These limitations may constrain or preclude certain uses of the results. Innovative approaches for assessing intra- and interpersonal competencies can address these limitations.

Recommendation 8: Federal agencies and foundations should support additional research, development, and validation of new intra- and interpersonal competency assessments that address the shortcomings of existing measures.

Fairness in Assessment

The *Standards* make clear that fairness to all individuals for whom an assessment is intended should be a driving concern throughout assessment development, validation, and use. Assessment development should minimize construct-irrelevant characteristics that would interfere with the ability of some individuals or subgroups to show their standing on a competency or lead to individual or subgroup differences in the meaning of test scores.

Whenever differences in subgroup scores are observed, follow-up research may be needed to examine the reasons, the potential sources of bias, and the comparability of score interpretations across individuals and subgroups in light of the intended uses of the assessment results. The committee applied these fairness principles in its review of current assessments of the eight identified competencies.

Conclusion: Despite the ever-increasing diversity of undergraduate student populations, attention to fairness for diverse populations is often inadequate in the development, validation, and use of current assessments of the eight identified competencies.

Recommendation 9: Researchers and practitioners in higher education should consider evidence on fairness during the development, selection, and validation of intra- and interpersonal competency assessments.

Consideration of Contextual Factors

Self-, peer, or instructor ratings of an intrapersonal competency such as conscientiousness or an interpersonal competency such as teamwork may vary depending on local norms (e.g., reference group effects). In addition, contextual variables may mediate or moderate the relationships between intra- and interpersonal competencies and educational outcomes. For example, of an intervention intended to develop sense of belonging may be effective only for disadvantaged student groups (see Chapter 2).

Conclusion: Appropriate interpretation of the results of intra- and interpersonal competency assessments requires consideration of contextual factors such as student background, college climate, and department or discipline.

Recommendation 10: Higher education researchers and assessment experts should incorporate data on context (e.g., culture, climate, discipline) into their analyses and interpretations of the results of intra- and interpersonal competency assessments.

Implementing this recommendation will require that higher education researchers use appropriate statistical analyses that incorporate data on context when examining assessment results. Such analyses include use of multilevel statistical models, measurement invariance analyses, application of differential item functioning, and mediator and moderator analyses. These analyses can enhance understanding of the complex interactions and processes entailed in students' individual competencies and of features of higher education contexts that contribute to students' persistence and success. Multiple measures also can be used to minimize the possibility that inferences about a student's intra- and interpersonal competencies are due to a particular measurement approach.

4

Assessment Uses and Stakeholders within Higher Education

The conceptual bases and empirical findings of the previous chapters indicate that eight intrapersonal competencies show some evidence of predicting success in higher education. In addition, as discussed further in Chapter 5, some intra- and interpersonal competencies have been identified as desired outcomes for 2- and 4-year college graduates. In this report, the committee recommends further research to better understand each group of competencies and their relationships to students' college success, research that reflects one vital use of assessment. In addition, if research confirms these relationships, assessments of these competencies can provide data useful for informing college decision making and buttressing improvements in teaching, learning, and co-curricular support services. Such uses of assessment are the focus of this chapter. Following a brief discussion of the growth of assessment in higher education, the chapter describes the current and potential future uses of assessments of intra- and interpersonal competencies. The third section details the potential users of such assessments—the various stakeholders who may use and apply the resulting data on student competencies. The fourth section focuses on how these data can be used to support improvement and explores factors that facilitate or inhibit such uses of the data. Next is a set of cases illustrating how some colleges and universities are already using assessments of intra- and interpersonal competencies to enhance college readiness and success. The chapter ends with conclusions and recommendations.

THE GROWTH OF ASSESSMENT IN HIGHER EDUCATION

In response to demands for accountability and improvement in higher education, many colleges, universities, and fields of study have begun to identify specific learning outcomes for all college graduates and to assess students' attainment of these outcomes. Two recent surveys provide insights into this trend. Kuh and colleagues (2014) report findings from a survey of provosts at 1,200 regionally accredited undergraduate institutions, including a mix of 2- and 4-year public and private institutions; the survey received a 43 percent response rate. Large majorities of respondents reported that their institution had identified specific student learning outcomes and that assessment of these outcomes had increased since 2009. Hart Research Associates (2016) reports on an online survey of chief academic officers at 1,001 member institutions of the Association of American Colleges and Universities (AAC&U). The 325 respondents were representative of the association's membership, including 2- and 4-year public and private institutions ranging from regional state colleges to research-intensive universities. Their responses were similar to those reported by Kuh and colleagues (2014). A large majority (85 percent) of responding institutions had established a set of common learning outcomes for all

undergraduates, up from 78 percent in 2008. The learning outcomes included cognitive competencies (e.g., inquiry and analysis), intrapersonal competencies (e.g., lifelong learning) and interpersonal competencies (e.g., oral communication skills). The proportion of AAC&U member institutions that reported assessing these learning outcomes across the curriculum had grown to 89 percent, from 72 percent in 2009.

Interest in assessing postsecondary students' learning outcomes continues to grow, driven partly by concerns of policy makers and the broader public about college effectiveness (Arum and Roksa, 2011; Carey, 2014; Guttenplan, 2014; Kaminer, 2013). These concerns often include questions about whether the outcomes of higher education justify its costs (Bennett and Wilezol, 2013). A focal point of higher education policy since the 1980s, assessment and program evaluation for the purpose of improving students' learning today is a movement of notable strength (Astin, 2012; Ewell, 2008).

The growing implementation of assessment tools and measures not only reflects postsecondary institutions' growing interest in—and responsibility for—demonstrating the attainment of student learning outcomes, but also gives rise to a growing body of evidence on which to base institutional decision making. Furthermore, transparency in the documentation of learning outcomes has naturally fueled interest in how such outcomes can be improved and in what role data can and should play in improvement efforts.

ASSESSMENT USES

Researchers describe a number of ways in which student assessment data can be used in higher education. Rowntree (2015), for example, distinguishes among the following uses: selection, maintaining standards (i.e., quality control), student motivation, feedback to students, feedback to instructors, and preparation for life. Falchikov (2013) discusses assessment for measurement, procedure, inquiry, accountability, and quality control. The broader literature contains several similar conceptualizations, reinforcing the idea that assessment in higher education serves a range of uses, goals, and stakeholders (e.g., Hughes, 2014; Lambert and Lines, 2013). This section examines four major, interrelated uses of assessment:

- selection and placement of individual students;
- formative improvement of local educational processes, practices, and programs;
- research and evaluation supporting knowledge generation; and
- accountability.

Attention to two of these uses—formative improvement and accountability—is justified by their prominence in the higher education literature (Ewell, 2002, 2008, 2009). According to Ewell (2002), these two purposes entail distinct values and actors, with improvement focusing on internal stakeholders (faculty, staff, administrators, students within one institution) using data to change educational practices, and accountability focusing on documenting the level of institutional effectiveness for external audiences (accreditors, policy makers). This tension between the two uses in the literature is apparent in a recent survey on assessment of student learning outcomes (Kuh et al., 2014). Provosts from a national sample of public and private 2- and 4-year institutions indicated that their greatest worry related to assessment was that external accountability mandates stretched limited assessment resources and dominated institutional conversations about assessment around an agenda of compliance rather than improvement.

Nevertheless, improvement and accountability uses of assessment can be mutually reinforcing. Research shows that accountability serves to signal important educational goals and motivate action to improve teaching and learning (Dougherty and Reddy, 2011; National Research Council, 2011), so that assessments initially used for accountability may also serve to stimulate improvement. Based on the recent survey of provosts, Kuh and colleagues (2014) reached a similar conclusion, stating that assessment of student learning outcomes "is no longer primarily an act of compliance but—more appropriately and promisingly—is driven by a balance of compliance and institutional desire to improve" (p. 5).

Beyond these two uses, assessments for college selection and placement typically meet the needs of internal stakeholders (e.g., admissions officers, developmental education instructors) (see, for example, Atkinson and Geiser, 2009; Linn, 2009; Zwick, 2007). Finally, assessments for evaluation and research are designed to build generalizable knowledge for wider audiences about the nature of the competencies (e.g., Center for Advanced Research on Language Acquisition, 2016), their role in college success, and effective programs and strategies for supporting them.

Selection and Placement of Individual Students

College admissions officers use letters of recommendation as informal representations of intra- and interpersonal competencies, and formal measures also have been developed for this purpose. At both Oregon State University and DePaul University, the admissions process includes assessments of a set of competencies that includes planning and goal setting (behaviors related to conscientiousness) and positive self-concept (related to academic self-efficacy) (Sedlacek, 2006, 2011). Applicants to Oregon State University are required to submit six short essays related to these competencies (Oregon State University, n.d.) along with transcripts and standardized test scores, while applicants to DePaul University may choose not to submit standardized test scores and instead write short essays focusing on the same set of competencies[1] (Sedlacek, 2011; Cortes et al., 2014). Over a 5-year period, Tufts University experimented with using an optional assessment of four clusters of cognitive, intrapersonal, and interpersonal competencies (one cluster reflected prosocial goals and values; Sternberg, 2010). In a final example, some graduate programs used scores from the Personal Potential Index (PPI) (Kyllonen, 2008), formerly offered by the Educational Testing Service (ETS), when considering candidates for admission.

Augmenting existing admissions tests, which focus primarily on cognitive knowledge and skills and which show disparities in scores across different racial/ethnic and socioeconomic groups, with such assessments of students' intra- and interpersonal competencies could potentially improve admission decisions and reduce differences in scores across subgroups (see Klieger et al., 2014). However, research to date has not conclusively demonstrated that the use of assessments of intra- and interpersonal competencies can reduce racial disparities in selection (Foldes et al., 2008).

Assessments designed to measure intra- and interpersonal competencies also can be used to inform decisions about the selection and placement of students after they have been admitted. This use of assessment has grown in recent years, as researchers and test developers have created a variety of instruments targeting competencies that are thought, based on correlational research, to be related to student success and may or may not be malleable in response to interventions

[1]At both universities, the essays are scored based on a rubric.

(e.g., Cengage Learning, 2015; Noel-Levitz, Inc., 2011; Pickering et al., 1992). These instruments typically assess a range of competencies, including a few of those identified by the committee, along with other competencies.

Some institutions are using these new instruments to identify incoming students who are at risk of dropping out. The University of North Texas, for instance, required all incoming students in fall 2008 and 2009 to complete the ACT Student Readiness Inventory, now known as ACT Engage (ACT, Inc., 2012). The instrument measures some of the eight competencies identified by the committee (e.g., behaviors related to conscientiousness, aspects of sense of belonging) along many other competencies (Le et al., 2005). The university used the results to select at-risk students for intervention, inviting them to a one-on-one meeting with a counselor to discuss the test results and, more important, to establish a relationship with the student and refer her or him to campus resources. An unpublished quasi-experimental study of the approach showed promising results: at the end of the first semester, 74 percent of the treatment group remained in good academic standing, compared with 63 percent of the control group (Tampke, 2011).

Allen and colleagues (2009) conducted a study illustrating the potential benefits of assessing intra- and interpersonal competencies. The authors modeled the effects on student persistence of assessing incoming students' intra- and interpersonal competencies for the purpose of selecting at-risk students for intervention. They drew on prior studies that found that scores on assessments of certain competencies were correlated with indicators of college success (Robbins et al., 2004, 2006). They also drew on a prior meta-analysis of the effects of interventions on retention, clustering the interventions into four categories as follows: academic skill (r = .15), self-management (r = .29), socialization (r = .11), and first-year experience (r = .10) (Robbins et al., 2009[2]). Using a "typical" institution with a first-year academic failure rate of 24 percent and first-year dropout rate of 32 percent, they modeled different scenarios of the proportion of students identified as at risk and the effectiveness of the intervention to estimate the proportion of students that would be saved from dropping out by assessing these competencies. Positing that the effectiveness of an intervention potentially could be increased by targeting it to the students who need it the most, as indicated by assessments of students' intra- and interpersonal competencies, the researchers created scenarios in which the effect size of the intervention increased by 0 percent, 10 percent, and 20 percent as a result of the assessment. At the low end, the additional number of students saved by assessment of intra- and interpersonal competencies in addition to traditional academic predictors was 1.5 per 5,000, assuming that 10 percent of students were identified as "at risk" and received an intervention with an effect size of 0.20, and that the assessment produced no increase in effect size. At the high end, 140.5 students per 5,000 were saved, assuming that 50 percent of students were identified as "at risk" by the assessment and received an intervention with an effect size of 0.80, and that the assessment increased the effect size by 20 percent. The authors conclude that all of the factors that influence the practical benefits of measuring these competencies will vary across institutions, making it difficult for individual institutions to draw firm conclusions. Thus they suggest that colleges and universities

[2]This meta-analysis included interventions using a variety of research designs in contrast to the committee's approach, which entailed focusing only on experimental intervention studies with random assignment. Robbins and colleagues (2009) note three limitations of the study: (1) In several instances, they had to impute missing meta-analytic effects; (2) they calculated difference scores from different study designs on a common metric; and (3) they did not take into account institutional characteristics, system-level factors, and other variables that might influence student academic performance and retention.

conduct their own local research studies to understand the potential benefits.

Formative Improvement of Educational Processes, Practices, and Programs

Formative improvement is a process by which individuals (such as faculty and students), organizations (such as departments, counseling centers, and student affairs offices), or whole institutions use assessment data to drive strategic change efforts (Suskie and Banta, 2009). This improvement process is internal to the institution, engaging stakeholders who use assessment and/or evaluation data in real time to monitor and improve educational processes, practices, and programs in pursuit of desired goals. Local stakeholders often carry out a formative evaluation cycle that includes planning, gathering and interpreting evidence from student assessments, along with other evaluation evidence, and using the evidence to inform educational improvement (Banta and Blaich, 2011; Maki, 2010). Kuh and colleagues (2014) report that institutions frequently assess student learning outcomes for internal improvement purposes, including an institutional commitment to improving, faculty or staff interest in improving student learning, president and/or governing board direction, and concerns about the effectiveness and value of education.

One example of an improvement-oriented assessment process is that of Alverno College in Milwaukee, Wisconsin, which engaged administrators, faculty, and students in conversations over the course of several years to reimagine its curriculum. The college then adopted a unique curriculum focused on developing eight abilities, including interpersonal competencies, such as communication and social interaction; intrapersonal competencies, such as aesthetic engagement; and cognitive competencies, such as problem solving (Alverno College, 2016). Students' development of these competencies within academic disciplines is assessed by both the faculty and the students themselves and recorded in a diagnostic digital portfolio. Although these assessment data ultimately are used for both for formative and summative purposes (including graduation), the Alverno assessment process began as an effort to improve student learning on campus. Accordingly, Alverno uses assessment data in an ongoing process to improve educational practices and increase student learning, including learning in these intra- and interpersonal domains (Mentkowski et al., 2000).

These formative improvement processes often occur within regular learning cycles and are focused on determining what strategies work well in specific contexts for specific students and on what can be changed to work better. As an ongoing process concurrent with the implementation of instructional approaches or support services, assessment used for formative improvement contrasts with assessment for summative purposes, which is used to gauge the overall effectiveness of a course or program at its end or in comparison with other options. As Ewell (2008, p. 9) describes, formative "assessment is accomplished directly by practitioners (faculty and administrators) acting within the parameters of the teaching and learning process." In addition, students can act directly as agents of their own formative improvement. Seal and colleagues (2015), for example, report on an intra-/interpersonal assessment designed for use by college students in their own self-development of four dimensions of competence: (1) self-awareness, (2) consideration of others, (3) connection to others, and (4) influence orientation. As noted above, improvement-oriented assessment processes focus on progress within a particular institution, program, or course or on the growth of a particular individual (Suskie and Banta, 2009), and are internal to the institution or program.

Research and Evaluation Supporting Knowledge Generation

In contrast to formative improvement, which typically involves institutions collecting student assessment data, analyzing the data, and applying them internally, the third major use of assessment—research and evaluation—focuses on collecting and analyzing data to generate knowledge for a wider audience. Academic researchers and/or institutional research officers collect student assessment data throughout or at the end of a course or program. They study these data using more sophisticated study designs and analysis methods than those typically applied for formative improvement purposes. Researchers who assess student learning outcomes ask such questions as, "How well are people and programs performing?" "What are the best practices and programs to implement for student success?" and "What generalizable knowledge can be developed to share with others?" Such research and evaluation can provide the faculty, staff, and students of colleges and universities with vital information to inform the design or selection and implementation of new strategies and programs that support students' success in higher education.

Although assessment of intra- and interpersonal competencies for research and evaluation purposes is growing, the available research to date is limited, as noted in Chapter 2, and there have been calls for greater rigor. In a review of research on co-curricular interventions to develop motivation and other competencies thought to support and retain students in science, technology, engineering, and mathematics (STEM), for example, Estrada (2013) calls for the use of stronger research designs. Estrada suggests use of longitudinal designs and, when possible, randomized controlled trials to investigate the relationship between interventions focused on these competencies and student outcomes (e.g., cumulative grade point averages [GPAs], graduation rates). Further, as noted in Chapter 2 and as will also be reinforced in Chapter 5, additional research is needed to demonstrate more clearly which intra- and interpersonal competencies contribute most strongly to students' persistence and success in college. And as noted in Chapter 3, further research and development of assessments also is needed to define these competencies more clearly and measure them more accurately and to guide assessment users in drawing valid inferences from the resulting data.

Accountability

The committee defines accountability as a summative process through which a person, program, or institution is judged against some standard in a way that is comparable across individuals, programs, and/or institutions (Ewell, 2008; Suskie and Banta, 2009). In higher education, accountability processes often are developed and conducted by central bodies (such as accrediting agencies or federal or state government agencies) for such purposes as benchmarking of relative institutional performance; accreditation or certification; and decision making about such matters as rewards, sanctions, and funding (Alexander, 2000; Bender and Schuh, 2002; Burke, 2005; Dougherty and Hong, 2005). Many states, for example, have adopted performance-based accountability systems under which they allocate a portion of higher education funding to each institution based on measures of its success on such outcomes as graduation and retention (Dougherty and Reddy, 2011).

Most accountability occurs at the institutional level. Colleges and universities, for instance, are accountable to regional accreditation boards (e.g., the Middle States Commission on Higher Education) that are approved by the U.S. Department of Education (2016) to measure and

approve institutional quality. The accreditation process is high-stakes, used in determining whether students attending the institution may receive federal financial aid. According to the Middle States Commission on Higher Education (2015), Standard 5 for Accreditation,

> Assessment of student learning and achievement demonstrates that the institution's students have accomplished educational goals consistent with their program of study, degree level, the institution's mission, and appropriate expectations for institutions of higher education (p. 10).

In the K-12 context, state educational content standards and performance benchmarks guide the interpretation of student assessment data. In higher education, by contrast, accrediting agencies allow institutions to self-define their own standards for student learning based on their unique missions. The Middle States Commission on Higher Education (2015), for example, requires institutions to demonstrate

> consideration and use of assessment results for the improvement of educational effectiveness…[for such uses as] improving key indicators of student success, such as retention, graduation, transfer, and placement rates (p. 10).

Disciplinary accrediting bodies also use assessment for accountability. The Accrediting Board for Engineering and Technology (ABET), for example, has established student learning outcomes as part of its accreditation process. In engineering, programs are required to demonstrate that students have "an ability to function on multi-disciplinary teams" (Accreditation Board for Engineering and Technology, 2015, p. 3). This requirement has spurred research and development of assessments of teamwork competencies, as discussed in Chapter 5.

An individual college or university may also hold individual faculty members or counselors accountable for the progress of students under their care. This accountability may take the form of monitoring the extent to which an educator's students or advisees pass courses, make regular progress toward a degree, or even improve specific competencies—although such approaches can encourage faculty resistance to assessment (Kuh et al., 2014; see further discussion below).

For students, accountability may take the form of graduation standards that require certain levels of credits, achievement, and performance on assessments of intra- and interpersonal competencies. In response to the ABET accreditation requirements, some undergraduate engineering students are currently receiving grades and course credit based on assessments of such competencies as ethics and teamwork (see Chapter 5 for further discussion). This form of accountability is also in place at a few institutions that have targeted the development of intra- and interpersonal competencies as key goals. Returning to the example of Alverno College, it created a pervasive culture of improvement-oriented assessment on campus (including assessment of cognitive, intrapersonal, and interpersonal competencies) that was sufficient to fulfill accreditation goals (at the institution level) and graduation requirements (for individual students). Assessments of these competencies could be used more broadly by other colleges and universities to evaluate effectiveness at the institutional or individual level to the extent that the assessments accord with institutional missions and the quality of their measures is adequate.

Stakeholders in higher education may focus on different cognitive, intrapersonal, or interpersonal competencies when planning to assess them for different purposes. Competencies assessed at the institutional level for accountability purposes need to reflect the broad mission and learning outcomes of the institution, but faculty members would place higher priority on assessing competencies that advance the learning goals of their specific courses (Suskie, 2009). Assessments of these competencies with respect to a single course may be idiosyncratic to that course, but institutional measures must be common to or at least possible to aggregate across courses, departments, and schools. The challenge of such aggregation of the assessment data for purposes of program or institutional accountability is a potential barrier to wider assessment of intra- and interpersonal competencies (Kuh et al., 2014).

An additional challenge to the use of intra- and interpersonal competency assessments for accountability is the possibility of unintended consequences from high-stakes assessments that may not take appropriate account of the experience of underrepresented student groups. When considering data from assessments of student engagement and effort as a measure of institutional effectiveness, for example, Dowd and colleagues (2011) observe that the assessments may unintentionally benefit campuses with greater percentages of racial majority students. Such assessments typically do not account for contextual pressures faced by racial/ethnic groups that are in the minority at predominantly white campuses. For example, institutions participating in the voluntary National Survey of Student Engagement (NSSE) ask students to report on their level of engagement in various "best practice" learning and personal development activities provided on campus. However, Dowd and colleagues (2011) note that

> engagement benchmarks are based on indicators of educational "best practices" without consideration of the racialized "bad practices" that minoritized students experience as harmful to their self-worth (p. 19).

In essence, underrepresented minority students who experience a hostile climate may have to exert "intercultural effort" that may detract from their ability to engage in educational "best practices" (Dowd et al., 2011). By contrast, Dowd and colleagues (2011) propose that measuring other constructs that account for the campus climate for students of color, such as sense of belonging, may better represent the educational experiences of such students and therefore reward institutions that serve and support them. In sum, any assessments of intra- and interpersonal competencies for accountability purposes need to pay particular attention to the equity concerns that may manifest among these constructs and in their measurement.

Evidentiary Demands of Different Assessment Uses

It is important to recognize differences among the evidentiary demands of the four purposes of assessment discussed above. As noted in Chapter 3, the higher the stakes associated with the use of an assessment (e.g., for purposes of selection or accountability), the stronger the evidence must be that the results are valid for that purpose (Borden and Young, 2008; Dougherty and Hong, 2005; McCormick and McClenney, 2012; Suskie, 2009). If the assessment results will influence important decisions that will have critical consequences for individuals or institutions, there must be strong evidence that the assessment supports valid, reliable, and fair inferences to inform decision making. The evidence standards may be lower for other purposes, such as formative improvement or placement, where the consequences are less serious, particularly when

multiple sources of evidence are in play and decisions are not as permanent or binding (see further discussion in Chapter 3). In a study of an assessment for admitted students, for example, Markle and colleagues (2013a) observe that the instrument's intended purposes—selection of students for additional support and intervention after admission—are relatively low-stakes.

Furthermore, as discussed in the previous chapter, threats to validity can vary with an assessment's purpose. For example, individuals may have little reason to fake their responses in a research setting, but a strong incentive to do so when taking a high-stakes test for college admission or other accountability purposes. Consequently, it is important to establish the validity of an assessment for each of its potential purposes. For formative improvement purposes, for example, evidence is needed to establish that a particular construct is relevant to improvement (e.g., sense of belonging or growth mindset matters in a particular context), that the construct is well measured (the assessment is valid, reliable, and fair), that the construct is malleable by institutions.

STAKEHOLDERS

A variety of higher education stakeholder groups potentially could benefit from access to high-quality data on competencies that are related to college success.[3] The focus here is on six different groups that could be interested in the results of assessments of the eight competencies identified in Chapter 2 or other competencies shown to be related to college success. The groups have different needs, priorities, and expectations: families, K-12 educators, college and university faculty, college admissions and student affairs staff, college and university leaders and administrators, and policy makers and state and national regulators. In each case, this section briefly describes their chief concerns and how information from competency assessments might be relevant to them.

Families

In the assessment context, students and their parents and relatives are concerned mainly about individual students' preparation for college (i.e., development of the necessary skills and competencies) and admission to college (i.e., identification of the right institutions, demonstration of the needed skills and competencies). After admission, this group also is concerned about retention, graduation, and future employment. The common focus of families in all of these cases is individual-level performance, that is, the status or progress of a particular individual. Their interests encompass both one-time events (selection and/or placement, getting good grades) and longer-term development and support. In the former case, if higher education institutions were able to clearly articulate a set of desired student competencies based on research demonstrating their malleability and relationship to college success, then measures of individual status would assist families and students in demonstrating competencies needed for successful admission to these institutions. In the latter case, assessment information might help families and students focus on and students develop the competencies they would need for admission and for their development and success in college.

Currently, many families support their students' participation in high school sports, clubs, and leadership activities partly because college admissions officers consider such activities to be

[3]For an example of K-12 assessment stakeholders, see http://www.cal.org/flad/tutorial/impact/5stakeholdersmap.html [July 2016].

important indicators of an applicant's intra- and interpersonal competencies.[4] Students also represent those competencies through statements of intent and biographical information submitted as part of their college applications. In an example of more structured assessment of one such competency, Harackiewicz and colleagues (2012) conducted an experimental study of high school students and their families in which parents were provided a brochure and access to a website outlining the importance of STEM education. Students of the group of parents receiving the information enrolled, on average, in about one STEM class more in their last 2 years of high school relative to the students whose parents did not receive the information. In this example, the parents supported their students' perception of the "utility value" of STEM (i.e., their belief that science and mathematics are useful in everyday life and for a variety of careers) as a way to encourage their interest and persistence in STEM. The researchers assessed the participating students' perception of utility value through a survey asking such questions as "In general, how useful is what you learned in math classes?" If assessments of this or other competencies were widely available, families might welcome the resulting information for use in college preparation and admission. Families also could use this information about their students' intra- and/or interpersonal competencies in evaluating choices among institutions, selecting degree programs and specific courses and instructors, enhancing academic performance, identifying strategies that lead to persistence, and focusing on competencies that may lead to career and life success after graduation.

K-12 Educators

K-12 teachers, counselors, and school and school district administrators may be interested in information about students' intra- and interpersonal competencies at both at the individual and aggregate levels. They may use this information to inform both short-cycle individual actions (e.g., guiding students in college selection) and longer-cycle programmatic changes, such as programs or strategies for helping students improve on these competencies. These stakeholders can use assessment information specifically to focus K-12 programs on (1) developing cognitive, intrapersonal, and interpersonal competencies; (2) improving the efficacy of their college and career counseling efforts; (3) identifying competencies that need further development in individual students or student groups; and (4) documenting the value of their diplomas in preparing students for future college and workforce success (Bialik et al., 2016; Dilley et al., 2015; Plucker et al., 2015; Trilling and Fadel, 2009). The growing number of schools that are currently focusing on "21st-century" competencies are already using assessments of these competencies for all four of these purposes, as demonstrated in a recent review of schools identified by the Partnership for 21st Century Learning as "exemplar schools" (Brown, 2014). In another example, a federally funded program operating across school districts in southern Texas used the ACT Engage assessment, described earlier in this chapter, to identify traditionally underrepresented students low in the competencies thought to be related to college success for the purpose of targeting extra support to enhance their college readiness.

College and University Faculty

[4]For example, see https://bigfuture.collegeboard.org/get-started/outside-the-classroom/extracurriculars-matter-to-you-and-to-colleges [July 2016].

College and university faculty members are responsible for delivering and improving the education provided to students, both individually (through advising and mentoring) and in the aggregate (through teaching and program and course design). Thus, they are concerned both about formative improvement within ongoing courses and research and evaluation that can inform more fundamental redesign of teaching and curriculum materials. They could potentially use competency assessment data to help design activities to foster the development of intra- and interpersonal competencies in the context of academic disciplines, to improve the efficacy of mentoring and advising efforts, to monitor and improve program design, and to evaluate and enhance instructional strategies and activities. For example, faculty members have used assessments of mathematics self-efficacy to understand how teaching and learning activities can be redesigned to foster self-efficacy and potentially enhance achievement (e.g., Hall and Ponton, 2005; Peters and Hortecamp, 2010). In another example of potential formative use, a professor could choose a competency that is relevant to a specific course, assess students' level of that competency at the beginning of the course, and plan for how to use the assessment data to modify the teaching approaches used in the course later in the semester to enhance that competency (Suskie, 2009). In addition, based on the suggestive evidence that academic self-efficacy and positive future self are related to college success (see Chapter 2), a faculty member might assess science self-efficacy and identity (i.e., imagining one's future self as a scientist) for his or her advisees or students and then target mentoring activities or research experiences as a way to increase these competencies, with the goal of retaining the students in science fields. This potential use of assessment data is suggested by research demonstrating that mentoring and research experiences can foster students' feelings of self-efficacy in science, identification with science, and commitment to science careers (Chemers et al., 2011; Estrada et al., 2011).

Faculty members also may assess these competencies for program and student accountability purposes. In response to the ABET accreditation requirement that programs demonstrate students' attainment of intra- and interpersonal competencies including ethics and teamwork, engineering administrators and faculty members have begun to assess these competencies (Lattuca et al., 2006; see Use Case 3 later in this chapter). Some undergraduate engineering faculty already use data from assessments of ethical reasoning to assign student grades and award academic credit (National Academies of Sciences, Engineering, and Medicine, 2016b; see Chapter 5 for further discussion).

Based on the national survey of provosts described above, Kuh and colleagues (2014) concluded that more faculty involvement is essential to sustain progress in assessment of student learning outcomes (including intra- and interpersonal outcomes). They found agreement among administrators, rank-and-file faculty members, and assessment scholars that faculty engagement in implementing assessment and interpreting the resulting data is essential to improve teaching and learning and enhance institutional effectiveness.

College Admissions and Student Affairs Staff

College admissions and student affairs staff focus their efforts on individual students or applicants, both in making important admission and placement decisions and in providing ongoing individual support and development. Student affairs staff also may be concerned with programmatic decisions that affect all students, selected subgroups of students (e.g., freshmen, at-risk students), or self-selected student groups (e.g., interest groups, student government representatives). As noted earlier in this chapter, both groups already are using assessments of

various intra- and interpersonal competencies. At Oregon State University, for example, admissions staff members consider data on intra- and interpersonal competencies when selecting students for admission, while student affairs staff use the data for academic advising, student services, and on and off campus referrals (Sedlacek, 2011).

College admissions staff could use data from assessments of competencies clearly related to college success for purposes of selection or gatekeeping, while student affairs staff could use the data primarily for improvement and research and evaluation purposes. Although central administrators, admissions personnel, and faculty could clearly benefit from using such data to guide and improve college programs and instructional approaches, many may need to be convinced about the efficacy of these competencies in increasing college students' success.

Student affairs officers in colleges and universities already deal extensively with students' development of intra- and interpersonal competencies and could use high-quality assessment data for purposes of program evaluation, design of support services, and related activities. As noted above, a few universities are already administering tests of self-efficacy, motivation, and related competencies to incoming students in order to select at-risk students for intervention (Fain, 2015). Illustrating the potential value of such efforts, Lent and colleagues (2003) studied 328 students in introductory engineering courses and found that creating an environment that provides support and removes barriers is associated with higher self-efficacy and (indirectly) with increased intent to persist in engineering.

College and University Leaders and Administrators

College administrators, such as presidents, deans, and department heads, are responsible for delivering and improving the education provided to students in the aggregate. Whereas college faculty care about short-cycle improvements in ongoing courses and course redesign to enhance student development and content knowledge, administrators are likely to be more concerned with long-term development of students' intra- and interpersonal competencies and the relationship of these competencies to retention, graduation, and success in careers in the aggregate.[5] In addition, assessment information on students' intra- and interpersonal competencies could be useful to administrators for the evaluation and improvement of degree programs, courses, instructors, co-curricular activities, and equity. Administrators at the University of Nevada, Las Vegas, for example, used data from the Multi-institutional Study of Leadership to examine student perceptions of the campus climate by race, and then used the results to develop new student programs designed to support positive and reduce negative perceptions (Early and Blevins, n.d.).

Policy Makers and State and National Regulators

Policy makers with potential interest in using data on students' intra- and interpersonal competencies include legislators, boards of education, accreditation boards, and disciplinary societies. Although this stakeholder category is quite diverse, all of these groups desire information that about the quality of institutions of higher education. They ask such questions

[5]Although evidence of differing administrator and faculty perspectives on assessment of intra- and interpersonal competencies is limited, a substantial literature exists on the differing perspectives of these two groups in general and with respect to other aspects of higher education, such as technology use (Campbell and Slaughter, 1999; Ehrenberg, 1999; Palm, 2006; Seidman, 1985; Stark et al., 1997).

about these institutions as "Are the academic programs 'good enough'?" "Are student graduation rates at expected levels?" "What can be done to improve quality (e.g., how can courses or programs be improved)?" and "What factors influence student success in the aggregate and for specific groups, and how are departments and programs working to address these factors?"

As noted above, accreditors are charged with ensuring institutional quality (U.S. Department of Education, 2016), and as part of this process, some are beginning to demand that colleges demonstrate that their students are developing competencies beyond those that are purely academic (see, for example, Spurlin et al., 2008 and Appendix B). In fact, the primary driver of additional assessment in higher education has been demands from accreditation agencies (Kuh et al., 2014). Further, state systems, accreditors, and professional associations and related consortia may be able to use measures of intra- and interpersonal competencies to improve institutions and advance system goals.

Examples of Uses of Assessment Data by Specific Stakeholders

Table 4-1 makes explicit the kinds of questions that different stakeholders could address through different uses of data on intra- and interpersonal competencies. This table is intended to illustrate but not to limit the range of possibilities for the use of such data. Note also that, although the table is organized by stakeholder group, the committee argues below that the use of data for institutional change and improvement is likely to be most effective when different stakeholders act in concert across levels, drawing on multiple sources of assessment data (Chatterji, 2005; Dowd and Tong, 2007).

TABLE 4-1 Example Questions about Intra- and Interpersonal Competencies That Are Relevant to Different Stakeholders and Different Purposes

Stakeholders (Users)	Purpose of Assessment (Uses)			
	Selection and Placement	Formative Improvement	Research and Evaluation	Accountability
Students/ Families	(Parents) To what extent does my child have the necessary competencies for success at this particular college/in this major?	(Students) How well am I improving my intra- and interpersonal competencies during college, and what do I need to continue to improve upon? What adjustments can I make in my intra- and interpersonal competencies so that I can succeed in college?	(Researchers) How does family support during college influence students' development of intra- and interpersonal competencies?	(Students) Are there ways to document my learning of intra- and interpersonal competencies for job search portfolios and graduate school applications?
K-12 Educators	(High school counselors) To what extent does this student have the intra- and interpersonal competencies necessary for college success? Do all of our students have access to programming that fosters the development of intra- and interpersonal competencies?	(Grade 12 teachers) What can I learn about my students' intra- and interpersonal competencies at the beginning of the year that will enable me to help them develop those competencies by graduation?	(Teachers) Does a specific curricular intervention support students in developing the intra- and interpersonal competencies necessary for college success?	(K-12 Administrators) What level of intra- and interpersonal competencies do students need to develop during high school in order to be successful in college? How can we document that students from our school achieve this level of competency?

College/ University Faculty	To what extent do students have the intra- and interpersonal competencies necessary for success in my course?	How are my students improving their intra- and interpersonal competencies in addition to learning course content in my class? If my students are not improving in these competencies by midsemester, can I help them in new ways to improve by the end of the term?	Does explicit teaching of a certain competency lead to improved outcomes for underrepresented students in my course? Does explicit teaching of a certain competency lead to student success (grades, completion, content learning) in my course?	How can I document students' change in these competencies during the course for the purposes of program accreditation? To what extent does my course contribute to the broader goals of my unit (program, school, college) regarding these competencies? Does attention to intra- and interpersonal competencies in my instruction improve the degree to which underrepresented students achieve program goals?
College/ University Administrators	What proportion of our students are deficient in these intra- and interpersonal competencies when they arrive on campus? What proportion of our underrepresented students are deficient in these competencies when they matriculate?	What extra support can we provide to students who are lower on these intra- and interpersonal competencies to help ensure that they are retained at and graduate from this college?	Does proficiency in certain intra- and interpersonal competencies lead to student success at our institution? Does proficiency in certain intra- and interpersonal competencies lead to improved outcomes for underrepresented student groups attending our institution?	To what extent do students improve in these competencies during their undergraduate career? Can we document this improvement for the purposes of institutional accreditation?
College Admissions and Student Affairs Staff	(Admissions Staff) To what extent do incoming students have the necessary intra- and interpersonal competencies for success at this particular college?	(Student Affairs Staff) Can resident assistants provide extra support for students who are lower in certain intra- and interpersonal competencies?	(Student Affairs Researchers) Which student affairs interventions are effective in helping students improve upon these competencies across institutional contexts?	To what extent does the student affairs division contribute to students' gaining intra- and interpersonal competencies that could be documented as meeting accreditation standards?

		What aspects of resident assistant support are most helpful for creating a sense of belonging for underrepresented student groups?		Do all students have access to programming designed to improve intra- and interpersonal competencies? To what extent are all students accessing these interventions?
Policy makers and Regulators	Not applicable	Do institutions have the necessary resources to assess students' intra- and interpersonal competencies throughout the student experience? Do institutions have the necessary resources to address any deficiencies in students' intra- and interpersonal competencies that are identified during the semester?	(State Consortia) How do state systems compare in terms of how students fare in these competencies? (Disciplinary Accreditors) Does proficiency in certain competencies lead to increased student success in disciplinary outcomes? (State Consortia) Are some state systems better than others at fostering specific intra- and interpersonal competencies among underrepresented student groups? If so, are they accomplishing this in replicable ways?	(Regional Accreditors) Can we include assessments of intra- and interpersonal competencies as a way for colleges to demonstrate that their students have developed proficiency in certain areas prior to graduation? (Disciplinary Accreditors) What is the level of proficiency in intra- and interpersonal competencies that students in our discipline achieve prior to graduation? Is this level sufficient for the skills needed in related careers?

SOURCE: Created by the committee.

Moreover, while this report focuses on the development and assessment of student competencies, students are not the only appropriate focus of assessment, evaluation, and change. Rather, intra- and interpersonal competencies develop and exist in interaction with local context that may support or diminish them in various ways. It is critical, for example, that programs and strategies designed to support students' sense of belonging take account of the college context and climate. In essence, it is necessary not only to assess how students feel about the climate (whether they feel they belong), but also to evaluate whether the environment poses structural or normative barriers that inhibit a sense of belonging, particularly for underrepresented minority students (Hurtado et al., 1998). Similarly, the intra- and interpersonal competencies of faculty members, counselors, instructors, and other college staff influence students' success. Data from assessments of these practitioners' competencies could complement student data to further understanding of the relationship between the two and support the formulation of plans for improvement (Bensimon, 2007; Dowd, 2015).

Many researchers argue that uses of assessment and evaluation data for formative improvement are most important for improving college coursework and programs and increasing student success (Suskie and Banta, 2009). The third column of Table 4-1 gives examples of the kinds of formative questions that could be answered by specific stakeholders if they had access to assessment data on students' intra- and interpersonal competencies. The top of this column poses improvement questions that an individual student might ask (e.g., "How well am I improving my intra- and interpersonal competencies during college?"). The answers to these questions might lead the student to seek out counseling or advice from faculty or other mentors, participate in support services or extracurricular activities, consider new career options, or take other actions to meet his or her individual goals.

The middle rows of Table 4-1 illustrate the kinds of improvement questions that faculty and staff in different roles could answer with data from intra- and interpersonal competency assessments. For example, student affairs staff might use the data to determine the need for special supports or services, to focus on interventions that can strengthen specific competencies, or to assess whether existing strategies are having their intended effects or need improvement. Faculty, working alone or in disciplinary departments, might use the data to improve instructional strategies by integrating competency development into coursework. In general, policy makers may be less concerned with specific improvement decisions with respect to individual courses or strategies within their institutions, but their policies—for example, investment of available resources, incentives—can directly influence faculty and staff engagement in and commitment to formative improvement. Policy makers' direct concerns are more salient in other rows of the table, such as in taking action through policy and practice to ensure that their institutions are accredited.

CREATING IMPROVEMENT: VISION AND REALITY

To this point, this chapter has described the intersection of assessment uses and stakeholders as important potential contexts for assessment practices that can support student success in higher education. The discussion has emphasized the distinctions among these contexts and provided illustrative questions to illustrate how the assessment of specific intra- or interpersonal competencies can provide important information to people (such as students and faculty) who need that information for specific purposes (such as course planning, course

selection, counseling, and accreditation) and ultimately to improve higher education and STEM retention and success.

Using Assessment Data to Serve Multiple Purposes

As noted above, assessments of intra- and interpersonal competencies may be able to serve the needs of both internal and external stakeholders simultaneously. Assessments undertaken for accountability purposes—to document success to external stakeholders, such as accreditors—can generate useful information that internal institutional stakeholders can apply to guide improvement efforts. At the K-12 level, for example, new federal legislation (the Every Student Succeeds Act, Public Law 114-95) highlights the role of intra- and interpersonal competencies as indicators of school quality in accountability systems. Theorists contend that, to the extent that improving these competencies is a valued outcome of elementary and secondary education, using results from assessments of these competencies as an indicator of school quality can lead directly to school improvement (e.g., Darling-Hammond et al., 2014).

At the higher education level, institutions' use of assessment data for purposes of internal institutional improvement can support accreditation and accountability functions. Institutions that have used assessment data internally to strengthen teaching approaches and student learning outcomes have then used these improvements to document for accreditors that they focus sufficiently on learning for accreditation (Mentkowski et al., 2000). Institutions that effectively weave assessment of learning into their institutional culture begin with using assessment for improvement and end with using it for accreditation.

Moreover, external accountability pressures on institutions of higher education can fuel internal improvement processes. The learning outcomes assessment movement, which evolved in response to state and federal accountability demands (Campbell, 2015), provides an example. As discussed in Appendix B, this movement has incorporated intra- and interpersonal competencies along with academic skills and knowledge as valued goals of higher education (Association of American Colleges and Universities, 2007). Likewise, the ABET accreditation requirement that undergraduate engineering programs demonstrate students' acquisition of teamwork, ethics, and other intra- and interpersonal competencies has catalyzed an array of improvement efforts in instruction, curriculum design, internship programs, and assessment (e.g., National Academies of Sciences, Engineering, and Medicine, 2016b; Lattuca et al., 2006).

Despite the possibilities for cross-fertilization between the improvement and accountability paradigms, however, research suggests that the improvement paradigm results in broader buy-in from institutions of higher education and is more successful in promoting change relative to externally imposed accountability (Blaich and Wise, 2010; Ewell, 2008; Kuh et al., 2014; Suskie and Banta, 2009). Ewell (2008) and Blaich and Wise (2010) warn of difficulties in using data derived from an accountability framework for institutional change. For example, externally imposed assessment processes (such as accreditation and system-level accountability efforts) often incur lower faculty investment in understanding or applying the resulting data as compared with processes focused on improvement. Change processes that are externally imposed also tend to focus on standardization across contexts instead of situating assessments within individual institutions, departments, and courses, which have their own educational goals and norms that are important for creating change on college campuses. In essence, it may be more efficacious to use improvement data to document a culture of evidence-based improvement for

accountability than to use accountability data to garner the buy-in necessary for systematic improvement in a college setting.

Barriers to Use of Assessment Data for Improvement

The committee found a number of examples of institutions using student assessment data for improvement purposes, but the prevailing reality is less cogent. In working with dozens of colleges and universities that participated in the Wabash National Study of Liberal Arts Education, Blaich and Wise (2010) observed that "although many campuses gather mounds of evidence, few use it to get better at promoting student learning" (p. 77). Kuh and colleagues (2014, p. 4) similarly observe that "although more assessment evidence is now available, its use is not nearly as pervasive as it must be to guide institutional actions toward improving student outcomes."

One barrier to the use of assessment data to support improvement is that some faculty, staff, and administrators are unfamiliar with basic principles of educational assessment. They may lack access to and familiarity with high-quality data systems, and they have little experience with using assessment data—or other types of data—to inform and improve teaching and student support practices. This group of stakeholders may be sophisticated in other forms of research but less familiar with the contexts for use of assessment data in higher education. As a result, they will need a great deal of support, including training in how to interpret and use the data, when provided with assessment data on students' competencies. This challenge is illustrated by a recent national survey in which provosts at 2- and 4-year institutions were asked about assessment of student learning outcomes (Kuh et al., 2014). When asked how the use of assessment could be advanced at their institutions, 64 percent called for more professional development for faculty, 63 percent said they wished that more faculty were using assessment results, and 56 percent responded that additional financial or staff resources were key. One provost at a master's degree-granting institution commented (Kuh et al., 2014, p. 28): "Many faculty struggle with determining how to conduct a proper assessment and then how to use the results." Additionally, some stakeholders may be resistant to the use of data from assessments of students' competencies given norms, assumptions, and values about assessment (Blaich and Wise, 2010). Provosts reported, for example, that progress on assessment of student learning outcomes may be slowed by faculty members' worry that the results will be used in performance reviews (Kuh et al., 2014).

That different stakeholders have different information needs and may be interested in different constructs, measurement methods, and questions (Ewell, 2008, Campbell, 2015) also can be a barrier to effective use of assessment data. In efforts to promote institutional improvement, it is rarely the case that one stakeholder group can opt for one assessment for one purpose. Instead, the incorporation of any new assessments, including those of intra- and interpersonal competencies, will have to be negotiated and coordinated among many stakeholders, each with different interests and priorities. Higher education leaders and administrators will likely need to work with faculty, student affairs staff, and institutional research and assessment experts collectively to (1) identify the competencies that are most germane to the overall institutional mission at large, (2) agree on measures that will be acceptable to all stakeholders in terms of quality and practicality, (3) collect the data, and then (4) allocate resources and develop processes for changing practices across curricula and support structures (Dowd and Tong, 2007). Institutions will need to plan ahead to build campus support

and capacity across multiple stakeholders (administrative leaders, faculty, student affairs staff, and students) (Blaich and Wise, 2010). Focusing specifically on assessment of intra- and interpersonal competencies among entering undergraduates, Allen and colleagues (2009) call for such an approach. They argue that the value of measuring these student competencies depends on the effectiveness of institutional support programs for high-risk students and recommend research into integrated systems for identifying and intervening with such students.

Supporting Use of Assessment for Institutional Improvement

The committee's review of research in higher education and organizational change indicates that any change in higher education, including increased use of assessments (whether of one or more of the eight competencies identified in Chapter 2 or other cognitive, intrapersonal, or interpersonal competencies), requires careful planning and consideration of multiple factors, as discussed below.

Motivating Stakeholders to Assess Intra- and Interpersonal Competencies

A robust literature base on organizational change culled from several disciplines (e.g., economics, sociology, psychology, anthropology, business, and education) suggests that organizations can play a role in the motivation of individuals. Further, there is a growing base of literature that examines organizational change, specifically, on college campuses (e.g., Bess and Dee, 2008; Birnbaum, 1989; Kezar, 2001). These organizational change and motivation theories may be relevant as institutions consider how to help faculty implement and use assessments of intra- and interpersonal competencies. Although this literature base is robust and varied, certain theories have been directly mirrored in what practitioners in higher education have done to foster assessment use on college campuses.

Vroom's (1964) landmark expectancy theory, drawn from management literature, proposes that incentives can motivate individual behaviors under certain conditions: when individuals can expect to achieve high performance (expectancy), when they can depend on a reward for achieving high performance (instrumentality), and when they care sufficiently about the reward (valence). Other economic theories describe how the use of financial incentives can motivate both desired and unanticipated behaviors if the incentives are tied to the wrong measures (Gibbons, 1998; Williamson, 1975, 1985). Beyond incentives, anthropological theories applied to organizations (broadly) and colleges (specifically) have found that institutional and disciplinary cultures play a strong role in individuals' adopting certain behaviors (Bess and Dee, 2008; Schein, 1992).

Certain strategies that practitioners have used to implement assessment practices successfully on different college campuses appear to mirror the organizational change and motivational theories. For example, Kuh and colleagues (2014) observe that faculty may not have sufficient understanding of assessment practices to expect a payoff for these behaviors (i.e., expectancy in Vroom's theory). Given that assessment practices often are not tied to reward structures in higher education, faculty also may not see instrumentality in practicing assessment. Thus if there were substantial evidence linking the eight intra- and interpersonal competencies described in this report to college success, college administrators could look to incentives, such as changing reward structures, to motivate faculty to adopt such assessment practices. However, survey by Kuh and colleagues (2014) also revealed that one barrier to faculty use of assessments

was faculty members' concern that assessment results would be linked to evaluation. Therefore, any incentives created (including use of assessment in reward structures) would need to be carefully balanced to address such faculty concerns.

Integrating Assessment with Institutional Culture

Beyond incentives, literature on higher education assessment suggests that a culture of assessment throughout the fabric of the institution may foster buy-in from faculty (Dowd and Tong, 2007; Dwyer et al., 2006;). Kuh and colleagues (2014) argue that "culture, climate, context, and language all matter deeply." Cultures (long-standing) and climates (more immediate) that emphasize innovation, improvement, and evidence-based decision making all contribute to the use of assessment results by faculty. For example, institutions might link assessment use to already established internal procedures. Given the competing priorities in faculty roles, if assessment practices are integrated into the work of the institution (for example, through teaching and learning centers) and integrated into the established commitments of faculty (for example, in curricular reform initiatives), faculty may view assessment practices as more integral and less burdensome. This literature and other research suggest that college and university leaders and administrators are unlikely to accomplish the goal of using assessment data to support long-term improvement without the active involvement and commitment of multiple stakeholder groups.

A burgeoning literature describing the conditions under which assessment processes lead to institutional improvement in higher education underscores the importance of cross-stakeholder coordination and collaboration. Although this literature is based on research on the use of assessments to measure cognitive competencies, its principal findings appear to be relevant to the assessment of intra- and interpersonal competencies as well. Effecting change in higher education requires acting on several institutional levels concurrently toward similar goals, triangulating multiple forms of data, and using results in feedback loops serving the needs of participating stakeholders (Dowd and Tong, 2007; Ewell, 2008).

The effective collaboration of multiple stakeholders, including students, appears to be key to the success of improvement efforts: when the values and goals of the institution and stakeholders are incorporated and when relevant stakeholders actively support the assessment and improvement efforts, greater institutional improvement is observed (Blaich and Wise, 2010; Cistone and Bashford, 2002; Dowd and Tong, 2007; Suskie, 2009). Local structures and processes that involve stakeholders in developing and agreeing on assessment goals, purposes, and procedures and that provide time for stakeholders to fully discuss the assessment results can help build ownership and understanding and combat resistance. The Brockford College example described later in this chapter illustrates this process. A multistakeholder committee took time to fully understand the theory behind prosocial leadership, select appropriate measurement instruments, and ultimately make changes to the institution's educational programming based on the assessment data.

The inclusion of practitioners in the process appears to be particularly important, as they are the ones who ultimately will be responsible for change in the institution or program (Dowd and Tong, 2007; Dowd et al., 2011; Middaugh, 2009; Welsh and Metcalf, 2003). To promote a culture of assessment use, especially for improvement, Baker (2012) recommends the following elements: assessment champions, central committees or other capacity to guide effective assessment and use of its results, disciplinary departments or programs that can serve as centers

of assessment excellence, and institutional support for improving faculty and staff assessment practices.

Some researchers have suggested that institutions conduct advance planning for how assessment data, once collected and analyzed, will translate into and support subsequent change and improvement. They also propose that institutions commit themselves fully to the goals of their assessment. Beyond simply disseminating assessment results, they argue, institutional leaders should set aside appropriate resources for the development of pathways that will facilitate discussion of the data and the formulation of options for actually using the data to change practices and improve student competencies (Blaich and Wise, 2010; Cistone and Bashford, 2002).

As noted earlier, a recent survey of the status of assessment in U.S. colleges and universities found increasing attention to the assessment of student learning outcomes (Kuh et al., 2014) and progress in supporting the use of assessment. Responding provosts cited as the most prevalent and important supports for assessment explicit institutional policy on assessing student learning, faculty engagement and involvement in assessment, and increased centralized capacity for assessment work. Relatively less prevalent were student participation in assessment activities and significant involvement of student affairs staff. Moreover, nearly two-thirds of the responding provosts identified as pressing needs more professional development in assessment for faculty and more faculty using assessment results. Case studies, for example, reveal the help faculty and staff may need in mapping the connections between their priority questions and available assessment data (Blaich and Wise, 2010).

USE CASES: EXAMPLES OF USING INTRA- AND INTERPERSONAL COMPETENCY ASSESSMENTS

The use cases presented in this section illustrate how a variety of institutions have taken a systematic approach to effective use of data from currently available assessments of intra- and interpersonal competencies for different purposes in local contexts. The use cases also offer ideas for potential additional uses of assessment data.

Use Case 1: Assessing Competencies for Placement and Intervention

The University of New Mexico is a flagship research university that enrolls many low-income, first-generation students, some of whom are underprepared for college. College leaders wanted to be able to determine which admitted students would need the most support to succeed and graduate. They recognized that, in addition to academic skills, a student's level of certain intra- and interpersonal competencies could be an important indicator of whether the student would be at risk for dropping out. To identify and provide support for the most at-risk students, administrators decided to supplement information from traditional academic measures (high school GPA, SAT, ACT) with SuccessNavigator, an ETS test that measures four clusters of academic and intra- and interpersonal competencies: tools and strategies for academic success, commitment, self-management, and social support. As noted in Chapter 3, the instrument measures a few of the competencies identified in Chapter 2, along with a range of other competencies. For example, the "tools and strategies" cluster includes measures of organization, a behavior related to conscientiousness, and "self-management" includes measures of academic self-efficacy. Correlational analyses of data gathered from 2- and 4-year institutions showed that

test scores predicted GPA, persistence, and course grades, even after controlling for standardized test scores and high school GPA (Markle et al., 2013a).

In fall 2015, University of New Mexico administrators required all incoming students who were first-generation, scholarship recipients, enrolling in STEM programs, and athletes (approximately 1,500 in total) to take the test (Fain, 2015). The university decided to include athletes because they must balance the time demands of training and competition with their academic work, with high stakes attached—the possibility of losing their scholarship if they fall behind in their classes. The office of student affairs used the test results to offer students identified as at risk extra academic supports, such as regular meetings with tutors or academic advisers. The student affairs office also shared the test results with advisers for their use in recommending specific courses for particular students and developing and managing the "success plans" that some students were required to create. Based on individual students' assessment scores, student affairs staff required some students to meet monthly with their advisers to review their success plans and others to meet with their tutors regularly to check on their grades (Fain, 2015).

This use case demonstrates the connection among administrative priorities, institutional context, and the role of intra- and interpersonal competencies within broader student-success initiatives. The University of New Mexico's mission to serve a diverse student body with varying degrees of preparation for college provides the backdrop for how the assessment was used—in this case, for student risk assessment and selection for special support and advising programs, with the goal of improving retention and graduation rates. Extending beyond that use, the University of New Mexico or other universities could use this or a similar test for another purpose—evaluation. In this case, the university would randomly assign new students to an experimental or a control group. The experimental group would take the test and be assigned to special support and advising services, while the control group would not take the test or be assigned to these services. After the first semester or the first year, researchers would gather evidence of success for both groups (e.g., GPA, fraction returning for the second semester or year) and analyze the results to evaluate the effectiveness of the intervention (including the use of the assessment).

In a related example, Iowa Western Community College used the same test to inform placement of students in developmental courses (Fain, 2015). Although 80 percent of Iowa Western students were placed in developmental math classes, college leaders hypothesized that students with strong motivation and related intrapersonal competencies might succeed in college-level math courses despite their low placement test scores. Using this assessment to complement placement tests and high school GPAs allowed the college to bypass remediation for these students (Fain, 2015). Similarly, an observational study of 3,647 students at four campuses within a large, urban community college system found no statistically significant difference in passage rates for students who were placed in math courses based on their academic placement scores alone and those whose placement was accelerated based on their SuccessNavigator test scores (Rikoon et al., 2014). (Accelerated students were within 1 standard deviation below the cutoff score recommended for college-level course placement, and comparison students were those just above the cutoff score.)

Use Case 2: Accountability Driving Assessment for Improvement at Pennsylvania State University

External accountability pressures can sometimes catalyze the use of assessment data for improvement in courses and programs of study. In 1996, ABET introduced the new Engineering Criteria 2000 (EC2000), requiring undergraduate engineering programs to demonstrate students' progress toward specific learning outcomes, including intra- and interpersonal competencies (e.g., the ability to work in interdisciplinary teams). In 2002, the Pennsylvania State University (PSU) engineering college hired a team of researchers (Lattuca et al., 2006) to evaluate the effects of the new criteria on student learning outcomes and educational and organizational policies and practices. The authors used a pre-post design to gather information from 1994 program graduates (before the new criteria were in place) and 2004 graduates (after the new criteria were in place); they also gathered information from administrators, faculty, and employers using one-time surveys that asked about perceptions of change following implementation of the new criteria. Based on these various sources of evidence, the authors found that engineering programs placed greater emphasis on learning outcomes (also referred to as engineering professional skills) and active learning, rather than simply lecturing, than they had prior to the new accreditation requirements. Surveys also identified high levels of faculty support for continuous improvement. More than 75 percent of department chairs estimated that the majority of their faculty members supported continuous improvement efforts, and more than 60 percent of chairs reported moderate to strong support for the assessment of student learning. Faculty corroborated this finding: nearly 90 percent of the faculty respondents reported some personal effort in assessment, and more than half reported moderate to significant levels of personal effort. For the most part, moreover, faculty members did not perceive their assessment efforts to be overly burdensome; nearly 70 percent described their level of effort as "about right."

These changes in teaching practices and curriculum improvements appeared to influence student learning outcomes positively. Compared with their 1994 counterparts, and after taking differences in graduates' and institutional characteristics into account, 2004 graduates reported

- more active engagement in their own learning,
- more interaction with instructors,
- more instructor feedback on their work,
- more time spent studying abroad,
- more international travel,
- more involvement in engineering design competitions, and
- more emphasis in their programs on openness to diverse ideas and people.

Although they tended to be small, 7 of 10 statistically significant differences between pre- and post-EC2000 graduates persisted even after adjusting for an array of graduate and institutional characteristics.

Use Case 3: Assessing Prosocial Values for Improvement at the College at Brockport[6]

In 2009, the College at Brockport, a college in the State University of New York (SUNY) system, began discussing how to develop student leaders on campus. Brockport is a residential, public 4-year institution with approximately 7,000 students. College leaders convened a committee of faculty, students, and staff to develop a certificate program in leadership. To understand leadership on the campus, the committee decided to use the Social Change Model of leadership development, focusing extensively on prosocial goals and values. The committee then decided to collect data so they could better understand the educational practices that facilitated prosocial leadership development on this specific campus. As the basis for its assessment, the committee chose to participate in a national survey, the Multi-Institutional Study of Leadership, mentioned earlier, that uses the Social Change Model of leadership development.

Analysis of the survey results led administrators to believe that several activities, including community service, internships, mentoring relationships, and attendance at leadership conferences, facilitated the growth and development of student leadership on their campus. As a result, they developed a new leadership program structured around these existing activities and allocated appropriate resources to allow more students to engage in the new program, including the component activities. Beyond using the campus-specific survey data to develop the leadership program, they used the national data from this same survey to benchmark the program's success against that of other programs. They found that students involved in their program participated relatively more in "high-impact practices" in leadership development. The institution viewed this as a great success and integrated the leadership program into its strategic plan.

This use case demonstrates how leaders in higher education can use assessment data in multiple ways to inform multiple stakeholders for several interrelated purposes. The data collected were initially used largely for formative purposes, for improvement—specifically by the faculty and student affairs staff that were developing and administering the new leadership program. These stakeholders were interested primarily in data that could help them identify those conditions and educational opportunities that facilitated leadership development on Brockford's campus. To these individuals, the context of the campus mattered in understanding how the survey results would influence the development of specific supports for student leadership experiences. Secondarily, as noted above, central administrators used the data to benchmark the success of the program against that of other programs nationally in a summative way for integration into broader strategic plans. This benchmarking capability is a benefit of using assessment data that apply locally but link to a national sample. This case also highlights how colleges see intra- and interpersonal competencies themselves as important goals for students. In essence, Brockford did not assess prosocial leadership development only because it leads to graduation, but because it is of value for Brockford's students. This view resonates with the outcomes movement in higher education discussed in Chapter 5, which focuses on developing such competencies as teamwork, ethical responsibility, and leadership as critical outcomes for all college students.

[6]See http://leadershipstudy.net/reports-publications/#campus-spotlight-series [August 2016].

CONCLUSIONS AND RECOMMENDATIONS

Addressing its charge to prioritize the uses of assessments of intra- and interpersonal competencies, the committee reviewed research on how higher education institutions are using assessments of cognitive, intrapersonal, and interpersonal competencies. Based on this review, it identified four major uses in higher education:

- selection and placement of individual students;
- formative improvement of local educational processes, practices, and programs;
- research and evaluation supporting knowledge generation; and
- accountability.

Assessments of intra- and interpersonal competencies for these four purposes are carried out by a variety of stakeholders, including families, K-12 schools, faculty members, college administrators, accreditors, and state and federal policy makers. To understand how these stakeholders presently and potentially could use data resulting from these assessments, the committee reviewed relevant higher education literature and reports on current practice.

Assessment Processes Supporting Student Success

Individual stakeholders in higher education have differing needs for data resulting from assessments of intra- and interpersonal competencies at different levels of aggregation, depending on the immediacy of those needs, the purposes to be served by the data, and their assessment-related knowledge and skills (Blaich and Wise, 2010; Dowd and Tong, 2007; Ewell, 2008). These variations in uses of the data necessitate different measures, different levels of evidence, and different kinds of buy-in for the assessment process and its uses. It is important to consider these contextual aspects of the assessment process when implementing an intra- and interpersonal competency assessment in practice.

Conclusion: Assessments of intra- and interpersonal competencies in higher education are most valuable for supporting student success when their selection, design, analysis, and interpretation are guided by stakeholder information needs, intended uses, and users

Recommendation 11: Leaders in higher education should select, design, analyze, and interpret data from assessments of intra- and interpersonal competencies based on stakeholder information needs, intended uses, and users.

The research literature contains convincing evidence that institutions of higher education can benefit from using assessments for both institutional improvement and accountability purposes, and these uses can ultimately be mutually reinforcing. However, assessment processes that emphasize improvement tend to garner more institutional support, including faculty buy-in, relative to those emphasizing accountability (Dowd and Tong, 2007; Ewell, 2008). Indeed, some administrators are concerned that external accountability mandates may focus institutional

conversations about assessment on bottom-line compliance rather than institutional improvement, especially given limited assessment resources (Kuh et al., 2014). College stakeholders also tend to be more receptive to assessment processes when they are internally derived, sensitive to specific institutional and disciplinary contexts, and driven by a belief that the assessment can serve the goal of improving student learning outcomes. Therefore, institutional improvement requires planning for needed resources and putting systems in place to support moving assessments from data collection to improvement processes.

Conclusion: Assessments are more likely to be implemented and used by stakeholders to improve student success when they are motivated by internal institutional improvement purposes than when they are motivated by accountability purposes.

Research has highlighted the need for multiple stakeholders across levels (staff, students, faculty members, administrators) to work together in an assessment process to effect pervasive change on a college campus (Dowd and Tong, 2007). The example of Brockford College described above illustrates how assessment results can be used to catalyze an improvement process in which multiple stakeholders work together toward a shared goal (in this case, improving students' leadership abilities). Complementary local- and institutional-level applications of the assessment data made it possible to incorporate the leadership goals that motivated the process into broader institutional strategic initiatives, ensuring that the improvements realized were pervasive across the institution. In the University of New Mexico example, advisers and student affairs staff used assessment data individually with students to tailor support services, while central administrators saw the data as useful for broader strategic initiatives aimed at retaining diverse and underprepared students to graduation.

Conclusion: Assessments are more likely to contribute to student retention and completion if efforts to use their results involve stakeholders at multiple levels of the organization (e.g., student support services, faculty, diversity officers, administrators) as opposed to involving individual stakeholders acting alone.

Support for Stakeholders' Assessment Capacity

Administrators and faculty in institutions of higher education may not have specialized training or expertise in educational assessment with regard to instrument design and selection, test administration, data analysis, or the best uses of assessment data. Yet while some stakeholders on campus, such as institutional researchers and assessment experts, can help with educating the broader campus community about assessment, they may not be familiar enough with the specific issues involved in intra- and interpersonal competency assessment. Therefore, training targeted at specific stakeholders may be necessary for the full value of these assessments to be realized. In addition, although data on intra- and interpersonal competencies can potentially add substantial value to efforts to enhance the success of underrepresented groups, faculty may not be familiar with this particular use of the data.

Conclusion: Some stakeholders in higher education will require support and training to develop the knowledge and skills needed to select, use, and interpret

data from assessments to improve student success in higher education. Such training can also help stakeholders understand how these assessments can contribute to the success of underrepresented student groups in particular and how to engage stakeholders who are resistant to assessment in general.

Research Needs

Research has yielded preliminary evidence of the importance of the eight competencies identified in Chapter 2 to success in college, and case studies of the use of cognitive assessment data by institutions of higher education for purposes of institutional and instructional improvement also are widely available (e.g., Astin, 2012; Blaich and Wise, 2010; Borden and Young, 2008). By contrast, evidence on how data from assessments of the eight competencies or other intra- or interpersonal competencies can be used for these purposes is relatively sparse. As additional assessments of these competencies take place on college campuses, they may yield a more robust understanding of how such assessments can lead to improvement within specific institutional, disciplinary, and student contexts.

Conclusion: Limited evidence is available from an organizational science perspective on how stakeholders in higher education can use data on intra- and interpersonal competencies for improvement and evaluation purposes.

Recommendation 12: To broaden understanding of how assessments of intra- and interpersonal competencies can lead to greater student retention and success, institutions of higher education should study and report on their use of these assessments for improvement purposes (e.g., enhancing student support services, developing underrepresented students' sense of belonging, improving courses, identifying effective programs).

5

Intra- and Interpersonal Competencies as College Outcomes

Chapter 2 of this report focuses on competencies showing some evidence of a relationship to undergraduate persistence and success, as measured by such indicators as persistence from year to year, grade point average (GPA), and graduation. This chapter responds to the growing interest among higher education policy makers in assessing and developing intra- and interpersonal competencies to prepare students for professional employment and community and family life after college. It identifies and defines a set of such college outcomes, examines their relationship to college completion, and considers the opportunities and challenges related to assessing them.

INTRA- AND INTERPERSONAL COLLEGE OUTCOMES

As discussed in Chapter 4, demands for accountability and improvement in higher education have led many colleges, universities, and fields of study to identify specific learning outcomes that their graduates should achieve and to assess students' development of these outcomes. As part of this trend, higher education leaders have identified intra- and interpersonal competencies as important college outcomes.

Frameworks of College Outcomes

The current interest in learning outcomes builds on work done over the past two decades to expand the long-term goals of higher education to include capacities that graduates will need for success in life and work. The Accreditation Board for Engineering and Technology (ABET), the accrediting organization that oversees engineering, computer science, and engineering technology programs, pioneered these efforts in 1996 when it adopted the *Engineering Criteria 2000* (Accreditation Board for Engineering and Technology, 1996). Responding to employer concerns about graduates' lack of professional skills, ABET moved from basing its accreditation on inputs (e.g., courses offered, student supports) to basing it on learning outcomes (Lattuca et al., 2006). These criteria, which are still in effect today (Accreditation Board for Engineering and Technology, 2015), specify 11 learning outcomes and require programs to assess and demonstrate students' progress toward each. The outcomes include the following intra- and interpersonal competencies:

- an ability to function on multidisciplinary teams,
- an understanding of professional and ethical responsibility,

- an ability to communicate effectively, and
- a recognition of the need for and an ability to engage in lifelong learning.

Several other individuals and organizations have since developed frameworks of college outcomes. Oswald and colleagues (2004) analyzed themes across a wide range of college mission statements and stated institutional objectives to identify 12 dimensions of college success, organized around three high-level categories:

- intellectual behaviors;
- interpersonal behaviors (e.g., communicating and dealing effectively with others, multicultural appreciation); and
- intrapersonal behaviors (e.g., ethics, career orientation, perseverance).

Around the same time, in a separate effort, leaders from 15 community colleges agreed on an initial framework of student competencies for the knowledge economy that included a set of interpersonal skills and a set of personal skills (Miles and Wilson, 2004).

Building on and extending these efforts, in 2005 the Association of American Colleges and Universities (AAC&U) launched the Liberal Education and America's Promise (LEAP) initiative to define more clearly and promote a 21st-century liberal education for all students, in both 2- and 4-year institutions, regardless of their field of study. Based on dialogue with employers and colleges and universities, analysis of reports from the business community and accreditation agencies, and research on how people learn (National Research Council, 2000), the initiative developed a framework of 16 "essential learning outcomes." This framework is organized around four high-level dimensions (Association of American Colleges and Universities, 2007):

- knowledge of human cultures and the physical and natural world,
- intellectual and practical skills,
- personal and social responsibility, and
- integrative and applied learning.

As defined by the LEAP initiative, a 21st-century liberal education also includes three strategies to help students attain these learning outcomes: (1) high-impact educational practices (e.g., first-year programs, collaborative assignments, service learning); (2) authentic assessments (discussed further below); and (3) students' signature work (e.g., capstone, internship, field work) (Schneider, 2015b).

By incorporating both acquisition of content knowledge and skills and application of learning, the AAC&U (2007) vision of 21st-century higher education is similar to the definition of "21st-century competencies" in a prior study related to the present one (National Research Council, 2012b, pp. 5-6):

> Through deeper learning…the individual develops expertise in a particular domain of knowledge and/or performance. The product of deeper learning is transferable knowledge, including content knowledge in a domain and knowledge of how, why, and when to apply this knowledge to answer questions and solve

problems. We refer to this blend of both knowledge and skills as "21st-century competencies."

Markle and colleagues (2013b) synthesized several of these frameworks to identify seven critical domains of competence for college graduates. Most recently, the Lumina Foundation (2015) built on the AAC&U (Association of American Colleges and Universities, 2007) framework to develop the Degree Qualifications Profile, outlining what graduates should know and be able to do at the associate's, bachelor's, and master's levels. A beta version of that framework was released in 2011, and it was revised based on feedback from more than 400 2- and 4-year colleges and universities, four of the seven regional accrediting agencies, and several higher education associations. The framework is organized around five high-level dimensions similar to those in the AAC&U (Association of American Colleges and Universities, 2007) framework:

- specialized knowledge,
- broad and integrative knowledge,
- applied and collaborative learning,
- civic and global learning, and
- intellectual skills.

Identifying and Defining Key Outcomes

The committee reviewed all of these frameworks and other reports on the goals of higher education, searching for those intra- and inter-personal competencies that appeared most frequently. Through this process, the committee identified the following six competencies for college graduates:

- ethics,
- lifelong learning/career orientation,
- intercultural/diversity competence,
- civic engagement/citizenship,
- communication, and
- teamwork.

These six competencies are summarized within selected outcomes frameworks in Table 5-1, and they are defined briefly below.

[TABLE 5-1]

Ethics

The reports and frameworks examined by the committee offer several practical definitions of the ethics competencies to be developed by 2- and 4-year institutions (see Table B-1); the authors of these reports set aside the centuries of debate among philosophers and religious leaders about the meaning of ethics and how to promote ethical behavior. For example, AAC&U

(Association of American Colleges and Universities, 2007) includes "ethical reasoning and action" as an essential learning outcome within the "personal and social responsibility" dimension, whereas the Lumina Foundation (2015) include "ethical reasoning" within the "intellectual skills" dimension, as follows (p. 17):

> Ethical reasoning thus refers to the judicious and self-reflective application of ethical principles and codes of conduct resident in cultures, professions, occupations, economic behavior and social relationships to making decisions and taking action.

Disciplinary accrediting organizations view ethics as an important component of preparing students to work within the discipline. The guidelines of the American Chemical Society (2015), for example, state (p. 17):

> Ethics should be an intentional part of the instruction in a chemistry program. Students should be trained in the responsible treatment of data, proper citation of others' work, and the standards related to plagiarism and the publication of scientific results.

Lifelong Learning/Career Orientation

The AAC&U's (Association of American Colleges and Universities, 2007) personal and social responsibility dimension includes "foundations and skills for lifelong learning," whereas Oswald and colleagues' (2004) intrapersonal dimension of college success includes "career orientation." The League for Innovation identifies "learning to learn" as a key outcome for community college graduates in the knowledge economy (Miles and Wilson, 2004). The ABET accreditation criteria for undergraduate engineering programs include "ability to engage in lifelong learning." In a similar vein, Markle and colleagues' (2013b) synthesis of frameworks of higher education outcomes identifies "self-directed learning" as a key competency within the "life skills" category.

Intercultural/Diversity Competence

The frameworks and reports examined by the committee use a cluster of related terms to refer to intercultural competence (see Table B-1). AAC&U (Association of American Colleges and Universities, 2007), for example, identifies "intercultural competence and knowledge" as an essential learning outcome, while Oswald and colleagues (2004) identify "multicultural tolerance and appreciation" as one of the 12 dimensions of college success, falling within the "interpersonal behaviors" category. Oswald and colleagues (2004, p. 189) define multicultural interpersonal behaviors as follows: "Showing openness, tolerance, and interest in a diversity of individuals (e.g., by culture, ethnicity, or gender). Actively participating in, contributing to, and influencing a multicultural environment." Within the "intellectual skills" dimension of the Lumina Foundation (2015) framework, engaging diverse perspectives is identified as a key outcome for 2- and 4-year graduates (see Table B-1), with a note that it is also relevant to two other dimensions—"applied and collaborative learning" and "civic and global learning." Finally, the Educational Testing Service (ETS) is currently conducting research to define more clearly

"intercultural competency and diversity" and to develop an assessment framework for this competency (Griffith et al., 2016).

Civic Engagement/Citizenship

The frameworks reviewed by the committee use various terms for civic engagement and citizenship, including "civic knowledge and engagement—local and global"; "social responsibility, citizenship, and involvement"; and simply "citizenship" (see Table B-1). Adding to these earlier, brief definitions, Torney-Purta and colleagues (2015) conducted an extensive review of existing frameworks, definitions, and assessments of civic-related constructs in higher education to develop a more comprehensive framework for assessing this complex competency. The proposed framework divides civic learning into two broad domains—civic competency and civic engagement—each of which includes the dimensions of civic competency (civic knowledge, analytic skills, and participatory and involvement skills); and civic engagement (motivations, attitudes and efficacy, and democratic norms).

Communication

Most frameworks of college learning outcomes include oral and written communication. Research suggests that this competency involves both cognitive and interpersonal skills as the individual receives and interprets messages from others and formulates appropriate responses (Levy and Murnane, 2004; National Research Council, 2012b). The frameworks reviewed by the committee highlight various dimensions of communication as critical for 2- and 4-year graduates, including "oral communication" (Association of American Colleges and Universities, 2007) and "the ability to communicate effectively" (Accreditation Board for Engineering and Technology, 2015). Miles and Wilson (2004) define communication skills simply as reading, writing, speaking, and listening, whereas the American Chemical Society (2015, p. 17) states:

> Effective communication is vital to all professional chemists. Speech and English composition courses alone rarely give students sufficient experience in oral and written communication of technical information. The chemistry curriculum should include critically evaluated writing and speaking opportunities so students learn to present information in a clear and organized manner, write well-organized and concise reports in a scientifically appropriate style, and use relevant technology in their communications. Because chemistry is a global enterprise, knowledge of one or more foreign languages or an international experience can be a valuable asset to chemistry students and add greatly to a student's ability to communicate with other chemists worldwide.

The Degree Qualifications Profile of the Lumina Foundation (2015) describes "communicative fluency" as follows (p. 18):

> The use of messages to achieve shared understanding of meaning depends on effective use of language, intentional engagement of audience, cogent and coherent iteration and negotiation with others, and skillful translation across

multiple expressive modes and formulations, including digital strategies and platforms.

Teamwork

Teamwork, a complex competency involving communication skills, is frequently identified as a critical outcome for college graduates. The AAC&U (Association of American Colleges and Universities, 2014) teamwork rubric defines teamwork as encompassing five types of behaviors under the control of the individual team members: (1) contributes to team meetings, (2) facilitates the contributions of team members, (3) individual contributions outside team meetings, (4) fosters constructive team climate, and (5) responds to conflict. The rubric describes four levels of performance along each of the five dimensions. Markle and colleagues (2013b) identify four slightly different dimensions of teamwork skills and behaviors, which are shared across several competency frameworks (Markle et al., 2013b, p. 15): (1) fulfill roles within a team; (2) treat group members with respect; (3) motivate group members; and (4) possess leadership skills. As discussed further below, defining and measuring teamwork is challenging, and further research is needed to define the concept more clearly and develop valid, reliable measures.

RELATIONSHIP TO COLLEGE COMPLETION

When examining these six competencies in light of its charge, the committee considered whether and to what extent any of them might also be related to traditional measures of success *during* college, such as GPA, persistence from year to year, and/or graduation. The committee reasoned that perhaps these six competencies have their effects on eventual career outcomes in part by contributing to the academic success of college students. An initial search of the literature produced little evidence related to this conjecture, one way or the other. As discussed later in this chapter, further research is needed to define more clearly and assess the competencies that have been identified as valued outcomes of college and to explore possible areas of conceptual and empirical overlap with the focal competencies identified in Chapter 2.

Literature Review

The committee conducted a literature search to identify publications examining possible relationships between the six college outcomes discussed above and college success (see Appendix A). The search yielded no rigorous research. Other than the published work by Bowman (2014), the committee found little high-quality research on the association between these six competencies and college success.

The committee was nevertheless able to find a few isolated studies relevant to the questions of interest. Teamwork, for example, is one of the six focal competencies for college graduates, and Fortenberry and colleagues (2007) provide evidence from a single institution that engaging students in team projects increased their persistence in engineering. An emerging body of "discipline-based education research" suggests that carefully designed group learning activities can support learning in science, technology, engineering, and mathematics (STEM) disciplines (National Research Council, 2012a), but this research focuses on acquisition of STEM concepts and skills and does not assess teamwork competencies. Further, while

instructors may assign team projects, they are unprepared to facilitate students' development of teamwork competencies (Borrego et al., 2013).

In the future, conceptual and technological advances in assessment, research, and instructional design may lead to improvements in teaching and assessing teamwork. To the extent that explicit teaching and grading of teamwork result, data may be available for use in future studies to examine in-depth relationships between teamwork competencies and GPA, persistence, and graduation. Similarly, as discussed further below, undergraduate engineering programs are beginning to teach and grade ethics explicitly. Here, too, future studies may yield understanding of additional relationships between ethics and GPA, persistence, and graduation.

Commissioned Research

Given the lack of published research on the possible contributions of the six outcome competencies to college persistence and success, the committee commissioned two original data analyses. At the committee's December 2015 workshop, education researcher Nicholas Bowman (2014) shared his study focused on a construct closely related to intercultural competence—openness to diversity and challenge. Bowman found that openness to diversity and challenge was statistically significantly related to college experiences and, critically, to first-year GPA. It was also a marginally statistically significant predictor of first-to-second-year student retention. Bowman's (2014) study was based on analysis of data from the Wabash Study of Liberal Arts Education. To address its questions, the committee commissioned Bowman to conduct further analysis of these data. At the same workshop, economist David Deming presented his academic research drawing on the National Longitudinal Survey of Youth (NLSY) to investigate the labor market rewards for social skills (Deming, 2015). The committee commissioned Deming to conduct further analysis of those data.

Commissioned Analysis No. 1

The committee asked Bowman to draw once again on the Wabash Study data (a sample of 8,475 students) to reanalyze data on the relationships among openness to diversity and challenge, college experiences, and student success and to identify possible differences in these relationships for various student groups (genders, underrepresented minorities). In addition, the committee asked him to analyze (both overall and for different subgroups) the relationships among the following four competencies, college experiences, and student success:

- ethics (moral reasoning),
- intercultural/diversity competence,
- civic engagement/citizenship, and
- teamwork.

Bowman examined four of the six competencies discussed in this chapter, omitting lifelong learning/career orientation and communication, for which there were no relevant data in the Wabash Study dataset. He used measures of the four competencies available in the Wabash data (see Box 5-1).

[BOX 5-1]

Bowman (2016) reports that with the exception of ethics, all of the other competencies he examined (teamwork, intercultural/diversity competence, and civic engagement/citizenship) were statistically significant predictors of college engagement. These competencies were all measured at college entrance, while college engagement (experiences) was measured near the end of the first year, obviating the possibility that the college engagement influenced the status of the competencies (see Box 5-1). However, his results for GPA were quite different: only ethics was a statistically significant predictor of GPA in years 1 and 4 ($r = .103$-$.108$, $p < .001$), while the effects of the other competencies were not consistently statistically significant in both years. Bowman suggests that ethics is likely a cognitive competency, which would account for its ability to predict grades but not engagement. Finally, turning to retention, the only consistently significant predictor of retention was civic engagement ($r = .207$, $p < .001$; $r = .073$; $p < .05$; and $r = .115$, $p < .001$ in years 2, 3, and 4, respectively). In the paper describing his analysis and results, Bowman (2016) acknowledges that there is no particular theoretically grounded explanation for why civic engagement might predict retention.

Bowman also estimated these equations across a variety of subgroups, including race/ethnicity, sex, first-generation status, and standardized test scores. While he found some significant effects of different competencies on different outcomes, he observed no clear or consistent pattern in the findings.

In his presentation to the committee, Bowman remarked that even given his extensive familiarity with the higher education literature, he could find no strong empirical basis on which to develop hypotheses or otherwise ground his analyses. These comments reinforced the committee's perception that research is lacking on the possible relationships between college outcomes and competencies related to college success.

Commissioned Analysis No. 2

To explore its questions about the possible roles of teamwork and communication in college success, the committee commissioned Deming to extend his research on social and cognitive skills (Deming, 2015), which draws on data from the NLSY. In that study, Deming used a social skills index with two components (see Box 5-2):

- data from two sociability items (How sociable are you now? How sociable were you at age 6?—extremely shy/somewhat shy/somewhat outgoing/extremely outgoing); and
- data on the number of clubs in high school and (yes/no) participation in team sports in high school.

The committee asked Deming to reexamine these data to explore whether social skills (as measured by the same social skills index) might be related to college success, both among the general student population and for different subgroups of students.

[BOX 5-2]

Deming reported to the committee that NLSY respondents with greater levels of social skills had significantly greater levels of completed schooling. A 1 standard deviation (SD) increase in social skills was correlated with an increase of about 0.78 years of completed education ($p < .05$). Adding controls first for race, sex, and age, and second for geography (census division, metro area, and urbanicity fixed effects) reduced the correlations only slightly ($r = 0.775$ and $r = 0.882$, $p < .05$). Including a measure of cognitive skills (the Armed Forces Qualification Test [AFQT]) reduced the size of the association by somewhat over 50 percent, but it remained relatively large and statistically significant ($r = 0.376$, $p < .05$). Relative to the AFQT measure of cognitive skills, however, social skills had only about a fourth of the effect on years of schooling. When the attainment of a bachelor's degree was the dependent variable, Deming's results generally replicated these findings ($r = 0.101$-0.046 as the same controls were added, $p < .05$). Deming found no differences between men and women in the effects of social skills on years of schooling and receipt of a bachelor's degree, and small and mixed effects for race. Like Bowman, Deming reported to the committee that he was unaware of any rigorous prior research on the questions of interest to the committee.

The committee noted the limitations of Deming's social skills index for measuring the focal competencies identified in this chapter—teamwork and communication—and raised questions about the two components of the index. When Deming estimated two additional models, each using one of the two different components of the social skills index, he found that the same patterns held for both models (see Box 5-2).

Findings from the Literature Review and Commissioned Analyses

To date, in summary, little high-quality research has been conducted on the relationship between the six college outcomes discussed in this chapter and college success. The commissioned analyses by Bowman and Deming provide some evidence that these outcomes contribute to college success, but their findings need to be interpreted as only suggestive rather than definitive. The few published studies discussed above also provide limited evidence that some of the six outcomes contribute to college success, but again are only suggestive.

ASSESSMENT OF INTRA- AND INTERPERSONAL COLLEGE OUTCOMES

As noted earlier, colleges and universities are beginning to assess a broader set of student learning outcomes, including intra- and interpersonal outcomes, and are using a variety of methods for those assessments. Provosts responding to a survey in 2014 indicated they were using a wider range of assessment instruments than they had in 2009 (Kuh et al., 2014). National surveys, such as the National Survey of Student Engagement, were the most popular assessment instrument (used by 85% of respondents), followed by rubrics (69%) and classroom-based assessments that are aggregated to the institutional level (66%). Other methods included alumni surveys, incoming student placement exams, locally developed surveys, capstone projects or papers, and locally developed knowledge and skill measures. Similarly, Hart Research Associates (2016) found that the proportion of AAC&U member institutions assessing learning outcomes had increased from 6 years earlier, both in general education and more broadly at the institutional level. The respondents reported using a variety of measurement tools; prime among

them for general education was the use of rubrics applied to samples of student work and capstone projects. Forty-two percent of those assessing outcomes in general education said they used AAC&U's VALUE (Valid Assessment of Learning in Undergraduate Education) rubrics. Among those using locally created rubrics, well over half (58%) reported that the VALUE rubrics informed the development of these local rubrics.

Described below are selected examples of current efforts to measure the six competencies identified earlier. Some of these efforts focus on one specific competency, while others are more general in nature, designed to assess multiple competencies. As examples, they are not fully representative of all such efforts.

VALUE Rubrics

VALUE rubrics are available for all six competencies discussed in this chapter. The rubrics were designed to measure authentic student work so that colleges and universities could determine whether and how well students were attaining the 16 essential learning outcomes. In 2007-2009, teams of faculty and administrators from a range of institutional types located across the country analyzed and synthesized existing campus rubrics and college mission statements, consulted experts in relevant fields, and obtained feedback from faculty to develop the initial set of VALUE rubrics (Maki, 2015). Other faculty at more than 100 campuses tested the use of the rubrics for scoring local student work samples in three rounds of drafting, campus testing, and revision. At 12 campuses, the rubrics were used to evaluate e-portfolios (Maki, 2015). Reflecting AAC&U's (Association of American Colleges and Universities, 2007) emphasis on the application as well as acquisition of knowledge, the rubrics were developed for use in assessing a variety of student products, ranging from traditional papers to signature projects (such as capstone projects, service learning, internships, and other applied activities).

Each rubric includes a definition of the competency, framing language, a glossary, and benchmarks for assessing five dimensions of the competency at four levels, from benchmark (level 1) to capstone (level 4) (see Table 5-2). All are available at the AAC&U website.[1]

[TABLE 5-2]

Maki (2015) proposes that the process used to develop the VALUE rubrics—drafting and validation by faculty who are closest to student learning and outcomes assessment, together with pilot testing—helps ensure the face and content validity of these measures. In addition, a test of interrater reliability showed relative convergence in the scores of 44 faculty members from across disciplines that had independently scored student work online using the critical thinking, integrative learning, and civic engagement VALUE rubrics. By December 2015, the VALUE rubrics had been accessed by about 42,000 individuals from more than 2,800 colleges and universities (Schneider, 2015b). In addition, they have been used across institutions, beginning in 2013 when seven colleges and universities in Massachusetts jointly tested protocols for collecting and scoring samples of student work (Maki, 2015). Faculty submitted 350 student work samples representing written communication, quantitative literacy, and critical thinking. To establish interrater agreement, the samples were distributed for independent scoring, followed by group discussion to build shared understanding of performance levels for each competency. The

[1] https://www.aacu.org/value-rubrics [February 2017].

results of individual scoring of institutional samples were entered into a spreadsheet that identified remaining areas of disagreement between the two scorers assigned to evaluate each piece of work—areas that would need to be addressed in future scoring sessions. This group of institutions then offered regional workshops to support a larger pilot project.

Building on the Massachusetts initiative, the State Higher Education Officers Association and AAC&U developed the Multi-State Collaborative to Advance Learning Outcomes Assessment, including state higher education systems from 12 states. In its initial phase, the collaboration engaged faculty raters in examining student work from 68 2- and 4-year colleges and universities in nine state systems. The Bill & Melinda Gates Foundation then provided funding to expand this collaborative model in a demonstration/implementation year beginning in September 2015. With this grant, the collaborative plans to develop a methodology for ensuring a high level of reliability and validity of the results obtained using the VALUE rubrics, test the process for identifying and gathering samples of student work, and develop a national VALUE database for benchmarking student learning. The committee looks forward to these developments, including public release of information on the reliability and validity of the rubrics.

Assessing Ethics in Engineering

In the two decades since the ABET *Engineering Criteria 2000* document was released, engineering education experts, faculty members, and researchers have developed a variety of methods for assessing the intra- and interpersonal competencies required by these accreditation standards (e.g., Shuman et al., 2005). Some recent efforts are described in a National Academies of Sciences, Engineering, and Medicine (2016b) report on *Infusing Ethics into the Development of Engineers.* Based on submissions from 44 programs, the committee that developed that report identified 25 exemplary programs that included some type of assessment of student learning outcomes. Although these programs reported using a wide variety of informal and formal assessment methods, they did not report on the reliability or validity of any of these methods. In other words, the insights offered by the exemplary programs are commendable, but empirical work on these programs is sorely needed. That said, several programs assessed outcomes based at least partly on student feedback, whether through comments or emails, more formal student rating systems (for example, the IDEA Student Ratings of Instruction system[2]), or faculty-designed questionnaires.

Faculty at Texas State University, a Hispanic-serving institution, and the University of Texas at Tyler, whose student population is 60 percent women, developed two modular courses on ethical, health, and safety issues related to nanotechnology for undergraduates in engineering and engineering technology. The modules, developed in collaboration with industry, were infused into nontechnical introductory courses and more technical courses from the sophomore through senior years. Student learning outcomes were assessed through in-class assignments and separate interval and end-of-term assessments, focused on student understanding, engagement, and satisfaction. Assessment also included student ratings of the courses (on a five-point scale, from poor to excellent) and a series of focus groups. Student retention was higher than it had

[2]According to the IDEA website, the IDEA system is designed to improve teaching and includes carefully formulated questions designed to elicit students' thoughts about their own learning, rather than simply opinions about the instructor.

been prior to the new, modular courses, and student feedback was used to revise the modules further.

At Northeastern University, faculty developed an engineering ethics program including courses in engineering students' second and fourth years, which drew on faculty members' professional engineering experiences and integrated students' internship learning experiences. Assessment included students' anonymous evaluations of the junior-year course and their scores on an independent test—the Fundamentals of Engineering (FE) examination, a required assessment for certification as a licensed professional engineer. The FE exam includes questions on the topic of ethics and business practices, and the test administrator provides a separate score on this topic. Over the period 2005-2013, a group of about 400 Northeastern students who had completed the ethics program scored 4.7 percent higher than the national average on the ethics and business practices section of the test.

Two other exemplary programs used the Defining Issues Test-2 (DIT-2[3]) to assess student learning. The DIT-2, a selected-response assessment, presents students with five scenarios to assess their understanding of ethical issues. One program reported using a newly developed moral reasoning instrument, the Engineering Ethical Reasoning Instrument (Zoltowski et al., 2013). In a separate publication, the developers of that instrument reported that it was still in the scale and construct validation stages (Zoltowski et al., 2013).

Assessing Teamwork

Employers invariably identify teamwork as one of the most important competencies required of 2- and 4-year college graduates, but both conceptualizing and measuring this construct in the classroom environment can be difficult. The intra- and interpersonal aspects of teamwork are intimately related to a team's task and environment, team processes (e.g., member attitudes, interactions, communication patterns), and team outcomes (products or services created by the team, along with outcomes for the team members) (Kozlowski and Ilgen, 2006). In undergraduate education, group and team learning activities are growing, but instruction and assessment typically have focused on only one aspect of team outcomes—the team's product, such as a paper, presentation, or video (Britton et al., 2015; Hughes and Jones, 2011). Such assessment taps cognitive knowledge and skills but provides little information on team processes, including the development of interpersonal teamwork competencies.

In an important early study focused on team processes, Stevens and Campion (1994) conducted a review of the literature on teams to develop a comprehensive taxonomy of individual teamwork knowledge, skills, and attitudes (KSAs). The taxonomy includes two major factors: *interpersonal KSAs*, comprising conflict resolution, collaborative problem solving, and communication, and *self-management KSAs*, comprising goal setting/performance management, planning, and task coordination. Stevens and Campion (1999) then developed the Teamwork KSA test to measure these constructs. The test items present hypothetical teamwork scenarios along with alternative responses for selection by the test taker. This test has been widely used for employee selection and has also been incorporated into research in higher education.

Chen and colleagues (2004), for example, used the Teamwork KSA test as a focus for the development and evaluation of a semester-long teamwork training course for undergraduates that included both didactic reading and lecture and experiential learning in teams. A quasi-experimental evaluation showed that the course had a statistically significant effect on increasing

[3]See http://ethicaldevelopment.ua.edu/dit-and-dit-2 [June 2016].

the levels of students' teamwork knowledge and skills, but not on improving teamwork-related attitudes and self-efficacy. More recently, Bedwell and colleagues (2014) incorporated elements of Stevens and Campion's (1999) taxonomy into a broader taxonomy of interpersonal skills, and these authors propose approaches for integrating the development and assessment of interpersonal skills into the master of business administration curriculum.

Although Stevens and Campion's (1999) taxonomy and test have informed the assessment of teamwork in higher education, O'Neill and colleagues (2012, p. 1) recently identified statistical limitations of the test. Reviewing the literature, they found "…an average criterion validity of .20 for the Teamwork-KSA Test," which they characterize as low, although a more thorough examination of the criterion measures tied to the dimensions of the taxonomy, along with a comparison of teamwork skills measures, would be needed to conduct a more solid assessment of validity. That said, O'Neill and colleagues (2012) were unable to locate any research on the item properties, factor structure, or subscale reliabilities.

Researchers have moved away from paper-and-pencil tests of teamwork and toward the use of team member ratings, which may reduce biases associated with self-report ratings and better represent dynamic team processes. These team member rating approaches are based on definitions of teamwork that focus on individual contributions to team success, as discussed further below, and draw on the literature on performance management in organizations (Ohland et al., 2012). They offer several advantages for assessing and developing teamwork relative to other measurement methods. Team members are well positioned to evaluate their own and their peers' contributions to a team, and they can learn about teamwork through the process of rating themselves and their peers based on research-based definitions of individual contributions to team success (Ohland et al., 2012). Studies of business students have found that requiring student team members to rate themselves and their peers reduces "social loafing" (doing little or no work while other team members carry out the task) and is associated with higher perceived grade fairness and more positive attitudes toward teamwork (Aggarwal and O'Brien, 2008; Chapman and van Auken, 2001; Erez et al., 2002, cited in Ohland et al., 2012).

At the same time, however, self- and peer-rating approaches also face challenges. For example, in a study of "team-based learning" (TBL), an instructional strategy used in medical education, Thompson and colleagues (2007) found that many students were resistant to the peer evaluation that was used to inform a portion of their course grades. One medical school dropped the peer evaluation component of TBL because of student hostility, and in other schools, students gamed the rating system by giving every team member the same rating. When rating themselves, students may have an inflated sense of their own contributions to the team and unrealistic expectations of their teammates' contributions (Ohland et al., 2012). Another challenge is that these rating approaches typically focus on a single team experience, but a student who displays strong teamwork competencies in a particular team may or may not be able to transfer those competencies to other team contexts. Additional research is needed on the consistencies and inconsistencies of students' teamwork and team performance across teams, tasks, and time.

An additional challenge is the potential for various forms of bias to influence peer ratings. Haynes and Heilman (2013) report on a series of studies examining how women and men allocated credit for joint success in performing a task. Women gave more credit to their male teammates and took less credit themselves unless their role in bringing about the successful performance was irrefutably clear, or they were given explicit information about their likely competence before completing the task. However, women did not credit themselves less when

their teammate was female. Collectively, the studies showed that women working in gender-diverse teams tended to devalue their contributions to the collaborative work.

In a separate study, supported by the National Science Foundation, Joshi (2014) assembled and analyzed data across more than 60 science and engineering research teams. Results from this analysis indicated that recognition and utilization of the expertise of male and female scientists and engineers were influenced by the gender and gender identification of the rater, the team's gender composition, and female faculty representation in the discipline in which the teams were embedded. Relative to male team members, female team members evaluated educated female and male team members more positively. Male team members who identified strongly with their gender evaluated highly educated female colleagues more negatively than less educated female colleagues. In male-dominated teams, the expertise of highly educated men was used to a greater extent than was the case in teams dominated by women. Finally, teams with a greater proportion of highly educated women were significantly more productive (in terms of research publications) in disciplines with greater female faculty representation.

Such challenges notwithstanding, several investigators have developed rating systems for assessing teamwork. Loughry and colleagues (2007) reviewed the literature on organizational teams to identify ways in which an individual can contribute to a team and translated these findings into a large pool of potential test items. After surveying students, they reduced the pool to 29 specific types of contributions clustered into five broad categories: contributing to the team's work; interacting with teammates; keeping the team on track; expecting quality; and having relevant knowledge, skills, and abilities.

The authors then created the Comprehensive Assessment of Team Member Assessment (CATME) instrument in both a long version with 87 items and a short version with 33 items. Students rate their peers using Likert scales. Even the short version, however, requires students to read 33 items and rate each of their teammates on each item. This represents a nontrivial burden on students, as well as on the instructor who tries to draw inferences from the large number of ratings. To address this problem, Ohland and colleagues (2012), with support from the National Science Foundation, developed CATME-B, a web-based instrument that collects and analyzes confidential self- and peer-evaluation data. Instead of Likert rating scales, these authors developed behaviorally oriented rating scales that students can use to identify three levels of performance across the five categories of team member contributions cited above. Three tests of the instrument demonstrated psychometric characteristics equivalent to those of the much longer initial version of CATME (Loughry et al., 2007), high convergence with another peer-evaluation scale created by Van Duzer and McMartin (2000), and a statistically significant relationship with final course grades in a course requiring a high level of team interaction.

SPARK—the Self and Peer Assessment Resource Kit (Freeman and McKenzie, 2002)—is another system for peer and self-evaluations of teamwork competencies. It provides a template that allows faculty to customize the evaluation criteria according to specific disciplines or project goals.

More recently, Kulturel-Konak and colleagues (2014) introduced PEAR—the Peer Evaluation and Assessment Resource—also developed with support from the National Science Foundation. In addition to simplifying the collection and analysis of student ratings, the instrument is intended to measure students' developmental progression in teamwork over the course of the undergraduate years. The PEAR assessment framework is based on the model of domain learning, which posits that the development of expertise in a domain proceeds in three progressive stages: acclimation, competency, and proficiency. The model also proposes that the

nature of domain knowledge, strategic processing abilities, and interests differ across these three stages. Thus, the web-based tool incorporates three types of rubric items (knowledge, interest, strategic processing), mapped against the three developmental stages of the model of domain learning. PEAR also differs from CATME in its flexibility, as it allows instructors to create their own custom rubrics. A pilot study demonstrated the feasibility of the framework and data gathering instrument. According to Kulturel-Konak and colleagues (2014), the instrument was in the alpha testing stage, to be followed by evaluation of the reliability of the instrument and the validity of the framework.

Thus, although progress is being made in assessing teamwork, assessment of this competency is impeded by the lack of a robust conceptual model of teamwork processes and outcomes that is directly tied to an assessment framework. The teamwork literature to date shows a wide variety of models and definitions of teamwork competencies and a wide variety of new assessment instruments (e.g., online self- and peer-rating systems). Teamwork researchers continue to grapple with new and different frameworks and applications. In developing a teamwork assessment for midwifery students, for example, Hastie and colleagues (2014) chose not to build on CATME-B because the web-based measure could not be altered or customized, and because they viewed the CATME items as not well defined and even likely to impede students' understanding of expected teamwork skills and behaviors. Instead, the authors created a new rubric based on a revision of the VALUE teamwork rubric (Association of American Colleges and Universities, 2007). Most recently, Britton and colleagues (2015) drew on a taxonomy of teamwork dimensions derived from surveys used in health care settings (Valentine et al., 2012), along with the rubric created by Hastie and colleagues (2014), to develop yet another rubric for self- and peer evaluations of teamwork—the Team-Q. Evaluating this new tool, Britton and colleagues (2015) found that it had high internal consistency, interrater reliability was within an acceptable range, and factor analyses provided evidence of convergent and discriminant validity. The authors also obtained preliminary evidence that teamwork skills improved over time when taught and assessed.

As noted above, undergraduates increasingly are required to complete team projects and engage in group learning activities. However, the development of expertise in any domain, including teamwork, requires feedback as well as practice (National Research Council, 2000, 2012b), so simply working in teams without direct instruction, assessment, and feedback will not necessarily develop students' teamwork competencies. The current lack of valid, reliable assessments of teamwork impedes the development of effective instructional approaches for teaching teamwork along with subject matter content during group learning activities (National Research Council, 2015a; Kulturel-Konak et al., 2014).

The Engineering Professional Skills Assessment

With support from the National Science Foundation, Zhang and colleagues (2015) developed and tested a new performance assessment designed to measure five engineering professional skills identified by ABET as critical student learning outcomes: (1) understanding of professional and ethical responsibility; (2) ability to communicate effectively; (3) broad understanding of the impact of engineering solutions in global, economic, environmental, and cultural/social contexts; (4) recognition of the need for the ability to engage in lifelong learning; and (5) knowledge of contemporary issues. Unlike all of the other assessment examples discussed in this section, this assessment focuses on measuring group rather than individual

performance. It uses a prompting scenario to present students with a contemporary engineering issue lacking a clear-cut solution and an analytical scoring rubric with five dimensions corresponding to the skills cited above. Groups of four to six students were instructed to engage in a 45-minute discussion around the issue presented by the scenario, and trained raters used the scoring rubric to assess the group's performance on each of the five skills. The raters did not rate individual students but rather the groups as a whole.

In a small, exploratory study, Zhang and colleagues (2015) collected and analyzed data from 20 discussion groups at three engineering colleges to determine whether the scenario used affected performance scores. Although based on a small sample, the findings tentatively suggested that scores on the assessment discriminated among student groups and that using different scenarios appeared to have a minimal effect on student group scores. In addition, student groups varied in their measured proficiency across the five ABET outcomes, suggesting the need for a stronger emphasis on developing these outcomes in the undergraduate engineering curriculum.

Assessing Civic Competency and Engagement

In response to widespread interest in developing students' civic competency and engagement (also referred to as citizenship), ETS convened a research team to define this construct more clearly and to develop an assessment framework. As discussed earlier, Torney-Purta and colleagues (2015) conducted an extensive review of the literature on defining and assessing civic engagement to identify challenges and opportunities for designing and implementing assessments of this complex competency. They identified two dimensions of this competency: 1) participatory and involvement skills; and (2) civic engagement, which comprises motivations, attitudes, and efficacy, democratic norms and values, and participation and activities. They also considered which item formats and task types would be most likely to ensure fair and reliable scoring for a future assessment of civic competency and engagement. Those efforts were part of a suite of higher education assessments in various stages of research and development at ETS, referred to as HEIghten™. Reflecting higher education leaders' growing interest in assessing intra- and interpersonal competencies, the suite includes two of the six competencies that are the focus of this chapter—civic engagement/citizenship and intercultural/diversity competence. Following on the publication and dissemination of the Torney-Purta et al. (2015) study, the ETS team (Liu et al., 2015) planned the following further development activities: prototyping and cognitive interviews, feedback from user audiences, revision of the test blueprint, item writing and test development, pilot study, validation studies, and operational testing.

CONCLUSIONS AND RECOMMENDATIONS

The intra- and interpersonal competencies of ethics, lifelong learning/career orientation, intercultural/diversity competence, civic engagement/citizenship, communication, and teamwork have been identified as valued outcomes of college education. Although it might seem intuitive that these competencies would predict academic success, there is little evidence to date that these desired outcomes for graduates are actually improving in college itself and also contribute to persistence, GPA, and graduation. There simply are too many large gaps in the research literature

and in the available data to say with any certainty that these competencies do or do not matter for students' success in college.

Conclusion: To date, only limited research has been conducted on the intra- and interpersonal competencies that have been identified as important learning outcomes for college graduates. Therefore, little is known about whether and under what conditions these competencies are related to persistence and success in college.

This gap in the research comes at a time when educational policy makers are pursuing two potentially complementary aims related to intra- and interpersonal competencies: (1) increasing students' persistence to graduation, and (2) developing students' competencies for life and work after graduation. Research is needed to explore whether, and to what extent, these aims are complementary.

Recommendation 13: Federal agencies and foundations should invest in research examining whether, and under what conditions, the intra- and interpersonal competencies identified as outcomes for college graduates may also be related to students' persistence and success in college.

Research Needs

The six competencies discussed in this chapter have the potential to significantly broaden understanding of the noncognitive determinants of student postsecondary success. The committee focused on these particular competencies because all of them have been identified by blue ribbon panels and top educational researchers as desirable outcomes of higher education. Many observers view these six competencies as critical to the workplace success of the next generation of college graduates, and indeed a growing body of research shows that collectively, these competencies are valuable in the labor market and other aspects of life (e.g., Deming, 2015).

Very little is known about the empirical associations between the six competencies that are the focus of this chapter and the ability and process of students moving successfully through college. This paucity of knowledge opens up a host of research opportunities. The committee believes a sensible research agenda will require going well beyond calculation of correlation coefficients between, on the one hand, ethics, lifelong learning/career orientation, intercultural/diversity competence, civic engagement/citizenship, communication, and teamwork and, on the other hand, college completion, GPA, and other indicators of performance. Three particular issues concerning a needed research agenda are worth noting.

Assessment Issues

First, the state of measurement of most of these competencies is still markedly underdeveloped, as is the case for the eight competencies identified in Chapter 2. In contrast with the century-long history of measuring cognitive competencies, researchers have only in the last decade devoted a concerted effort to defining intra- and interpersonal competencies and designing and testing scales and items appropriate to their assessment. Furthermore, as discussed

in Chapter 3, many new and rapidly evolving psychometric and technological advances are now available to assist in the measurement of these complex competencies. These advances open up new opportunities to conduct much more test development work, reinforcing the committee's recommendation for such work (Recommendation 8).

Theoretical and Conceptual Issues

Second, much theoretical and conceptual work remains to be done before statistical analysis of these competencies is undertaken to explore potential areas of conceptual overlap between college outcomes and predictors of college persistence. This research would examine such questions as the following:

- What are the theoretical reasons (if any) to expect that civic engagement, for example, is empirically (even causally) associated with GPA and/or college persistence?
- What theoretical models or frameworks would guide investigation of this potential relationship?

The committee notes that some researchers have found ethical behavior to be associated with both conscientiousness (e.g., Gensler, 1996; Kalshoven et al., 2011) and responsibility, which, in turn, is related to conscientiousness (Jackson and Roberts, 2015). In another example, the college outcomes of intercultural/diversity competency and civic engagement/citizenship discussed in this chapter appear to overlap conceptually with the competency of prosocial values and goals discussed in Chapter 2 (Wolniak et al., 2012). A process-focused elaboration of these theoretical frameworks—and a concomitant measurement framework to capture these processes—will be required if research on these competencies and postsecondary success is to proceed in a cumulative and informative manner.

Finally, the committee endorses a multimethods research agenda aimed at better understanding the role of these competencies in student success. The commissioned analyses of Bowman and Deming each demonstrate the potential of survey-based regression analysis to begin to explore these questions. Additionally, other researchers might favor studies of particular institutions, and might adopt experimental, quasi-experimental, experience-sampling, field research, multiobserver, or qualitative case study methods rather than relying heavily on surveys. The committee believes the state of the literature is such that investment in the development of a conceptual framework of intra- and interpersonal competencies tied to a multiyear, multi-institutional research agenda holds particular promise.

BOX 5-1
Data and Research Methods Used in Commissioned Analysis No. 1

The 2006-2009 Wabash Study of Liberal Arts Education was a large-scale study designed to explore teaching practices, programs, and institutional structures that support liberal arts education and to examine how such experiences and student outcomes can be measured effectively. Colleges and universities were selected to participate in the Wabash Study on the basis of their strong commitment to this form of education. The overall sample contained 49 private and public institutions (30 liberal arts colleges, 16 universities, and 3 community colleges), which included religiously affiliated, single-sex, and minority-serving schools; these institutions also exhibited a wide range of selectivity, tuition costs, and geographic diversity.

The Wabash data are of exceptionally high quality. First-year students were sampled in fall 2006, 2007, and 2008 (Time 1). The respondents provided demographic information and completed both a questionnaire on various high school experiences, interests, attitudes, and values and a battery of assessments. At the end of their first year (Time 2), students who took part in the initial assessment were invited to participate in a second round of data collection. They completed the same battery of assessments, along with questionnaires that asked about their college experiences, interests, attitudes, and values. A total of 8,475 students at 4-year institutions participated in the second wave, yielding a retest response rate of 50.7 percent. Institutional data were obtained on whether students reenrolled in the fall of their second, third, and fourth years of college. The sample was carefully weighted to be representative of the incoming classes of the participating institutions. For this analysis, Bowman (2016) removed the three community colleges from the analytic sample because there were relatively few of these students and institutions, and retention outcomes have a different meaning for this group.

Using constructs within the Wabash dataset, Bowman measured ethics based on a score for postconventional moral reasoning on the Defining Issues Test 2 ($\alpha = 0.77$-0.81; Bebeau and Thoma, 2003; Rest et al., 1999). He measured intercultural/diversity with overall score on the 15-item Miville-Guzman Universality-Diversity Scale ($\alpha = 0.85$; Fuertes et al., 2000; Miville et al., 1999). Civic engagement was measured with a single item on frequency of volunteering during high school senior year (1 = never to 5 = very often), and teamwork was measured through 28 items representing group dimensions within the Socially Responsible Leadership Scale ($\alpha = 0.86$; Dugan, 2006). Bowman analyzed the relationships of these competencies with college persistence, using institutional data on reenrollment in the second, third, and fourth years of college and self-reported cumulative college grade point average (GPA). Although college engagement (experiences) was not a focal indicator for the committee, Bowman included it in his analysis, as he had done in his published study (Bowman, 2014), because of its long-demonstrated relationship with college retention in the higher education literature.

Bowman conducted a relatively simple but rigorous statistical analysis. He estimated a series of linear regression models predicting college engagement, grades, and retention. His models included both institutional fixed effects and a variety of precollege control variables, including standardized test scores and high school GPA. He also estimated these equations across a variety of subgroups, including race/ethnicity, sex, first-generation status, and standardized test scores.

BOX 5-2
Data and Research Methods Used in Commissioned Analysis No. 2

Deming used the National Longitudinal Study of Youth (NLSY) 1979 survey to test the relationship between social skills and educational attainment. The U.S. Bureau of Labor Statistics created the NLSY 1979 using three independent probability samples to represent the entire population of youth aged 14 to 21 as of December 31, 1978, residing in the United States on January 1, 1979. The survey was conducted yearly from 1979 to 1993 and then biannually from 1994 through 2012; it included detailed measures of prelabor market skills, schooling experience, employment, and wages.

Deming estimated a series of ordinary least squares regression models, using as his dependent variables both years of education completed and the attainment of a bachelor's degree (the latter coded dichotomously). He did not use grade point average (GPA) as a dependent variable because the NLSY lacks a measure of college grades. He controlled for several other measures of cognitive and noncognitive skills and an assortment of standard covariates. As fixed effects, his controls included race, sex, age, census division, metro area, and urban/rural membership, along with two widely used intrapersonal skill measures—the Rotter Locus of Control and the Rosenberg Self-Esteem Scale. Although Deming's data and measures are not exactly the same as those used by Bowman in his commissioned analysis, the two analyses are broadly compatible.

In response to the committee's questions about the social skills index, Deming estimated two additional models. The first used only the two sociability items to construct the social skills index, while the second used only the number of clubs in high school and (yes/no) participation in team sports to construct the index. In both cases, Deming normalized the index to have mean zero and standard deviation 1. With both models, social skills were significantly correlated with educational attainment even after controlling for demographics, cognitive skills, and intrapersonal skills. However, the magnitude of the coefficients was much larger in the second model, when social skills were measured based on participation in high school clubs and sports.

TABLE 5-1 Intra- and Interpersonal Competencies within Selected Outcomes Frameworks

Competency	Accreditation Board for Engineering and Technology, 2014	Oswald et al., 2004	Association of American Colleges and Universities, 2007	Markle et al., 2013b	Lumina Foundation, 2015
		Intrapersonal Competencies			
Ethics	*Professional and ethical responsibility*	*Ethics,* within intrapersonal behaviors	*Ethical reasoning and action,* within personal and social responsibility	*Ethics and integrity,* within citizenship	*Ethical reasoning,* within intellectual skills
Lifelong Learning/Career Orientation	*A recognition of the need for and an ability to engage in lifelong learning*	*Career orientation and perseverance,* within intrapersonal behaviors; *continuous learning,* within intellectual behaviors	*Foundations and skills for lifelong learning,* within personal and social responsibility	*Self-directed learning,* within life skills	Refers to students' learning and engagement throughout their academic careers and beyond, within broad and integrative knowledge
		Intra- and Interpersonal Competencies			
Intercultural/Diversity Competence		*Multicultural appreciation*, within interpersonal behaviors	*Intercultural knowledge and competence,* within personal and social responsibility	*Respect for others,* within citizenship	*Civic and global learning; engaging diverse perspectives*
Civic Engagement/Citizenship		*Citizenship,* within interpersonal behaviors	*Civic knowledge and engagement—local and global,* within personal and	*Citizenship*	*Civic and global learning*

			social responsibility		
Interpersonal Competencies					
Communication	*The ability to communicate effectively*	*Communicating and dealing well with others,* within interpersonal behaviors	*Written and oral communication*, within intellectual and practical skills	*Effective communication*	*Communicative fluency,* within intellectual skills
Teamwork	*Ability to function on multidisciplinary teams*	*Leadership (showing skills in a group),* within interpersonal behaviors	*Teamwork and problem solving,* within intellectual and practical skills	*Teamwork*	*Applied and collaborative learning*

SOURCE: Created by the committee.

TABLE 5-2 Sample VALUE Rubric: Foundations and Skills for Lifelong Learning

Selected Dimension	Capstone Level
Initiative	Completes required work, generates and pursues opportunities to expand knowledge, skills, and abilities.
Independence	Educational interests and pursuits exist and flourish outside classroom requirements. Knowledge and/or experiences are pursued independently.
Reflection	Reviews prior learning (past experiences inside and outside of the classroom) in depth to reveal significantly changed perspectives about educational and life experiences, which provide foundation for expanded knowledge, growth, and maturity over time.

SOURCE: Schneider, 2015b. Reprinted with permission.

6

Next Steps

This chapter highlights next steps toward realizing the potential of intra- and interpersonal competencies to support students' persistence and success in higher education. In so doing, it summarizes the committee's conclusions and highlights its recommendations.

COMPETENCIES FOR COLLEGE SUCCESS

Based on its review of the evidence, the committee concluded that only limited research has been conducted to date on the potential relationships between various intra- and interpersonal competencies and students' college success.

Recommendation 1: Federal agencies and foundations should invest in research exploring the possible relationships between various intra- and interpersonal competencies and students' college success. To address gaps in the research base, these investments should include support for research examining

- **how interpersonal competencies may be related to student success in 4-year colleges;**
- **how intra- and interpersonal competencies may be related to student success in community colleges; and**
- **how intra- and interpersonal competencies may be related to students' success in 2- and 4-year STEM programs and majors.**

Based on the limited intervention studies conducted to date, the committee found promising evidence that the three competencies of sense of belonging, growth mindset, and utility values are related to college success and malleable in response to interventions. The committee found more modest evidence that five other competencies are similarly related to college success and malleable, yielding the following list:

- behaviors related to conscientiousness,
- sense of belonging,
- academic self-efficacy,
- growth mindset,
- utility goals and values,
- intrinsic goals and interest,

- prosocial goals and values, and
- positive future self.

Although the evidence for sense of belonging, growth mindset, and utility values is encouraging, further research is needed to replicate and extend this evidence and to better understand the potential impact of the other five competencies. The following recommendations are based on the committee's belief that colleges and universities could play an important role in facilitating this research.

Recommendation 2: Federal agencies and foundations should invest in intervention research using random assignment and research employing a range of other methods to understand better the competencies identified in this report, their relationship to college success, and the mechanisms through which they operate to improve college success. Research focused on supporting the college success of underrepresented student groups should be a priority.

Recommendation 3: Colleges and universities should support the intervention research proposed in Recommendation 2 by facilitating the implementation and evaluation of random-assignment interventions, thereby gaining valuable information about their students and building the knowledge base on effective interventions needed to increase student retention and success.

Issues of race, ethnicity, gender, social class, and culture need to be primary considerations when educators, administrators, researchers, and policy makers think about competencies and their contribution to college success. Based on the literature, the committee concluded that certain competencies develop and function differently for different groups and within different cultural and institutional contexts.

Recommendation 4: To help reduce disparities in college success among student groups, institutions of higher education should evaluate and improve their social and learning environments to support the development of the eight identified competencies, especially among underrepresented student groups.

ASSESSMENT METHODS FOR COLLEGE COMPETENCIES

The committee reviewed the nature and quality of existing assessments of the eight competencies identified above, together with research and professional standards related to the general process of the development, validation, implementation, interpretation, and evaluation of the results of intra- and interpersonal competency assessments. Noting that various professional and legal standards apply to assessing these competencies, the committee concluded that most current assessments of the eight identified competencies are uneven in quality, providing only limited evidence to date that they meet professional standards of reliability, validity, and fairness. Considering potential high-stakes uses of these current assessments or of other assessments of

intra- and interpersonal competencies, the committee concluded that the development and validation of such assessments for high-stakes purposes is a rigorous, time-consuming, and expensive process that depends critically on assessment and psychometric expertise. Validity, reliability, and fairness are necessary considerations in evaluating assessment quality.

Recommendation 5: When developing and validating intra- and interpersonal competency assessments to be used for high-stakes purposes, stakeholders in higher education (e.g., faculty, administrators, student services offices) should comply with professional standards, legal guidelines, and best practices to enable appropriate interpretations of the assessment results for particular uses.

Recommendation 6: Institutions of higher education should not make high-stakes decisions based solely on current assessments of the eight identified competencies, given the relatively limited research to date demonstrating their validity for predicting college success.

The committee determined that, when used for low-stakes purposes, assessments need not meet the high evidentiary requirements associated with high-stakes individual student assessments. Even when assessments are not used for high-stakes purposes, however, they need to be sensitive to the competencies they are intended to measure. The committee concluded that even low-stakes uses of intra- and interpersonal competency assessments require attention to validity, reliability, and fairness.

Recommendation 7: Those who develop, select, or use intra- and interpersonal competency assessments should pay heed to, and collect evidence of, validity, reliability, and fairness as appropriate for the intended high-stakes or low-stakes uses.

This report notes several limitations of self-report measures. Regardless of whether an assessment is high- or low-stakes, for example, respondents may be motivated to present themselves in a favorable light. In recent years, various methods have been shown to mitigate these limitations. Ratings by others, for instance, avoid some of the problems of self-ratings and have been shown to yield more reliable and predictive data in many contexts. Based on its analysis, the committee concluded that most existing assessments of the eight identified competencies, as well as many assessments of other intra- and interpersonal competencies, use self-report measures, whose well-documented limitations may constrain or preclude certain uses of the assessment results, and that many innovative approaches for assessing competencies address these limitations.

Recommendation 8: Federal agencies and foundations should support additional research, development, and validation of new intra- and interpersonal competency assessments that address the shortcomings of existing measures.

Professional standards make clear that fairness to all individuals for whom an assessment is intended should be a driving concern throughout assessment development, validation, and use. Based on its analysis of research and current practice, the committee concluded that, despite the ever-increasing diversity of undergraduate student populations, attention to fairness for diverse populations often is inadequate in the development, validation, and use of current assessments of the eight identified competencies. Because these fairness standards apply broadly to assessments of all types of competencies, the committee makes the following recommendation.

Recommendation 9: Researchers and practitioners in higher education should consider evidence on fairness during the development, selection, and validation of intra- and interpersonal competency assessments.

Local norms and contexts may influence self-, peer, or other ratings of an intra- or interpersonal competency and may also mediate or moderate the relationships between these competencies and educational outcomes. The committee concluded that appropriate interpretation of the results of intra- and interpersonal competency assessments requires consideration of such contextual factors as student background, college climate, and department or discipline.

Recommendation 10: Higher education researchers and assessment experts should incorporate data on context (e.g., culture, climate, discipline) into their analyses and interpretations of the results of intra- and interpersonal competency assessments.

ASSESSMENT USES AND STAKEHOLDERS

A variety of stakeholders, including families, K-12 schools, faculty members, college administrators, accreditors, and state and federal policy makers, use assessment in higher education for four major purposes:

- selection and placement of individual students;
- formative improvement of local educational processes, practices, and programs;
- research and evaluation supporting knowledge generation; and
- accountability.

Based on its analysis of the relevant literature, the committee concluded that in higher education, assessments are more likely to be valuable for supporting student success when their selection, design, analysis, and interpretation are guided by stakeholder information needs, intended uses, and users

The committee observed that different higher education stakeholders have different needs for assessments of intra- and interpersonal competencies, depending on the immediacy of those needs, the purposes to be served by the data, and stakeholders' assessment-related knowledge and skills. Having concluded that the selection of an assessment and appropriate use and interpretation of its results vary depending on the intended uses and users, the committee makes the following recommendation.

Recommendation 11: Leaders in higher education should select, design, analyze, and interpret data from assessments of intra- and interpersonal competencies based on stakeholder information needs, intended uses, and users.

Based on its review of the research on uses of assessment in higher education, the committee concluded further that assessments of intra- and interpersonal competencies are more likely to be implemented and used by stakeholders to improve student success when they are motivated by internal institutional improvement purposes than when they are motivated by accountability purposes.

After reviewing research on the assessment process, the committee concluded that assessments of intra- and interpersonal competencies are more likely to contribute to improved college success among students if efforts to use the results involve stakeholders at multiple levels of the organization, as opposed to individual stakeholders acting alone. In addition, some higher education stakeholders will require additional support and training to develop the knowledge and skills needed to select, use, and interpret assessments of intra- and interpersonal competencies to improve student success in higher education. Considering the available research, moreover, the committee concluded that, relative to the limited research on the relationship between intra- and interpersonal competencies and college success, less evidence is available from an organizational science perspective on how stakeholders in higher education can use data on intra- and interpersonal competencies for improvement and evaluation purposes.

Recommendation 12: To broaden understanding of how assessments of intra- and interpersonal competencies can lead to greater student retention and success, institutions of higher education should study and report on their use of these assessments for improvement purposes (e.g., enhancing student support services, developing underrepresented students' sense of belonging, improving courses, identifying effective programs).

INTRA- AND INTERPERSONAL COMPETENCIES AS COLLEGE OUTCOMES

The intra- and interpersonal competencies of ethics, lifelong learning/career orientation, intercultural/diversity competence, civic engagement/citizenship, communication, and teamwork have been identified as valued outcomes of college education. The committee concluded that to date, only limited research has been conducted on these competencies, so that little is known about whether and under what conditions these competencies are related to students' persistence and success in college.

Recommendation 13: Federal agencies and foundations should invest in research examining whether, and under what conditions, the intra- and interpersonal competencies identified as outcomes for college graduates may also be related to students' persistence and success in college.

CONCLUSION

In conclusion, the committee believes that implementation of its recommendations would provide the research base and practical guidance needed to identify more clearly, develop, and assess those intra- and interpersonal competencies that support student persistence and success in undergraduate education.

References

Accreditation Board for Engineering and Technology. (1996). *Engineering Criteria 2000.* Baltimore, MD: Author.

Accreditation Board for Engineering and Technology. (2015). *Criteria for Accrediting Engineering Programs, 2016-2017*. Baltimore, MD: Author. Available: http://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2016-2017 [November 2015].

Achieve. (2010). *Strategies for K-12 and Postsecondary Alignment*. Washington, DC: Author. Available: http://www.achieve.org/files/StrategiesforK-12andPostsecondaryAlignment.pdf [August, 2016].

ACT, Inc. (2012). *Engage Case Study: University Implements Engage to Increase Success of First-Semester College Students*. Iowa City, IA: Author. Available: http://www.act.org/content/dam/act/unsecured/documents/ENGAGE_Case_Study_U_North_Texas.pdf [May 2016].

Aggarwal, P., and O'Brien, C.L. (2008). Social loafing on group projects structural antecedents and effect on student satisfaction. *Journal of Marketing Education, 30*(3), 255-264.

Alan, S., Boneva, T., and Ertac, S. (2015). *Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit*. Working Paper No. 2015-009. Chicago, IL: Human Capital and Economic Global Working Group.

Alexander, F.K. (2000). The changing face of accountability. *Journal of Higher Education, 71*(4), 411-431.

Allen, J., Robbins, S.B., and Sawyer, R. (2009). Can measuring psychosocial factors promote college success? *Applied Measurement in Education, 23*(1)*,* 1-22.

Alverno College. (2016). *Our Unique Curriculum*. Milwaukee, WI: Author. Available: http://www.alverno.edu/academics/ouruniquecurriculum [May 2016].

American Chemical Society. (2015). *Undergraduate Professional Education in Chemistry: 2015 ACS Guidelines for Bachelor's Degree Programs.* Washington, DC: Author. Available: http://www.acs.org/content/dam/acsorg/about/governance/committees/training/2015-acs-guidelines-for-bachelors-degree-programs.pdf [November 2015].

American Educational Research Association, American Psychological Association, and National Council of Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.

American Management Association. (2010). *AMA 2010 Critical Skills Survey*. Temple Hills, MD: Author. Available: http://www.amanet.org/training/articles/3727.aspx [January 2016].

Aronson, J., Fried, C.B., and Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*(2)*,* 113-125.

Arum, R., and Roksa, J. (2011). *Academically Adrift: Limited Learning on College Campuses.* Chicago, IL: University of Chicago Press.

Association of American Colleges and Universities. (2007). *College Learning for the New Global Century.* Washington, DC: Author. Available: https://www.aacu.org/sites/default/files/files/LEAP/GlobalCentury_final.pdf [December 2016].

Astin, A.W. (2012). *Assessment for Excellence: The Philosophy and Practice of Assessment and Evaluation in Higher Education*. Plymouth, UK: Rowman and Littlefield Publishers.

Atkinson, R.C., and Geiser, S. (2009). Reflections on a century of college admissions tests. *Educational Researcher*, *38*(9), 665-676.

Autor, D. (2014). Skills, education, and the rise of earnings inequality among the "other 99 percent." *Science*, *344*(6186), 843-851. Available: http://science.sciencemag.org/content/344/6186/843 [January 2016].

Autor, D., Levy, F., and Murnane, R. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, *118*(4), 1279-1333.

Bago d'Uva, T., Lindeboom, M., O'Donnell, O., and van Doorslaer, E. (2011). Slipping anchor? Testing the vignettes approach to identification and correction of reporting heterogeneity. *Journal of Human Resources, 46*(4), 872-903.

Bailey, D., Duncan, G.J., Odgers, C., and Yu, W. (2015). *Persistence and Fadeout in the Impacts of Child and Adolescent Interventions.* Life Course Centre Working Paper No. 2015-27. Brisbane, Australia: The University of Queensland. Available: http://www.lifecoursecentre.org.au/wp-content/uploads/2015/11/2015-27-LCC-Working-Paper-Bailey-et-al.1.pdf [August 2016].

Baker, G. (2012). *Texas A&M International University: A Culture of Assessment INTEGRATEd*. Champaign, IL: National Institute for Learning Outcomes Assessment. Available: http://www.learningoutcomeassessment.org/documents/TAMIU.pdf [February 2017].

Bandura, A., and Schunk, D.H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology, 41*(3)*,* 586-598.

Banta, T.W., and Blaich, C.F. (2011). Closing the assessment loop. *Change: The Magazine of Higher Learning, 43*(1), 22-27.

Barrick, M.R., and Mount, M.K. (1993). Autonomy as a moderator of the relationship between the Big Five personality dimensions and job performance. *Journal of Applied Psychology, 78*(1)*,* 111-118.

Beaty, J.C., Nye, C.D., Borneman, M.J., Drasgow, F., Kantrowitz, T.M., and Grauer, E. (2011). Proctored vs. unproctored Internet tests: Are unproctored non-cognitive tests as

predictive of job performance? *International Journal of Selection and Assessment, 19*(1), 1-10.

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

Bedwell, W.L., Fiore, S.M., and Salas, E. (2014). Developing the future workforce: An approach for integrating interpersonal skills in the MBA classroom. *Academy of Management Learning & Education, 13*(2), 171-186.

Beigman Klebanov, B., Burstein, J., Harackiewicz, J., Priniski, S., and Mulholland, M. (2016). Enhancing STEM motivation through personal and communal values: NLP for assessment of utility value in student writing. *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications,* 199-205. Available: https://www.semanticscholar.org/paper/Enhancing-STEM-Motivation-through-Personal-and-Klebanov-Burstein/890444c3308b493fe5b855f0651b57fb4d0b84c0/pdf [August 2016].

Bender, B.E., and Schuh, J.H. (Eds.). (2002). *Using Benchmarking to Inform Practice in Higher Education*. San Francisco, CA: Jossey Bass.

Bennett, W.J., and Wilezol, D. (2013). *Is College Worth It? A Former United States Secretary of Education and a Liberal Arts Graduate Expose the Broken Promise of Higher Education.* Nashville, TN: Thomas Nelson.

Bensimon, E. M. (2007). The underestimated significance of practitioner knowledge in the scholarship on student success. *The Review of Higher Education, 30* (4), 441–469.

Bess, J.L., and Dee, J.R. (2008). *Understanding College and University Organization: Dynamics of the System* (Vol. 2). Sterling, VA: Stylus Publishing, LLC.

Bettinger, E.P. and Baker, R.B. (2014). The effects of student coaching: An evaluation of a randomized experiment in student advising. *Educational Evaluation and Policy Analysis, 36*(1), 3-19. Available: http://epa.sagepub.com/content/early/2013/09/17/0162373713500523.full.pdf [August 2016].

Bialik, M., Martin, J., Mayo, M., and Trilling, B. (2016). *Evolving Assessments for a 21st Century Education*. Washington, DC: Assessment Research Consortium, Center for Curriculum Redesign.

Birnbaum, R. (1989). *How Colleges Work: The Cybernetics of Academic Organization and Leadership.* San Francisco, CA: Jossey-Bass.

Blackwell, L.S., Trzesniewski, K.H., and Dweck, C.S. (2007). Implicit theories of intelligence predict achievement scores across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*(1), 246-263.

Blaich, C.F., and Wise, K.S. (2010). Moving from assessment to institutional improvement. *New Directions for Institutional Research*, *2010*(S2), 67-78.

Bollen, K.A., and Hoyle, R.A. (1990). Perceived cohesion: A conceptual and empirical examination. *Social Forces, 69*(2), 479-504.

Bong, M., Lee, S.K., and Woo, Y-K. (2015). The role of interest and self-efficacy in pursuing mathematics and science. In K.A. Renninger, M. Nieswandt, and S. Hidi (Eds.), *Interest in Learning Mathematics and Science* (pp. 33-48). Washington, DC: American Educational Research Association.

Borden, V.M.H., and Young, J.W. (2008). Measurement validity and accountability for student learning. *New Directions for Institutional Research*, *2008*(S1), 19-37.

Borrego, M., Karlin, J., McNair, L.D., and Beddoes, K. (2013). Team effectiveness theory from industrial and organizational psychology applied to engineering student project teams—A review. *Journal of Engineering Education, 102*(4), 472-512.

Borum, V., and Walker, E. (2012). What makes the difference? Black women's undergraduate and graduate experiences in mathematics. *The Journal of Negro Education, 81*(4), 366-378.

Bouchard, T.J. (2004). Genetic influences on human psychological traits: A survey. *Current Directions in Psychological Science, 13*(4), 148-151.

Bowen, W.G., Chingos, M.M., and McPherson, M.S. (2011). *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton, NJ: Princeton University Press.

Bowman, N.A. (2014). Conceptualizing openness to diversity and challenge: Its relation to college experiences, achievement, and retention. *Innovations in Higher Education, 39*(4), 277-291.

Bowman, N.A. (2016). *Intrapersonal and Interpersonal Competencies Predicting College Engagement, Grades, and Retention.* Presentation prepared for the Committee on Assessing Intrapersonal and Interpersonal Competencies, National Research Council, Irvine, CA, February 10. Available: http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_172576.pdf [August 2016].

Bowman, N.A., Hill, P.L., Denson, N., and Bronkema, R. (2015). Keep on truckin' or stay the course? Exploring grit dimensions as differential predictors of educational achievement, satisfaction, and intentions. *Social Psychological and Personality Science, 6*(6), 639-645.

Brady, S.T., Reeves, S.L., Garcia, J., Purdie-Vaughns, V., Cook, J.E., Taborsky-Barba, S., and Cohen, G.L. (2016). The psychology of the affirmed learner: Spontaneous self-affirmation in the face of stress. *Journal of Educational Psychology, 108*(3)*,* 353-373.

Britton, E., Simper, N., Leger, A., and Stephenson, J. (2015). Assessing teamwork in undergraduate education: A measurement tool to evaluate individual teamwork skills. *Assessment and Evaluation in Higher Education,* 1-20.

Brown, C. (2014). *Patterns of Innovation: Showcasing the Nation's Best in 21st Century Learning*. Washington, DC: P21. Available: http://www.p21.org/storage/documents/exemplars/P21_Patterns_of_Innovation_Final.pdf [August 2016].

Brown, A., and Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, *71*(3), 460-502.

Brown, P., Lauder, H., and Ashton, D. (2010). *The Global Auction: The Broken Promises of Education, Jobs, and Income.* New York: Oxford University Press.

Burke, J.C. (2005). *Achieving Accountability in Higher Education: Balancing Public, Academic, and Market Demands*. San Francisco, CA: Jossey-Bass.

Burnette, J.L., O'Boyle, E.H., VanEpps, E.M., Pollack, J.M., and Finkel, E.J. (2013). Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychological Bulletin, 139*(3)*,* 655-701.

Byrd, C.M., and Chavous, T.M. (2011). Racial identity, school racial climate, and school intrinsic motivation among African American adolescents: The importance of person-context congruence. *Journal of Research on Adolescence, 21*(4), 849-860.

Byrd, C.M., and Chavous, T.M. (2012). The congruence between African American students' racial identity beliefs and their academic climates: Implications for academic motivation and achievement. In J. Sullivan and E. Ismail (Eds.), *African American Racial Identity: Racial and Cultural Dimensions of the Black Experience* (pp. 345-370). Lanham, MD: Lexington Books.

Cabrera, A.F., Nora, A., and Castaneda, M.B. (1993). College persistence: Structural equations modeling test of an integrated model of student retention. *Journal of Higher Education, 64*(2), 123-138.

Cabrera, A.F., Nora, A., Terenzini, P.T., Pascarella, E.T., and Hagedorn, L. (1999). Campus racial climate and the adjustment of students to college: A comparison between White students and African-American students. *Journal of Higher Education, 70*(2), 134-160.

Cahalan, M., Perna, L., Yamashita, M., Ruiz, R., and Franklin, K. (2016). *Indicators of Higher Education Equity in the United States: 2016 Historical Trend Report*. Washington, DC: Pell Institute for the Study of Opportunity in Higher Education, Council for Opportunity in Education and Alliance for Higher Education and Democracy of the University of Pennsylvania.

Campbell, C.M. (2015). Serving a different master: Assessing college educational quality for the public. In M.B. Paulsen (Ed.), *Higher Education: Handbook of Theory and Research* (p. 30). Switzerland: Springer.

Campbell, T.I.D., and Slaughter, S. (1999). Faculty and administrators' attitudes toward potential conflicts of interest, commitment, and equity in university-industry relationships. *Journal of Higher Education*, *70*(3), 309-352.

Carey, K. (2014, July 28). Building a better college ranking system. Wait, Babson beats Harvard? *The New York Times*. Available: https://www.nytimes.com/2014/07/29/upshot/building-a-better-college-ranking-system-babson-beats-harvard.html?_r=0 [December 2016].

Carlone, H., and Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching, 44*(8), 1187-1218.

Carnevale, A.P., Smith, N., and Melton, M. (2011). *STEM: Science, Technology, Engineering, and Mathematics*. Washington, DC: Georgetown University Center on Education and the Workforce. Available: https://cew.georgetown.edu/wp-content/uploads/2014/11/stem-complete.pdf [August 2016].

Casner-Lotto, J., and Barrington, L. (2006). *Are They Really Ready to Work?* Washington, DC: Conference Board, Partnership for 21st Century Skills, Corporate Voices for Working Families, and Society for Human Resource Management. Available: http://files.eric.ed.gov/fulltext/ED519465.pdf [August 2016].

Cattell, R.B. (1944). Psychological measurement: Normative, ipsative, and interactive. *Psychological Review, 51*(5)*,* 292-303.

Cengage Learning. (2015). *Cengage Learning CSFI 2.0*. Available: http://www.cengage.com/csfi/#outcomes [May 2016].

Center for Advanced Research on Language Acquisition. (2016). *Theory of Assessment*. Available: http://carla.umn.edu/assessment/vac/research/theory.html [August 2016].

Chamorro-Premuzic, T., and Furnham, A. (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality, 37*(4)*,* 319-338.

Chang, M.J., Eagan, M.K., Lin, M.H., and Hurtado, S. (2011). Considering the impact of racial stigmas and science identity: Persistence among biomedical and behavioral science aspirants. *The Journal of Higher Education*, *82*(5), 564-596.

Chang, M.J., Sharkness, J., Hurtado, S., and Neuman, C. (2014). What matters in college for retaining aspiring scientists and engineers from underrepresented racial groups. *Journal of Research on Science Teaching, 51*(5), 555-580.

Chapman, K.J., and van Auken, S. 2001. Creating positive group project experiences: An examination of the role of the instructor on students' perceptions of group projects. *Journal of Marketing Education, 23*(2), 117-127.

Chatterji, M. (2005). Evidence on "what works": An argument for extended-term mixed-method (ETMM) evaluation designs. *Educational Researcher, 34*(5), 14-24.

Chemers, M.M., Hu, L., and Garcia, B.F. (2001). Academic self-efficacy and first-year college student performance and adjustment. *Journal of Educational Psychology, 93*(1)*,* 55-64.

Chemers, M.M., Zurbriggen, E.L., Syed, M., Goza, B.K., and Bearman, S. (2011). The role of efficacy and identity in science career commitment among underrepresented minority students. *Journal of Social Issues*, *67*(3), 469-491.

Chen, G., Donahue, L.M., and Klimoski, R.J. (2004). Training undergraduates to work in organizational teams. *Academy of Management Learning and Education, 3*(1), 27-40.

Chen, L., Yoon, S.-Y., Leong, C.W., Martin, M., and Ma, M. (2014). *An Initial Study on Analyzing Structured Video Interviews by Using Multimodal Emotion Detection*. Proceedings of the Workshop of Emotion Representations and Modeling for HCI Systems (ERM4HCI), Istanbul, Turkey.

Cistone, P.J., and Bashford, J. (2002). Toward a meaningful institutional effectiveness plan: Learning from accreditation. *Planning for Higher Education, 30*(4), 15-23.

Cleary, A. (1966). *Test Bias: Validity of the Scholastic Aptitude Test for Negro and White Students in Integrated Colleges*. Research Bulletin RB-66-31. Princeton, NJ: Educational Testing Service.

Cohen, G.L., Steele, C.M., and Ross, L.D. (1999). The mentor's dilemma: Providing critical feedback across the racial divide. *Personality and Social Psychology Bulletin, 25*, 1302-1318.

Cohen, G.L. and Garcia, J. (2005). Negative stereotypes as collective threats. *Journal of Personality and Social Psychology, 89,* 566-582.

Connelly, B.S., and Ones, D.S. (2010). Another perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin, 136*(6)*,* 1092-1122.

Converse, P.D., Oswald, F.L., Imus, A., Hedricks, C., Roy, R., and Butera, H. (2008). Comparing personality test formats and warnings: Effects on criterion-related validity and test-taker reactions. *International Journal of Selection and Assessment, 16*(2)*,* 155-169.

Corker, K.S., Oswald, F.L., and Donnellan, M.B. (2012). Conscientiousness in the classroom: A process explanation. *Journal of Personality, 80*(4)*,* 995-1027.

Cortes, C., Klaas, C., and Stachler, S. (2014). *The DIAMOND Project: An Overview of Depaul's Introduction of Noncognitive Measures in Freshman Admission.* PowerPoint presentation prepared for Ritsumeikan Asia Pacific University. Available: https://irma.depaul.edu/irmadocs/DIAMONDVisit4-17-14final.pdf [November 2016].

Credé, M., Tynan, M.C., and Harms, P.D. (2016). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology* [Epub ahead of print].

Cronbach, L.J., Gleser, G.C., Nanda, H., and Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements*. New York: Wiley.

Darling, E., Molina, K., Sanders, M., Lee, F., and Zhao, Y. (2008). Belonging and achieving: The role of identity integration. In M. Maehr, S. Karabenick, and T. Urdan (Eds.), *Advances in Motivation and Achievement: Social Psychological Perspective on Motivation and Achievement* (Vol. 15). New York: Elsevier.

Darling-Hammond, L., Wilhoit, G., and Pittenger, L. (2014). Accountability for college and career readiness: Developing a new paradigm. *Education Policy Analysis Archives*, *22*(86), 1-38.

DeAngelo, L., Franke, R., Hurtado, S., Pryor, J.H., and Tran, S. (2011). *Completing College: Assessing Graduation Rates at Four-Year Institutions*. Los Angeles, CA: Higher Education Research Institute.

de Ayala, R.J. (2009). *The Theory and Practice of Item Response Theory.* New York: Guilford Press.

Deci, E.L., and Ryan, R.M. (1991). A motivational approach to self: Integration in personality. In R. Dienstbier and R.M. Ryan (Eds.), *Nebraska Symposium on Motivation: 1990* (pp. 237-288). Lincoln, NE: University of Nebraska Press.

DeLuca, S., and Rosenbaum, J.E. (2001). Individual agency and the life course: Do low SES students get less long-term payoff for their school efforts? *Sociological Focus, 34*(4), 357-376.

Deming, D. (2015). *The Growing Importance of Social Skills in the Labor Market*. NBER Working Paper 21473. Available: http://www.nber.org/papers/w21473 [January 2016].

De Young, C.G., Quilty, L.C., and Peterson, J.B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*(5)*,* 880-896.

Dilley, A., Fishlock, J., and Plucker, J.A. (2015). *What We Know about Communication* (P21 Research Series). Washington, DC: P21. Available: http://www.p21.org/our-work/4cs-research-series/communication [August 2016].

DiPrete, T.A., and Buchmann, C. (2013). *The Rise of Women: The Growing Gender Gap in Education and What It Means for American Schools.* New York: Russell Sage Foundation.

Dorans, N.J., and Holland, P.W. (1992). *DIF Detection and Description: Mantel-Haenszel and Standardization*. ETS Research Report No. RR-92-10. Princeton, NJ: Educational Testing Service.

Dougherty, K.J., and Hong, E. (2005). *State Systems of Performance Accountability for Community Colleges: Impacts and Lessons for Policymakers*. Achieving the Dream Policy Brief. Boston, MA: Jobs for the Future.

Dougherty, K.J., and Reddy, V. (2011). *The Impacts of State Performance Funding Systems on Higher Education Institutions: Research Literature Review and Policy Recommendations*. CCRC Working Paper No. 37. New York: Community College Research Center, Columbia University. Available: http://ccrc.tc.columbia.edu/media/k2/attachments/impacts-state-funding-higher-education.pdf [April 2016].

Dowd, A.C. (2015). *Accountability, Assessment, and Institutional Change.* Presentation prepared for the Committee on Assessing Intrapersonal and Interpersonal Competencies, Washington, DC, December 14. Available: http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_170240.pdf [August 2016].

Dowd, A.C., and Tong, V.P. (2007). Accountability, assessment, and the scholarship of "best practice." *Higher Education: Handbook of Theory and Research*, *22*, 57-119.

Dowd, A.C., Sawatzky, M., and Korn, R. (2011). Theoretical foundations and a research agenda to validate measures of intercultural effort. *The Review of Higher Education, 35*(1), 17-44.

Downing, S.M., and Haladyna, T.M. (Eds.). (2006). *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Duckworth, A.L., and Yeager, D.S. (2015). Measurement Matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44*(4), 237-251.

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

Duckworth, A.L., Gendler, T.S., and Gross, J.J. (2016). Situational strategies for self-control. *Perspectives on Psychological Science 2016, 11*(1) 35-55.

Dugan, J. (2006). Explorations using the social change model: Leadership development among college men and women. *Journal of College Student Development, 47*(2)*,* 217-225.

Duncan, G., and Magnuson, K. (2011). The nature and impact of early achievement skills, attention skills, and behavior problems. In G.J. Duncan and R.J. Murnane (Eds.), *Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances* (pp. 47-70). New York: Russell Sage.

Duncan, T.C., and McKeachie, W.J. (2005). The making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist, 40*(2), 117-128.

Durik, A.M., Schechter, O.G., Noh, M., Rozek, C.S., and Harackiewicz, J.M. (2015). What if I can't? Success expectancies moderate the effects of utility value information on situational interest and performance. *Motivation and Emotion, 39*(1)*,* 104-118.

Durlak, J.A., and Weissberg, R.P. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405-432.

Durlak, J.A., Weissberg, R.P., and Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology, 45*(3-4)*,* 294-309.

Dweck, C.S. (2006). *Mindset: The New Psychology of Success*. New York: Random House.

Dweck, C.S., and Molden, D.C. (2005). Self-theories: Their impact on competence motivation and acquisition. In A.J. Elliot and C.S. Dweck (Eds.), *Handbook of Competence and Motivation* (pp. 122-140). New York: Guilford Press.

Dwight, S.A., and Donovan, J.J. (2003). Do warnings not to fake reduce faking? *Human Performance, 16*(1)*,* 1-23.

Dwyer, C. A., Millett, C. M., & Payne, D. G. (2006). A culture of evidence: Postsecondary assessment and learning outcomes. Princeton, N.J.: Educational Testing Service.

Eagan, K., Hurtado, S, Figueroa, T., and Hughes, B. (2014). *Examining STEM Pathways among Students Who Begin College at Four-Year Institutions.* Paper prepared for the Committee on Barriers and Opportunities in Completing 2-Year and 4-Year STEM Degrees, National Research Council. Available: http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_088834.pdf [April 2015].

Early, S.L., and Blevins, S.L. (n.d.). *Multi-Institutional Study of Leadership Executive Summary.* Las Vegas, NV: University of Nevada, Las Vegas. Available: https://oit.unlv.edu/nwccu/sites/default/files/2013/S2/App%202.D.11.2%20Multi%20Institutional%20Study%20of%20Leadership%20Exec%20Summary.pdf [August 2016].

Eccles, J. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist, 44*(2)*,* 78-89.

Eccles, J., Adler, T.F., Futterman, R., Goff, S.B., Kaczala, C.M., Meece, J., and Midgley, C. (1983). Expectancies, values and academic behaviors. In J.T. Spence (Ed.), *Achievement and Achievement Motives: Psychological and Sociological Approaches* (pp. 75-146). San Francisco, CA: Freeman.

Educational Testing Service. (2010). *How ETS Works To Improve Test Accessibility*. Princeton, NJ: Author.

Educational Testing Service. (2014). *ETS Standards for Quality and Fairness*. Princeton, NJ: Author. Available: https://www.ets.org/s/about/pdf/standards.pdf [August 2016].

Educational Testing Service. (2015). *About the HEIghten Outcomes Assessment Suite*. Princeton, NJ: Author. Available: https://www.ets.org/heighten/about [November 2015].

Edwards, A.L. (1957). *The Social Desirability Variable in Personality Assessment and Research.* New York: Dryden.

Ehrenberg, R.G. (1999). Adam Smith goes to college: An economist becomes an academic administrator. *The Journal of Economic Perspectives*, *13*(1), 99-116.

Ercikan, K., and Oliveri, M.E. (2016). In search of validity evidence in support of interpretation and use of complex constructs: Discussion of research on assessing 21st century skills. *Applied Measurement in Education, 29*(4), 310-318.

Erez, A., LePine, J.A., and Elms, H. (2002). Effects of rotated leadership and peer evaluation on the functioning and effectiveness of self-managed teams: A quasi-experiment. *Personnel Psychology, 55*(4), 929-948.

Ericsson, K.A., and Simon, H.A. (1993). *Protocol Analysis: Verbal Reports as Data*. Boston, MA: MIT Press.

Erikson, E.H. (Ed.). (1963). *Youth: Change and Challenge*. New York: Basic Books.

Eskreis-Winkler, L., Young, V., Brunwasser, S.M., Shulman, E.P., Tsukayama, E., and Duckworth, A.L. (2016). Using wise interventions to motivate deliberate practice. *Journal of Personality and Social Psychology, 111*(5), 728-744.

Estrada, M. (2013). *Ingredients for Improving the Culture of STEM Degree Attainment with Co-Curricular Supports for Underrepresented Minority Students.* Available: http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_088832.pdf [August 2016].

Estrada, M., Woodcock, A., Hernandez, P.R., and Schultz, P.W. (2011). Toward a model of social influence that explains minority student integration into the scientific community. *Journal of Educational Psychology, 103*(1), 206-222.

Ewell, P.T. (2002). An emerging scholarship: A brief history of assessment. In T.W. Banta and Associates (Eds.), *Building a Scholarship of Assessment* (pp. 3-25). San Francisco, CA: Jossey-Bass.

Ewell, P.T. (2008). Assessment and accountability in America today: Background and context. *New Directions for Institutional Research*, *2008*(S1), 7-17.

Ewell, P.T. (2009). *Assessment, Accountability, and Improvement: Revisiting the Tension*. Champaign, IL: National Institute for Learning Outcomes Assessment.

Fain, P. (2015). Measuring motivation. *Inside Higher Education*, September 28. Available: https://www.insidehighered.com/news/2015/09/28/nonacademic-skills-test-ets-fills-blanks-students-likelihood-success [May 2016].

Falchikov, N. (2013). *Improving Assessment through Student Involvement: Practical Solutions for Aiding Learning in Higher and Further Education*. New York: Routledge.

Falk, C.F., and Cai, L. (2016). A flexible full-information approach to the modeling of response styles. *Psychological Methods, 21*(3), 328-347.

Fitch, T., Marshall, J., and McCarthy, W. (2012). The effect of solution-focused groups on self-regulated learning. *Journal of College Student Development, 53*(4)*,* 586-595.

Foldes, H.J., Duehr, E.E., and Ones, D.S. (2008). Group differences in personality: Meta-analyses comparing five U.S. racial groups. *Personnel Psychology, 61*(3)*,* 579-616.

Folger, W.A., Carter, J.A., and Chase, P.B. (2004). Supporting first generation college freshmen with small group intervention. *College Student Journal, 38*(3) 472-475.

Fortenberry, N.L., Sullivan, J.F., Jordan, P.N., and Knight, D.W. (2007). Engineering education research aids instruction. *Science, 317,* 1175-1176. Available: http://itll.colorado.edu/images/uploads/about_us/publications/Papers/SCIENCE%20PUBLISHED%20VERSION%202007Aug31.pdf [February 2016].

Fredericksen, N.O. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist, 39*(3), 193-202.

Freeman, M., and McKenzie, J. (2002). SPARK: A confidential web-based template for self and peer assessment of student teamwork: Benefits of evaluating across different subjects. *British Journal of Educational Technology*, *33*(5), 551-569. Available: http://onlinelibrary.wiley.com/doi/10.1111/1467-8535.00291/epdf [April 2016].

Freudenthaler, H., Spinath, B., and Neubauer, A. (2008). Predicting school achievement in boys and girls. *European Journal of Personality, 22*(3)*,* 231-245.

Freund, A.M., and Riediger, M. (2006). Goals as building blocks of personality and development in adulthood. In D.K. Mroczek and T. Little (Eds.), *Handbook of Personality Development* (pp. 353-372). Mahwah, NJ: Lawrence Erlbaum Associates.

Fuertes, J., Miville, M., Mohr, J., Sedlacek, W., and Gretchen, D. (2000). Factor structure and short form of the Miville-Guzman Universality-Diversity Scale. *Measurement and Evaluation in Counseling and Development, 33*(3)*,* 157-169.

Gelman, A., and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, UK: Cambridge University Press.

Gelman, A., and Loken, E. (2013). *The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There Is No "Fishing Expedition" or "P-Hacking" and the Research Hypothesis was Posited Ahead of Time.* Technical Report. New York:

Department of Statistics, Columbia University. Available: http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf [November 2016].

Gensler, H.J. (1996). *Formal Ethics*. London, UK and New York: Routledge.

Gerhards, L., and Gravert, C. (2015). *Grit Trumps Talent? An Experimental Approach.* Working Paper in Economics No. 629. Göteborg, Sweden: University of Gothenberg.

Gibbons, R. (1998). Incentives in organizations. NBER Working Paper No. 6695. Available: http://www.nber.org/papers/w6695 [February, 2017].

Gierl, M.J., and Haladyna, T.M. (Eds.). (2013). *Automatic Item Generation: Theory and Practice*. New York: Routledge.

Goldberg, L.R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*(1), 26-34. Available: http://projects.ori.org/lrg/PDFs_papers/Goldberg.Am.Psych.1993.pdf [August 2016].

Graham, M. J., Frederick, J., Byars-Winston, A., Hunter, A., and Handelsman, J. (2014). Increasing persistence of college students in STEM. *Science, 341*(6153), 1455-1456.

Griffith, R.L., Wolfeld, L., Armon, B.K., Rios, J., and Liu, O.L. (2016). *Assessing Intercultural Competence in Higher Education: Existing Research and Future Directions.* ETS Research Report, Early View Article. Available: http://onlinelibrary.wiley.com/doi/10.1002/ets2.12112/abstract;jsessionid=597A7A04A370988C4D4710C53A6885CC.f04t02 [September 2016].

Gruber, M.J., Gelman, B.D., and Ranganath, C. (2014). States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron, 84*(2), 486-496. Available: http://www.cell.com/neuron/abstract/S0896-6273(14)00804-6 [August 2016].

Guttenplan, D. (2014). Re-evaluating the college rankings game. *The New York Times,* June 1. Available: https://www.nytimes.com/2014/06/02/world/europe/re-evaluating-the-college-rankings-game.html?_r=0 [December 2016].

Hall, J.M., and Ponton, M.K. (2005). Mathematics self-efficacy of college freshmen. *Journal of Developmental Education*, *28*(3), 26-32. Available: http://files.eric.ed.gov/fulltext/EJ718579.pdf [May 2016].

Hall, N.C., Hladkyj, S., Perry, R. P., and Ruthig, J.C. (2004). The role of attributional retraining and elaborative learning in college students' academic development. *The Journal of Social Psychology, 144*(6), 591-612.

Hall, N.C., Perry, R. P., Chipperfield, J.G., Clifton, R.A., and Haynes, T.L. (2006). Enhancing primary and secondary control in achievement settings through writing-based attributional retraining. *Journal of Social and Clinical Psychology, 25*(4), 361-391

Hall, N.C., Perry, R.P., Goetz, T., Ruthig, J.C., Stupnisky, R.H., and Newall, N.E. (2007). Attributional retraining and elaborative learning: Improving academic development through writing-based interventions. *Learning and Individual Differences, 17*(3), 280-290.

Hamm, J.M., Perry, R.P., Clifton, R.A., Chipperfield, J.G., and Boese, G.D. (2014). Attributional retraining: A motivation treatment with differential psychosocial and performance benefits for failure prone individuals in competitive achievement settings. *Basic and Applied Social Psychology, 36*(3), 221-237.

Hanson, C. (Ed.). (2014). *In Search of Self: Exploring Undergraduate Identity*. San Francisco, CA: Jossey-Bass.

Harackiewicz, J.M., Barron, K.E., Tauer, J.M., Carter, S.M., and Elliot, A.J. (2000). Short-term and long-term consequences of achievement goals: Predicting interest and performance over time. *Journal of Educational Psychology*, *92*(2)*,* 316-330.

Harackiewicz, J.M., Barron, K.E., Tauer, M.M., and Elliot, A.J. (2002). Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation. *Journal of Educational Psychology, 94*(3)*,* 562-575.

Harackiewicz, J.M., Rozek, C.S., Hulleman, C.S., and Hyde, J.S. (2012). Helping parents to motivate adolescents in mathematics and science: An experimental test of a utility-value intervention. *Psychological Science, 23*(8), 899-906.

Harackiewicz, J.M., Canning, E.A., Tibbets, Y., Giffen, C.J., Blair, S.S., Rouse, D.I., and Hyde, J.S. (2014). Closing the social class achievement gap for first-generation students in undergraduate biology. *Journal of Educational Psychology, 106*(2)*,* 375-389.

Harackiewicz, J.M., Canning, E.A., Tibbetts. Y., Priniski, S.J., and Hyde, J.S. (2015). Closing achievement gaps with a utility-value intervention: Disentangling race and social class. *Journal of Personality and Social Psychology* [Epub ahead of print].

Harrison, L.A., Stevens, C. M., Monty, A.M., and Coakley, C.A. (2006). The consequences of stereotype threat on the academic performance of white and non-white lower income college students. *Social Psychology of Education, 9*(3)*,* 341-357.

Hart Research Associates. (2010). *Raising the Bar: Employers' Views on College Learning in the Wake of the Economic Downturn. A Survey among Employers Conducted on behalf of the American Association of Colleges and Universities.* Washington, DC: American Association of Colleges and Universities. Available: https://www.aacu.org/sites/default/files/files/LEAP/2009_EmployerSurvey.pdf [August 2016].

Hart Research Associates. (2015). *Falling Short? College Learning and Career Success*. Washington, DC: American Association of Colleges and Universities. Available: http://www.aacu.org/sites/default/files/files/LEAP/2015employerstudentsurvey.pdf [May 2016].

Hart Research Associates. (2016). *Trends in Learning Outcomes Assessment: Key Findings from a Survey among Administrators at AAC&U Member Institutions.* Washington, DC: American Association of Colleges and Universities.

Hastie, C., Fahy, K., and Parratt, J. (2014). The development of a rubric for peer assessment of individual teamwork skills in undergraduate midwifery students. *Women and Birth, 27*(3), 220-226.

Hausmann, L., Ye, R., Schofield, J., Woods, R. (2009). Sense of belonging and persistence in white and African American first-year students. *Research in Higher Education*, *50*(7), 649-669.

Haynes, M.C., and Heilman, M.E. (2013). It had to be you (not me)! Women's attributional rationalization of their contribution to successful joint work outcomes. *Personality and Social Psychology Bulletin*, *39*(7), 956-969.

Haynes, T.L., Daniels, L.M., Stupnisky, R.H., Perry, R.P., and Hladkyj, S. (2008). The effect of attributional training on mastery and performance motivation among first-year college students. *Basic and Applied Psychology, 30*(3), 198-207.

He, J., and van de Vijver, F.J.R. (2015). Self-presentation styles in self-reports: Linking the general factors of response styles, personality traits, and values in a longitudinal study. *Personality and Individual Differences, 81,* 129-134.

Heggestad, E.E., Morrison, M., Reeve, C., and McCloy, R.A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology, 91*(1), 9-24.

Henderson, C., Beach, A., and Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *Journal of Research in Science Teaching*, *48*(8), 952-984.

Hernandez, P.R., Schultz, P.W., Estrada, M., Woodcock, A., and Chance, R.C. (2013). Sustaining optimal motivation: A longitudinal analysis of interventions to broaden participation of under-represented minorities in STEM. *Journal of Educational Psychology, 105*(1).

Hidi, S., and Renninger, K.S. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111-127.

Hill, P.L., and Jackson, J. (2016). The invest-and-accrue model of conscientiousness. *Personality and Social Psychology Review, 20*(2), 141-154.

Hughes, G. (2014). *Ipsative Assessment: Motivation through Marking Progress*. London, UK: Palgrave Macmillan UK.

Hughes, K. (2013). *The College Completion Agenda 2012 Progress Report*. Princeton, NJ: The College Board. Available: http://media.collegeboard.com/digitalServices/pdf/advocacy/policycenter/college-completion-agenda-2012-progress-report.pdf [January 2016].

Hughes, R.L., and Jones, S.K. (2011). Developing and assessing college student teamwork skills. *New Directions for Institutional Research*, *2011*(149), 53-64.

Hulleman, C.S., Godes, O., Hendricks, B.L., and Harakciewicz, J.M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology, 102*(4)*,* 880-895.

Hurtado, S., and Carter, D.F. (1997). Effects of college transition and perceptions of the campus racial climate on Latino students' sense of belonging. *Sociology of Education, 70*(4)*,* 324-345.

Hurtado, S., Carter, D.F., and Kardia, D. (1998). The climate for diversity: Key issues for institutional self study. *New Directions for Institutional Research, 1998*(98), 53-63.

Hurtado, S., and Guillermo-Wan, C. (2013). *Diverse Learning Environments: Assessing and Creating Conditions for Student Success.* Final report to the Ford Foundation. Los Angeles, CA: Higher Education Research Institute, University of California, Los Angeles. Available: http://www.heri.ucla.edu/ford/DiverseLearningEnvironments.pdf [November 2016].

Hurtado, S., and Ruiz Alvarado, A. (2016). *Theory and Research on Persistence and Sense of Belonging.* PowerPoint prepared for the Committee on Assessing Intrapersonal and Interpersonal Competencies, Irvine, CA, February 10-11.

Hurtado, S., Han, J.C., Sáenz, V.B., Espinosa, L., Cabrera, N., and Cerna, O. (2007). Predicting transition and adjustment to college: Biomedical and behavioral science aspirants' and minority students' first year of college. *Research in Higher Education*, *48*(7), 841-887.

International Test Commission. (2014). ITC guidelines on quality control in scoring, test analysis, and reporting of test scores*. International Journal of Testing, 14*(3)*,* 195-217.

Irvine, S.H., and Kyllonen, P.C. (2002). *Item Generation for Test Development.* Mahwah, NJ: Lawrence Erlbaum Associates.

Jackson, J.J., and Roberts, B.W. (2015). Conscientiousness. In T.A. Widiger (Ed.), *The Oxford Handbook of the Five Factor Model* (pp. 1-32). Oxford, UK: Oxford University Press.

Jang, K.L., Livesly, W.J., and Vernon, P.A. (1996). Heritability of the big five personality dimensions and their facets: A twin study. *Journal of Personality*, *64*(3), 577-591.

Johnson, D.R., Soldner, M., Leonard, J.B., Alvarez, P., Inkelas, K.K., Rowan-Kenyon, H.T., and Longerbeam, S.D. (2007). Examining sense of belonging among first-year undergraduates from different racial/ethnic groups. *Journal of College Student Development, 48*(5)*,* 525-542.

Johnstone, C.J., Thompson, S.J., Bottsford-Miller, N.A., and Thurlow, M.L. (2008). Universal design and multimethod approaches to item review. *Educational Measurement: Issues and Practice, 27*(1)*,* 25-36.

Joshi, A. (2014). By whom and when is women's expertise recognized? The interactive effects of gender and education in science and engineering teams. *Administrative Science Quarterly*, *59*(2), 202-239.

Kalshoven, K., Hartog, D.N.D., and De Hoogh, A.H.B. (2011). Ethical leader behavior and big five factors of personality. *Journal of Business Ethics, 100*(2)*,* 349-366.

Kaminer, A. (2013). Lists that rank colleges' value are on the rise. *The New York Times,* October 27.

Kane, M.T. (2006). Validation. In R.L. Brennan (Ed.), *Educational Measurement* (4th Edition) (pp. 17-64). Westport, CT: American Council on Education/Praeger Publishers.

Kapteyn, A., Smith, J.P., van Soest, A., and Vonkova, H. (2011). *Anchoring Vignettes and Response Consistency.* RAND Working Paper WR-840. Available: https://www.rand.org/content/dam/rand/pubs/working_papers/2011/RAND_WR840.pdf [July 2016].

Karabenick, S.A., and Urdan, T.C. (Eds.). (2014). *Motivational Interventions: Advances in Motivation and Achievement* (Vol. 18). Bingley, UK: Emerald Group Publishing Limited.

Kell, H., Martin-Raugh, M., and Carney, L. (in press). Applicant interviews: A meta-analysis of predictive features.

Kelley, T.L. (1927). *Interpretation of Educational Measurements.* Yonkers-on-Hudson, NY: World Book Company.

Kena, G., Hussar W., McFarland J., de Brey C., Musu-Gillette, L., Wang, X., Zhang, J., Rathbun, A., WilkinsonFlicker, S., Diliberti M., Barmer, A., Bullock Mann, F., and Dunlop Velez, E. (2016). *The Condition of Education 2016* (NCES 2016-144). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Available: http://nces.ed.gov/pubs2016/2016144.pdf [August 2016].

Kezar, A. (2001). Understanding and facilitating organizational change in the 21st century. *ASHE-ERIC Higher Education Report*, *28*(4), 147.

King, G., and Wand, J. (2007). Comparing incomparable survey responses: New tools for anchoring vignettes. *Political Analysis*, *15*(1), 46-66.

King, G., Murray, C.J.L., Salomon, J.A., and Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review, 98*(1)*,* 191-207.

Klieger, D.M., Cline, F.A., Holtzman, S.L., Minsky, J.L., and Lorenz, F. (2014). *New Perspectives on the Validity of the GRE® General Test for Predicting Graduate School*

Grades. GRE Board Research Report No. 14-03, ETS Research Report No. RR-14-26. Princeton, NJ: Educational Testing Service.

Klieger, D.M., Ezzo, C., Bochenek, J.L., and Cline, F.A. (2015). The predictive validity of non-cognitive skills for graduate and professional student success: Some initial findings. In K. Marino (Chair), *Validity and Fairness Issues in Assessing Higher Education Students*. Symposium conducted at the annual meeting of the American Education Research Association (AERA), Chicago, IL.

Kozlowski, S.W.J., and Ilgen, D.R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest, 7*(3), 77-124.

Kuh, G.D., Jankowski, N., Ikenberry, S.O., and Kinzie, J. (2014). *Knowing What Students Know and Can Do: The Current State of Learning Outcomes Assessment at U.S. Colleges and Universities*. Champaign, IL: National Institute for Learning Outcomes Assessment. Available: http://www.learningoutcomeassessment.org/documents/2013%20Abridged%20Survey%20Report%20Final.pdf [March 2016].

Kulterel-Konak, S., Konak, A., Kremer, G.E.O., Esparragoza, I., and Yoder, G. (2014). Peer Evaluation and Assessment Resource (PEAR) to assess students' professional skills. In Y. Guan and H. Liao (Eds.), *Proceedings of the 2014 Industrial and Systems Engineering Research Conference,* Montréal, Canada.

Kuncel, N.R., Kochevar, R.J., and Ones, D.S. (2014). A meta-analysis of letters of recommendation in college and graduate admissions: Reasons for hope. *International Journal of Selection and Assessment, 22*(1)*,* 101-107.

Kyllonen, P.C. (2008). *The Research Behind the ETS Personal Potential Index (PPI)*. Princeton, NJ: Educational Testing Service.

Kyllonen, P.C., and Bertling, J.P. (2013). Innovative questionnaire assessment methods to increase cross-country comparability. In L. Rutkowski, M. von Davier, and D. Rutkowski (Eds.), *Handbook of International Large-Scale Assessment Data Analysis*. London, UK: Chapman and Hall/CRC.

Kyoung Ro, H., and Loya, K. (2015). The effect of gender and race intersectionality on student learning outcomes in engineering. *Review of Higher Education, 38*(3), 359-396.

Laitusis, C.C. (2007). Validity and accommodations: A variety of approaches to accessible assessments. In C.C. Laitusis, and L.L. Cook (Eds.), *Large-Scale Assessments and Accommodations: What Works?* (pp. 73-81). Washington, DC: Council for Exceptional Children.

Lambert, D., and Lines, D. (2013). *Understanding Assessment: Purposes, Perceptions, Practice*. New York: Routledge.

Landivar, L.C. (2013). *Disparities in STEM Employment by Sex, Race, and Hispanic Origin.* Report Number ACS-24. Available: http://www.census.gov/library/publications/2013/acs/acs-24.html# [June 2016].

Lane, S., Raymond, M.R., and Haladyna, T.M. (2015). *Handbook of Test Development* (2nd Edition). New York: Routledge.

Landau, M.J., Oyserman, D., Keefer, L.A., and Smith, G.C. (2014). The college journey and academic engagement: How metaphor use enhances identity-based motivation. *Journal of Personality and Social Psychology, 106*(5)*,* 679-698.

Lattuca, L.R., Terenzini, P.T., and Volkwein, J.F. (2006). *Engineering Change: A Study of the Impact of EC 2000*. Baltimore, MD: ABET, Inc. Available: http://www.abet.org/wp-content/uploads/2015/04/EngineeringChange-executive-summary.pdf [March 2016].

Le, H., Casillas, A., Robbins, S.B., and Langley, R. (2005). Motivational and skills, social, and self-management predictors of college outcomes: Constructing the student readiness inventory. *Educational and Psychological Measurement, 65*(3), 482-508.

Leighton, J., Cui, Y., and Kor, M.K. (2009). Testing expert-based and student-based cognitive models: An application of the attribute hierarchy methods and hierarchy consistency index. *Applied Measurement in Education, 22*(3)*,* 229-254.

Lent, R.W., Brown, S.D., Schmidt, J., Brenner, B., Lyons, H., and Treistman, D. (2003). Relation of contextual supports and barriers to choice behavior in engineering majors: Test of alternative social cognitive models. *Journal of Counseling Psychology*, *50*(4), 458-465.

Lent, R.W., Brown, S.D., Sheu, H.-B., Schmidt, J., Brenner, B.R., Gloster, C.S., Wilkins, G., Schmidt, L.C., Lyons, H., and Treistman, D. (2005). Social cognitive predictors of academic success and goals in engineering: Utility for women and students at historically black universities. *Journal of Counseling Psychology, 52*(1)*,* 84-92.

Levy, F.J., and Murnane, R.J. (2004). *The New Division of Labor: How Computers Are Creating the Next Job Market.* Princeton, NJ: Princeton University Press.

Levy, F.J., and Murnane, R.J. (2013). *Dancing with Robots: Human Skills for Computerized Work*. Washington, DC: Third Way. Available: http://www.thirdway.org/report/dancing-with-robots-human-skills-for-computerized-work [January 2016].

Linn, R.L. (2009). Comments on Atkinson and Geiser: Considerations for college admissions testing. *Educational Researcher*, *38*(9), 677-679.

Liu, O.L., Bridgeman, B., and Adler, R.M. (2012). Measuring learning outcomes in higher education: Motivation matters. *Educational Researcher, 41*(9)*,* 352-362.

Liu, O.L., Torney-Purta, J., and Roohr, K. (2015). *Innovations in Conceptualizing and Assessing Civic Competency in Higher Education*. Forum presentation at Educational Testing Service, Washington, DC, November 4.

Locks, A.M., Hurtado, S., Bowman, N.A., and Oseguera, L. (2008). Extending notions of campus climate and diversity to students' transition to college. *The Review of Higher Education, 31*(3), 257-285.

Lord, F.M. (1955). Estimating test reliability. *Educational and Psychological Measurement, 16,* 325-336.

Loughry, M.L., Ohland, M.W., and Moore, D.D. (2007). Development of a theory-based assessment of team member effectiveness. *Educational and Psychological Measurement*, *67*(3), 505-524.

Lowell, B.L. and Salzman, H. (2007). *Into the Eye of the Storm: Assessing the Evidence on Science and Engineering Education, Quality, and Workforce Demand.* Washington, DC: The Urban Institute. Available: http://www.urban.org/research/publication/eye-storm [September 2016].

Luecht, R. (2013). Assessment engineering task model maps, task models and templates as a new way to develop and implement test specifications. *Journal of Applied Testing Technology*, *14*(S1). Available: http://www.testpublishers.org/jatt-volume-14 [December 2016].

Lumina Foundation. (2015). *The Degree Qualifications Profile: A Learning-Centered Framework for What College Graduates Should Know and Be Able to Do to Earn the Associate's, Bachelor's or Master's Degree.* Indianapolis, IN: Lumina Foundation. Available: https://www.luminafoundation.org/files/resources/dqp.pdf [November 2015].

Lundberg, S. (2013). The college type: personality and educational inequality. *Journal of Labor Economics, 31*(3), 421-441.

Luzzo, D.A., Hasper, P., Albert, K.A., Bibby, M.A., and Martinelli, Jr., E.A. (1999). Effects of self-efficacy enhancing interventions on the math/science self-efficacy and career interests, goals, and actions of career undecided college students. *Journal of Counseling Psychology, 46*(2), 233-243.

Maki, P.L. (2010). *Assessing for Learning: Building a Sustainable Commitment across the Institution* (2nd Edition). Sterling, VA: Stylus Publishing.

Maki, P.L. (2015). *Assessment that Works: A National Call, a Twenty-First-Century Response.* Washington, DC: American Association of Colleges and Universities.

Magidson, J.F., Roberts, B.W., Collado-Rodriguez, A., and Lejuez, C.W. (2014). Theory-driven intervention for changing personality: Expectancy value theory, behavioral activation, and conscientiousness. *Developmental Psychology, 50*(5)*,* 1442-1450.

Malcom, L., and Malcom, S. (2011). The double bind: The next generation. *Harvard Educational Review, 81*(2), 388-389.

Mann, A., and DiPrete, T.A. (2013). Trends in gender segregation in the choice of science and engineering majors. *Social Science Research, 42*(15), 19-41.

Markle, R., Olivera-Aguilar, M., Jackson, T., Noeth, R., and Robbins, S. (2013a). *Examining Evidence of Validity, Reliability, and Fairness for the SuccessNavigator Instrument*. Princeton, NJ: Educational Testing Service. Available: http://www.ets.org/research/policy_research_reports/publications/report/2013/jqob [May 2016].

Markle, R., Brenneman, M., Jackson, T., Burrus, J. and Robbins, S. (2013b). *Synthesizing Frameworks of Higher Education Learning Outcomes.* RR-13-22. Princeton, NJ: Educational Testing Service. Available: http://www.ets.org/research/policy_research_reports/publications/report/2013/jrkk [November 2015].

Markus, H., and Nurius, P. (1986). Possible selves. *American Psychologist, 41*(9)*,* 954-961.

Martens, A., Johns, M., Greenberg, J., Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology, 42*(2), 236-243. Available: http://www.ir.canterbury.ac.nz/bitstream/handle/10092/507/stereotype_threat_martens.pdf?sequence=2&isAllowed=y [August 2016].

Massey, D.S., Charles, C.Z., Lundy, G.F., and Fischer, M.J. (2003). *The Source of the River: The Social Origins of Freshmen at America's Selective Colleges and Universities*. Princeton, NJ: Princeton University Press.

Matsudaira, J. (2015). *Defining and Measuring Institutional Quality in Higher Education*. Paper prepared for the National Academies of Sciences, Engineering, and Medicine Workshop on Quality in the Undergraduate Experience, Washington, DC, December 14-15.

Matten, K.D., and Shaw, E.J. (2010). A look beyond cognitive predictors of academic success: Understanding the relationship between academic self-beliefs and outcomes. *Journal of College Student Development, 51*(6)*,* 665-678.

Mattern, K.D., Patterson, B.F., Shaw, E.J., Korbin, J.L., and Barbuti, S.M. (2008). *Differential Validity and Prediction of the SAT.* Research Report No. 2008-4. New York: The College Board.

McAbee, S.T., and Oswald, F.L. (2013). The criterion-related validity of personality measures for predicting GPA: A meta-analytic comparative-validity competition. *Psychological Assessment, 25*(2)*,* 532-544.

McAdams, D.P. (2015). *The Art and Science of Personality Development*. New York: Guilford Press.

McConnell, D.A., and van der Hoeven Kraft, K. (2011). Affective domain and student learning in the geosciences. *Journal of Geoscience Education, 59*(3), 106-110.

McConnell, D.A., Stempien, J.A., Perkins, D., van der Hoeven Kraft, K.J., Vislova, T., Wirth, K.R., Budd, D.A., Bykerk-Kauffman, A., Gilbert, L.A., and Matheney, R.K. (2010). The little engine that could—less prior knowledge but high self-efficacy is equivalent to greater prior knowledge and low self-efficacy. *The Geological Society of America Abstracts with Programs*, *42*(5), 191.

McCormick, A.C., and McClenney, K. (2012). Will these trees ever bear fruit?: A response to the special issue on student engagement. *The Review of Higher Education, 35*(2), 307-333.

McCrae, R.R., and Costa, P.T., Jr. (2008). The five-factor theory of personality. In O.P. John, R.W. Robins, and L.A. Pervin (Eds.), *Handbook of Personality: Theory and Research* (3rd Edition) (pp. 159-180). New York: Guilford Press.

McDaniel, M.A., Whetzel, D.L., Schmidt, F.L., and Maurer, S.D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology, 79*(4)*,* 599-616.

McDaniel, M.A., Morgeson, F.P., Finnegan, E.B., Campion, M.A., and Braverman, E.P. (2001). Predicting job performance using situational judgment tests: A clarification of the literature. *Journal of Applied Psychology, 86,* 730-740.

McDonald, R. (1999). *Test Theory: A Unified Treatment.* Mahwah, NJ: Lawrence Erlbaum Associates.

McFarland, L. (2003). Warning against faking on a personality test: Effects on applicant reactions and personality test scores. *International Journal of Selection and Assessment, 11*(4)*,* 265-276.

Megivern, D., Pellerito, S., and Mowbray, C. (2003). Barriers to higher education for individuals with psychiatric disabilities. *Psychiatric Rehabilitation Journal, 26*(3)*,* 217-231.

Menec, V. H., Perry, R.P., Struthers, C.W., Schonwetter, D.J., Hechter, F.J., and Eichholz, B.L. (1994). Assisting at-risk college students with attributional retraining and effective teaching. *Journal of Applied Social Psychology, 24,* 675-701.

Mentkowski, M., Rogers, G., Doherty, A., Loacker, G., Hart, J.R., Rickards, W., O'Brien, K., Riordan, T., Sharkey, S., Cromwell, L., Diez, M., Bartels, J., and Roth, J. (2000). *Learning That Lasts: Integrating Learning, Development, and Performance in College and Beyond.* San Francisco, CA: Jossey-Bass.

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice, 14*(4), 5-8.

Middaugh, M.F. (2009). Closing the loop: Linking planning and assessment. *Planning for Higher Education. 37*(3)*,* 4-14.

Middle States Commission on Higher Education. (2015). *Standards for Accreditation and Requirements of Affiliation* (13th Edition). Philadelphia, PA: Author. Available: http://www.msche.org/publications/RevisedStandardsFINAL.pdf [August 2016].

Miles, C.L., and Wilson, C. (2004). Learning outcomes for the twenty first century: Cultivating student success for college and the knowledge economy. *New Directions for Community Colleges, 2004*(126). Available: http://onlinelibrary.wiley.com/doi/10.1002/cc.157/pdf [November 2015].

Mislevy, R.J., and Riconscente, M.M. (2006). Evidence-centered assessment design. In S.M. Downing and T.M. Haladyna (Eds.), *Handbook of Test Development* (pp. 61-90). Mahwah, NJ: Lawrence Erlbaum Associates.

Mislevy, R.J., Almond, R.G., and Lukas, J.F. (2003). *A Brief Introduction to Evidence-Centered Design.* Princeton, NJ: Educational Testing Service.

Mislevy, R.J., Oranje, A., Bauer, M.I., von Davier, A.A., Hao, J., Corrigan, S., Hoffman, E., DiCerbo, K., and John, M. (2014). *Psychometric Considerations in Game-Based Assessment*. Redwood City, CA: Glass Lab Games.

Miville, M., Gelso, C., Pannu, R., Liu, W., Touradji, P., Holloway, P., and Fuertes, J. (1999). Appreciating similarities and valuing differences: The Miville-Guzman Universality-Diversity Scale. *Journal of Counseling Psychology, 46*(3)*,* 291-307.

Miyake, A., Kost-Smith, L.E., Finkelstein, N.D., Pollock, S. J., Cohen, G. L., and Ito, T.A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science, 330*(6008)*,* 1234-1237.

Moore, J., Casillas, A., and Way, J. (2015). *Academic Achievement, Behaviors, and College Completion.* ACT Research and Policy Data Byte 2015-6. Iowa City, IA: ACT. Available: http://www.act.org/content/dam/act/unsecured/documents/2015-6.pdf [November 2016].

Morisano, D., Hirsh, J.B., Peterson, J.B., Pihl, R.O., and Shore, B. M. (2010). Setting, elaborating, and reflecting on personal goals improves academic performance. *Journal of Applied Psychology, 95*(2), 255-264.

Murphy, S.C., Klieger, D.M., Borneman, M.J., and Kuncel, N.R. (2009). The predictive power of personal statements in admissions: A meta-analysis and cautionary tale. *College and University*, *84*(4), 83-86, 88.

National Academies of Sciences, Engineering, and Medicine. (2016a). *Barriers and Opportunities for 2-Year and 4-Year STEM Degrees: Systemic Change to Support Students' Diverse Pathways*. Washington, DC: The National Academies Press.

National Academies of Sciences, Engineering, and Medicine. (2016b). *Infusing Ethics into the Development of Engineers: Exemplary Education Activities and Programs.* Washington, DC: The National Academies Press.

National Research Council. (2000). *How People Learn: Brain, Mind, Experience and School (Expanded Edition)*. Committee on Developments in the Science of Learning with additional material from the Committee on Learning Research and Educational Practice; Board on Behavioral, Cognitive, and Sensory Sciences; Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press. Available: http://www.nap.edu/catalog/9853/how-people-learn-brain-mind-experience-and-school-expanded-edition [March 2016].

National Research Council. (2011). *Incentives and Test-based Accountability in Education*. M. Hout and S.W. Elliott (Eds.), Committee on Incentives and Test-based Accountability in

Higher Education. Washington, DC: The National Academies Press. Available: http://www.nap.edu/catalog/12521/incentives-and-test-based-accountability-in-education [August 2016].

National Research Council. (2012a). *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering.* S.R. Singer, N.R. Nielsen, and H.A. Schweingruber (Eds.), Committee on the Status, Contributions, and Future Directions of Discipline-Based Education Research; Board on Science Education; Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. (2012b). *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century.* J.W. Pellegrino and M.L. Hilton (Eds.), Committee on Defining Deeper Learning and 21st Century Skills, Center for Education, Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. (2015a). *Enhancing the Effectiveness of Team Science*. N.J. Cooke and M.L. Hilton (Eds.), Committee on the Science of Team Science; Board on Behavioral, Cognitive, and Sensory Sciences; Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council (2015b). *Measuring Human Capabilities: An Agenda for Basic Research on the Assessment of Individual and Group Performance Potential for Military Accession*. Committee on Measuring Human Capabilities: Performance Potential of Individuals and Collectives; Board on Behavioral, Cognitive, and Sensory Sciences; Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Science Board. (2012). *Science and Engineering Indicators 2012* (NSB 12-01). Arlington, VA: National Science Foundation. Available: https://www.nsf.gov/statistics/seind12 [September 2016].

National Science Board. (2016). *Science and Engineering Indicators 2016* (NSB 16-01). Arlington, VA: National Science Foundation. Available: https://www.nsf.gov/statistics/2016/nsb20161/#/report/chapter-3/s-e-labor-market-conditions [September 2016].

National Science Foundation. (2013). *Women, Minorities, and Persons with Disabilities in Science and Engineering 2013*. Arlington, VA: Author. Available: http://www.nsf.gov/statistics/wmpd/2013/pdf/nsf13304_digest.pdf [July 2016].

Noel-Levitz, Inc. (2011). *The College Student Inventory Implementation Guidelines*. Available: https://www.ruffalonl.com/upload/Student_Retention/RMS/CSI_Implementation.pdf [May 2016].

Noftle, E.E., and Robins, R.W. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology, 93*(1), 116-130.

Nora, A., and Crisp, G. (2009). Hispanics and higher education: An overview of research, theory, and practice. In J.C. Smart (Ed.) *Higher Education: Handbook of Theory and Research,* Vol. 24 (pp. 321-358). New York: Springer Publishing.

Nordstrom, D.E. (2012). *Motivational, Parental, and Cultural Influences on Achievement and Persistence in Basic Skills Mathematics at the Community College.* Ed.D. Dissertation, University of Southern California.

Obama, B.H. (2009). *Remarks of President Barak Obama-Address to Joint Session of Congress*. Available: https://www.whitehouse.gov/the-press-office/remarks-president-barack-obama-address-joint-session-congress [July 2016].

OECD. (2013a). *Country Note: Education at a Glance: OECD Indicators 2012, United States*. Available: http://www.oecd.org/unitedstates/CN%20-%20United%20States.pdf [February 2016].

OECD. (2013b). *PISA 2012 Results: Ready to Learn: Students' Engagement, Drive and Self-Beliefs* (Vol. III). Paris: OECD Publishing.

Ohland, M.W., Sheppard, S.D., Lichtenstein, G. Eris, O., Chachra, D., and Layton, R.A. (2008). Persistence, engagement, and migration in engineering, *Journal of Engineering Education, 97*(3), 259-278.

Ohland, M.W., Loughry, M.S., Woehr, D.J., Bullard, L.G., Felder, R.M., Finelli, C.J., Layton, R. A., Pomeranz, H.R., and Schmucker, D.G. (2012). The comprehensive scale for self- and peer-evaluation. *Academy of Management Learning and Education, 11*(4), 609-630. Available: http://amle.aom.org/content/11/4/609.full.pdf+html [May 2014].

Okun, M.A., and Finch, J.F. (1998). The Big Five personality dimensions and the process of institutional departure. *Contemporary Educational Psychology, 23*(3)*,* 233-256.

O'Neill, T.A., Goffin, R.D., and Gellatly, I.R. (2012). The knowledge, skill, and ability requirements for teamwork: Revisiting the teamwork-KSA test's validity. *International Journal of Selection and Assessment, 20*(1), 36-52.

Ong, M., Wright, C., Espinosa, L.L. and Orfield, G. (2011). Inside the double bind: A synthesis of empirical research on undergraduate and graduate women of color in STEM. *Harvard Educational Review, 81*(2), 172-207.

Oregon State University. (n.d.). *Insight Resume*. Available: http://oregonstate.edu/senate/agen/reports/insight.html [November 2016].

Oseguera, L., and Rhee, B.S. (2009). The influence of institutional retention climates on student persistence to degree completion: A multilevel approach. *Research in Higher Education, 50*(6), 546-569.

Oswald, F.L., Schmitt, N., Kim, B.H., Ramsey, L.J., and Gillespie, M.A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology, 89*(2), 187-207.

Pace, V.L., and Borman, W.C. (2006). The use of warnings to discourage faking on noncognitive inventories. In R. Griffith (Ed.), *A Closer Examination of Applicant Faking Behavior* (pp. 281-302). Greenwich, CT: Information Age Publishing, Inc.

Palm, R. (2006). Perspectives from the dark side: The career transition from faculty to administrator. *New Directions for Higher Education*, *2006*(134), 59-65.

Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions.* New York: Oxford University Press.

Paulsen, M.B., and St. John, E.P. (2002). Social class and college costs: Examining the nexus between college choice and persistence. *Journal of Higher Education*, *73*(2), 189-236.

Pedhazur, E.J., and Schmelkin, L.P. (1991). *Measurement, Design, and Analysis: An Integrated Approach.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Pentland, A. (2008) *Honest Signals: How They Shape Our World*. Cambridge, MA: M.I.T. Press.

Pentland, A. (2015). *Social Physics*. PowerPoint prepared for the Committee on Assessing Intrapersonal and Interpersonal Competencies, Washington, DC, December 14. Available:

http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_170244.pdf [January 2016].

Perry, R. P. and Magnusson, J. (1989). Causal attributions and perceived performance: Consequences for college students' achievement and perceived control in different instructional conditions. *Journal of Education Psychology, 81*, 164-172.

Perry, R.P., Stupnisky, R.H., Hall, N.C., Chipperfield, J.G., and Weiner, B. (2010). Bad starts and better finishes: Attributional retraining and initial performance in competitive achievement settings. *Journal of Social and Clinical Psychology, 29*(6), 668-700.

Peters, M.L., and Kortecamp, K. (2010). Rethinking undergraduate mathematics education: The importance of classroom climate and self-efficacy on mathematics achievement. *Current Issues in Education*, *13*(4). Available: http://www.bu.edu/sed/files/2013/11/Kortecamp-Rethinking-undergraduate-mathematics-education-1.pdf [May 2016].

Pickering, J.W., Calliotte, J.A., and McAuliffe, G.J. (1992). The effect of non-cognitive factors in the retention and success of entering college freshmen. *Journal of the Freshman Year Experience*, *4*, 7-30.

Pintrich, P.R., Smith, D.A.F., García, T., and McKeachie, W.J. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ).* Ann Arbor, MI: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning. Available: http://files.eric.ed.gov/fulltext/ED338122.pdf [November 2016].

Pintrich, P.R., Smith, D.A.F., García, T., and McKeachie, W.J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement, 53*(3)*,* 801-813.

Pittman, L.D., and Richmond, A. (2008). University belonging, friendship quality, and psychological adjustment during the transition to college. *The Journal of Experimental Education, 76*(4), 343-361.

Plucker, J.A., Kennedy, C., and Dilley, A. (2015). *What We Know About Collaboration* (P21 Research Series). Washington, DC: Partnership for 21st Century Skills. Available: http://www.p21.org/storage/documents/docs/Research/P21_4Cs_Research_Brief_Series_-_Collaboration.pdf [August 2016].

Power, R.A., and Pluess, M. (2015). Heritability estimates of the Big Five personality traits based on common genetic variants. *Translational Psychiatry*, *5*, e604. Available: http://www.nature.com/tp/journal/v5/n7/full/tp201596a.html [May 2016].

Poropat, A.E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*(2)*,* 322-338.

President's Council of Advisors on Science and Technology. (2012). *Report to the President Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering and Mathematics.* Available: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_2-25-12.pdf [February, 2017].

Raju, N.S., Laffitte, L.J., and Byrne, B.M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 87*(3)*,* 517-529.

Greiff, S. and Kyllonen, P.C. (2016). Contemporary assessment challenges: The measurement of 21st century skills. *Applied Measurement in Education, 29* (4), 243-244.

Renninger, K.A., and Hidi, S.E. (2011). Revisiting the conceptualization, measurement, and generation of interest. *Educational Psychologist, 46*(3)*,* 168-184.

Renninger, K.A., and Hidi, S.E. (2016). *The Power of Interest for Motivation and Engagement*. New York: Routledge.

Rest, J., Narvaez, D., Thoma, S., and Bebeau, M. (1999). DIT2: Devising and testing a revised instrument of moral judgment. *Journal of Educational Psychology, 91*(4)*,* 644-659.

Richardson, M., Abraham, C., and Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2)*,* 353-387.

Ridgell, S.D., and Lounsbury, J.W. (2004). Predicting academic success: General intelligence, "Big Five" personality traits, and work drive. *College Student Journal, 38*(4)*,* 607-619.

Rikoon, S., Liebtag, T., Olivera-Aguilar, M., Robbins, S., and Jackson, T. (2014). *A Pilot Study of Holistic Assessment and Course Placement in Community College: Findings and Recommendations*. Princeton, NJ: Educational Testing Service. Available: http://www.ets.org/research/policy_research_reports/publications/report/2014/jtra [May 2016].

Robbins, S., Lauver, K., Le, H., Davis, D., Langley, R., and Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130*(2)*,* 261-288.

Robbins, S.B., Allen, J., Casillas, A., Peterson, C., and Le, H. (2006). Unraveling the differential effects of motivational and skills, social, and self-management measures from traditional predictors of college outcomes. *Journal of Educational Psychology, 98*(3), 598-616.

Robbins, S., Oh, I.-S., Le, H., and Button, C. (2009). Intervention effects on college performance and retention as mediated by motivational, emotional, and social control factors: Integrated meta-analytic path analyses. *Journal of Applied Psychology, 94*(5)*,* 1163-1184.

Roberts, B.W., and DelVecchio, W. (2000). The rank-order consistency of personality from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126*(1)*,* 3-25.

Roberts, B.W., Walton, K.E., and Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132*(1)*,* 1-25.

Roberts, B.W., Kuncel, N., Shiner, R., Caspi, A., and Goldberg, L.R. (2007). The power of personality: The comparative validity of personality traits, socio-economic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science, 2*(4)*,* 313-345.

Roberts, B.W., Krueger, R.F., Lejuez, C., Richards, J.M., and Hill, P.L. (2014). What is conscientiousness and how is it assessed? *Developmental Psychology, 50*(5), 1315-1330.

Robins, R.W., and Pals, J.L. (2002). Implicit self-theories in the academic domain: Implications for goal orientation, attributions, affects, and self-esteem change. *Self and Identity, 1*(4)*,* 313-336.

Robins, R.W., Tracy, J.L., and Sherman, J.W. (2007). What kinds of methods do personality psychologists use? A survey of journal editors and editorial board members. In R.W. Robins, R.C. Fraley, and R.F. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 673-678). London, UK: Guilford Press.

Roeser, R.W., Peck, S.C., Nasir, N.S. (2006). Self and identity processes in school motivation, learning, and achievement. In P.A. Alexander and P.H. Winne (Eds.), *Handbook of Educational Psychology* (2nd Edition) (pp. 392-393). Mahwah, NJ: Lawrence Erlbaum Associates.

Rosenberg, M. (1965). *Society and the Adolescent Self-Image.* Princeton, NJ: Princeton University Press.

Rothwell, J. (2013). *The Hidden STEM Economy.* Washington, DC: Metropolitan Policy Program at Brookings Institute. Available: http://www.brookings.edu/~/media/research/files/reports/2013/06/10-stem-economy-rothwell/thehiddenstemeconomy610.pdf [April 2015].

Rowntree, D. (2015). *Assessing Students: How Shall We Know Them?* New York: Routledge.

Ruiz-Primo, M.A. (2015). Cognitive labs. In R. Gunstone (Ed.), *Encyclopedia of Science Education* (pp. 167-171). New York: Springer.

Ruthig, J.C., Perry, R.P., Hall, N.C., and Hladkyj, S. (2004). Optimism and attributional retraining: Longitudinal effects on academic achievement, test anxiety, and voluntary course withdrawal in college students. *Journal of Applied Social Psychology, 34*(4), 709-730.

Sackett, P.R. (2012). Faking in personality assessments: Where do we stand? In M. Zieger, C. MacCann, and R.D. Roberts (Eds.), *New Perspectives on Faking in Personality Assessment* (pp. 330-344). New York: Oxford.

Salgado, J.F., and Táuriz, G. (2014). The Five-Factor Model, forced-choice personality inventories and performance: A comprehensive meta-analysis of academic and occupational validity studies. *European Journal of Work and Organizational Psychology, 23*(1)*,* 3-30. Available: https://home.ubalt.edu/tmitch/642/Articles%20syllabus/FFM%20meta%20validity%20Salgadoa%20214.pdf [August 2016].

Salzer, M.S. (2012). A comparative study of campus experiences of college students with mental illnesses versus a general college sample. *Journal of American College Health, 60*(1)*,* 1-7.

Sansone, C., Thoman, D., and Fraughton, T. (2015). The relation between interest and self-regulation in mathematics and science. In K.A. Renninger, M. Nieswandt, and S. Hidi, *Interest in K-16 Mathematics and Science Learning* (pp. 111-132). Washington, DC: American Educational Research Association.

Sato, E., Patelis, T., and Geisinger, K.F. (2015). A series on fairness and testing noncognitive factors. *National Council on Measurement in Education Newsletter, 23*(3), 6-11.

Schechter, O.G., Durik, A. M., Miyamoto, Y., and Harackiewicz, J.M. (2010). The role of utility value in achievement behavior: The importance of culture. *Personality and Social Psychology Bulletin*, *37*(3), 303-317.

Schechter, O.G., Durik, A.M., Miyamato, Y., and Harackiewicz, J.M. (2011). The role of utility value in achievement behavior: The importance of culture. *Personality and Social Psychology Bulletin, 37*(3)*,* 303-317.

Schein, E.H. (1992). *Organizational leadership and culture.* San Francisco, CA: Jossey-Bass.

Schmeiser, C.B., and Welch, C.J. (2006). Test development. In R.L. Brennan (Ed.), *Educational Measurement* (4th Edition) (pp. 307-353). Washington, DC: American Council on Education.

Schmitt, N., and Kunce, C. (2002). The effects of required elaboration of answers to biodata questions. *Personnel Psychology, 55*(3)*,* 569-587.

Schmitt, N., Oswald, F.L., Kim, B.H., Gillespie, M., Ramsay, L.J., and Yoo, T.-Y. (2003). Impact of elaboration on socially desirable responding and the validity of biodata measures. *Journal of Applied Psychology, 88*(6), 979-988.

Schneider, C.G. (2015a). *Assessing Intrapersonal and Interpersonal Competencies: The VALUE Strategy*. PowerPoint prepared for the Committee on Assessing Intrapersonal and Interpersonal Competencies, Washington, DC, December 14. Available: http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_169552.pdf [January 2016].

Schneider, C.G. (2015b). *Assessing Intrapersonal and Interpersonal Competencies: The VALUE Strategy*. Handout prepared for the Committee on Assessing Intrapersonal and Interpersonal Competencies, Washington, DC, December 14. Available: http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_171740.pdf [April 2016].

Schwartz, S. J., Kurtines, W. M., and Montgomery, M.J. (2005). Facilitating identity exploration processes in emerging adults: An exploratory study. *Journal of Adolescent Research, 20,* 309-345.

Seal, C.R., Miguel, K., Alzamil, A., Naumann, S.E., Royce-Davis, J., and Drost, D. (2015). Personal-interpersonal competence assessment: A self-report instrument for student development. *Research in Higher Education Journal*, *27*. Available: http://www.aabri.com/manuscripts/142053.pdf [April 2016].

Sedlacek, W.E. (2006). *Beyond the Big Test: Noncognitive Assessment in Higher Education.* Hoboken, NJ: John Wiley and Sons.

Sedlacek, W.E. (2011). Using noncognitive variables in assessing readiness for higher education. *Readings on Equal Education, 25,* 187-205.

Segal, C. (2008). *Motivation, Test Scores, and Economic Success*. Working Paper No. 1124. Barcelona, Spain: Department of Economics and Business, Universitat Pompeu Fabra.

Seidman, E. (1985). *In the Words of the Faculty: Perspectives on Improving Teaching and Educational Quality in Community Colleges (Jossey-Bass Higher and Adult Education)*. San Francisco, CA: Jossey-Bass.

Sellers, R. M., Rowley, S. A. J., Chavous, T. M., Shelton, J. N., & Smith, M. A. (1997). Multidimensional Inventory of Black Identity: A preliminary investigation of reliability and construct validity. *Journal of Personality and Social Psychology, 73*, 805–815.

Seymour, E., and Hewitt, N. (1997). *Talking about Leaving: Why Undergraduates Leave the Sciences.* Boulder, CO: Westview Press.

Shapiro, D., Dundar, A., Wakhungu, P.K., Yuan, X., Nathan, A., and Hwang, Y. (2015). *Completing College: A National View of Student Attainment Rates—Fall 2009 Cohort*. Signature Report No. 10. Herndon, VA: National Student Clearinghouse Research Center. Available: https://nscresearchcenter.org/wp-content/uploads/SignatureReport10.pdf [February 2016].

Sherman, D.K., Cohen, G.L., Nelson, L.D., Nussbaum, A.D., Bunyan, D.P., and Garcia, J. (2009). Affirmed yet unaware: Exploring the role of awareness in the process of self-affirmation. *Journal of Personality and Social Psychology, 97*(5)*,* 745-764.

Shuman, L.J., Besterfield-Sacre, M., and McGourty, J. (2005). The ABET "professional skills": Can they be taught, can they be assessed*? Journal of Engineering Education*, *94*(1), 41-55.

Simmons, J.P., Nelson, L.D.., and Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359-1366.

Singell, L.D. (2004). Come and stay a while: Does financial aid effect enrollment and retention at a large public university? *Economics of Education Review*, *23*(5), 459-471.

Smith, G.C., and Oyserman, D. (2015). Just not worth my time? Experienced difficulty and time investment. *Social Cognition, 33*(2)*,* 1-18.

Snyder, C.R., Harris, C., Anderson, J.R., Holleran, S.A., Irving, L.M., Sigmon, S.T., Yoshinobu, L., Gibb, J., Langelle, C., and Harney, P. (1991). The will and the ways: Development and validation of an individual-differences measure of hope. *Journal of Personality and Social Psychology, 60*(4)*,* 570-585.

Society for Industrial and Organizational Psychology. (2003). *Principles for the Validation and Use of Personnel Selection Procedures* (4th Edition). Available: http://www.siop.org/_principles/principles.pdf [August 2016].

Specht, J., Bleidorn, W., Dennissen, J.J.A., Hennecke, M., Huttermann, R., and Kandler, C. (2014). What drives adult personality development? A comparison of theoretical perspectives and empirical evidence. *European Journal of Personality, 28*(3)*,* 218-230.

Spurlin, J., Rajala, S., and Lavelle, J. (Eds.). (2008). *Designing Better Engineering Education Through Assessment: A Practical Resource for Faculty and Department Chairs on Using Assessment and ABET Criteria to Improve Student Learning*. Sterling, VA: Stylus Publishing.

Stark, J.S., Lowther, M.A., Sharp, S., and Arnold, G.L. (1997). Program-level curriculum planning: An exploration of faculty perspectives on two different campuses. *Research in Higher Education*, *38*(1), 99-130.

Steele, C.M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613-629.

Stephens, N.M., Hamedani, M.D., and Destin, M. (2014). Closing the social-class achievement gap: A difference-education intervention improves first-generation students' academic performance and all students' college transition. *Psychological Science, 25*(4), 943-953.

Sternberg, R.J. (2010). *College Admissions for the 21st Century.* Cambridge, MA: Harvard University Press.

Stevens, M.J., and Campion, M.A. (1994). The knowledge, skill, and ability requirements for teamwork: Implications for human resource management. *Journal of Management*, *20*(2), 503-530.

Stevens, M.J., and Campion, M.A. (1999). Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management*, *25*(2), 207-228.

Strayhorn, T.L. (2012). *College Student's Sense of Belonging: A Key to Educational Success for All Students*. New York: Routledge.

Stricker, L.J. (1963). Acquiescence and social desirability response styles, item characteristics, and conformity. *Psychological Reports, 12*(2)*,* 319-341.

Suskie, L. (2009). *Assessing Student Learning: A Common Sense Guide*. San Francisco, CA: Jossey-Bass.

Suskie, L., and Banta, T.W. (2009). *Assessing Student Learning: A Common-Sense Guide* (2nd Edition). San Francisco, CA: Jossey-Bass.

Tampke, D.R. (2011). *Psychosocial Factor Modeling and Retention Outcomes: Exploring the Efficacy of Early Intervention*. Presentation to the 17th Annual National Conference on Students in Transition, Houston, TX, November 15. Available: http://sc.edu/fye/events/presentation/sit/2010/files/81-C.pdf [May 2016].

Thompson, B.M., Schneider, V.F., Haidet, P., Levine, R.E., McMahon, K.K., Perkowski, L.C., and Richards, B.F. (2007). Team-based learning at ten medical schools: Two years later. *Medical Education, 41*(3)*,* 250-257. Available: http://medicina.fm.usp.br/cedem/did/seminarios/TBL_ten_med_school_2y_later_MedEduc2007.pdf [October 2014].

Thurlow, M.L., Thompson, S.J., and Lazarus, S.S. (2006). Considerations for the administration of tests to special needs students: Accommodations, modifications, and more. In S.M. Downing and T.M. Haladyna (Eds.), *Handbook of Test Developmen*t (pp. 653-673). Mahwah, NJ: Lawrence Erlbaum Associates.

Tinto, V. (1994). *Leaving College: Rethinking the Causes and Cures of Student Attrition*. Chicago, IL: University of Chicago.

Titus, M. (2004). An examination of the influence of institutional context on persistence at four-year colleges and universities: A multilevel approach. *Research in Higher Education*, *45*(7), 673-699.

Torney-Purta, J.V., Cabrera, J., Roohr, K.C., Liu, O.L., and Rios, J.A. (2015). *Assessing Civic Competency and Engagement in Higher Education: Research Background, Frameworks, and Directions for Next-Generation Assessment*. ETS RR-15-34. Princeton, NJ: Educational Testing Service. Available: http://www.ets.org/research/policy_research_reports/publications/report/2015/jvdz [February 2015].

Trapmann, S., Hell, B., Hirn, J.W., and Schuler, H. (2007). Meta-analysis of the relationship between the Big Five and academic success at university. *Zeitschrift für Psychologie, 215*(2)*,* 132-151.

Trautwein, U., Ludtke, O., Nagy, N., Lenski, A., Niggli, A., and Schnyder, I. (2015). Using individual interest and conscientiousness to predict academic effort: Additive, synergistic, or compensatory effects? *Journal of Personality and Social Psychology, 109*(1)*,* 142-162.

Trilling, B., and Fadel, C. (2009). *21st Century Skills: Learning for Life in Our Times*. San Francisco, CA: John Wiley and Sons.

Turner, E.A., Chandler, M., and Heffer, R.W. (2009). The influence of parenting styles, achievement motivation, and self-efficacy on academic performance in college students. *Journal of College Student Development, 50*(3)*,* 337-346.

U.S. Department of Education. (2015). *Digest of Education Statistics.* Table 326.10. National Center for Education Statistics. Available: http://nces.ed.gov/programs/digest/d15/tables/dt15_326.10.asp [August 2016].

U.S. Department of Education. (2016). *Overview of Accreditation in the United States*. Washington, DC: Author. Available: http://www2.ed.gov/admins/finaid/accred/accreditation.html#Overview [August 2016].

Valentine, M.A., Nembhard, I.M., and Edmondson, A.E. (2012). Measuring teamwork in health care settings: A review of survey instruments. *Medical Care, 53*(4), 16-30.

Van Duzer, E., and McMartin, F. (2000). Methods to improve the validity and sensitivity of a self/peer assessment instrument. *IEEE Transactions on Education, 43*(2)*,* 153-158.

Vandenberg, R.J., and Lance, C.E. (2002). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1)*,* 4-70. Available: http://journals.sagepub.com/doi/pdf/10.1177/109442810031002 [February 2017].

van der Liden, W.J. (2005). *Linear Models for Optimal Test Design.* New York: Springer.

Vansteenkiste, M., Simons, J., Lens, W., Soenens, B., Matos, L., and Lacante, M. (2004a). Less is sometimes more: Goal content matters. *Journal of Educational Psychology, 96*(4), 755-764.

Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K.M., and Deci, E.M. (2004b). Motivating learning, performance, and persistence: The synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality and Social Psychology, 87*(2), 246-260.

Vars, F. and Bowen, W.G. (1998). SAT scores, race and academic performance in academically selective colleges and universities. In C. Jenks and M. Phillips (Eds.), *The Black-White Test Score Gap* (pp. 401-430). Washington, DC: Brookings Institution Press.

von Davier, A.A., Zhu, M., and Kyllonen, P.C. (Eds.). (2017). Innovative Assessment of Collaboration. *Methodology of Educational Measurement and Assessment*. Switzerland: Springer International Publishing.

Vroom, V.H. (1964). *Work and Motivation*. New York: John Wiley & Sons.

Walpole, M.B., Burton, N.W., Kanyi, K., and Jackenthal, A. (2002). *Selecting Successful Graduate Students: In-Depth Interviews with GRE Users*. GRE Board Research Report No. 99-11R. Princeton, NJ: Educational Testing Service.

Walton, G. M., and Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology, 92*(1)*,* 82-96.

Walton, G.M., and Cohen, G.L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science, 331*(6023)*,* 1447-1451.

Walton, G.M., Cohen, G.L., Cwir, D., and Spencer, S.J. (2012). Mere belonging: The power of social connections. *Journal of Personality and Social Psychology, 102*(3), 513-532.

Walton, G.M., Peach, J.M., Logel, C., Spencer, S.J., and Zanna, M.P. (2015). Two brief interventions to mitigate a "chilly climate" transform women's experience, relationships, and achievement in engineering. *Journal of Educational Psychology, 107*(2), 468-485.

Weinberger, C.J. (2014). The increasing complementarity between cognitive and social skills. *The Review of Economics and Statistics, 96*(5), 849-861.

Welsh, J.F., and Metcalf, J. (2003). Faculty and administrative support for institutional effectiveness activities: A bridge across the chasm? *The Journal of Higher Education, 74*(4), 445-468.

Wigfield, A., and Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review, 30*(1)*,* 1-35.

Wiley, A., Wyatt, J., and Camara, W.J. (2010). *The Development of a Multidimensional College Readiness Index* (College Board Research Report No. 2010-3). New York: The College Board. Available: http://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2010-3-development-multidimensional-college-readiness-index.pdf [November 2016].

Williamson, O. (1975). *Markets and Hierarchies*. New York: The Free Press.

Williamson, O. (1985). *The Economic Institutions of Capitalism*. New York: The Free Press.

Willingham, W.W., and Breland, H. (1982). *Personal Qualities and College Admissions.* New York: College Entrance Examination Board.

Willingham, W.W., Young, J.W., and Morris, M.M. (1987). Success in college: The role of personal qualities and academic ability. *The Journal of Higher Education, 58*(1), 120-122.

Willms, J.D. (2003). *Student Engagement at School: A Sense of Belonging and Participation: Results from PISA 2000*. Paris: OECD.

Wilson, T.D. and Linville, P.W. (1982). Improving the academic performance of college freshmen: Attribution retraining revisited. *Journal of Personality and Social Psychology, 42,* 367-376.

Wolniak, G.C., Mayhew, M.J., and Engberg, M.E. (2012). Learning's weak link to persistence. *Journal of Higher Education, 83*(6)*,* 795-823.

Xie, Y., and Killewald, A.A. (2012). *Is American Science in Decline?* Cambridge, MA: Harvard University Press.

Xie, Y., Fang, M., and Shauman, K. (2015). STEM education. *Annual Review of Sociology, 41*(1), 331-357.

Yeager, D.S., Henderson, M.D., Paunesku, D., Walton, G.M., D'Mello, S., Spitzer, B.J., and Duckworth, A. (2014). Boring but important: A self-transcendent purpose for leaning fosters academic self-regulation. *Journal of Personality and Social Psychology, 107*(4)*,* 559-580.

Yeager, D.S., Walton, G.M., Brady, S.T., Akcinar, E.N., Paunesku, D., Keane, L., Kamentz, D., Ritter, G., Duckworth, A.L., Urstein, R., Gomez, E.M., Markus, H.R., Cohen, G.L., and Dweck, C.S. (2016). Teaching a lay theory before college narrows achievement gaps at scale. *Proceedings of the National Academy of Sciences of the United States of America*, E3341-E3348. Available: http://www.pnas.org/content/113/24/E3341.full.pdf [August 2016].

Yoshida, E., Peach, J.M., Zanna, M.P., and Spencer, S.J. (2012). Not allautomatic associations are created equal: How implicit normative evaluations are distinct from attitudes and uniquely predict meaningful behavior. *Journal of Experimental Social Psychology, 48*(3)*,* 694-706.

Young, J.W, and Kobrin, J.L. (2001). *Differential Validity, Differential Prediction, and College Admission Testing: A Comprehensive Review and Analysis*. Research Report 2001-6. New York: College Entrance Examination Board.

Zerbe, W.J., and Paulhus, D.L. (1987). Socially desirable responding in organizational behavior: A reconception. *Academy of Management Review, 12*(2)*,* 250-254.

Zhang, M., Kranov, A.A., Byerlein, S.W., McCormack, J.P., Pedrow, P.D., and Schmeckpeper, E. (2015). *Investigating a Scenario-Based Performance Assessment of Engineering Professional Skills*. Proceedings of the 5th IEEE Integrated STEM Conference, Princeton, NJ, March 7.

Zoltowski, C.B., Buzzanell, P.M., and Oakes, W.C. (2013). *Utilizing an Engineering Ethical Reasoning Instrument in the Curriculum*. Paper ID #7607. Paper presented at the 120th ASEE Annual Conference and Exhibition, Atlanta, GA, June 23-26.

Zwick, R. (2007). College admissions in twenty-first-century America: The role of grades, tests, and games of chance. *Harvard Educational Review*, *77*(4), 419-429, 529.

Appendix A

Literature Searches Commissioned by the Committee

The committee commissioned six literature searches from the National Academies' research center to aid in the writing of this report.

PARAMETERS

Date: 2010-present
Language: English
Country: United States, International

DATABASES

ERIC (Ovid)
PsycINFO + PsycArticles (ProQuest)
Scopus
Web of Science

SEARCH DATE: SEPTEMBER 8, 2015

The committee first commissioned a search to explore potential relationships between intra- and interpersonal competencies and academic achievement and persistence for various student groups that may face disadvantages, such as minorities, low-income students, and those with disabilities. Search terms included intrapersonal competency, interpersonal competency, habit, agency, conscientiousness, grit, participation, teamwork, communication skills, academic mindset, productive disposition and college persistence or college success. They also included academic identity, STEM identity, identity, underrepresented groups, minority students, racial minorities, ethnic minorities, first generation students, disabilities, low-income students, LGBTQ students, nontraditional students, risk, resilience, and academic achievement or academic persistence or outcomes of education.

SEARCH DATE: SEPTEMBER 15, 2015

Next, the committee commissioned a search to explore potential relationships between various intra- and interpersonal competencies and college persistence. The search focused on skills and abilities that are clustered within the intra- and interpersonal domains in a prior

National Academies study (National Research Council, 2012b). Search terms included teamwork, communication, motivation, self-regulation, metacognition, and college persistence.

SEARCH DATE: OCTOBER 6, 2015

To inform deliberations about mental health and its relationship with college persistence, the committee commissioned another search focused on various intrapersonal skills and abilities that are clustered under the umbrella concept of "positive core self-evaluation" in the prior National Academies study (National Research Council, 2012b). The search terms included positive self-evaluation, neurosis/neuroticism, emotional stability, self-monitoring, self-reinforcement, psychological health, college completion, college persistence, and academic persistence.

SEARCH DATE: NOVEMBER 11, 2015

Next, the committee commissioned a search to explore potential relationships between interpersonal competencies and college persistence. The search focused on various interpersonal skills and abilities are clustered under the umbrella concepts of "teamwork and collaboration" and "leadership" in the prior National Academies study (National Research Council, 2012b). Search terms included communication, collaboration, teamwork, cooperation, coordination, interpersonal skills, empathy, perspective taking, trust, service orientation, conflict resolution, negotiation, leadership, responsibility, assertive communication, self-presentation, social influence, and college persistence or college completion.

SEARCH DATE: JANUARY 12, 2016

To inform its discussions about intra- and interpersonal competencies that college leaders have identified as important outcomes for all graduates, the committee commissioned a search to explore potential relationships between such competencies and college persistence. Search terms included ethics, ethical reasoning, ethical action, intercultural knowledge, intercultural competence, intercultural sensitivity, diversity knowledge, lifelong learning, self-directed learning, civic engagement, teamwork, oral communication, and college persistence or college completion.

SEARCH DATE: MARCH 9, 2016

To follow up on the January search, the committee commissioned a search to explore what is known about assessing competencies that college leaders have identified as important outcomes for all graduates. Search terms included ethics, ethical reasoning, ethical action, intercultural knowledge, intercultural competence, intercultural sensitivity, diversity knowledge, lifelong learning, self-directed learning, civic engagement, teamwork, oral communication, interpersonal skills, and assessment or assess or measure and college or undergraduate or university and college student.

SEARCH DATE: SEPTEMBER 19, 2016

Finally, the committee commissioned a search to explore literature on assessing the eight competencies identified in Chapter 2. Search terms included assessing or assessment or measuring or measurement along with terms reflecting behaviors related to conscientiousness (e.g., study habits, goal setting, time management, grit, persistence); sense of belonging (belonging, connectedness, campus climate, school climate, social integration); academic self-efficacy (academic self-efficacy, self-confidence, school success, self-concept); and the other competencies identified by the committee, including growth mindset, utility goals, utility values, intrinsic goals, intrinsic interests, inherent interests, intrinsic motivation, prosocial goals, prosocial values, altruism, social responsibility, social commitment, positive future self, imagined future identity, and perceived future success.

Appendix B

Intervention Study Table

To address its charge to identify intra- and interpersonal competencies that not only are related to success during college but also are malleable in response to interventions, the committee searched the social-psychological intervention literatures. This appendix provides a table summarizing the results of that search, available in the Resources tab on the report home page: https://www.nap.edu/catalog/24697/supporting-students-college-success-the-role-of-assessment-of-intrapersonal. From the Resources tab, click on "Intervention Table from Appendix B."

The search process involved several steps. The committee started with all of the references in chapters in *Motivational Interventions* (Karabenick and Urdan, 2014), a recent edited volume summarizing work on the kinds of interventions that were the focus of this study. The committee also conducted searches in Google Scholar and of websites for researchers known to be conducting social-psychological interventions. Whenever relevant intervention studies were found, the committee reviewed their reference lists for additional studies. All told, the committee discovered 49 articles addressing a total of 61 studies that met the following criteria: the intervention (1) sought to manipulate one of the competencies identified in Chapter 2, (2) consisted of clearly defined treatment and control/comparison groups comprising college students or individuals who were about to matriculate in a college, (3) included at least 10 subjects per group, (4) incurred less than 50 percent attrition between the time it was initiated and the time the outcome was measured, and (5) employed random assignment. The search produced the following number of studies for each competency: conscientiousness (7), sense of belonging (10), academic self-efficacy (2), growth mindset (17), utility goals and values (15), intrinsic goals and values (3), prosocial goals and values (2), and positive future self (5). Among these 61 studies, 29 were conducted after 2010.

Appendix C
Assessments Used in the Intervention Literature

TABLE C-1

Studies of Relationship with College	Studies Assessing the Competency	Constructs Assessed in These Studies	Assessment Instruments	Studies Reporting Reliability (for 1 or more instruments)
		Behaviors Related to Conscientiousness		
7	15	Self-regulated learning/effort Motivation Intended learning behaviors	19 (13 self-report, selected response [SRSR])	Reliability: 5 (range 0.67-0.98)
	Duckworth et al., 2016	Self-regulated learning/effort	1. Instrument with 2 items measuring success at achieving study goal (SRSR) 2. Checklist instrument on level of temptation by various distractions (SRSR)	
	Fitch et al., 2012	Self-regulated learning	3. Instrument with 44 of 81 original items from the "Motivated Strategies for Learning Questionnaire (MSLQ)" (SRSR)	
	Haynes et al., 2008	Motivation	4. Instrument with 4 items on mastery motivation from the MSLQ, α = .67-.75 (SRSR) 5. Instrument with 4 items on performance motivation from the MSLQ, α = .77 (SRSR)	✓
	Landau et al., 2014, Study 1	Intended learning behaviors	6. Instrument with 5 items on interest in attending a workshop on preparing for final exams (SRSR) 7. Number of participants who tore off a slip of paper listing a website at which to access online guides for successful studying	

Studies of Relationship with College	Studies Assessing the Competency	Constructs Assessed in These Studies	Assessment Instruments	Studies Reporting Reliability (for 1 or more instruments)
	Landau et al., 2014, Study 2	Self-regulation	8. Instrument with 50 easily solvable addition problems (academic effort measured by the number correctly solved within 5 minutes)	
	Landau et al., 2014, Study 3	Intended academic behaviors	9. Instrument presenting 1-hour time slots over exam weekend, with three response options: coursework, socializing, or solitary leisure time (SRSR)	
	Liu et al., 2012	Motivation	10. Student Opinion Scale with 10 items, α = 0.84-0.85 (SRSR)	✓
	Morisano et al., 2010	Motivation	11. Concluding questionnaire with 15 scaled items (SRSR)	
	Schecter et al., 2011, Study 1	Self-regulation	12. Time spent on practice problems following an intervention to learn a new math technique (Participants were asked to practice the technique on some multiplication problems, with no time frame specified. The experimenter surreptitiously recorded the number of seconds spent practicing.)	
	Senko and Harackiewicz, 2005	Motivation	13. Instrument with 9 items, α = .73-.88 (SRSR)	✓
	Struthers and Perry, 1996	Motivation	14. An SRSR instrument, α = .73-.98	✓

Studies of Relationship with College	Studies Assessing the Competency	Constructs Assessed in These Studies	Assessment Instruments	Studies Reporting Reliability (for 1 or more instruments)
	Vansteenkiste et al., 2004a, Study 1	Self-regulation	15. Instrument with 4 items on superficial processing (e.g., "I skipped parts of the text that I did not understand very well", α = .80) and 4 items on deep processing, α = .84 (SRSR)	✓
	Walton et al., 2012	Motivation-related behaviors in math	16. An SRSR instrument 17. Time spent (persistence) on an insoluble math puzzle	
	Yeager et al., 2014, Study 3	Self-regulation	18. Time spent on each problem in an online exam review environment: number of milliseconds that each problem was displayed before participants submitted a correct answer (Measures behavior that could signify intent to truly learn.)	
	Yeager et al., 2014, Study 4	Self-regulation	19. Time spent on math problems vs. consuming online media	
		Academic Self-Efficacy		
	2	2	8 (7 SRSR)	Reliability: 2 (range 0.76-0.95)
	Betz and Schifano, 2000	Self-efficacy in "realistic" activities (e.g., using tools)	1. Instrument with 15 items on interest in realistic activities, α = .76 (SRSR) 2. Instrument with 30 items—3 10-item subscales of the "Skills Confidence Inventory" (SRSR) 3. "Occupational Self-Efficacy Scale" with 20 items (SRSR)	✓
	Luzzo et al., 1999	Math/science self-efficacy	4. Math/science course self-efficacy scale (slightly revised version of the "Mathematics Self-Efficacy	✓

Studies of Relationship with College	Studies Assessing the Competency	Constructs Assessed in These Studies	Assessment Instruments	Studies Reporting Reliability (for 1 or more instruments)
			College Courses Scale," α = .95 [SRSR]) 5. "Self-Efficacy for Scientific and Technical Fields," 15-items, α = .90 (SRSR) 6. Math/science occupational self-efficacy (revised version of the "Mathematics Career Self-Efficacy Scale"): 10-items, α = .95 (SRSR) 7. Math/science career interests: 15 items, α = .95 (SRSR) 8. Survey of courses, majors, and career aspirations: two research assistants coded responses; interrater reliability was 92 percent	
			Growth Mindset	
17	12	Malleability beliefs Academic expectations Attributions Perceived performance	20 (20 SRSR)	Reliability: 2 (range 0.63-0.88)
	Aronson, et al., 2002	Malleability beliefs	1. Postintervention questionnaire with 2 items on malleability of intelligence (SRSR) 2. Phone interview conducted 9 weeks later that included the same 2 questions (SRSR)	
	Boese et al., 2013	Academic expectations Attributions	3. Instrument with 1 item on course grade expectations (select from a range from 1 = F to 8 = A+) and 1 item on course responsibility (i.e., attribution) (SRSR) 4. Questionnaire with 12 items on attributions (3 each on luck, effort, ability, teacher effectiveness) (SRSR)	

Studies of Relationship with College	Studies Assessing the Competency	Constructs Assessed in These Studies	Assessment Instruments	Studies Reporting Reliability (for 1 or more instruments)
	Cohen et al., 1999	Malleability beliefs	5. Instrument with 6 items on perceived ability and motivation to improve writing and identification with writing (SRSR)	
	Hall et al., 2004	Attributions; perceived performance	6. Instrument with 24 items on perceived control, α = .78 (SRSR) 7. Instrument with 4 items on perceived success, α = .88 (SRSR)	✓
	Hall et al., 2006	Academic expectations attributions	8. Instrument with 2 items on perceived success in the course (current and expected) (SRSR) 9. Instrument(s) with 1 item on causal attributions, 2 items on controllable attributions (α = .63), and 4 items on uncontrollable attributions (α = .63) (SRSR)	✓
	Hamm et al., 2014	Attributions and related emotions	10. Instrument with 4 items on causal attributions (SRSR) 11. Instrument with 2 items on attribution-related emotions (SRSR)	
	Menec et al., 1994, Study 1	Expectations	12. Instrument with 2 items on expectations for (a) next test grade (on a scale of 0 to 100) and (b) final course grade (on a scale of 0 to 100) (SRSR)	
	Menec et al., 1994, Study 2	Attributions	13. Instrument on locus of control with 3 items on internal attributions (ability and skill) and 3 items on external attributions (luck) (SRSR)	

Studies of Relationship with College	Studies Assessing the Competency	Constructs Assessed in These Studies	Assessment Instruments	Studies Reporting Reliability (for 1 or more instruments)
		Expectations	14. Instrument with 2 items: (a) expected performance on next test (scale from not at all successful = 1 to very successful = 10), and (b) expected course grade (one item, scale from 1 = F to 10 = A+). (SRSR)	
	Perry and Magnusson, 1989	Perceived performance	15. Instrument with 6 items on perceived performance (ability, control, emotions, test difficulty, and overall performance) (SRSR)	
	Perry et al., 2010	Attributions	16. Instrument with 1 item on causal attributions (SRSR)	
		Attribution-related emotions	17. Instrument with 6 items on attribution-related and performance-related emotions (SRSR)	
	Ruthig et al., 2004	Perceived performance	18. Instrument with 37 true-false items on test anxiety from the Test Anxiety Scale (SRSR)	
	Wilson and Linville, 1982	Perceived performance	19. Instrument with 4 items on the freshman year and their own performance (SRSR)	
		Expectations	20. Expected performance improvement: expected grade point average (GPA) for current semester, next semester, at graduation (SRSR)	
			Intrinsic Goals/Interest	
3	2	Intrinsic motivation/interest	3 (2 SRSR)	Reliability: 1 (range 0.72)
	Hamm et al., 2014	Intrinsic motivation	1. "MAACH Intrinsic Motivation Scale", α = .72 (SRSR)	✓

Studies of Relationship with College	Studies Assessing the Competency	Constructs Assessed in These Studies	Assessment Instruments	Studies Reporting Reliability (for 1 or more instruments)
	Vansteenkiste et al., 2004a, Study 1	Intrinsic motivation	2. Instrument with 16 items on reasons for engaging in the learning task (external, introjected, identified, intrinsic) (SRSR)	
		Intrinsic interest	3. Number of visits to the library and/or the recycling center to learn more about recycling	
			Positive Future Self	
5	3	Domain identification Identity processes Possible future self	5 (5 SRSR)	Reliability: 3 (range 0.60-0.92)
	Harrison et al., 2006	Domain identification	1. "Domain Identification Measure" with 7 items on identification with English, α = .88, and 10 items on identification with math," α = .88 (SRSR)	✓
	Schwartz et al., 2005	Self-constructive identity processes	2. "Identity Style Inventory" with 40 items, including 11 on exploration-oriented style (α = .61), 9 on closure-oriented style (α = .70), and 10 on avoidance-oriented style (α = .70) (SRSR)	✓
		Self-discovery identity processes	3. " Critical Problem Solving Scale": performance-based measure of critical thinking in response to two life choice dilemmas (one personal, one interpersonal) generated by the participants (For each dilemma, participants (a) list potential alternative solutions, (b) identify the best alternative, (c) identify the worst alternative, (d) list pros and cons of the best alternative, (e) list pros and cons of the worst alternative, and (f) identify the alternative they would select) (SRSR?)	
		Self-actualization	4. Self-actualization Scale with 6 items, α = .60	

Studies of Relationship with College	Studies Assessing the Competency	Constructs Assessed in These Studies	Assessment Instruments	Studies Reporting Reliability (for 1 or more instruments)
			(SRSR)	
	Walton and Cohen, 2007, Study 2	Possible future self	5. Instrument with 3 Socially Responsible Leadership Scale (SRLS) items on possible academic selves (e.g., "In the future I could see myself being successful at [school name].") (α = .84)	✓
			Prosocial Goals/Values	
2	0*	0		
			Sense of Belonging	
8	8	Collective threat Sense of belonging/academic fit Social integration	25 (21 SRSR)	Reliability: 6 (range 0.63-0.93)
	Brady et al., 2016	Sense of belonging	1. Instrument with 10 items, α = .82 (SRSR)	✓
	Cohen and Garcia, 2005	Collective threat (opposite end of spectrum from Sense of Belonging)	2. Instrument with 5 items on self-esteem/ belonging uncertainty, α = .75 (SRSR) 3. Instrument with 14 items on stereotype distancing (SRSR) 4. Instrument with 40-word fragment completion tasks targeting words related to (a) negative racial stereotypes and (b) race; designed to measure racial stereotype activation 5. Instrument with 1 item on perceived exposure to evaluative scrutiny (SRSR) 6. Racial identification scale with 1 item (SRSR): "Prior to the experiment, the scale's reliability and	✓

Studies of Relationship with College	Studies Assessing the Competency	Constructs Assessed in These Studies	Assessment Instruments	Studies Reporting Reliability (for 1 or more instruments)
			validity were established. In a sample of 42 undergraduates (16 Black, 26 White), the test–retest reliability (over 2–8weeks) proved satisfactory, r (41) = .80, p < .01. In another sample of 35 Black undergraduates, evidence of convergent validity was found in the form of a strong correlation with the 'Race Centrality' subscale of the 'Multidimensional Inventory of Black Identity' (Sellers et al., 1997), r(34) = .79, p <.01.'" (Cohen and Garcia, 2005, pp. 571-572)	
	Hausmann et al., 2009	Sense of belonging/academic fit	7. "Sense of Belonging Scale," α = .94 (SRSR)	✓
	Stephens et al., 2014	Social/academic integration	8. Instrument with 15 items on behaviors: 3 on tendency to seek college resources (α = .56), 9 on activism (α = .85), and 3 yes-no items on joining extracurricular activities (SRSR) 9. Instrument with 2 items on well-being (α = .77) and 6 items on academic fit (α = .82) (SRSR) 10. Instrument with 7 items on how frequently different sources of social support are available (α = .90) and 2 items on time spent maintaining relationships with friends/family from home (SRSR)	✓
		Collective threat	11. Instrument with 2 items on social identity threat, α = .63 (SRSR) 12. "Psychological Distress Scale" presents respondents with 10 items (symptoms of psychological distress) and asks them to rate (on a 1 to 5 scale) how much of the time they felt that way in the past 30 days (α = .90) (SRSR)	

Studies of Relationship with College	Studies Assessing the Competency	Constructs Assessed in These Studies	Assessment Instruments	Studies Reporting Reliability (for 1 or more instruments)
	Walton and Cohen, 2007, Study 2	Sense of belonging/academic fit	13. Academic/social fit scale with 17 items (α = .79-.84) (SRSR) 14. Instrument with 2 items on academic identification, 3 on enjoyment of academic work, 2 on self-efficacy, 1 on potential to succeed in college, 1 on possible future selves (α = .84) and 3 on evaluative anxiety (e.g., "How anxious would you be about asking a question or making a comment in a large lecture class?" [p. 89]) (SRSR) 15. Daily online survey with the 17-item academic/social fit scale and one item on social fit on adverse days (Respondents listed negative and positive events of the day, rated each event's level of negativity on a 1 to 10 scale, and then rated the day's overall level of negativity [α = .87].) (SRSR) ("This measurement procedure was informed by research suggesting that the validity of subjective reports of well-being can be enhanced by having respondents review specific events in their day" [p. 89]).	✓
	Walton and Cohen, 2011	Collective threat	16. Instrument with items on belonging uncertainty (SRSR) 17. Racial stereotype activation: word completion task	
		Sense of belonging/academic fit	18. Daily online/email instrument on social belonging including (a) 17 items on social fit, (b) 2 items on self-efficacy, and (c) 1 item on potential to succeed in college (SRSR) 19. Daily online/email instrument with one item on social fit on adverse days (Respondents listed negative and positive events of the day, rated each	

Studies of Relationship with College	Studies Assessing the Competency	Constructs Assessed in These Studies	Assessment Instruments	Studies Reporting Reliability (for 1 or more instruments)
			event's level of negativity on a 1 to 10 scale, and then rated the day's overall level of negativity.) (SRSR)	
	Walton et al., 2012, Study 1	Sense of belonging in math	20. Instrument with items on social connectedness to math (α = .79-0.93) (SRSR) 21. Ratings of participant-generated reasons for success/failure in math (SRSR)	✓
	Walton et al., 2015	Sense of belonging among women in engineering	22. Instrument on attitudes toward engineering including items on (a) evaluation of current experience, (b) sense of belonging, (c) self-efficacy, and (d) expected success (Combined Index α = .87 (SRSR) 23. "Modified Implicit Association Test" (behavioral measure of biased attitudes) 24. Brief daily online survey (SRSR) including (a) importance of daily adversity (Instrument asks respondents to list negative and positive events of the day and rate each one's level of negativity and importance.), (b) sources of stress (Instrument lists potential sources of personal and academic stress and asks respondents to rate the level of stress from each source that day.), and (c) confidence in handling stress (Instrument presents respondents with 20 pairs of adjectives from which to choose, describing how they felt (e.g., "bad-good," "ashamed-proud") and asks them to rate their level of each feeling.) 25. Instrument asks participants to list 5 closest friends on campus, with gender and major.	

Studies of Relationship with College	Studies Assessing the Competency	Constructs Assessed in These Studies	Assessment Instruments	Studies Reporting Reliability (for 1 or more instruments)
			Utility Goals/Values	
15	4	Utility value Situational Interest	7 (7 SRSR)	Reliability: 3 (range 0.78-0.93)
	Durik et al., 2015, Study 1	Situational Interest	1. Instrument with 3 items on situational interest in using math technique (e.g., "The left-to-right technique is interesting," α = .88) (SRSR)	✓
	Hulleman et al., 2010, Study 1	Utility Value Situational Interest	2. Scale with 3 items on perceived utility value, α = .84 (SRSR) 3. Instrument with 5 items on situational interest, α = .89, and 1 yes/no item on maintained interest in using the math technique ("Do you think you will use the technique you learned today on your own in the future?" [p. 4].) (SRSR)	✓
	Hulleman et al., 2010, Study 2	Utility value Situational Interest	4. Scale with 3 items on perceived utility value, α = .78-.88 (SRSR) 5. Scale with 5 items on situational interest in the field of psychology, α = .93, and 1 yes/no item on interest in majoring in psychology (SRSR)	✓
	Schechter et al., 2011	Utility value Situational Interest	6. Scale with 2 items on utility value (SRSR) 7. Instrument with 4 items on situational interest (SRSR)	

NOTES: Instruments shown in quotations are existing instruments developed by a prior author(s). SRSR=self-report, selected response format. No studies reported evidence of fairness analysis. Only one study reports evidence of validity (Cohen and Garcia, 2005).
✓ = study reports reliability.
*Although none of the 61 studies meeting the committee's criteria assessed prosocial goals/values, Yeager and colleagues (2014) includes a study (Study 1) that assessed this competency using an instrument with SRSR format including three items on self-transcendent motives for attending college (α = .75), three items on intrinsic-oriented motives (α = .70), and three items on self-oriented motives (α = .50).

TABLE C-2 Format Totals from Table C-1

Competency	Instruments	SRSR	Other Formats
Behaviors related to conscientiousness	19	13	1. Behavioral measure—number taking info. on study skills workshop 2. Behavioral measure of time spent on addition problems 3. Time spent on practice problems 4. Time spent on insoluble math puzzle 5. Time spent on each problem in online math review 6. Time spent on math problems vs. consuming online media
Academic Self-efficacy	8	7	Survey of courses, majors, and career aspirations: two research assistants coded responses; interrater reliability was 92 percent
Growth Mindset	20	20	
Intrinsic Goals/Interest	3	2	Behavioral measure-Number of visits to the library and/or the recycling center to learn more about recycling.
Positive Future Self	5	4	"Critical Problem Solving Scale": performance-based measure of critical thinking about one's identity in response to two life choice dilemmas (one personal, one interpersonal) generated by the participants
Prosocial Goals/Values	0	0	0
Sense of Belonging	25	21	1. Word Fragment Completion Task 2. Word Fragment Completion Task (used in 2 studies) 3. Modified Implicit Attitude Test 4. List name, gender, and major of 5 closest friends (to assess female engineering majors' social networks)
Utility Value	7	7	
Totals	87	74	13

Appendix D

Biographical Sketches of Committee Members

JOAN HERMAN *(Chair)* is co-director emerita of the National Center for Research on Evaluation, Standards, and Student Testing at the University of California, Los Angeles, where she currently serves as senior research scientist. Her research has explored the effects of testing on schools and the design of assessment systems to support school planning and instructional improvement. Her recent work focuses on the validity and utility of teachers' formative assessment practices and the assessment of deeper learning. She also has wide experience as an evaluator of school reform. Dr. Herman is past president of the California Educational Research Association and has held a variety of leadership positions in the American Educational Research Association, National Organization of Research Centers, and Knowledge Alliance. She is an elected member of the National Academy of Education and fellow of the American Educational Research Association. Dr. Herman is current editor of *Educational Assessment*, served on the Joint Committee for the Revision of Standards for Educational and Psychological Testing, and is a member of the National Academies' Board on Testing and Assessment. She received her B.A. in sociology from the University of California, Berkeley; was awarded an M.A. and Ed.D. in Learning and Instruction from the University of California, Los Angeles; and is a member of Phi Beta Kappa.

DAVID BILLS is associate dean for academic affairs and graduate programs at the University of Iowa College of Education and professor of the sociology of education. He holds a secondary appointment in the Department of Sociology. Dr. Bills' research interests are in social stratification, education and the workplace, labor markets, technological and organizational change, educational demography, and social inequality. His current work focuses on the growing role of labor market intermediaries and employers' use of algorithms for personnel decisions. Dr. Bills has held short-term appointments at the Wissenschaftszentrum Berlin für Sozialforschung and the Institute for the Study of Social Inequality in Amsterdam. He is collaborating on various research projects with colleagues from Brazil, Germany, Ukraine, Italy, the Netherlands, Luxembourg, and Albania. He holds a Ph.D. in sociology from the University of Wisconsin, Madison.

CORBIN M. CAMPBELL is assistant professor of higher education at Teachers College, Columbia University. Her research focuses on understanding the organizational contexts that support learning and growth for students and faculty in higher education. Her recent work focuses on the assessment of quality in higher education and how the conceptualization and measurement of quality may serve to incubate (or alternatively result in pulling away from) college teaching and learning. Her secondary work explores the organizational contexts that

facilitate faculty's sense of agency in their careers. Dr. Campbell received a National Academy of Education/Spencer Foundation fellowship and subsequent funding from the Spencer Foundation for her work on assessing college educational quality. She serves on the editorial boards of *Review of Higher Education*, *Research in Higher Education*, and the *Journal of College Student Development*. She holds a Ph.D. in higher education from the University of Maryland, an M.A. from The Ohio State University, and a B.A. in psychology from the University of Virginia.

TABBYE CHAVOUS is professor of education and psychology and associate dean for academic programs and initiatives at the Rackham Graduate School, University of Michigan (UM). Her research interests and projects center on social identity processes (around race/ethnicity, gender, and social class) among ethnic minority adolescents and young adults in secondary and postsecondary education contexts and implications for students' academic identity development (including academic engagement and motivation), subsequent academic performance and persistence, and psychological adjustment. Her work also focuses on measurement and impacts of diversity and multicultural climates within secondary and higher education settings. Dr. Chavous is a principal investigator and co-director of UM's Center for the Study of Black Youth in Context. She currently has an NSF grant for a project examining race, gender, and academic identification processes among college students pursuing academic pathways in science, technology, engineering, and mathematics fields. She received her Ph.D. in community psychology from the University of Virginia.

GREG J. DUNCAN is distinguished professor of education at the University of California, Irvine. He has published extensively on neighborhood effects on the development of children and adolescents and other issues involving welfare reform, income distribution, and its consequences for children and adults. His recent research has shifted from these environmental influences to the comparative importance of the skills and behaviors developed during childhood. In particular, he seeks to understand the relative importance of early academic skills, cognitive and emotional self-regulation, and health in promoting children's eventual success in school and the labor market. Dr. Duncan was elected president of the Society for Research in Child Development for 2009-2011, was awarded the Klaus J. Jacobs Research Prize 2013, and received the Society for Research in Child Development Award for Distinguished Contributions to Public Policy and Practice in Child Development in 2015. Dr. Duncan is member of the National Academy of Sciences. He received his Ph.D. in economics from the University of Michigan.

SYLVIA HURTADO is a professor and served more than a decade as director of the Higher Education Research Institute at University of California, Los Angeles (UCLA). She has published numerous articles and books related to her primary interest in student educational outcomes, campus climates, college impact on student development, and diversity in higher education. She has served on a number of editorial boards for journals in education and served on boards including that of the Higher Learning Commission, and is past president of the Association for the Study of Higher Education. In 2015, she was awarded the Exemplary Research Award from the Postsecondary Division of the American Educational Research Association. Dr. Hurtado has coordinated several national research projects, including a U.S. Department of Education-sponsored project on how colleges are preparing students to achieve

the cognitive, social, and democratic skills needed to participate in a diverse democracy. She is engaged in two National Institutes of Health projects on the preparation of underrepresented students for biomedical and behavioral science research careers. She obtained her Ph.D. in education from UCLA, Ed.M. from Harvard Graduate School of Education, and A.B. in sociology from Princeton University.

PATRICK C. KYLLONEN is senior research director of the Center for Academic and Workforce Readiness and Success at the Educational Testing Service in Princeton, New Jersey Center scientists conduct innovative research on higher education assessment; workforce readiness; international large-scale assessment research; and assessment of 21st-century skills assessment, such as creativity, collaborative problem solving, and situational interviews. Dr. Kyllonen is a fellow of the American Psychological Association and the American Educational Research Association. He is co-author (with A. von Davier and M. Zhu) of *Innovative Assessment of Collaboration* and has served on several National Academies committees. Dr. Kyllonen received his B.A. from St. John's University and Ph.D. from Stanford University.

DAN P. MCADAMS is Henry Wade Rogers professor of psychology and professor of human development and social policy at Northwestern University. Currently, he also serves as chair of the Psychology Department. Professor McAdams works in the areas of personality and life-span developmental psychology. His theoretical and empirical writings focus on concepts of self and identity in contemporary American society and on themes of power, intimacy, redemption, and generativity across the adult life course. He is the winner of the Henry A. Murray Award from the American Psychological Association for research on personality and the study of lives, the 2006 Theodore Sarbin Award for contributions to theoretical and philosophical psychology, and the 2012 Jack Block Award from the Society for Personality and Social Psychology for career contributions to personality psychology. He is a fellow of the American Psychological Association (Division 8) and the American Psychological Society, has served on the Executive Committee of the Society for Personality and Social Psychology, and is a founding member of the Association for Research in Personality. Professor McAdams received his B.S. from Christ College, Valparaiso University in 1976, and his Ph.D. in psychology and social relations from Harvard University in 1979.

FREDERICK L. OSWALD is professor of industrial-organizational psychology in the Department of Psychology at Rice University. His substantive expertise deals with psychological testing and personnel selection within organizational, education, and military settings. Substantively, his work focuses on defining, modeling, and predicting societally relevant outcomes (e.g., job performance, academic performance, satisfaction, turnover) from psychological measures that are based on cognitive and motivational constructs (e.g., cognitive abilities, personality traits, situational judgment tests, job knowledge and skill, and biographical data). His statistical work in meta-analysis, structural equation modeling, and adverse impact also informs psychological testing and personnel selection issues in the research, practice, and legal arenas. Dr. Oswald is president of the Society for Industrial and Organizational Psychology (American Psychological Association [APA] Division 14). He is also a fellow of Evaluation, Measurement, and Statistics (APA Division 5); APA; and the Association for Psychological Science. He received a B.A. in psychology from the University of Texas at Austin in 1992 and his Ph.D. in psychology from the University of Minnesota in 1999.

JONATHAN PLUCKER is Julian C. Stanley endowed professor of talent development at The Johns Hopkins University, Center for Talented Youth and School of Education. He was previously Raymond Neag endowed professor at the University of Connecticut and professor of educational psychology and cognitive science at Indiana University, where he was founding director of the Center for Evaluation and Education Policy. His work focuses on education policy and talent development. Recent work includes a research collaboration with the Partnership for 21st Century Skills and studies of creative and affective assessment. His work on defining and studying excellence gaps is part of a larger effort to reorient policy makers' and educators' thinking about how best to promote success and high achievement for all children. Dr. Plucker is a fellow of the American Psychological Association (APA) and the American Association for the Advancement of Science, and recipient of the 2012 Arnheim Award for Outstanding Achievement from APA and 2013 Distinguished Scholar Award from the National Association for Gifted Children. He received his B.S. in chemistry education and M.A. in educational psychology from the University of Connecticut and his Ph.D. in educational psychology from the University of Virginia.

K. ANN RENNINGER is Dorwin P. Cartwright professor of social theory and social action and chair of the Department of Educational Studies, Swarthmore College. Her research interests include the role of interest in learning and development; the relationship between interest and other motivational variables; change in the cognitive and affective functioning of learners; and links among theory, research, and practice as they pertain to changed understanding. Her research program focuses on the role of interest in learning and conditions that support the development and deepening of learner interest. She studies these questions across a variety of contexts both in and out of school, including children's play and students' work with expository text, mathematical word problems, and science. Her studies are typically undertaken in collaboration with practitioners. A former Spencer Fellow of the National Academy of Education, she received a B.A. from the University of Pennsylvania and her Ph.D. in education and child development from Bryn Mawr College.

BRIAN STECHER is a senior social scientist at the RAND Corporation, an associate director of RAND Education, and a professor at the Pardee RAND Graduate School. His research focuses on measuring educational quality and evaluating education reforms, with particular emphasis on assessment and accountability systems. He has directed prominent national and state evaluations of No Child Left Behind, mathematics and science systemic reforms, and class size reduction. His measurement-related expertise includes test development (prototype performance assessments for teacher certification, hands-on science tasks for middle school students), test validation (the quality of portfolio assessments in Vermont and Kentucky), and the use of assessments for school improvement (formative and interim assessments, the quality of classroom assessments, and measures of inter- and intrapersonal competencies). Dr. Stecher has presented findings to policy makers at the state and national levels, to practitioners, and to the public. He has published widely in professional journals and is currently a member of the editorial board of *Educational Assessment*. He received his Ph.D. from the University of California, Los Angeles.

Appendix E
Workshop Agenda

Committee on Assessing Intrapersonal and Interpersonal Competencies
Workshop on Assessing Intrapersonal and Interpersonal Competencies
December 14, 2015
National Academy of Sciences Building
2101 Constitution Avenue, NW
Washington, DC 20418
Lecture Room

OPEN SESSION

8:00 AM **Informal Introductions**

8:30 AM **Welcome and Introduction to the Workshop**

- ***Heidi Schweingruber,*** *Director, Board on Science Education*
- ***Joan Herman,*** *Co-Director Emeritus, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Committee Chair*

8:45 AM-10:30 AM **Session 1: Setting the Context**
Moderator: ***Joan Herman*****,** *CRESST, Committee Chair*

8:50-9:10 **The Growing Importance of Social Skills in the Workplace** (20 minutes)

- ***David Deming,*** *Harvard Graduate School of Education* (by teleconference)

9:10-9:35 **Assessing Intrapersonal and Interpersonal Competencies: The VALUE Strategy** (25 minutes)

- ***Carol Geary Schneider,*** *American Association of Colleges and Universities*

9:35-9:55 **Accountability, Assessment, and Institutional Change** (20 minutes)

- ***Alicia Dowd,*** *University of Southern California*

9:55-10:05 **Reflections on the Presentations** (10 minutes)

- ***Corbin Campbell,*** *Columbia University, Committee Member*

10:05-10:30 Questions from the Committee and the Audience[1] (25 minutes)

10:30 Break

10:45 AM-12:30 PM Session 2: Exploring Key Constructs

Moderator: *Jonathan Plucker,* *University of Connecticut, Committee Member*

10:50-11:10 Conscientiousness (20 minutes)

- ***Joshua Jackson,*** *Washington University in St. Louis*
  - The extent to which conscientiousness is amenable to environmental influence
  - The power of conscientiousness as a predictor of success in many fields
  - How conscientiousness works
  - How conscientiousness develops over the life course
  - What kinds of interventions might be done to enhance conscientiousness

11:10-11:30 Motivation and Interest (20 minutes)

- ***Judith Harackiewicz,*** *University of Wisconsin*
  - Defining motivation/achievement goals
  - How motivation develops in adolescence and the transition to adulthood
  - The relationship between motivation and academic achievement
  - Interventions to increase students' motivation in science, technology, engineering, and mathematics (STEM)

11:30-11:50 Openness to Diversity and Challenge (20 minutes)

- ***Nicholas Bowman,*** *University of Iowa*
  - Defining openness to diversity and challenge
  - How openness develops in adolescence and the transition to adulthood
  - The relationship between openness to diversity and college persistence
  - What kinds of interventions might enhance openness to diversity?

11:50-12:10 Reflections on the Presentations (10 minutes each)

- ***Dan P. McAdams,*** *Northwestern University, Committee Member*
- ***K. Ann Renninger,*** *Swarthmore College, Committee Member*

12:10-12:30 Questions from the Committee and the Audience (20 minutes)

12:30-1:30 Lunch

[1]Committee members taking web audience questions: Brian Stecher, Greg Duncan.

1:30 PM-2:45 PM **Session 3: Measurement Opportunities and Challenges**
Moderator: ***David Bills,*** *University of Iowa, Committee Member*

1:35-1:55 **Designing Assessments of Intrapersonal and Interpersonal Competencies** (20 minutes)
- ***Pat Kyllonen,*** *Educational Testing Service, Committee Member*

1:55-2:15 **Sociometric Measures of Teamwork Competencies** (20 minutes)
- ***Alex (Sandy) Pentland,*** *MIT Media Lab* (by teleconference)

2:15-2:25 **Reflections on the Presentations** (20 minutes)
- ***Fred Oswald,*** *Rice University, Committee Member*

2:25-3:00 **Questions from the Committee and the Audience** (35 minutes)

3:00 Break

3:15 PM-4:15 PM **Session 4: Reflections and Next Steps**
Moderator: ***Joan Herman,*** *National Center for Research on Evaluation, Standards, and Student Testing, Committee Chair*

3:15-3:50 **Workshop Reflections, Committee Perspectives (**5-10 minutes each**)**
- ***Tabbye Chavous,*** *University of Michigan* **(**perspectives on diversity and inclusion)
- ***Greg Duncan,*** *University of California, Irvine* (perspectives on malleability and importance of constructs)
- ***Sylvia Hurtado,*** *University of California, Los Angeles* (perspectives on assessment purposes and audiences)
- ***Brian Stecher,*** *RAND* (perspectives on measurement)

3:50-4:15 **Final Reflections and Questions**
- ***Joan Herman,*** *CRESST*
- Questions and comments from the audience

4:15 Adjourn Workshop