Listening. Learning. Leading.®

Research Report
ETS RR–14-22

Assessing Quantitative Literacy in Higher Education: An Overview of Existing Research and Assessments With Recommendations for Next-Generation Assessment

Katrina Crotts Roohr

Edith Aurora Graf

Ou Lydia Liu

September 2014

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

Assessing Quantitative Literacy in Higher Education: An Overview of Existing Research and Assessments With Recommendations for Next-Generation Assessment

Katrina Crotts Roohr, Edith Aurora Graf, & Ou Lydia Liu

Educational Testing Service, Princeton, NJ

Quantitative literacy has been recognized as an important skill in the higher education and workforce communities, focusing on problem solving, reasoning, and real-world application. As a result, there is a need by various stakeholders in higher education and workforce communities to evaluate whether college students receive sufficient training on quantitative skills throughout their postsecondary education. To determine the key aspects of quantitative literacy, the first part of this report provides a comprehensive review of the existing frameworks and definitions by national and international organizations, higher education institutions, and other key stakeholders. It also examines existing assessments and discusses challenges in assessing quantitative literacy. The second part of this report proposes an approach for developing a next-generation quantitative literacy assessment in higher education with an operational definition and key assessment considerations. This report has important implications for higher education institutions currently using or planning to develop or adopt assessments of quantitative literacy.

Keywords Quantitative literacy; quantitative reasoning; mathematics; numeracy; student learning outcomes; higher education; next-generation assessment

doi:10.1002/ets2.12024

Literacy is defined as “the ability to read and write” or “knowledge that relates to a specified subject” (Merriam-Webster, 2014, para. 1–2). Building from this definition, quantitative literacy has been defined as the ability to interpret and communicate numbers and mathematical information throughout everyday life (e.g., Organisation for Economic Co-Operation and Development [OECD], 2012b; Rhodes, 2010; Sons, 1996; Steen, 2001). Sharing many common characteristics with other related constructs, such as numeracy, quantitative reasoning, and mathematical literacy, quantitative literacy emphasizes skills related to problem solving, reasoning, and real-world application (Mayes, Peterson, & Bonilla, 2013; Steen, 2001). Unlike traditional mathematics and statistics, quantitative literacy is a “habit of mind” (Rhodes, 2010, p. 25; Steen, 2001, p. 5), focusing on certainty rather than uncertainty and data from the empirical world rather than the abstract (Steen, 2001, p. 5).

Quantitative literacy can be considered an essential element in society, especially in relation to many duties of citizens, such as the “allocation of public resources, understanding media information, serving on juries, participating in community organizations, and electing public leaders” (Steen, 2004, p. 28). The importance of quantitative literacy in society has been recognized by the higher education community (Rhodes, 2010). For instance, 91% of the member institutions of the Association of American Colleges and Universities (AAC&U) identified quantitative reasoning as an important learning outcome (AAC&U, 2011). Employers have also recognized the need for quantitative skills, insisting that all college graduates have quantitative skills regardless of their intended career path (National Survey of Student Engagement [NSSE], 2013a). In a recent online survey conducted by Hart Research Associates (2013), among the 318 employers surveyed about necessary skills for a successful college graduate in today’s economy, 90% stated that higher education institutions should continue to emphasize or increase the emphasis on a students’ ability to work with numbers and understand statistics (Hart Research Associates, 2013). Similarly, Casner-Lotto and Barrington (2006) found that among 400 surveyed employers, 64.2% identified mathematics as a very important basic knowledge skill for 4-year college graduates to be successful in today’s workforce. The authors also noted that basic mathematical skills underpin applied skills such as critical thinking and problem solving.

Corresponding author: K. C. Roohr, E-mail: kroohr@ets.org

Although the importance of quantitative literacy is recognized both in higher education and the workforce, many students do not feel prepared to use quantitative reasoning skills in the workplace. A survey conducted by McKinsey and Company (2013) was administered to 4,900 former Chegg (a textbook rental company) customers, which included a mix of 2- and 4-year college students graduating between 2009 and 2012. Among the students surveyed, 24% of 4-year college students and 34% of 2-year college students felt underprepared to use quantitative reasoning skills upon graduating college (McKinsey & Company, 2013). The underpreparedness of 2- and 4-year college students may be linked to the lack of student engagement in quantitative reasoning tasks in either a student's freshman year or student's senior year of college. For instance, the 2013 NSSE found that 49–63% of freshman (NSSE, 2013b) and 46–56% of senior (NSSE, 2013c) students either never or only sometimes reached conclusions based on their own analysis of numerical information, used numerical information to examine real-world problems, or evaluated other people's conclusions from numerical information. Results also found that students in fields other than science, technology, engineering, and mathematics (STEM; e.g., social science, education, communication, arts, and humanities) engaged in quantitative activities less often than their peers in STEM majors (NSSE, 2013a). Given the mismatch between college students' preparedness in quantitative literacy and the demands from stakeholders, there is an urgent need by various stakeholders in higher education and workforce communities to evaluate whether students receive sufficient training in quantitative skills in college.

Results from the Program for the International Assessment for Adult Competencies (PIAAC) also showed the underpreparedness of students' quantitative skills. PIAAC Numeracy measures adults' mathematical skills in real-world contexts. When focusing on adults aged 16 to 65 with bachelor's degrees, results showed that only 18% of US adults with a bachelor's degree scored in the top two proficiency levels (out of five) on the Numeracy measure, which was below an international average of 24% (Goodman et al., 2013). These results point to the critical need to understand why adult Americans are behind in quantitative literacy skills. Actions should be taken to delineate the various components underlying quantitative literacy, and quality assessments should be developed to identify students' strengths and weaknesses in quantitative literacy when they enter college.

The purposes of this report are to review and synthesize existing frameworks, definitions, and assessments of quantitative literacy, quantitative reasoning, numeracy, or mathematics and to propose an approach for developing a next-generation quantitative literacy assessment. We first examine how quantitative literacy is defined throughout the literature by various stakeholders with a focus in higher education. We then review existing assessments of quantitative literacy, quantitative reasoning, numeracy, or mathematics, considering both the structural and psychometric quality of those assessments. Following this review, we discuss challenges and issues surrounding the design of a quantitative literacy assessment. After reviewing and synthesizing the existing frameworks, definitions, and assessments, we propose an approach for developing a next-generation quantitative literacy assessment with an operational definition, framework, item formats, and task types. The goal of this article is to provide an operational framework for assessing quantitative literacy in higher education while also providing useful information for institutions developing in-house assessments. The next-generation assessment development should involve collaboration between institutions and testing organizations to ensure that the assessment has instructional value and meets technical standards.

Existing Frameworks, Definitions, and Assessments of Quantitative Literacy

Existing Frameworks and Definitions

Various terms have been used to represent the use of quantitative skills in everyday life, such as quantitative literacy, quantitative reasoning, numeracy, mathematical literacy, and mathematics (Mayes et al., 2013; Steen, 2001). These various terms have subtle differences in their definitions (Steen, 2001). Vacher (2014) attempted to decipher these subtle differences using WordNet, an online lexical database for English, and also found that the terms *numeracy*, *quantitative literacy*, and *quantitative reasoning* have subtle differences in their meaning, even though they are commonly treated as synonymous terms. Using WordNet, Vacher proposed four core components that correspond to these terms including: (a) "skill with numbers and mathematics," (b) "ability to read, write and understand material that includes quantitative information," (c) "coherent and logical thinking involving quantitative information," and (d) "disposition to engage rather than avoid quantitative information" (p. 11). The author proposed that numeracy includes (a), (b), and (d); quantitative literacy includes (b), (c), and (d); and quantitative reasoning includes (c) and (d) (Vacher, 2014). Note that these categorizations are also arbitrary.

With various terms being used, there has been some disagreement among faculty in higher education institutions about how quantitative literacy is defined (Steen, 2004). Despite this disagreement, definitions of quantitative literacy and similar constructs throughout the literature have many commonalities, as shown in Vacher (2014). Recognizing these commonalties is critical to develop a concrete definition of quantitative literacy. Definitions throughout the literature have been developed either for understanding what it means to be quantitatively literate or for developing assessments and curricula. This section describes frameworks and definitions of quantitative literacy and synonymous terms or constructs (e.g., quantitative reasoning, numeracy) identified in the literature by national and international organizations, workforce initiatives, higher education institutions and researchers, and K–12 theorists and practitioners.

Frameworks by National and International Organizations

AAC&U's Liberal Education and America's Promise (LEAP) and Lumina's Degree Qualifications Profile (DQP) are two higher education initiatives developed by national organizations that identify quantitative skills as an element of their frameworks. The LEAP initiative was launched in 2005 and emphasizes the importance of a 21st century liberal education (AAC&U, 2011). Similarly, the DQP tool was developed with the intent of transforming US higher education by clearly identifying what students should be expected to know and do upon earning an associate's, bachelor's, or master's degree (Adelman, Ewell, Gaston, & Schneider, 2011). Both initiatives discuss important educational outcomes at the college level, with LEAP focusing on outcomes for every college student (AAC&U, 2011) and DQP focusing on outcomes for college students at specific degree levels, regardless of student major (Adelman et al., 2011). As part of the LEAP initiative, a set of Valid Assessment of Learning in Undergraduate Education (VALUE) rubrics was developed for each learning outcome, including quantitative literacy. In defining quantitative literacy, both quantitative reasoning and numeracy are recognized as synonymous terms to quantitative literacy (Rhodes, 2010). The rubric identified six important skills of quantitative literacy: interpretation, representation, calculation, application/analysis, assumptions, and communication, each defined in terms of proficiency level (Rhodes, 2010). Alternatively, the DQP uses the term *quantitative fluency* and breaks down quantitative fluency into different categories based on degree level, discussing different skills such as interpretation, explanation of calculations, creation of graphs, translation of problems, construction of mathematical arguments, reasoning, and presentation of results in various formats (Adelman et al., 2014).

Similar efforts in defining quantitative literacy have been made by the American Mathematical Association of Two-Year Colleges (AMATYC; Cohen, 1995), the Mathematical Association of America (MAA; Sons, 1996), and the OECD (2012b). The AMATYC developed a clear set of standards for introductory college mathematics intended for college students obtaining either an associate's or a bachelor's degree, similar to the DQP. However, instead of describing various quantitative skills for students across degree levels, a framework for mathematics standards was developed, focusing on students' intellectual development, instructors' pedagogical practices, and curricular content in higher education. The OECD (2012a) developed a framework with four facets of numeracy—contexts, responses, mathematical content/information/ideas, and representations—as well as a list of enabling factors and processes, such as the integration of mathematical knowledge and conceptual understanding of broader reasoning, problem-solving skills, and literacy skills. Alternatively, the MAA simply provided a list of five skills that every college student should have to be quantitatively literate, emphasizing skills such as interpretation, representation, problem solving, and estimation (Sons, 1996). These mathematical skills are similar to those enumerated by other national and international associations. Quantitative literacy definitions from these national and international organizations can be found in Table 1.

Frameworks by Workforce Initiatives

The US federal government and workforce initiatives have also recognized the importance of student learning outcomes but have focused on the term mathematics. The Employment and Training Administration's Industry Competency Model, developed by the US Department of Labor (USDOL), models essential skills and competencies for the workplace, specifically, economically important industries in the health, technology, and science fields (USDOL, 2013). This model, unlike the models developed by national and international organizations, is represented by stacked building blocks with more general competencies at the bottom building to more narrow competencies at the top. The second block from the bottom, the academic block, defines mathematics in terms of quantification, computation, measurement and estimation, and

Table 1 Definitions of Quantitative Literacy From National and International Organizations

Framework	Definition
AAC&U's Liberal Education and America's Promise	"Quantitative literacy (QL)—also known as numeracy or quantitative reasoning—is a 'habit of mind,' competency, and comfort in working with numerical data. Individuals with strong QL skills posses the ability to reason and solve quantitative problems from a wide array of authentic contexts and everyday life situations. They understand and can create sophisticated arguments supported by quantitative evidence and they can clearly communicate those arguments in a variety of formats (using words, tables, graphs, mathematical equations, etc., as appropriate)" (Rhodes, 2010, p. 25).
Lumina's Degree Qualifications Profile 2.0	"The student [at the bachelor's level] translates verbal problems into mathematical algorithms as to construct valid arguments using the accepted symbolic system of mathematical reasoning and presents the resulting calculations, estimates, risk analyses or quantitative evaluations of public information in papers, projects or multimedia presentations. The student constructs mathematical expressions for complex issues most often described in non-quantitative terms" (Adelman et al., 2014, p. 22).
Mathematical Association of America (MAA)	"A college student who is considered quantitatively literate should be able to: 1. Interpret mathematical models such as formulas, graphs, tables, and schematics, and draw inferences from them. 2. Represent mathematical information symbolically, visually, numerically, and verbally. 3. Use arithmetical, algebraic, geometric and statistical methods to solve problems. 4. Estimate and check answers to mathematical problems in order to determine reasonableness, identify alternatives, and select optimal results. 5. Recognize that mathematical and statistical methods have limits" (Sons, 1996, Part II, para. 6).
Organisation for Economic Co-Operation and Development (OECD)	"The ability to access, use, interpret and communicate mathematical information and ideas in order to engage in and manage the mathematical demands of a range of situations in adult life. To this end, numeracy involves managing a situation or solving a problem in a real context, by responding to mathematical content/information/ideas represented in multiple ways" (OECD, 2012b, p. 20).

application, defining important content within each skill area (USDOL, 2013). Another workforce-based definition for mathematics was developed by Capital Workforce Partners (2014), with a list of career competency standards based on a range of interviewed employees. These standards include quantitative skills such as a person's basic ability to do mathematics, apply mathematics to business, create tables and graphs, integrate information, and use mathematical functions (Capital Workforce Partners, 2014).

Frameworks by Higher Education Institutions and Researchers

In addition to the higher education initiatives by the AAC&U and Lumina Foundation, institutions have developed in-house frameworks that guide quantitative literacy or quantitative reasoning assessments and coursework. Many of these in-house frameworks are similar in structure to the AAC&U VALUE rubric with a list of skills at different quantitative literacy proficiency levels (e.g., Samford University, 2009; University of Kentucky, 2012). Other institutions, such as Michigan State University, have developed standards for students at three different stages of quantitative literacy development, a similar approach to Lumina's DQP (Estry & Ferrini-Mundy, 2005). Alternatively, like the MAA, some institutions simply list the skills required of a quantitatively literate individual (e.g., Mount St. Mary's College, 2013). Compared with the large-scale higher education frameworks and those of national organizations such as the MAA, the definitions of quantitative literacy show considerable overlap, including skills such as application, evaluation of arguments, quantitative expression, interpretation, reasoning, and problem solving.

Like higher education institutions, researchers have also attempted to construct definitions, frameworks, and standards for quantitative literacy. For example, Steen (2001) identified 10 quantitative literacy elements such as confidence with mathematics, interpreting data, logical thinking, mathematics in context, and number and symbol sense. Likewise, Mayes et al. (2013) developed a framework for quantitative reasoning in the context of science, focusing on components such as quantification act (i.e., identifying objects, observing attributes, and assigning measures), and quantitative literacy, interpretation, and modeling. Among the definitions developed by researchers, many have defined quantitative literacy in terms of application to real-world problems (Hollins University, 2013; Kirsch, Jungeblut, Jenkins, & Kolstad, 2002; National Numeracy Network [NNN], 2013; OECD, 2000; Ward, Schneider, & Kiper, 2011), or in terms of reasoning skills (J. Bennett & Briggs, 2008; Hollins University, 2013; Langkamp & Hull, 2007; NNN, 2013; Steen, 2004).

Frameworks and Standards by K–12 Experts and Practitioners

The most well-known K–12 standards relevant to quantitative literacy are the Common Core State Standards for Mathematics developed by the Council of Chief State Officers (CCSSO) and the National Governors Association (NGA) for Best Practices. Although developed for K–12 with a focus on standards for mathematics that should be taught in school, the Common Core State Standards for Mathematics were constructed to help improve students' college and career readiness in terms of quantitative knowledge and skills, identifying specific mathematical content areas and competencies students need to master, such as problem solving, reasoning, modeling, and expression, within the content areas of number and quantity, algebra, functions, modeling, geometry, and statistics and probability (NGA & CCSSO, 2010). These various skills identified in the Common Core State Standards for Mathematics are highly related to many of the higher education and workforce definitions of quantitative literacy and quantitative reasoning.

The American Diploma Project (Achieve, Inc., 2004) also linked K–12 education to postsecondary education and careers. This project established a set of English and mathematical skills and benchmarks that high school graduates should master to be successful in their future endeavors. Mathematics benchmarks were organized into four content strands: (a) number sense and numerical operations, (b) algebra, (c) geometry, and (d) data interpretation, statistics, and probability. The American Diploma Project also noted that mathematical skills are crosscutting and involve a student's ability to blend knowledge and skills when problem solving, to connect new information with existing knowledge, and to access and assess knowledge from a variety of sources (Achieve, Inc., 2004), which are common skills identified within quantitative literacy. These mathematical skills and benchmarks in both the Common Core State Standards and American Diploma Project are comparable to many of the skills identified within higher education and workforce initiatives.

Another set of K–12 frameworks, focusing more on noncognitive skills within core subject areas, includes the Partnership for 21st Century Skills (P21) Math Map (Saltrick et al., 2011). This framework differs from other frameworks by focusing on mathematical content knowledge and mathematical processes integrated with 21st century skills such as creativity and innovation, critical thinking and problem solving, communication and collaboration, and other noncognitive skills. This framework is intended to make teaching and learning of mathematics more engaging, relevant, and rigorous for students (Saltrick et al., 2011).

Existing Assessments Measuring Quantitative Literacy Skills

A number of tests and subtests assess the quantitative literacy, quantitative reasoning, numeracy, or mathematics skills of students in higher education. Most of these assessments are multiple-choice tests administered on a computer. Table 2 summarizes these existing college-level and adult-level assessments, which include the three assessments approved by the Voluntary System of Accountability Program (VSA) to provide evidence of student learning outcomes in colleges and universities: the Collegiate Assessment of Academic Proficiency (CAAP), Collegiate Learning Assessment+ (CLA+), and the *ETS*® Proficiency Profile (EPP; VSA, 2013). Other assessments measuring quantitative literacy, quantitative reasoning, numeracy, or mathematics include the College-Level Examination Program (*CLEP*®), Graduate Management Admissions Test (GMAT), the *GRE*® General Test, National Assessment of Adult Literacy (NAAL), PIAAC, and two assessments developed by Insight Assessment, including Quant Q and the Test of Everyday Reasoning—Numeracy (TER-N).

In addition to these widely used assessments measuring quantitative literacy skills, many institutions have developed their own quantitative assessments. For example, the University of Cambridge developed an essay assessment called the Sixth Term Examination Papers in Mathematics (STEP) to evaluate pure mathematics, mechanics, and probability and

Table 2 Existing Assessments Measuring Quantitative Literacy Skills

Test	Developer	Format	Delivery	Length	# Items	Themes/topics
College-Level Examination Program (CLEP) College Mathematics	College Board	Multiple choice; multiple-selection multiple choice; numeric entry	Computer	90 min	60 items (not all items contribute to final score)	Measures the examinees' ability to solve routine, straightforward problems, and nonroutine problems that require an understanding of concepts and application of skills and concepts. Topics on the assessment include sets, logic, real number system, functions and their graphs, probability and statistics, algebra, and geometry (College Board, 2012).
Collegiate Assessment of Academic Proficiency (CAAP) Mathematics	ACT	Multiple choice	Paper/pencil	40 min	35 items	Assesses proficiency in solving mathematical problems encountered in many postsecondary curricula, emphasizing quantitative reasoning rather than memorization of formulas. The content areas tested include (a) pre-algebra, (b) elementary, intermediate, and advanced algebra, (c) coordinate geometry, and (d) trigonometry (CAAP Program Management, 2012).
Collegiate Learning Assessment+ (CLA+) Scientific and Quantitative Reasoning (SQR)	Council for Aid to Education (CAE)	Multiple choice	Computer	30 min (not a distinct subtest—SQR items are within the 30 min period)	10 SQR items (out of 26 total items on the full CLA+)	A set of 10 multiple-choice items all attached to documents that emulate real-world scenarios or problems in a work or academic environment. Documents include reference sources such as data tables or graphs, a newspaper article, research report, etc. These multiple-choice items require careful analysis and evaluation of information by examinees (Zahner, 2013).
ETS Proficiency Profile (EPP) Mathematics	Educational Testing Service (ETS)	Multiple choice	Computer and paper/pencil	Approximately 30 min (full test is 2 hours)	27 items (standard form)	Assesses the ability to "recognize and interpret mathematical terms; read and interpret tables and graphs; evaluate formulas; order and compare large and small numbers; interpret ratios, proportions, and percentages; read scientific measuring instruments; recognize and use equivalent mathematical formulas or expressions" (ETS, 2010, p. 4).

Table 2 Continued

Test	Developer	Format	Delivery	Length	# Items	Themes/topics
Graduate Management Admissions Test (GMAT) Quantitative	Graduate Management Admission Council (GMAC)	Multiple choice	Computer	75 min	37 items	Measures the ability to reason quantitatively, solve quantitative problems, and interpret graphical data. Both problem-solving and data-sufficiency questions are used and require the knowledge of arithmetic, elementary algebra, and commonly known concepts of geometry (GMAC, 2013a).
Graduate Record Examinations® (GRE) Quantitative Reasoning Measure	ETS	Multiple choice; multiple-selection multiple choice; numeric entry	Computer and paper/pencil	70 min	40 items	Measures the ability to interpret and analyze quantitative information and use mathematical skills in arithmetic, algebra, geometry, and data interpretation to solve problems (ETS, 2013b).
National Assessment of Adult Literacy (NAAL) Quantitative Literacy	US Department of Education	Open-ended/short answer	Paper/pencil	Untimed	47 items	Assesses the ability to "identify, describe, or perform an arithmetic operations (addition, multiplication, subtraction, and division) either in prose or document materials" (Institute of Educational Statistics, n.d., para. 5).
Programme for the International Assessment of Adult Competencies (PIAAC) Numeracy	Organisation for Economic Co-Operation and Development (OECD)	Multiple choice; clicking/selecting objects; numeric entry; highlighting objects	Computer and paper/pencil	Around 60 min (but is not timed)	56 items	Measures the ability to solve problems in real contexts (everyday life, work, society, further learning) by responding (identify, locate or access; act upon and use: order, count, estimate, compute, measure, model; interpret; evaluate/analyze; communicate) to mathematical content (quantity and number; dimension and shape; pattern, relationships and change; data and chance), represented in multiple ways (objects and pictures; numbers and symbols; formulae; diagrams and maps, graphs, tables; texts; technology-based displays) (OECD, 2012b).

Table 2 Continued

Test	Developer	Format	Delivery	Length	# Items	Themes/topics
Quant Q	Insight assessment	Multiple choice	Computer and paper/pencil	50 min	28 items	Assesses basic mathematical knowledge and integration of critical thinking skills, as well as quantitative reasoning. Score reports indicate the test taker's skills in pattern recognition, probability combinatorics, geometry and optimization, and out-of-the-box algebra (i.e., items involving algebra or other more basic mathematical techniques) (Insight Assessment, 2013a).
Test of Everyday Reasoning—Numeracy (TER-N)	Insight assessment	Multiple choice	Computer and paper/pencil	Not available	40 items	Measures quantitative reasoning in addition to critical thinking skills. Score reports indicate the test taker's skills in numeracy and reasoning skills, including overall general reasoning, analysis, interpretation, evaluation, inference, explanation, induction, and deduction (Insight Assessment, 2013b).

statistics (Admissions Testing Service, 2013). The STEP Mathematics is also used across other institutions in the United Kingdom such as the University of Warwick (University of Warwick, 2013). Similarly, the Center for Assessment and Research Studies at James Madison University developed a multiple-choice assessment called the Quantitative Reasoning Test, Version 9 (QR-9). This assessment measures learning objectives such as the use of different mathematical methods to analyze, organize, and interpret different phenomena. The assessment also evaluates a student's ability to discriminate between association and causation (Sundre, 2008). Other measures developed to assess quantitative literacy or quantitative reasoning include the Quantitative Literacy Skills Assessment by Colby-Sawyer College (Steele & Kilic-Bahi, 2008), the Quantitative Literacy Assessment by Miami University (Ward et al., 2011), the Quantitative Reasoning Assessment by Wellesley College (Wellesley College, n.d.), and Carleton College's Quantitative Inquiry, Reasoning, and Knowledge (QuIRK) initiative (Carleton College, 2013).

Similarly, a number of K–12 assessments target aspects of quantitative literacy such as the Programme for International Student Assessment (PISA) Mathematics, an international assessment measuring mathematical literacy for 15-year-old students, and PISA Financial Literacy, an international assessment for 15-year-olds measuring student knowledge and application of both financial concepts and risks (OECD, 2013). In addition to international assessments, national K–12 accountability mathematics assessments have been built using the Common Core State Standards, such as the Partnership for Assessment of Readiness for College and Careers (PARCC, 2014), and the Smarter Balanced Assessment Consortium (SBAC, n.d.). Likewise, a research and development initiative is being conducted at the Educational Testing Service (ETS) on a K–12 accountability measure called the Cognitively-Based Assessment *of*, *for*, and *as* Learning (*CBAL*™), with one of the content areas being mathematics. The goal of this initiative is to unify three main components: accountability, formative assessment, and professional support (ETS, 2014a).

In the following sections we discuss the test content, contexts, item types, calculator use, test reliability, and validity evidence, including convergent, concurrent, and predictive validity evidence.

Test Content and Contexts

The existing assessments measuring quantitative literacy skills assess a variety of content areas and contexts. Content is defined as the mathematical knowledge and skills needed to answer a question, and context is defined as the setting described in the question (Dwyer, Gallagher, Levin, & Morley, 2003). The assessed content is identified for all assessments except the CLA+ Scientific and Quantitative Reasoning (SQR), with the most commonly identified content consisting of geometry and measurement, algebra, probability and statistics, number sense, arithmetic, and pre-algebra (see Table 3). Additionally, items across assessments are written both to pure mathematical contexts and to applied contexts. Pure mathematical contexts include items that assess strict mathematical content such as solving an algebraic expression. Existing assessments with a proportion of test items written to a pure mathematical context include CAAP Mathematics, CLEP Mathematics, EPP Mathematics, GMAT Quantitative, and the GRE Quantitative Reasoning measure. Applied contexts vary across assessments and include contexts such as real-world scenarios (GRE Quantitative Reasoning), accompanying documentation (e.g., newspaper articles, data tables, emails; CLA+ SQR), problems encountered in postsecondary curricula (CAAP Mathematics), and specific disciplines (e.g., humanities, social sciences, and natural sciences; EPP Mathematics). PIAAC Numeracy has the most clearly defined contexts, with items written to work-related, personal, society and community, and education and training contexts (OECD, 2012b).

Item Format

Single- and multiple-selection multiple-choice items are the most commonly used item formats throughout the 10 assessments measuring quantitative literacy skills. For a single-selection multiple-choice item, the answer key consists of only one correct choice, while for a multiple-selection multiple-choice item, the answer key consists of one or more choices that satisfy the conditions specified in the question. Single-selection multiple-choice items are used across all of the assessments, and multiple-selection multiple-choice items are used by both the GRE Quantitative Reasoning and the PIAAC Numeracy. Some assessments (e.g., CLA+ SQR) also group multiple-choice items together using a common stimulus such as a table, graph, or other data presentation.

Table 3 Test Content on Existing Assessments Measuring Quantitative Literacy Skills

	Geometry and measurement	Algebra	Probability and statistics	Number sense	Arithmetic	Pre-algebra	Pattern recognition	Trigonometry
Collegiate Assessment of Academic Proficiency (CAAP) Mathematics	X	X				X		X
College-Level Examination Program (CLEP) Mathematics	X	X	X	X				
ETS Proficiency Profile (EPP) Mathematics	X	X	X	X	X	X		
Graduate Management Admissions Test (GMAT) Quantitative	X	X	X	X	X	X		
GRE Quantitative Reasoning	X	X	X	X	X	X		
National Assessment of Adult Literacy (NAAL) Quantitative Literacy					X			
Program for the International Assessment for Adult Competencies (PIAAC) Numeracy	X	X	X	X			X	
Quant Q	X	X	X				X	
Test of Everyday Reasoning—Numeracy (TER-N)					X			
Total	7	7	6	5	5	4	2	1

Multiple-choice format lends itself to various task types such as the quantitative comparison task found in the GRE Quantitative Reasoning measure and the data-sufficiency task in the GMAT Quantitative section. The quantitative comparison task involves the comparison of two quantities and asks the examinee to determine whether one of the quantities is greater than, less than, or equal to the other quantity, or whether the relationship is indeterminable based on the information provided (ETS, 2013c). The data-sufficiency task involves two statements and asks whether the statements provide sufficient information to answer a given question (GMAC, 2013a).

Computer delivery of an assessment allows for variations on traditional multiple-choice response formats, such as items involving the clicking or highlighting of objects, which are used by PIAAC Numeracy. For instance, for a multiple-selection multiple-choice item, instead of selecting multiple checkboxes, an examinee could click on multiple bars on a bar graph. In addition to multiple-choice items, another common item format across assessments is numeric entry, where an examinee enters a numeric value as the response rather than selecting one from a list of choices. Numeric entry is used by CLEP Mathematics, GRE Quantitative Reasoning, and PIAAC Numeracy.

Calculator Use

An important consideration with any assessment of mathematics is whether a calculator will be permitted. Existing higher education quantitative assessments such as the EPP and CAAP allow calculators but stress that all problems can easily be solved without a calculator. The PIAAC Numeracy measure also permits calculator use, recognizing that calculators are easily available when conducting quantitative tasks throughout everyday life (PIAAC Numeracy Expert Group, 2009). Similarly, the GRE Quantitative Reasoning measure allows an examinee to use a calculator to help shorten the time it takes to perform computation; however, it is noted that the calculator is provided solely as a supplement to the assessment and does not replace the examinee's knowledge of mathematics (ETS, 2014b).

The use of a calculator can be advantageous for a quantitative literacy assessment and can improve construct validity by allowing the examinee to focus on problem-solving skills rather than strict computation of a test item (Bridgeman, Harvey, & Braswell, 1995). Calculator use has also been found to improve mathematical problem-solving strategies and positively influence students' attitudes toward mathematics (Ellington, 2003). It is important to think about the impact of having a calculator on a quantitative literacy assessment. Although Bridgeman et al. (1995) found that construct validity can be improved with the use of a calculator, the authors also found that in some cases construct validity can decrease. A major advantage to having a computer-based assessment is that developers can easily include some items that allow a calculator and some items that do not allow a calculator (e.g., questions on estimation) while also controlling the calculator features (e.g., basic vs. scientific vs. graphing).

Test Reliability

Reliability refers to the consistency of measures (American Educational Research Association [AERA], American Psychological Association, & National Council on Measurement in Education, 1999; Traub & Rowley, 1991). Methods for estimating reliability include parallel form, test–retest, split-half, internal consistency, and interrater reliability. Both parallel form and test–retest reliability estimates require multiple test administrations, whereas split-half and internal consistency (e.g., coefficient α) estimates are derived from items within a single test administration. To estimate reliability on human-scored constructed-response items, interrater reliability is estimated by calculating the score agreement across multiple raters. Test length is highly related to reliability, with tests with a larger number of items typically yielding higher reliability estimates than tests with a smaller number of items (Traub & Rowley, 1991). For the same reason, a multiple-choice test typically has higher reliability than a constructed-response test, as more multiple-choice items can be administered than constructed-response items within the same time frame.

As discussed previously, many of the existing tests measuring quantitative literacy skills use multiple-choice items and have published results on test or subscale reliability. For instance, satisfactory reliability estimates have been found on the EPP Mathematics standard form with estimates around .85 (ETS, 2010; Lakin, Elliott, & Liu, 2012), on the CLEP College Mathematics with estimates around .90 (College Board, 2012), and on the CAAP Mathematics with estimates of .95 and .93 for freshman and senior students, respectively (Klein et al., 2009). Satisfactory reliability estimates have also been found for both the GRE Quantitative Reasoning measure and GMAT Quantitative section with reliability estimates of .95 and .90, respectively (ETS, 2013a; GMAC, 2013b). For both assessments with constructed-response items (i.e., NAAL Quantitative Literacy and PIAAC Numeracy), no information was found on the internal consistency of those measures; however, information on interrater reliability was reported. For example, the 2003 NAAL Quantitative Literacy showed high percent agreement between raters ranging from 92.6% to 100% (Baldi, 2009), and PIAAC Numeracy's high percent agreement was 99.1% within-country and 96.7% across countries (Tamassia, Lennon, & Yamamoto, 2013).

Convergent Validity Evidence

Convergent validity evidence looks at the relationship between scores across tests measuring similar constructs (AERA et al., 1999). Klein et al. (2009) examined the relationship among the three approved VSA measures and found a strong relationship between EPP and CAAP Mathematics with a student-level correlation of .76 and a school-level correlation of .98. These results provide evidence that both the EPP and CAAP Mathematics sections are measuring a similar construct. At the time of this study, the CLA did not have an equivalent quantitative literacy section to examine this relationship.

Concurrent Validity Evidence

Concurrent validity refers to the relationship between a predictor and a criterion measured at the same time rather than at a later time (AERA et al., 1999). Concurrent validity has been evaluated for EPP Mathematics by examining the relationship between student performance on EPP Mathematics and grade point average (GPA), finding that across a 10-year period, students with higher GPA consistently yielded higher EPP Mathematics scores (Liu & Roohr, 2013). A similar relationship was found between test takers' EPP Mathematics scores and the number of college credit hours they had taken (Liu & Roohr, 2013). These results suggest that indicators of students' success, such as GPA and the number of credit hours completed, are strongly associated with the level of performance on EPP Mathematics.

Predictive Validity Evidence

Predictive validity refers to how well particular outcomes of interest measured at a later time (e.g., first-year graduate student GPA) are predicted from test scores on an assessment that purports to measure relevant constructs (e.g., GRE;

AERA et al., 1999). Although some of the existing assessments (i.e., CAAP, CLA+, EPP) measure certain aspects of college learning outcomes, results from these assessments may also predict other college-level outcomes. To date, predictive validity for the assessments measuring quantitative literacy skills has been examined by looking at a variety of school-level outcomes. For example, moderate correlations ranging from .23 to .48 have been found between GPA or grades and test scores on CAAP Mathematics, GMAT Quantitative, and GRE Quantitative Reasoning (CAAP Program Management, 2012; GMAC, 2013b; Kuncel, Hezlett, & Ones, 2001). Small to moderate correlations have also been found between EPP Mathematics scores and credit hours or courses completed (Lakin et al., 2012; Marr, 1995). Other investigated school-level outcomes have included faculty ratings, comprehensive exam scores, publication citation count, degree attainments, time to complete, and research productivity. Operational predictive validity evidence (i.e., correlations with corrections for range restriction and criterion unreliability) has ranged from .11 to .47 between these additional school-level outcomes and GRE Quantitative Reasoning test scores (Kuncel et al., 2001).

It is evident that much of the existing predictive validity evidence has focused on the prediction of college-level outcomes; however, more predictive validity evidence is needed after students leave college. Essentially, future research should consider using a next-generation quantitative literacy assessment to predict long-term life outcomes. For instance, future research should evaluate the relationship between the assessment scores and whether a student can make sound quantitative decisions in life, such as making a decision between renting or buying a property. Making sound financial decisions was identified as a critical content area for college graduates in the workforce (Casner-Lotto & Barrington, 2006), so obtaining this evidence could help to predict whether students will have those skills related to financial decisions and other related quantitative skills that are common in the workforce.

Broad Issues in Assessing Quantitative Literacy in Higher Education

In developing a new assessment, it is important to consider challenges and broad issues in assessing that construct. Recognizing these challenges and issues during test development can help to ensure a reliable, valid, and fair assessment for examinees that is commensurate with the stakes of the assessment. This section describes a set of issues pertaining to the assessment of quantitative literacy in higher education.

Mathematics Versus Quantitative Literacy

When assessing quantitative literacy, it is important to understand the difference between quantitative literacy and traditional mathematics. Steen (2001) clearly addressed this difference, stating that mathematics typically focuses on a "Platonic realm of abstract structures," whereas quantitative literacy is more "anchored in data derived from and attached to the empirical world" (p. 5). Steen also noted that there is a difference between quantitative literacy and statistics. He stated that statistics is "primarily about uncertainty," whereas quantitative literacy is mainly about the "logic of certainty" (p. 5).

Quantitative literacy is distinctively different from mathematics and involves solving problems using mainly primary- and secondary-level mathematics within a particular context, such as the workplace, focusing on the student's ability to use reasoning skills to address those context-specific real-world problems (Steen, 2001, 2004). Another distinction between quantitative literacy and mathematics is that mathematics is typically practiced on its own as a discipline, whereas quantitative literacy is typically employed alongside other literacies (Ewell, 2001), such as reading and writing. The difference between quantitative literacy and mathematics can also be found across various assessments. For example, items on the *SAT*® and ACT are typically decontextualized and focus on strict mathematical content (Steen, 2004). Alternatively, quantitative literacy questions can be made very difficult using basic mathematical content and increasing the complexity of mathematical reasoning processes to reach a solution (Dwyer et al., 2003). Recognizing and understanding these differences between quantitative literacy and mathematics is critical when developing a quantitative literacy assessment.

General Versus Domain Specific

The quantitative skills a student is expected to master can vary based on the student's major, and this raises the following question: Should an assessment of quantitative literacy be domain specific or more general? There is no question that students pursuing a mathematics or science degree will take more quantitative courses than students pursuing an English or history degree. However, regardless of a student's major or career path, all students should receive context-rich quantitative literacy as part of their college education (Ewell, 2001; NSSE, 2013a) and should be able to "draw information from charts, graphs, and geometric figures," and have the "ability to complete straightforward estimations and calculations" (Rhodes, 2010, p. 25). In the 2005 study sponsored by the Conference Board of the Mathematical Sciences (CBMS2005; Lutzer, Rodi, Kirkman, & Maxwell, 2007), quantitative requirements were examined across 4-year institutions focusing on the requirements within a college at the institution (e.g., College of Arts and Sciences) with a mathematics department. Results indicated that 9 of 10 of these colleges at 4-year institutions had a quantitative requirement in Fall 2005 with courses such as calculus, elementary statistics, college algebra, pre-algebra, and special general education courses satisfying this requirement (Lutzer et al., 2007). These results show that all students, regardless of major, are being required to take quantitative courses.

Quantitative skills are not only an important part of a college education but are also needed in many aspects of society, such as dealing with money, working with schedules, or reading mainstream media. Developing an assessment to target more general quantitative literacy skills will allow for a broader representation of competencies, as well as those that are transferrable across majors. In the redesign of the DQP, Adelman et al. (2014) recognized the need for a framework to represent "broad, integrative studies and crosscutting proficiencies that graduates need for continued learning in complex and changing environments" (p. 12). The authors recognized that students change their jobs many times throughout their life, so having these general skills is important. Additionally, given that approximately 27% of students do not end up pursuing a career related to their major field of study (Abel & Deitz, 2013), it may be advantageous for students to take a more general quantitative literacy assessment rather than a domain-specific assessment.

Total Score Versus Subscores

Another challenge in developing a quantitative literacy assessment is whether to report subscores to examinees. The advantage of reporting subscores is that it provides score users with diagnostic information about their strengths and weaknesses; however, this advantage only holds if the subscores are sufficiently reliable and distinct from each other. As previously discussed, reliability is impacted by test length (Sinharay, Puhan, & Haberman, 2011; Traub & Rowley, 1991), and a short assessment is likely to yield lower reliability estimates. This phenomenon was found when examining the reliability of the CLA+ SQR section. With only 10 multiple-choice items, reliability estimates were quite low at .42 and .62 on Forms A and B, respectively (Zahner, 2013). These results suggest that caution needs to be used when reporting subscores based on a small number of test items.

It is not only important to consider the number of items within a subscore, but also to determine what diagnostic information the subscores should represent. A quantitative literacy assessment measures mathematical skills and content within varying real-world contexts. These skills and content areas are likely to overlap, making it difficult to distinguish which items belong to which specific subscore. Score users may also have different preferences for how the score information should be presented. For example, an employer may value a holistic score to evaluate a job candidate's overall quantitative competency, whereas a higher education institution may prefer subscores so actions may be taken to improve the areas that showed deficiencies. As a result, research should be conducted to determine the appropriateness of subscore reporting for a next-generation quantitative literacy assessment.

Students' Test-Taking Motivation

Students' motivation in taking low-stakes assessment is a major challenge to the validity of the test scores. In a recent study by Liu, Bridgeman, and Adler (2012), the authors found significant differences in performance among motivated and unmotivated students taking a low-stakes learning outcomes assessment in higher education. Without considering the role of motivation, invalid conclusions could be made about a higher education institution in relation to different learning outcomes. Liu et al. (2012) showed that nonfinancial incentives could successfully increase an individual student's motivation and therefore increase test scores. When developing a next-generation quantitative literacy assessment, a testing program or sponsor could consider offering a certificate of achievement to increase the relevance of the test to students so they may take it more seriously. This certificate could also be presented to future employers to show proficiency in quantitative literacy skills.

Proposing a Next-Generation Quantitative Literacy Assessment Framework

Using the information from the above review of definitions, frameworks, and assessments, we propose a next-generation quantitative literacy framework designed to measure the degree to which students in 2- and 4-year higher education institutions are able to identify and solve mathematical problems of practical importance. The framework is intended for all college students, regardless of major, and will inform the development of a quantitative literacy assessment that could be used for college accreditation, instructional improvement within the institution, and assessing individual students' proficiency levels.

Operational Definition

Development of an operational or construct definition is "essential to validity and fairness to the scientific integrity of the inferences drawn from assessments" and helps to aid "decision-making about assessment design, development, and interpretation" (Dwyer et al., 2003, p. 1). Based on a synthesized review of the frameworks and definitions by national and international organizations, workforce initiatives, higher educational institutions and researchers, and K–12 theorists and practitioners, as well as existing assessments measuring quantitative literacy skills, we propose an operational definition for a next-generation quantitative literacy assessment. The proposed definition is three-dimensional, addressing problem-solving skills, content, and contexts. Problem-solving skills are central to many of the frameworks and definitions of quantitative literacy, quantitative reasoning, and numeracy. Similarly, mathematical content is central to many of the assessments measuring quantitative literacy skills. Lastly, a clear distinction between mathematics and quantitative literacy is that quantitative literacy involves the application of primary and secondary-level mathematics to real-world problems, thus stressing the importance of contexts. As a result of our synthesized review, we define quantitative literacy as follows:

> Quantitative literacy is the comprehension of mathematical information in everyday life, and the ability to detect and solve mathematics problems in authentic contexts across a variety of mathematical content areas. Solving these applied mathematical problems includes (a) interpreting information, (b) strategically evaluating, inferring, and reasoning, (c) capturing relationships between variables by mapping, interpreting, and modeling, (d) manipulating mathematical expressions and computing quantities, and (e) communicating these ideas in various forms.

The three quantitative literacy dimensions—problem-solving skills, content, and contexts—are shown in Tables 4, 5, and 6, respectively. Each of these aspects of the proposed quantitative literacy framework is aligned with definitions, frameworks, and assessments throughout the relevant quantitative literacy literature.

Mathematical Problem-Solving Skills

Employers need people "who can figure out how to solve a problem; that's more than just knowing how to plug numbers into a calculator in the right order" (Stumbo & Lusi, 2005, p. 1). When solving problems, examinees should be able to pose questions, organize information, draw diagrams, analyze situations, draw conclusions, and communicate and interpret results (Cohen, 1995). The five most common mathematical problem-solving skills identified among the existing quantitative literacy frameworks, definitions, and assessments are (a) interpretation, (b) strategic knowledge and reasoning, (c) modeling, (d) computation, and (e) communication. Table 4 defines each of these five skills briefly and describes foci for assessing each skill. For example, communication involves the presentation of mathematical concepts, data, procedures, and solutions in a variety of forms (e.g., written, graphic, or tabular format). Like other skills in the framework, communication is aligned with existing quantitative frameworks (i.e., AAC&U's Quantitative Literacy VALUE Rubric, AMATYC, MAA, and P21 Math) and assessments (i.e., CBAL, PIAAC Numeracy, and PISA Mathematics). These five problem-solving skills are not mutually exclusive but, instead, are dynamically interrelated. For instance, to appropriately interpret a graphical display, an examinee may be required to use strategic knowledge and reasoning skills to solve a problem.

Mathematical Content

Dwyer et al. (2003) noted that to solve quantitative reasoning problems, mathematical content knowledge is needed. After examining and synthesizing the quantitative literacy literature and existing assessments, we identified four content areas

Table 4 Description of Mathematical Problem-Solving Skill Areas for the Proposed Quantitative Literacy Framework

Problem-solving skill	Brief description	Focus of assessment
(a) Interpretation	The understanding and explanation of mathematical information, such as the ability to understand data, read graphs, draw conclusions, and recognize sources of error	*Understand mathematical terms and representational devices* Read and interpret basic mathematical notation, concepts, and terminology, such as percentage and average, as well as relationships between quantities expressed in terms of equations, formulas, or data representations, such as tables, graphs, and other diagrams
(b) Strategic knowledge and reasoning	The formulation and evaluation of mathematics problems using heuristics, and the ability to recognize relationships about mathematical concepts and situations	*Build and develop mathematical strategies* Construct and explore mathematical strategies and heuristics to solve problems using inductive and deductive reasoning *Develop and test conjectures* Formulate mathematical hypotheses and evaluate their consequences *Evaluate the validity of mathematical strategies* Evaluate the accuracy of solutions and detect any potential flaws or improbable results *Draw appropriate inferences and conclusions* Explain and justify mathematical results in different mathematical forms
(c) Modeling	The process of capturing relationships present in the environment or in mathematical forms, and expressing the model in one or more mathematical representations	*Translate information into mathematical forms* Convert informal contextual information into equations, graphs, diagrams, tables, or mathematical text *Map mathematical relationships* Use tools such as equations, inequalities, diagrams, two-way tables, graphs, flow-charts, and formulas to express quantitative relationships (e.g., linear relationships, triangle inequality) *Apply mathematical models* Apply mathematical models and relationships to real-world contexts
	The evaluation and revision of a model for accuracy and applicability	*Determine reasonableness of a mathematical model* Use estimation methods to check a solution; interpret the results and reflect on whether a solution makes sense *Revise mathematical models* Adjust mathematical models to make improvements if a model has not served its purpose
(d) Computation	The process of identifying and performing appropriate algebraic manipulations and arithmetic calculations to solve a problem	*Identify appropriate computational strategies* Correctly determine or select mathematical operations or computational methods for solving a problem *Accurately calculate* Perform arithmetic or algebraic operations accurately to solve a problem
(e) Communication	The presentation of higher-level concepts and ideas (e.g., mathematical arguments and models) as well as solutions to problems and more standard procedures; communication may take various mathematical forms and is customized to the appropriate target audience	*Present mathematical concepts, data, procedures, and solutions in a variety of forms* Communicate procedures and results in written, graphical, or tabular format using correct mathematical terminology and notation

Table 5 Mathematical Content Areas for the Proposed Quantitative Literacy Framework

Content area	Brief description	Focus of assessment
(a) Number and operations	Real numbers, order properties, and physical quantities	• Understand fundamental types of real numbers, including positive and negative numbers, integers, fractions and decimals, even and odd integers, prime numbers, rational and irrational numbers • Understand the order properties of real numbers and the number line • Understand physical quantities as real numbers with units, such as time money, weight, temperature, distance, area, and volume
	Arithmetic operations on real numbers	• Add, subtract, multiply, and divide real numbers, as well as exponentiate and take roots • Understand the properties of arithmetic operations (i.e., commutative, distributive) as well as the role the operations have in defining fractions, decimals, factors, multiples, and remainders • Understand relationships between arithmetic operations and the ordering of real numbers (e.g., the product of two negative numbers is a positive number)
	Estimation	• Use estimation to approximate answers • Use estimation to judge reasonableness of answers
	Proportional reasoning	• Compute and interpret percents and percent change • Compute and interpret rates, ratios, and proportions
(b) Algebra	Variables, algebraic expressions, and their use in representing quantities	• Use variables to represent varying quantities • Use arithmetic operations on variables to form algebraic expressions • Manipulate and simplify algebraic expressions
	Functions, their types and properties, and their use in solving problems	• Understand the concept of a function, including domain and range, use function notation, and evaluate functions • Know various types of elementary functions, including linear, quadratic, polynomial, and exponential • Understand properties of various types of functions • Represent and interpret functions graphically in a coordinate plane • Use functions to model varying quantities in order to solve problems
	Equations, inequalities, and their use in solving problems	• Understand equations and inequalities as conditions that must be satisfied by varying quantities • Solve problems using algebraic representations by setting up equations or inequalities involving functions or algebraic expressions • Graph equations and inequalities in a coordinate plane • Solve equations or inequalities algebraically, graphically, or by ad hoc methods, such as inspection or repeated substitution • Interpret solutions of equations or inequalities to solve problems

Table 5 Continued

Content area	Brief description	Focus of assessment
(c) Geometry and measurement	Geometric figures in one, two, and three dimensions	• Understand lines and angles in a plane, including parallel and perpendicular lines • Know two-dimensional and three-dimensional geometric figures, such as triangles, circles, polygons, rectangular solids, cylinders, and spheres • Understand transformations, congruence, and similarity of two-dimensional figures • Graph geometric figures in a coordinate plane
	Geometric figure measurements (e.g., area, distance, length, volume, angles) for solving a problem	• Calculate area and perimeter/circumference of a two-dimensional object • Calculate volume and surface area of a three-dimensional object • Measure angles of polygons • Use the Pythagorean theorem to calculate the side lengths of a triangle • Use measurement formulas (e.g., volume, area) to solve problems
	Units and systems of measurement	• Understand units of measurement (e.g., time, money, weight, temperature, distance, area, volume) and when to apply them • Make conversions within a system of measurement (e.g., inches to feet, meters to kilometers) • Convert from one system of measurement to another (e.g., US customary units to metric system, Fahrenheit to Celsius)
(d) Statistics and probability	Data interpretation and representation	• Read and interpret data in graphical or tabular form to solve problems • Determine appropriateness of a table or graph used to represent a set of data (e.g., line graphs vs. bar graphs) • Compare alternative displays of the same data set or displays across multiple data sets (e.g., bar graphs and pie graphs) for similarities and differences • Create a table to organize frequency data, proportional quantities, or the relationship between two variables • Represent the frequency distribution of data using a dotplot, histogram, boxplot, or stem-and-leaf plot • Plot proportional quantities using a pie or bar graph • Create line charts or scatterplots to represent the relationship between two variables
	Descriptive statistics	• Interpret and calculate measures of central tendency (e.g., mean, median, and mode) for a distribution of data • Interpret and calculate measures of dispersion or spread (e.g., standard deviation, range, interquartile range) for a distribution of data
	Basic probability	• Understand random sampling with and without replacement, and equal probability for all outcomes • Calculate the probability of a single event using fractions and proportions (e.g., the probability of selecting an ace in a deck of cards) • Calculate the probability of two (or more) independent events (e.g., probability of a coin coming up tails after two coin tosses) • Understand and calculate conditional probability (e.g., probability of selecting an ace on the second draw after selecting an ace on the first draw)

Table 6 Description of Real-World Contexts for the Proposed Quantitative Literacy Framework

Context	Examples
(a) Personal/everyday life	Handling money and budgets; shopping; time management; personal travel; playing games of chance; sports statistics; reading maps; measurement in cooking, home repairs, or personal hobbies; calculating a tip; completing an order form; understanding and evaluating personal health; balancing a checkbook
(b) Workplace	Managing schedules, budgets, and project resources; using spreadsheets; making and recording measurements; payroll/accounting; completing purchase orders; tracking expenditures; predicting costs; job-related decision making
(c) Society	Population changes; unemployment rates; voting systems; public transportation; government; public policies; demographics; advertising; national statistics; economics; quantitative information in the media; raising funds for an organization; interpretation of research studies; environmental trends or issues

important for assessing quantitative literacy in a higher education context: (a) number and operations, (b) algebra, (c) geometry and measurement, and (d) probability and statistics (see Table 5). As shown in Table 3, these content areas are aligned with many of the existing assessments measuring quantitative literacy skills. These four content areas are also aligned with the four strands of American Diploma Project benchmarks. These mathematics benchmarks were based on a review of day-to-day experiences of people in the workplace and college classrooms, such as courses in college algebra and calculus, introductory chemistry, or introductory microeconomics (Achieve, Inc., 2004).

In a higher education context, these different content areas should have varying weights throughout an assessment. In the K–12 Common Core State Standards for Mathematics, a clear shift is apparent in content emphasis through the grade levels, with basic number and operations being emphasized in the earlier K–5 grade levels while geometry, probability, and statistics are emphasized more throughout Grades 6–12. The understanding of basic probability and statistics has been identified as a requirement in today's informational and technological age (Stumbo & Lusi, 2005). Similarly, the understanding of numbers and numerical operations serves as the foundation for many entry-level positions in the workforce, with many jobs also requiring some working knowledge of algebra, geometry, data interpretation, and probability and statistics (Stumbo & Lusi, 2005). For instance, occupations such as machine operators and licensed nurses require knowledge in number sense and numerical operations, algebra, and geometry. Other jobs, such as actuaries or manufacturing technicians, require knowledge in data interpretation, probability, and statistics (Achieve, Inc., 2004).

Even though the proposed next-generation quantitative literacy assessment is intended for students in higher education, mathematical content learned in K–12 can provide the foundational knowledge for using real-world applications on a higher education assessment. Steen (2004) stated that many of the underpinnings of quantitative literacy are typically mathematical topics from middle school. The Common Core State Standards for Mathematics echoed the importance of middle school mathematics recognizing that "some of the highest priority content for college and career readiness comes from Grades 6–8" (NGA & CCSSO, 2010, p. 84). In a review of mathematical competencies focal to the middle grades, Graf (2009) summarized research that supports the notion that the connection between numbers and operations and algebra is difficult for middle school students to grasp and that this difficulty persists throughout high school and college. Therefore, K–12 mathematical content, especially concepts learned up until Grade 10, should serve as a foundation for developing a next-generation quantitative literacy assessment. The construct of quantitative literacy focuses on problem solving in applied and authentic contexts, requiring advanced reasoning skills rather than just relying on memorization skills. Using mathematical content up to Grade 10 encompasses the mathematical knowledge required of all students, regardless of college major or intended career path, not just those students in a mathematical-focused college major.

Real-World Contexts

Quantitative literacy involves the application of mathematical skills to solve real-world problems in varying contexts, situations, or settings (Dwyer et al., 2003; Rhodes, 2010). In the employer survey conducted by Hart Research Associates (2013), 78% of employers stated that more emphasis needs to be placed on a student's ability to apply knowledge to real-world settings. The American Diploma Project noted that contexts in postsecondary institutions and in the workplace are very different from the high school classroom (Achieve, Inc., 2004). The context of a mathematical problem impacts the strategy employed (e.g., operations and procedures) and level of knowledge required to comprehend and solve it. To apply the next-generation quantitative literacy assessment to the real world, the proposed framework uses three authentic contexts: (a) personal/everyday life, (b) workplace, and (c) society. The personal/everyday life context addresses topics such as handling money and budgets, shopping, calculating a tip, and balancing a checkbook. The workplace context addresses topics across a wide array of jobs, including tasks such as using spreadsheets, managing schedules and budgets, and predicting costs. Lastly, the society context addresses topics such as population and demographic changes, unemployment rates, and quantitative information found in the media. Table 6 lists these topics and additional topics for these three contexts. The contexts should have a broad scope to ensure that the assessment is fair for all students, regardless of college major.

The contexts align with contexts described in AAC&U's Quantitative Literacy VALUE Rubric, Lumina's DQP, P21 Math, Common Core State Standards for Mathematics, PIAAC Numeracy, and PISA Mathematics and have also been identified in the workforce literature. For example, in a survey by Casner-Lotto and Barrington (2006), 71.5% of the over 400 employer respondents identified "exercise in personal finance and responsibility, e.g., balancing a checkbook, budgeting skills, retirement planning" (p. 52) as one of the "most critical" (p. 51) emerging content areas for college graduates entering the workforce. Similarly, American Diploma Project identified that mathematical skills are needed across a variety of job categories to conduct tasks such as measuring and/or computing the concentration of a solution, measuring volume and weight, computing the dose of a medication, or computing cost estimates or credit requests (Achieve, Inc., 2004).

Assessment Structure: Item Formats and Task Types

To measure problem-solving skills and mathematics content within different contexts, a variety of item formats and task types should be considered. These item formats and task types should provide an authentic and interactive test-taking experience for examinees. This section provides recommendations for both item formats and task types when developing a next-generation quantitative literacy assessment. It is recommended that all item formats employ the use of automated scoring because of its efficiency in providing instantaneous feedback to examinees (Zhang, 2013).

Item Formats

Most existing assessments measuring quantitative literacy skills in higher education use two item formats: selected-response and open-ended items. Examples of selected-response items include multiple-choice items or drop-down menu items. Similarly, examples of open-ended items include numeric entry or short answer items. Among the existing assessments, many assess students using selected-response items, specifically, multiple-choice items, because they are easy to score and typically yield higher reliability estimates. Typically, with selected-response items, more items can be administered in a shorter period of time, also allowing for more content coverage (R. E. Bennett, Morley, & Quardt, 2000). Existing assessments have also used some open-ended items such as numeric entry. Open-ended formats can eliminate random guessing and reduce measurement error, eliminate corrective feedback, and remove an examinee's ability to work backward from answer choices (Bridgeman, 1993). Table 7 shows a list of recommended item formats to be used in a next-generation quantitative literacy assessment that allows for automated scoring.

Selected-response items include both single- or multiple-selection items and computer-facilitated item formats such as clicking/selecting objects, drop-down menus, and drag and drop. For instance, a clicking/selecting objects item, similar to those found on PIAAC Numeracy, could be used to assess problem-solving skills such as interpretation or modeling. Additionally, a drop-down menu could be used to communicate information in graphical form. Drag and drop items could be used to correctly order a series of numbers. Open-ended items include numeric and expression entry, create/edit a graph/table, and short constructed response. Unlike numeric entry, expression entry (i.e., mathematical expression or a generating example item) allows an examinee to create a mathematical model. To avoid input errors, an expression item should have specific buttons that allow the examinee to develop the expression (R. E. Bennett et al., 2000). The create/edit a graph/table is similar to the graphical modeling item discussed in R. E. Bennett et al., where an examinee plots points and uses a tool to connect all of plotted points. These recommended selected-response and open-ended items can be technologically innovative and promote higher-level thinking.

Table 7 Next-Generation Quantitative Literacy Assessment Item Formats

Item format category[a]	Format response type	Description
Selected response	Single- and multiple-selection multiple choice	A question/stem with multiple answer choices of which one or more could be a correct response
	Drop-down menu	A variation of a traditional multiple-choice item, where one answer choice is selected via a drop-down menu; can use when customizing figures and tables
	Clicking/selecting objects	Select/click/highlight one or more parts of a table, figure, or text to answer the question
	Drag and drop	An examinee selects objects and places them in a specific location or order
Open ended	Numeric entry	A numerical answer must be entered rather than selected from a list of answer choices
	Expression entry	A mathematical expression must be entered using buttons that represent numbers and operators, variables, constants, and sub- and super-scripts
	Create/edit a graph/table	An examinee uses information in the question/stem to develop/edit a graphical display or table
	Short constructed response	Examinees must respond in their own words to a prompt based on text, graph, or other stimuli

[a] Items may be discrete or grouped together with a common figure or table.

It is important to understand that item format can impact both item difficulty and the reliability of the test. For example, Bridgeman (1993) found that when using the same underlying test questions in different formats (i.e., multiple-choice vs. numeric or simple formula entry), some items that were relatively easy in a multiple-choice format were quite difficult for examinees in a numeric entry format. The author also noted that the reliability of an assessment with either all multiple-choice or all numeric entry items can vary when the assessment has a fixed number of items or fixed testing time. For instance, with a fixed number of test items, an assessment with all numeric entry items may be more reliable because random guessing is removed (Bridgeman, 1993). However, it is important to note that within a fixed testing time, an assessment with all multiple-choice items is more likely to yield higher reliability estimates when compared to an assessment with all numeric entry items because the numeric entry item assessment would have fewer test items.

Task Types

Task types are used to address the foci of the assessment and can be used to enhance the previously described item formats. Table 8 provides a list of suggested task types for a next-generation quantitative literacy assessment. Similar tasks can be found on assessments such as CBAL Mathematics, GMAT Integrative Reasoning and GMAT Quantitative, CLA+ SQR, EPP Mathematics, PISA Mathematics, and GRE Quantitative Reasoning. Task types, in combination with various item formats, allow for more authenticity and examinee engagement. For instance, a task with a "clicking objects" response format (see Table 7) could assess whether an examinee can recognize an inconsistency among formulas in a spreadsheet (e.g., if a cell containing a formula for the sum of column totals and a cell containing a formula for the sum row totals from the same table do not match). The student could use the computer mouse to click in the cell or cells within the spreadsheet where the errors occur. This combination of item format and task type can recreate a real-world situation for examinees.

A Sample Item

Some of the considerations for a next-generation assessment framework can be clarified by a sample item that fits into each of three dimensions: problem-solving skill, content area, and context. Figure 1 shows a sample item written in the context of personal/everyday life (determining the data plan for a cell phone) and requires the examinee to use modeling, one of the problem-solving skills identified in Table 4. Part A of the problem requires the examinee to translate information into mathematical forms, and Part B requires the examinee to apply that mathematical model. In terms of content, this problem uses algebraic expressions and functions (see Table 5). The item response formats (Table 7) for this problem

Table 8 Next-Generation Quantitative Literacy Assessment Task Types

Task type	Description
Data accuracy	Statements are provided in reference to a question/stem. Examinee must determine whether the statements are accurate with respect to the stem
Data sufficiency	Statements are provided in reference to a question/stem. Examinee must determine whether the statements provide sufficient information to solve a problem or answer a question
Draw conclusions	Examinee reads data and draws conclusions or computes quantities
Evidence	Examinee provides or identifies types of information that must be sought in order to evaluate an argument, statement, or claim
Quantitative comparison	Examinee compares two presented quantities (less than, equal to, or greater than) or determines that there is not enough information to make the comparison
Recognize inconsistency	Examinee recognizes inconsistencies or errors in mathematical information (e.g., statements, equations, figures, or tables)
Representational equivalence	Examinee makes comparisons between two graphs, two tables, or a graph and a table and determines whether they are equivalent

Part A. When looking for a cheaper cell phone data plan, Andrea is considering a particular pay-per-use plan. The base cost of the plan is $25.00 per month, with a cost of $5.00 per 50 megabytes (MB) of data used.

Write a mathematical expression that Andrea can use to determine her monthly cost, C, in dollars, based on the amount of data used, D, in MB.

C= [ ]

Part B. Using the equation from Part A and the coordinate grid below, graph Andrea's monthly costs for 100, 200, 300, 400, and 500 MB of data used.

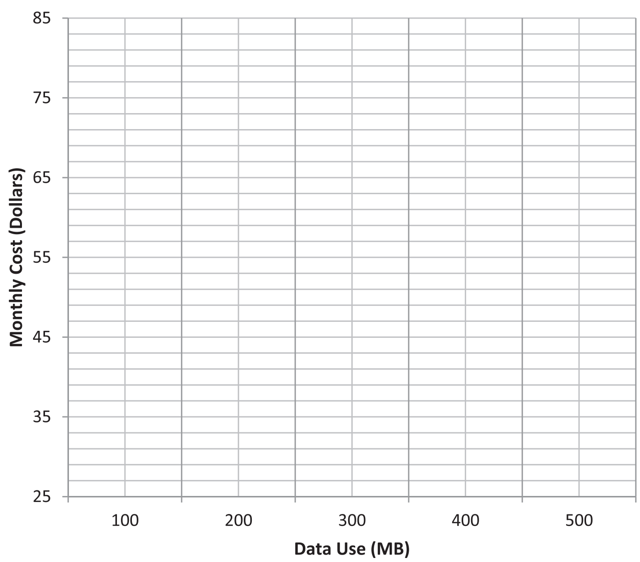

Figure 1 Sample expression entry item (Part A) and create a graph item (Part B).

include expression entry (Part A) and create a graph (Part B), and the item task type (Table 8) is draw conclusions. Other ways to represent this item without changing the construct, especially for Part B, include the use of drag and drop or clicking item formats. For instance, the examinee could drag a function line that is already created and place it on the graph or click on the function line. Both Parts A and B could also use multiple-choice items, but the use of alternative item formats makes the assessment more authentic. That said, because the item can use a variety of interaction methods, it is potentially more accessible for all students.

Potential Sources of Construct-Irrelevant Variance

When developing a next-generation quantitative literacy assessment in higher education, it is important to consider the accessibility of the assessment to all students, including students with limited English proficiency or students with disabilities. Developing an accessible assessment means careful consideration of multiple modes of delivery, as well as accessible methods for accessing test questions and entering responses. Technology-enhanced item formats may open doors for some students, while creating barriers for others, especially for students not used to completing quantitative problems on a computer. The choice of item formats (i.e., technology-enhanced or traditional) should depend on a number of factors such as construct representation, target population, scoring accuracy, scoring efficiency, and testing time. New item formats need to have clear directions on how to approach those items. For example, a clicking/selecting objects item could have the following instructions: Click two bars on the bar graph to answer the question below. The instructions would also need to indicate how the bar might change to show that it has been selected, and would need to indicate how to unselect the bar. A clear instruction is needed to ensure that examinees know how to appropriately approach the problem.

Another potential source of construct-irrelevant variance is the cognitive reading load on test questions. When assessing quantitative literacy, there is a certain level of assumed reading ability required to understand and answer test questions (Dwyer et al., 2003), especially as the items focus more on reasoning and interpretation skills. Research in K–12 educational assessment has investigated the impact of language factors on mathematics assessments. Abedi and Lord (2001) found that Grade 8 English language learners (ELLs) scored significantly lower than proficient English speakers on mathematical word problems and that ELLs performed better when taking the assessment with linguistically simplified test items. The authors noted that the complexity of text in mathematics problems was a larger issue for inexperienced and nonexpert problem solvers. Shaftel, Belton-Kocher, Glasnapp, and Poggio (2006) also evaluated the impact of linguistic features on mathematics items, finding a significant impact on student performance in Grades 4, 7, and 10. Interestingly, however, the effect size decreased as grade level increased, and Grade 10 item difficulty was most influenced by difficult mathematics vocabulary rather than other linguistic characteristics (Shaftel et al., 2006). It is likely that at the college level, mathematics vocabulary will also play a significant role in the difficulty of mathematical word problems. Therefore, the level of English proficiency required to answer a question should be considered in the development of test items.

Potential Advantages of the Proposed Framework and Assessment Considerations

The proposed framework and assessment considerations offer some distinct advantages over existing frameworks and assessments. A common theme throughout the literature on quantitative literacy and related constructs (e.g., quantitative reasoning) is that a quantitatively literate person needs to have the ability to interpret and manipulate mathematical information in many different contexts such as personal, work, and civic lives (Gillman, 2006). Many of the existing frameworks and assessments stress the importance of application to real-world problems but do not provide specific contexts (e.g., AAC&U's VALUE Rubric, Lumina's DQP, USDOL).The proposed framework has the advantage of defining three contexts for writing test items. The proposed framework also clearly defines both problem-solving skills and content areas, as both are important in developing a next-generation quantitative literacy assessment. Many of the existing frameworks and assessments (e.g., AAC&U's Quantitative Literacy VALUE Rubric) clearly define problem-solving skills but do not define mathematical content. Other frameworks and assessments (e.g., CAAP Mathematics) clearly define mathematical content and not the problem-solving skills.

Another advantage of the proposed framework is the alignment with existing frameworks, definitions, and assessments. The proposed framework carefully synthesizes problem-solving skills, content, and contexts identified by national and international organizations, workforce initiatives, higher education institutions and researchers, and K–12 theorists and practitioners. This synthesized approach is unique to the proposed framework. Additional advantages of the proposed framework and assessment considerations are the use of computer-facilitated item formats and the consideration of accessibility when developing the assessment.

Conclusion

Quantitative literacy has been identified as an important learning outcome to both higher education institutions and employers, making it critical to evaluate whether college students receive sufficient training on quantitative skills throughout their postsecondary education. A review of frameworks by various stakeholders and a review of existing assessments showed that many common themes exist when defining quantitative literacy or other related terms such as quantitative reasoning and numeracy. These common themes include overlap among mathematical skills, content areas, and contexts. This research also stressed that quantitative literacy involves more application and problem solving and less strict memorization of mathematical content. The existing research and assessments informed both the proposed framework and proposed assessment that involve applications and problem-solving skills across mathematical content areas. We hope this proposed framework provides clear guidance to stakeholders and assessment developers about the important aspects of the quantitative literacy construct. Using this framework to guide the development of a next-generation quantitative literacy assessment requires coordination between content experts, assessment developers, measurement experts, and potential score users at higher education institutions. This coordinated assessment development process requires both careful test development and psychometric evaluation to create a quality learning outcomes assessment for measuring quantitative literacy at the college level.

Acknowledgments

The authors would like to thank Daryl Ezzo for sharing his knowledge of mathematics assessment development and for assisting with the proposed framework and assessment considerations. The authors would also like to thank James Wyman Brantley for providing feedback and edits on this report during its development.

References

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, *14*(3), 219–234.

Abel, J. R., & Deitz, R. (2013). *Do big cities help college graduates find better jobs?* Retrieved from http://libertystreeteconomics.newyorkfed.org/2013/05/do-big-cities-help-college-graduates-find-better-jobs.html

Achieve, Inc. (2004). *Ready or not: Creating a high school diploma that counts*. Washington, DC: Author.

Adelman, C., Ewell, P., Gaston, P., & Schneider, C. G. (2011). *The degree qualifications profile*. Indianapolis, IN: Lumina Foundation.

Adelman, C., Ewell, P., Gaston, P., & Schneider, C. G. (2014). *The degree qualifications profile 2.0: Defining U.S. degrees through demonstration and documentation of college learning*. Indianapolis, IN: Lumina Foundation.

Admissions Testing Service. (2013). *About STEP*. Retrieved from http://www.admissionstestingservice.org/our-services/subject-specific/step/about-step/

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *The standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Association of American Colleges and Universities. (2011). *The LEAP vision for learning: Outcomes, practices, impact, and employers' view*. Washington, DC: Author.

Baldi, S. (2009). *Technical report and data file user's manual for the 2003 national assessment of adult literacy* (NCES 2009-476). Washington, DC: US Department of Education, National Center for Education Statistics.

Bennett, J., & Briggs, W. L. (2008). *Using and understanding mathematics: A quantitative reasoning approach* (4th ed.). Reading, MA: Addison.

Bennett, R. E., Morley, M., & Quardt, D. (2000). Three response types for broadening the conception of mathematical problem solving in computerized tests. *Applied Psychological Measurement*, *24*(4), 294–309.

Bridgeman, B. (1993). *A comparison of open-ended and multiple-choice question formats for the quantitative section of the graduate record examinations general test* (GRE Board Professional Report No. 88-13P). Princeton, NJ: Educational Testing Service.

Bridgeman, B., Harvey, A., & Braswell, J. (1995). Effects of calculator use on scores on a test of mathematical reasoning. *Journal of Educational Measurement*, *32*(4), 323–340.

CAAP Program Management. (2012). *ACT CAAP technical handbook 2011–2012*. Iowa City, IA: Author.

Capital Workforce Partners. (2014). *The career competencies system*. Retrieved from http://capitalworkforce.org/youth-resources/career-competencies-system/

Carleton College. (2013). *Carleton's quantitative inquiry, reasoning, and knowledge (QuIRK) initiative*. Retrieved from http://serc.carleton.edu/quirk/index.html

Casner-Lotto, J., & Barrington, L. (2006). *Are they really ready to work?: Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st century U.S. workforce*. New York, NY: Conference Board, Partnership for 21st Century Skills, Corporate Voices for Working Families, Society for Human Resource Management.

Cohen, D. (Ed.) (1995). *Crossroads in mathematics: Standards for introductory college mathematics before calculus*. Memphis, TN: American Mathematical Association of Two-Year Colleges.

College Board. (2012). *Test information guide: College-Level Examination Program® 2012–13 college mathematics*. New York, NY: Author.

Dwyer, C. A., Gallagher, A., Levin, J., & Morley, M. E. (2003). *What is quantitative reasoning? Defining the construct for assessment purposes* (Research Report No. RR-03-30). Princeton, NJ: Educational Testing Service.

Educational Testing Service. (2010). *ETS proficiency profile user's guide*. Princeton, NJ: Author.

Educational Testing Service. (2013a). *GRE guide to the use of scores 2013–2014*. Princeton, NJ: Author.

Educational Testing Service. (2013b). *GRE test content and structure*. Princeton, NJ: Author.

Educational Testing Service. (2013c). *Introduction to the quantitative reasoning measure*. Retrieved from http://www.ets.org/gre/revised_general/prepare/quantitative_reasoning

Educational Testing Service. (2014a). *The CBAL™ initiative: Innovation in K–12 assessment*. Retrieved from http://www.ets.org/research/topics/cbal/initiative/

Educational Testing Service. (2014b). *GRE: Using a calculator*. Retrieved from https://www.ets.org/gre/revised_general/prepare/quantitative_reasoning/calculator

Ellington, A. J. (2003). A meta-analysis of the effects of calculators on students' achievement and attitude levels in precollege mathematics classes. *Journal for Research in Mathematics Education*, *34*(5), 433–463.

Estry, D. W., & Ferrini-Mundy, J. (2005). *Quantitative literacy task force final report and recommendations*. Retrieved from http://acadgov.msu.edu/documents/QLFinalReport-1.pdf

Ewell, P. T. (2001). Numeracy, mathematics, and general education. In L. A. Steen (Ed.), *Mathematics and democracy: The case for quantitative literacy* (pp. 37–48). Princeton, NJ: Woodrow Wilson National Fellowship Foundation.

Gillman, R. (Ed.) (2006). *Current practices in quantitative literacy*. Washington, DC: Mathematical Association of America.

Goodman, M., Finnegan, R., Mohadjer, L., Krenzke, T., Hogan, J., Owen, E., & Provasnik, S. (2013). *Literacy, numeracy, and problem-solving technology-rich environments among U.S. adults: Results from the Program for the International Assessment of Adult Competencies 2012* (NCES 2014-008). Washington, DC: U.S. Department of Education, National Center for Educational Statistics.

Graduate Management Admission Council. (2013a). *Quantitative section*. Retrieved from http://www.mba.com/us/the-gmat-exam/gmat-exam-format-timing/quantitative

Graduate Management Admission Council. (2013b). *Validity, reliability, & fairness*. Retrieved from http://www.gmac.com/gmat/the-gmat-advantage/validity-reliability-fairness.aspx

Graf, E. A. (2009). *Defining mathematics competency in the service of cognitively based assessment for grades 3 through 8* (Research Report No. RR-09-42). Princeton, NJ: Educational Testing Service.

Hart Research Associates. (2013). *It takes more than a major: Employer priorities for college learning and student success*. Washington, DC: Author.

Hollins University. (2013). *Course goals and criteria document*. Retrieved from https://www.hollins.edu/registrar/documents/ESP_goalscriteria.pdf

Insight Assessment. (2013a). *Quant Q: Measure quantitative reasoning*. San Jose: California Academic Press.

Insight Assessment. (2013b). *Test of everyday reasoning-numeracy (TER-N)*. San Jose: California Academic .

Institute of Educational Statistics. (n.d.). *National Assessment of Adult Literacy (NAAL)*. Retrieved from http://nces.ed.gov/naal/si_questions.asp

Kirsch, I. S., Jungeblut, A., Jenkins, L., & Kolstad, A. (2002). *Adult literacy in America: A first look at the findings of the national adult literacy survey* (NCES 1993-275). Washington, DC: U.S. Department of Education.

Klein, S., Liu, O. L., Sconing, J., Bolus, R., Bridgeman, B., Kugelmass, H., … Steedle, J. (2009). *Test Validity Study (TVS) report*. Retrieved from https://cp-files.s3.amazonaws.com/26/TVSReport_Final.pdf

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, *127*(1), 162–181.

Lakin, J. M., Elliott, D. C., & Liu, O. L. (2012). Investigating ESL students' performance on outcomes assessments in higher education. *Educational and Psychological Measurement*, *72*(5), 734–753.

Langkamp, G., & Hull, J. (2007). *Quantitative reasoning and the environment*. Upper Saddle River, NJ: Pearson Education.

Liu, O. L., Bridgeman, B., & Adler, R. M. (2012). Measuring learning outcomes in higher education: Motivation matters. *Educational Researcher*, *41*(9), 352–362.

Liu, O. L., & Roohr, K. C. (2013). Investigating 10-year trends of learning outcomes at community colleges (Research Report No. RR-13-34). Princeton, NJ: Educational Testing Service.

Lutzer, D. J., Rodi, S. B., Kirkman, E. E., & Maxwell, J. W. (2007). *Statistical abstract of undergraduate programs in the mathematical sciences in the United States: Fall 2005 CBMS survey*. Providence, RI: American Mathematical Society.

Marr, D. (1995). *Validity of the academic profile*. Princeton, NJ: Educational Testing Service.

Mayes, R. L., Peterson, F., & Bonilla, R. (2013). Quantitative reasoning for learning progressions for environmental science: Developing a framework. *Numeracy*, *6*(1), Article 4. doi:10.5038/1936-4660.6.1.4

McKinsey & Company. (2013). *Voice of the graduate*. Philadelphia, PA: Author.

Merriam-Webster. (2014). Literacy. Retrieved from http://www.merriam-webster.com/dictionary/literacy

Mount St. Mary's College. (2013). *Definition of quantitative literacy*. Retrieved from http://www.msmc.la.edu/academics/quantitative-literacy-program/definition-of-quantitative-literacy.asp

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Washington, DC: Author.

National Numeracy Network. (2013). *What is numeracy/QL/QR?* Retrieved from http://serc.carleton.edu/nnn/resources/index.html

National Survey of Student Engagement. (2013a). *A fresh look at student engagement: Annual results 2013*. Bloomington: Indiana University Center for Postsecondary Research.

National Survey of Student Engagement. (2013b). *NSSE 2013 U.S. grand frequencies: Frequency distributions by Carnegie classification—First-year students*. Retrieved from http://nsse.iub.edu/2013_institutional_report/pdf/Frequencies/Freq%20-%20FY%20by%20Carn.pdf

National Survey of Student Engagement. (2013c). *NSSE 2013 U.S. grand frequencies: Frequency distributions by Carnegie classification—Seniors*. Retrieved from http://nsse.iub.edu/2013_institutional_report/pdf/Frequencies/Freq%20-%20SR%20by%20Carn.pdf

Organisation for Economic Co-Operation and Development. (2012a). *Literacy, numeracy and problem solving in technology-rich environments: Framework for the OECD survey of adult skills*. Retrieved from http://www.oecd.org/site/piaac/PIAAC%20Framework%202012--%20Revised%2028oct2013_ebook.pdf

Organisation for Economic Co-Operation and Development. (2012b). *The survey of adult skills: Reader's companion*. Paris, France: Author.

Organisation for Economic Co-Operation and Development. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving, and financial literacy*. Paris France: Author.

Organisation for Economic Co-Operation and Development and Statistics Canada. (2000). *Literacy in the information age. Final report of the international adult literacy survey*. Paris, France: Author.

Partnership for Assessment of Readiness for College and Careers. (2014). *The PARCC assessment*. Retrieved from http://www.parcconline.org/parcc-assessment

PIAAC Numeracy Expert Group. (2009). *PIAAC numeracy*: *A conceptual framework* (OECD Education Working Paper No. 35). Paris, France: OECD.

Rhodes, T. L. (Ed.) (2010). *Assessing outcomes and improving achievement: Tips and tools for using rubrics*. Washington, DC: Association of American Colleges and Universities.

Saltrick, S., Hadad, R., Pearson, M., Fadel, C., Regan, B., & Winn, J. (2011). *21st century skills map*. Washington, DC: Partnership for 21st Century Skills Retrieved from http://www.p21.org/storage/documents/P21_Math_Map.pdf

Samford University. (2009). *Quantitative literacy*. Retrieved from http://www.samford.edu/ctls/archives.aspx?id=2147484121

Shaftel, J., Belton-Kocher, E., Glasnapp, D., & Poggio, J. (2006). The impact of language characteristics in mathematics test items on the performance of English language learners and students with disabilities. *Educational Assessment*, *11*(2), 105–126.

Sinharay, S., Puhan, G., & Haberman, S. J. (2011). An NCME instructional module on subscores. *Educational Measurement: Issues and Practice*, *30*(3), 29–40.

Smarter Balanced Assessment Consortium. (n.d.). *Smarter balanced assessments*. Retrieved from http://www.smarterbalanced.org/smarter-balanced-assessments/

Sons, L. (Ed.) (1996). *Quantitative reasoning for college graduates: A complement to the standards*. Washington, DC: Mathematical Association of America.

Steele, B., & Kilic-Bahi, S. (2008). Quantitative literacy across the curriculum: A case study. *Numeracy*, *1*(2), 1–15.

Steen, L. A. (2001). The case for quantitative literacy. In L. A. Steen (Ed.), *Mathematics and democracy* (pp. 1–22). Princeton, NJ: Woodrow Wilson National Fellowship Foundation.

Steen, L. A. (2004). *Achieving quantitative literacy: An urgent challenge for higher education*. Washington, DC: Mathematical Association of America.

Stumbo, C., & Lusi, S. F. (2005). *Why isn't the mathematics we learned good enough for today's students? The imperative of mathematical literacy*. Washington, DC: Council of Chief State School Officers and Texas Instruments.

Sundre, D. L. (2008). *The Scientific Reasoning Test, Version 9 (SR-9) test manual*. Harrisonburg, VA: Center for Assessment and Research Studies.

Tamassia, C., Lennon, M. L., & Yamamoto, K. (2013). Scoring and reliability studies. In I. Kirsch and W. Thorn (Eds.),"*Technical report of the survey of adult skills (PIAAC)*. Paris, France: OECD.

Traub, R. E., & Rowley, G. L. (1991). An NCME instructional module on understanding reliability. *Educational Measurement: Issues and Practice*, *10*(1), 37–45.

U.S. Department of Labor. (2013). *Competency model clearinghouse: Mathematics*. Retrieved from http://www.careeronestop.org/competencymodel/blockModel.aspx?tier_id=2&block_id=9

University of Kentucky. (2012). *UK core quantitative reasoning rubric*. Retrieved from http://www.uky.edu/ukcore/sites/www.uky.edu.ukcore/files/QR%20Rubric.pdf

University of Warwick. (2013). *STEP papers and the advanced extension award in mathematics*. Retrieved from http://www2.warwick.ac.uk/fac/sci/maths/admissions/ug/aeastep/

Vacher, H. L. (2014). Looking at the multiple meanings of numeracy, quantitative literacy, and quantitative reasoning. *Numeracy*, *7*(2), 1–14.

Voluntary System of Accountability. (2013). *Voluntary system of accountability program*. Retrieved from http://www.voluntarysystem.org/slo_evaluation

Ward, R. M., Schneider, M. C., & Kiper, J. D. (2011). Development of an assessment of quantitative literacy for Miami University. *Numeracy*, *4*(2), Article 4. doi:10.5038/1936-4660.4.2.4

Wellesley College. (n.d). *Study packet for the quantitative reasoning assessment*. Wellesley, MA: Author.

Zahner, D. (2013). *Reliability and validity—CLA+*. New York, NY: Council for Aid to Education.

Zhang, M. (2013). Contrasting automated and human scoring of essays. *R&D Connections*, *21*, 1–11.

Suggested citation:

Roohr, K. C., Graf, E. A., & Liu, O. L. (2014). *Assessing quantitative literacy in higher education: An overview of existing research and assessments with recommendations for next-generation assessment* (ETS Research Report No. RR-14-22). Princeton, NJ: Educational Testing Service. doi:10.1002/ets2.12024.

Action Editor: Donald Powers

Reviewers: James Brantley and James Fife

Find other ETS-published reports by searching the ETS ReSEARCHER database at http://search.ets.org/researcher/

Listening. Learning. Leading.®

Research Report

ETS RR–14-37

Assessing Written Communication in Higher Education: Review and Recommendations for Next-Generation Assessment

Jesse R. Sparks

Yi Song

Wyman Brantley

Ou Lydia Liu

December 2014

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

Assessing Written Communication in Higher Education: Review and Recommendations for Next-Generation Assessment

Jesse R. Sparks, Yi Song, Wyman Brantley, & Ou Lydia Liu

Educational Testing Service, Princeton, NJ

Written communication is considered one of the most critical competencies for academic and career success, as evident in surveys of stakeholders from higher education and the workforce. Emphasis on writing skills suggests the need for next-generation assessments of writing proficiency to inform curricular and instructional improvement. This article presents a comprehensive review of definitions of writing proficiency from key higher education and workforce frameworks; the strengths and weaknesses of existing assessments; and challenges related to designing, implementing, and interpreting such assessments. Consistent with extant frameworks, we propose an operational definition including 4 strands of skills: (a) social and rhetorical knowledge, (b) domain knowledge and conceptual strategies, (c) language use and conventions, and (d) the writing process. Measuring these aspects of writing requires multiple assessment formats (including selected-response [SR] and constructed-response [CR] tasks) to balance construct coverage and test reliability. Next-generation assessments should balance authenticity (e.g., realistic writing tasks) and psychometric quality (e.g., desirable measurement properties), while providing institutions and faculty with actionable data. The review and operational definition presented here should serve as an important resource for institutions that seek to either adopt or design an assessment of students' writing proficiency.

Keywords Writing; writing assessment; higher education; written composition; communication; student learning outcomes

doi:10.1002/ets2.12035

Effective communication is fundamental to success in many aspects of life. Scholars such as Dewey (1938/1997) have acknowledged the importance of language as a primary medium through which learning takes place in educational and everyday experiences, asserting that "all human experience is ultimately social . . . it involves contact and communication" (p. 38). In order to interact successfully with others in academic, workplace, and community settings, individuals must be able to communicate — to convey or exchange information, knowledge, and ideas — clearly and effectively. Young learners begin to develop their communication skills in oral contexts, but as they progress through K–12, writing skills become increasingly important, shifting in emphasis from the development of foundational print literacy and transcription skills, to composing narratives about one's experiences, to expositions or analyses of phenomena, and ultimately to more sophisticated tasks, such as writing arguments or research reports.

The ability to write effectively using standard written English is particularly important in higher education, where proficiency with written communication is considered a critical student learning outcome (SLO). A survey conducted by the Association of American Colleges and Universities (AAC&U, 2011) found that 99% of the chief academic officers from 433 higher education institutions rated writing as one of the most important intellectual skills for their students. More recently, the Educational Testing Service (ETS, 2013a) conducted interviews with provosts or vice presidents of academic affairs from more than 200 institutions regarding the most commonly measured general education skills, finding that written communication was the most frequently mentioned competency considered by respondents as critical for both academic and career success. The focus on written communication is also apparent internationally. Notably, written communication is included as a *generic skill* expected of all students in the Assessment of Higher Education Learning Outcomes (AHELO) project, an effort to evaluate general learning outcomes of college students across nations, which is sponsored by the Organisation for Economic Co-operation and Development (OECD, 2012).

Reports from the workforce echo that of higher education. Written communication was among the most desired skills mentioned by a sample of 431 employers from various industries surveyed by the Conference Board (Casner-Lotto &

Corresponding author: J. R. Sparks, E-mail: jsparks@ets.org

Barrington, 2006); over 93% of respondents reported that written communication was "very important" (p. 41) for the workplace, yet 28% of respondents rated the writing skills of 4-year college graduates entering the workforce as "deficient" (p. 41). Further, 89% of 302 employers surveyed by the AAC&U (2011) said that colleges and universities should place more emphasis on communication skills, the highest endorsement of any skill included in the survey. Written communication skills are crucial for the workplace, yet many employers perceive college graduates as being underprepared for the writing tasks required at work. By contrast, college graduates report that learning to write effectively is one of the most important skills learned in their undergraduate career (e.g., Krahn & Silzer, 1995).

These discrepancies in perceptions across stakeholders underscore the need for valid, reliable assessments of written communication as a learning outcome that can provide institutions, employers, and individual students with meaningful information about students' skills. Recent calls for assessment reform also reflect the importance of designing assessments that have instructional relevance, provide feedback to teachers and students, and can be used to improve curriculum (Gordon Commission, 2013). A next-generation assessment of written communication competency at the higher education level could be used to inform revisions to curriculum and instruction in the service of developing students' writing skills, to make effective hiring decisions, and/or to provide students with feedback about their preparation for future academic or workforce pursuits. Such an assessment should be based on a precise definition of the written communication construct, which is supported by and consistent with current empirical research on writing in higher education.

Although there is general agreement that effective communication skills (both oral and written) are important, there is some ambiguity about how this competency should be defined. Markle, Brenneman, Jackson, Burrus, and Robbins (2013) reviewed definitions of effective communication from seven key frameworks of general education competencies in higher education. Based on this synthesis, the authors defined this competency as the ability to "effectively communicate multiple types of messages; communicate across multiple forms; and effectively deliver messages to varying audiences" (p. 16). This definition highlights three aspects of communication: the message's type (i.e., genre), form (i.e., medium), and recipient. Understanding these aspects of communication is important in both oral and written modalities. However, these aspects alone may not fully delineate the range of skills that specifically constitute proficiency with written communication.

The overwhelming emphasis on written communication among stakeholders suggests a need to examine existing frameworks, focusing on outcomes specific to writing. As with communication in general, definitions of writing skill vary across frameworks. Similarly, existing writing assessments vary in the extent to which they are designed to measure particular skills. For example, the writing component of the *TOEFL*® test is designed to measure writing in English as a second or foreign language, with particular attention to the integration of reading, writing, and listening skills, and use of particular rhetorical forms, such as summary or description (Cumming, Kantor, Powers, Santos, & Taylor, 2000); the particular configuration of writing skills assessed in the TOEFL test is consistent, but not completely overlapping, with the writing skills that might be targeted in an assessment of writing as an SLO. Thus, despite the apparent consensus on the importance of written communication as a critical competency, there are multiple definitions of what constitutes effective writing at the college level. For the purposes of designing and building next-generation assessments of written communication for higher education, a clear construct definition is needed. A primary goal of this article is to provide such a definition. A secondary goal is to identify and discuss the issues and challenges that must be considered when designing an assessment of written communication as a learning outcome.

In the first section of this report, we review existing definitions and frameworks of written communication in higher education. We also discuss models and theories from the field of writing research that can inform our definition of this construct. In the second part of this report, we review current assessments of written communication with respect to construct coverage, item formats, and reliability and validity evidence. We then discuss challenges in designing written communication assessments, including use of automated scoring techniques, and consider their relevance to curriculum and instruction. In the final section of this report, based on a synthesis of the frameworks reviewed, we propose an operational definition for a next-generation assessment of written communication; this definition is specifically intended to support the development of assessments of this particular SLO in higher education contexts. We also provide examples of item types designed to assess key writing skills.

In particular, the review of existing written communication assessments presented in the second part of this report is intended to aid higher education institutions in choosing among alternative assessments. Evidence suggests that the institutional emphasis on assessment of SLOs continues to increase, with learning institutions turning to a wide variety of assessments and approaches to meet demands for accountability (Kuh, Jankowski, Ikenberry, & Kinzie, 2014). Navigating the landscape of available instruments and assessment methods poses a challenge for higher education institutions, so the current synthesis is intended to serve as a helpful guide. We also hope that the approach to designing a next-generation written communication assessment described here will serve as a resource for institutions in developing their own writing assessments. Building collaborative partnerships between higher education institutions and testing organizations in the assessment design process can help ensure that SLO assessments meet standards of technical quality while maximizing instructional relevance.

Review of Existing Frameworks and Research

Definitions of Written Communication in Key Frameworks

Written communication involves the ability to effectively convey multiple types of messages, in multiple forms, to varying audiences, through a written medium (see Markle et al., 2013). However, writing is a multifaceted construct and is defined differently among various sources. Notions of what constitutes quality writing vary even among experts (Behizadeh, 2014). As emphasized by Murphy and Yancey (2008), arguments for the use of particular techniques for assessing students' writing are often based on competing theories about the nature of the writing construct—as a set of discrete skills, as a cognitive (or instructional) process that takes place over time, and more recently, as a meaning-making and highly social activity that varies across contexts and purposes for writing (p. 449). Since these various perspectives affect assessment design decisions, it is critical to determine consistencies among stakeholders' views of the underlying construct.

Table 1 presents definitions of written communication[1] drawn from nine key frameworks, including the Council of Writing Program Administrators (CWPA), National Council of Teachers of English (NCTE), and National Writing Project (NWP)'s *Framework for Success in Postsecondary Writing* (2011); the National Institutes of Health (NIH)'s definition of communication competency (OHR-NIH, 2014); the Quality Assurance Agency for Higher Education's Framework for Higher Education Qualifications (QAA-FHEQ); AAC&U's Liberal Education and America's Promise (LEAP) VALUE rubrics (Rhodes, 2010); Lumina's Degree Qualifications Profile (Adelman, Ewell, Gaston, & Schneider, 2011); the U.S. Department of Labor's Employment and Training Administration (US-DOL, 2014) Industry Competency Model Clearinghouse; European Commission's European Higher Education Area (EHEA) Competencies (i.e., the Bologna Framework; European Higher Education Area, 2005; González & Wagenaar, 2003); the Council for the Advancement of Standards in Higher Education (CAS)'s Framework for Learning and Development Outcomes (CAS, 2009); and the Assessment and Teaching of 21st-Century Skills KSAVE frameworks (Binkley et al., 2010).

Table 2 shows the correspondence between each framework reviewed and different dimensions of the writing construct mentioned within the various definitions. Definitions and learning outcomes across the various frameworks show some degree of consistency, but, interestingly, the configuration of features thought to underlie skilled writing at the college level varies such that no two frameworks define the construct in exactly the same way. Importantly for our present purposes, no single assessment of college writing has been designed on the basis of any of these frameworks or on a synthesis of them; these frameworks suggest learning and assessment targets but have not directly informed the development of specific large-scale assessments. We explore the relationships between existing assessments and aspects of the writing construct in the second part of this report.

Key Dimensions of Written Communication

Members of the CWPA, NCTE, and NWP (2011) collaborated to develop a framework describing the rhetorical and 21st century skills required for success in reading, critical thinking, and writing at the college level. This *Framework for Success in Postsecondary Writing* is intended to describe college readiness, or the expected knowledge, skills, and abilities of a student who has completed a first-year composition course in college and who demonstrates readiness to take on more advanced intellectual work in further academic or career settings. Specifically, the *Framework* organizes literacy skills into five dimensions: rhetorical knowledge (including understanding of various purposes, audiences, contexts, genres, and forms of writing), critical thinking (including analysis of reading materials, evaluating information sources' usefulness and reliability, and using research to support writing), writing processes (including planning, drafting, editing, revising, and responding to feedback), knowledge of conventions (including both surface-level grammatical conventions and more global concerns related to discourse content, organization, tone, and style), and composing in multiple environments (e.g.,

Table 1 Definitions of Written Communication From Current Frameworks of Learning Outcomes

Framework	Author/Sponsor	Written communication (or equivalent) definition
Framework for Success in Postsecondary Writing	Council of Writing Program Administrators, National Council of Teachers of English, and National Writing Project	***Rhetorical knowledge:*** The ability to analyze and act on understandings of audiences, purposes, contexts, genres, and forms in creating texts. This includes learning to compose a variety of texts for different disciplines and purposes. ***Critical thinking:*** The ability to analyze a situation or text and make thoughtful decisions based on that analysis. This includes conducting research from primary and secondary sources; evaluating those sources' credibility, bias, evidence, and reasoning; identifying and challenging writer's assumptions; and writing texts that are informed by one's research. ***Writing processes:*** The multiple strategies writers use to approach and undertake writing and research. This includes generating ideas, conducting research, drafting, revising, editing, and responding to feedback. ***Knowledge of conventions:*** The formal and informal guidelines that define what is considered correct (or appropriate) and incorrect (or inappropriate) in a piece of writing. This includes the surface features of a text, such as mechanics, spelling, and attribution of sources, as well as more global concerns, such as content, tone, style, organization, and evidence. Correct use of conventions is defined within specific genres and contexts. ***Composing in multiple environments:*** The ability to create writing using everything from traditional pen and paper to electronic technologies. This includes composing multiple forms, such as a traditional essay, a webpage or video, or a print brochure, and using electronic sources in those documents. http://wpacouncil.org/framework (CWPA et al., 2011)
Employment Competencies Dictionary - Communications	National Institutes of Health (NIH) Office of Human Resources	***Communications:*** Clearly and effectively conveys information; asks appropriate questions; organizes, expresses, and communicates ideas clearly in writing; asks clarifying questions and summarizes or paraphrases what others have said to verify understanding; uses analogies, visuals, and other techniques to tailor communications to specific audiences; identifies and uses effective communication channels and methods (e.g., presentations, electronic dissemination, social media); utilizes skill in presenting information, analysis, ideas, and positions in a clear, succinct, accurate, convincing manner, as is appropriate for the audience.http://hr.od.nih.gov/workingatnih/competencies/core/communication.htm (OHR-NIH, 2014)
Assessment & Teaching of 21st Century Skills (ATC21S)	Collaboration among Cisco, Intel, and Microsoft	*Ways of Working – Communication* ***Knowledge**:* Sound knowledge of basic vocabulary, functional grammar and style, functions of language; awareness of various types of verbal interaction (conversations, interviews, debates, etc.); understanding the main features of written language (formal, informal, scientific, journalistic, colloquial, etc.). ***Skills**:* Ability to communicate in written form and understand or make others understand various messages in a variety of situations and for different purposes; ability to write different types of texts for various purposes; to monitor the writing process (from drafting to proofreading); ability to formulate one's arguments in writing in a convincing manner and take full account of other viewpoints; skills needed to use aids (such as notes, schemes, maps) to produce or present complex texts in written form. ***Attitudes/Values/Ethics**:* Willingness to strive for esthetic quality in expression beyond the technical correctness of a word/phrase. (Binkley et al., 2010)
Degree Qualifications Profile (DQP)	The Lumina Foundation	*Intellectual Skills – Communication Fluency* ***At the associate level:*** Presents substantially error-free prose in both argumentative and narrative forms to general and specialized audiences. ***At the bachelor's level:*** Constructs sustained, coherent arguments and/or narratives and/or explications of technical issues and processes, in two media, to general and specific audiences. In a language other than English, in writing, conducts an inquiry with a non-English-language source concerning information, conditions, technologies, and/or practices in his or her major field. With collaborators, advances an argument or designs an approach to resolving a social, personal, or ethical dilemma. (Adelman et al., 2011, p. 14)

Table 1 Continued

Framework	Author/Sponsor	Written communication (or equivalent) definition
The Employment and Training Administration (ETA) Industry Competency Model Clearinghouse	U.S. Department of Labor (US-DOL)	***Organization and development*:** Creates documents such as letters, directions, manuals, reports, graphs, and flow charts; communicates thoughts, ideas, information, messages and other written information, which may contain technical material, in a logical, organized and coherent manner; ideas are well developed with supporting information and examples. ***Mechanics*:** Uses standard syntax and sentence structure; uses correct spelling, punctuation, and capitalization; uses appropriate grammar (e.g., correct tense, subject-verb agreement, no missing words). ***Tone:*** Writes in a manner appropriate for business; uses language appropriate for the target audience; uses appropriate tone and word choice (e.g., writing is professional and courteous). (US-DOL ETA, 2014)
European Higher Education Area Competencies (Bologna Framework)	European Commission: European Higher Education Area	***Generic competencies:*** Ability to communicate through the written word in one's native language; ability to communicate information, ideas, problems, and solutions to both specialist and nonspecialist audiences; ability to communicate in a second language. ***Specific competencies:*** Skills in presenting domain-relevant material and arguments in writing, to an informed audience; ability to write in one's own language or other languages using correctly various types of writing within one's discipline; ability to present arguments with clarity and accuracy in forms that are suitable for the audiences being addressed; receiving and responding to a variety of information sources (e.g., textual, numerical, verbal, graphical); communicating appropriately to a variety of audiences in written, verbal, and graphical forms (González & Wagenaar, 2003).
Framework for Higher Education Qualifications (QAA-FHEQ)	Quality Assurance Agency for Higher Education	Communicate the results of their study/work accurately and reliably, and with structured and coherent arguments; effectively communicate information, ideas, arguments, analysis, problems, and solutions in a variety of forms to both specialist and nonspecialist audiences, and deploy key techniques of the discipline effectively (QAA, 2008).
Framework for Learning and Development Outcomes (CAS Standards)	The Council for the Advancement of Standards in Education	***Effective communication:*** Conveys meaning in a way that others understand by writing coherently and effectively; writes after reflection; influences others through writing; effectively articulates abstract ideas; uses appropriate syntax and grammar; makes and evaluates presentations or performances (CAS, 2009); expresses thoughts and emotions through writing; writes essays or personal letters; writes in an organized fashion; moves from general to specific topics in writing; communicates in nontraditional forms (e.g., e-mail; CAS, 2006).
Liberal Education and America's Promise (LEAP)	Association of American Colleges and Universities	Written communication is the development and expression of ideas in writing. Written communication involves learning to work in many genres and styles. It can involve working with many different writing technologies and mixing texts, data, and images (Rhodes, 2010). Written communication involves five dimensions. ***Context of and purpose for writing,*** which includes considerations of audience, purpose, and the circumstances surrounding the writing task(s). ***Content development,*** the ways in which the text explores and represents its topic in relation to its audience and purpose. ***Genre and disciplinary conventions***, formal and informal rules inherent in the expectations for writing in particular forms and/or academic fields. ***Sources and evidence,*** use of high-quality, credible, relevant sources to develop one's ideas. ***Control of syntax and mechanics,*** appropriate use of language (Rhodes, 2010).

Table 2 Mapping of Written Communication Skills to Key Frameworks

Dimensions of writing construct	CWPA, NCTE, & WPA Framework	NIH-OHR	ATC21S	DQP	DOL-ETA	BOLOGNA	QAA	CAS	LEAP
Context and purpose	X	~	X	X	X	-	~	X	X
Audience awareness	X	X	~	X	X	X	X	~	X
Content development and organization	X	X	~	~	X	X	X	X	X
Genre conventions (text types/formats)	X	X	X	X	X	X	X	X	X
Disciplinary conventions (major/field)	X	-	~	X	-	X	X	-	X
Use of sources and textual evidence	X	~	-	X	~	X	-	-	X
Processes (planning, drafting, revision)	X	-	X	-	-	-	-	~	-
Modes and forms (multimedia, digital)	X	X	X	X	X	X	X	X	X
Word choice, tone, voice, and style of language	X	~	X	-	X	-	-	-	-
Language use, grammar, syntax, and mechanics	X	~	X	X	X	~	~	X	X

Note. X = mentioned as part of framework definition or rubric statements; ~ = indirectly mentioned as part of framework definition or rubric statements (i.e., statement that could be related to a dimension); - = not mentioned in the framework.

using traditional and digital production modes, and incorporating electronic sources in the written work; see Table 1). These five dimensions correspond nicely to the aspects of writing in higher education and workforce frameworks, and they appear to encompass all of the critical elements of written communication; accordingly, the following review is organized around these five dimensions. Importantly, this *Framework* includes critical engagement with and use of sources and emphasis on the writing process, two skills tha t are infrequently mentioned across the set of frameworks we reviewed. The *Framework* also highlights connections among reading, critical thinking, and the development of skilled writing; the interconnected nature of these literacy skills is widely acknowledged (e.g., NCTE-WPA, 2010).

Rhetorical Knowledge of Forms/Modes, Genres, and Disciplines

The most prevalent dimension across the frameworks we reviewed concerns skill in handling different forms of written products, with each of the nine frameworks including some attention to different types of communication. As defined in the AAC&U's LEAP VALUE rubrics, written communication "can involve working with many different writing technologies, and mixing texts, data, and images" (Rhodes, 2010, p. 1). Accordingly, frameworks emphasize that college graduates should be able to proficiently integrate multimedia (e.g., visual aids, charts, graphs, and images) to support comprehension of complex written material (Binkley et al., 2010); to "use effective communication channels and methods" (OHR-NIH, 2014, p. 1) including social media and electronic distribution; and to produce a variety of written forms, including letters, essays, e-mails, websites, reports, or presentations. This view of writing incorporates the notion of *multiliteracies* (Cope & Kalantzis, 2000), which emphasizes multilingual and multimodal literacy as critical in the 21st century. Thus, writing involves producing text using a variety of communication technologies, media, and dissemination channels.

With respect to genres of writing, the frameworks place particular emphasis on the genre of argument (mentioned in four of nine frameworks), which requires skill in presenting clear, coherent, and effective arguments that are convincing to an audience and that consider others' perspectives (Binkley et al., 2010). The genre of explanation is mentioned less often than argument. Explanations are called for explicitly in the DQP in the form of "explications of technical issues and processes" (Adelman et al., 2011, p. 14) and are more indirectly referenced in terms of "effectively [articulating] abstract ideas" (CAS, 2009, p. 46) in the CAS outcomes. Narrative, more common in K–12 settings, is mentioned in only one framework (Adelman et al., 2011). Other genres include directions, manuals, flow charts, and interviews.

In addition, adherence to disciplinary conventions (i.e., the forms and genres of expression that are valued within a major field or discipline) is mentioned in five of nine frameworks. Students are expected to be able to conduct inquiry within their discipline and to use correctly types and techniques of writing that are consistent with the values and expectations of the field. Genre and disciplinary considerations can be treated as part of a student's rhetorical knowledge (CWPA et al., 2011), but could also be considered a part of a student's conceptual knowledge of the discipline.

Rhetorical Knowledge of Context and Purpose

Attention to the context and purpose of a writing task is mentioned in a majority of frameworks (six of nine). Example purposes include advancing an argument to influence others or designing an approach to solve a problem. Writing should be appropriate for the purposes of the writing task, including use of appropriate tone and register (e.g., distinguish between formal and informal uses of language; write in a professional and courteous manner appropriate for business purposes). Context and purpose are closely related to genre and disciplinary considerations, and they are also a part of students' rhetorical knowledge (CWPA et al., 2011).

Rhetorical Knowledge of Audience

Audience awareness is directly mentioned in a majority of frameworks (seven of nine for audience and content) and is indirectly mentioned in the remainder. Audience design concerns a writer's attention to the knowledge, interests, and values of the recipient of a communication and skill in tailoring writing and expression to suit that audience (e.g., address experts and nonexperts in a specific field; address general and specific audiences). Some frameworks only indirectly mention audience awareness as part of the writing construct, in that writers should "convey meaning in a way that others understand" and should "write to influence others" (e.g., CAS, 2009, p. 46). These statements imply the notion of an audience (i.e., others should understand and be influenced by what is written), but they do not explicitly mention the term, nor do they indicate what *kinds* of others the writer might reasonably be expected to address.

Development and Organization of Content

Content development and organization is mentioned in seven of nine frameworks and can be defined as "the ways in which the text explores and represents its topic in relation to its audience and purpose" (Rhodes, 2010, p. 2). Organization involves producing prose that is logical, well structured, and coherent by, for example, moving from general topics to more specific ideas and details (CAS, 2009). Content development refers to the extent to which the writer effectively articulates abstract ideas and uses adequate supporting details (US-DOL ETA, 2014). When students are engaged in writing about something, their skill in developing and organizing that content in a coherent manner is critical for the communication to be successful.

Adherence to Language Conventions

Attention to language conventions is mentioned in six of nine frameworks. Statements relating to conventions (including syntax, grammar, and usage) underscore the idea that, by college, students should be fluent with text production skills and be able to compose "substantially error free prose" (Adelman et al., 2011, p. 14) with appropriate syntax and mechanics, spelling, grammar, and so forth. This includes knowledge of vocabulary, stylistic conventions, and the functions of language, both at surface and global levels (CWPA et al., 2011).

Writing From Sources and the Writing Process

The frameworks reviewed give relatively little attention to two features of written communication emphasized by the higher education writing community: (a) critical analysis and use of sources and (b) attention to the writing process. Using sources to support writing is included as a major dimension of the LEAP rubrics, suggesting attention to evaluating the relevance, quality, and credibility of those sources[2] (Rhodes, 2010); in contrast, the Bologna framework (González & Wagenaar, 2003) suggests that students should "receive and respond to a variety of information sources" (p. 144) in visual, oral (i.e., auditory), and textual formats, while the DQP states that students will be able to conduct inquiry from non-English language sources[3] (Adelman et al., 2011). The *Framework for Success in Postsecondary Writing* (CWPA et al., 2011) deals with writing from sources as an aspect of critical thinking and analysis of text materials, a process of conducting research from sources, knowledge of source attribution conventions, and incorporating electronic sources in multimedia productions; including elements related to use of sources in four of five dimensions of writing skill suggests that source use is particularly important for higher education.

With respect to the process dimension, skill in monitoring the writing process "from drafting to proof-reading" (Binkley et al., 2010, p. 22) is an important aspect of writing in ATC21S and is a major dimension of the *Framework for Success in Postsecondary Writing* (CWPA et al., 2011), but no other frameworks address this issue. In fact, the LEAP VALUE rubrics (Rhodes, 2010) specifically exclude notions of writing processes or strategies from their framework for student learning outcomes. However, as underscored by the *Framework* (CWPA et al., 2011), these strategies and processes are a critical aspect of writing at the college level and, thus, should be included in any comprehensive definition of written communication.

To summarize, based on the review of frameworks presented here, it is clear that in defining written communication, we must consider facility with multiple types (i.e., genres), forms (i.e., media), and audiences, in addition to the importance of the context and purpose for writing, and the importance of skill in manipulating both conceptual content (i.e., development and organization of ideas, critical analysis, and use of sources) and linguistic information (language, syntax, and mechanics; tone of voice and register) to suit the current communicative goals. The writing process (planning, drafting, and revision) is also critical.

Theoretical Perspectives on Writing From the Research Literature

It is evident from the frameworks reviewed in the previous section that writing is a complex skill, involving multiple dimensions, and that different perspectives on writing may differentially emphasize some of those dimensions. In this section, we explore theoretical perspectives that underlie these various dimensions of written communication competency and the importance of these dimensions for becoming a skilled writer. This survey of extant research literature is intended to enrich our definition of the construct and to suggest which dimensions might be more or less critical for higher education. Consistent with previous efforts to summarize the writing construct (Cumming et al., 2000), the work surveyed here suggests that writing involves cognitive processes situated within particular rhetorical or social contexts.

Sociocognitive Perspectives on Writing

Both social and cognitive perspectives on writing converge on the notion that writing is, by definition, social and purpose-driven (e.g., Bereiter & Scardamalia, 1987; Graham & Perin, 2007; Zimmerman & Risemberg, 1997; also see Deane, 2011). Genres of writing, for example, serve specific social goals and purposes (Bazerman, 2004), and those rhetorical goals shape and constrain the types and methods through which information should be recorded and shared with others when writing within a particular genre. In higher education, the focus is typically on *transactional* writing (i.e., writing to communicate or exchange information, ideas, or arguments with others in order to achieve particular purposes, such as to inform, persuade, or explain information to others; Burstein et al., 2014). Therefore, writers must consider the nature and needs of their audience(s) in order for communication to be successful (cf. Clark & Murphy, 1982). Sociocultural perspectives also emphasize that cultural conventions and social situations impact literacy practices (Perry, 2012), such that attention to the social context for writing is critical for both assessment and learning (cf. Behizadeh & Engelhard, 2011). This is consistent with work in the learning sciences suggesting that learning to write is best conceptualized as a process of socialization into a literate community of practice, whereby writers are guided by expert practitioners to gradually take increasing responsibility for producing the forms and genres of writing that are valued within a discipline or research community (Barab & Duffy, 2000; Bereiter & Scardamalia, 1987; Brown, Collins, & Duguid, 1989; Lave & Wenger, 1991). Instruction should strive to make writing socially meaningful to students (Alvermann, 2002).

This sociocognitive perspective has been applied to support the design of competency models and assessments. As an example, a model of ELA literacy incorporating social, conceptual, and linguistic dimensions has been developed under the Cognitively Based Assessment *of, for,* and *as* Learning (*CBAL*™) research initiative at ETS (Bennett, 2010). The CBAL

ELA competency model (Deane, 2011; Deane, Sabatini, & O'Reilly, 2011; Sabatini, O'Reilly, & Deane, 2013) specifies the reading, writing, and critical thinking skills that are necessary to learn in order to participate in key literacy practices (e.g., learning from informational text, engaging in argumentation, conducting inquiry and research). It is possible to conceptualize the various levels of knowledge and skill required for participation in literate activities as dealing with different types of knowledge representations (i.e., social, conceptual, and linguistic—including discourse, verbal, and print[4]—levels; Deane, 2011). Broadly, expert writing can be considered to involve a set of receptive skills (processing and comprehending information from source texts), expressive skills (synthesizing information from source texts and translating one's ideas into written words), and deliberative skills (applying appropriate strategic and meta-cognitive knowledge), which rely on the social, conceptual, and linguistic representations. A written product is the result of interactions among complex cognitive processes, as well as the knowledge of and skill in adapting one's production to meet the social and rhetorical constraints on what kind of writing must be produced to achieve one's purposes (Behizadeh & Engelhard, 2011). Thus, writing can be appropriately conceptualized as a set of sociocognitive practices (Behizadeh, 2014; Deane, 2011; Deane et al., in press), which experts can deploy strategically to achieve particular goals.

The CBAL ELA model specifies how the skills that support participation in various literate practices may develop, from novice to expert-level performances, by positing a set of hypothesized learning progressions (LPs) for the skills that constitute and contribute to performance of those key practices. These LPs can be used to support the design of assessments that target specific component skills (e.g., distinguishing between primary and secondary sources, making cross-text synthesizing inferences), while modeling the integrated practices required of professionals (e.g., writing a research report). At the most advanced levels of practice, writers are expected to take into account their purpose, audience, and disciplinary knowledge; in conducting research and inquiry, for example, writers are expected to present and support an original synthesis, review and evaluate evidence from relevant literatures (including seminal sources within the discipline), and to articulate how one's work contributes to and extends current knowledge and discourse about the issue (Sparks & Deane, 2014). These types of performances may not yet be achieved by the time students enter college, but are consistent with those expected in advanced undergraduate, graduate study, or professional practice. For more information on this effort, see http://elalp.cbalwiki.ets.org/.

Cognitive Processes of Writing

As described above, skilled writing requires the deployment and coordination of complex cognitive processes. A prevailing cognitive model, the Hayes-Flower model of writing (Hayes & Flower, 1980), specifies writing as consisting of interactions between the task environment (i.e., features of the writing assignment, such as the topic, audience, and context or purpose, and any text one has produced so far), the writer's long-term memory (i.e., knowledge of the topic, knowledge of the audience, and general plans for writing), and the writing process (i.e., planning, translating, and reviewing). Each aspect of the writing process is goal-directed and requires self-regulation. In the planning process, the writer retrieves relevant knowledge from long-term memory, evaluates the usefulness of the retrieved information, selects the most useful information, and organizes the information into a writing plan. Then the writer translates all these operations into sentences that can be understood by others. In the reviewing process, the writer reads or rereads the existing text and revises it when writing goals have not been satisfied (e.g., "I should address this counterargument to persuade the audience" or "I need to explain this complex idea in simple words"). These processes are recursive and interactive, as planning, translating, and reviewing can be triggered by one's goals.

Expert writers demonstrate qualitatively different writing processes compared to novices. Hayes and Flower (1980) found that skilled writers typically established their main writing goals and subgoals early in the writing process, while unskillful writers spent little time planning. Attention to one's goals for revision similarly explains observed differences in revising behaviors between expert and novice writers (Fitzgerald, 1987). First, expert writers typically spend substantial time and effort in revising their drafts (e.g., Holland, Rose, Dean, & Dory, 1985), but novice writers ignore the revision process or have little idea about how to do it well (Graves & Murray, 1980). Second, expert writers revise their work to improve its overall quality and to clarify the ideas that they want to convey to their audience (Hayes & Flower, 1986), while novice writers view revision as a task to correct grammar, spelling, diction, and punctuation (Faigley & Witte, 1981; MacArthur, Schwartz, & Graham, 1991; Sommers, 1980). Novice writers have an impoverished understanding of the revision process, resulting in revisions that are irrelevant to the meaning of the text, unconnected to genre considerations, and insufficient to help improve the quality of writing. While the ability to revise develops over time (Fitzgerald & Markman, 1987), many college students cannot perform this task adequately (Kinsler, 1990), suggesting that revision skill may differentiate more expert from less skilled writers.

Knowledge-Telling Versus Knowledge-Transformation

Another key difference between the writing practices of experts and novice students is in their approach to and conceptualization of the writing task with respect to content development and organization. In complex writing situations, where a writer must maintain and work toward achieving multiple goals, it is challenging for novice writers to handle all of the writing constraints without any support. Therefore, novices tend to approach the writing task as simply telling what is known about the topic (Bereiter & Scardamalia, 1987). In fact, many students, including some in college, compose using this *knowledge-telling* approach, because knowledge-telling may help reduce the burden of other cognitive processes, such as planning and revising, which makes the task of producing text manageable. However, students using this approach often overlook their rhetorical goals, the needs of the audience, the organization of the text, and the writing genre (Bereiter & Scardamalia, 1987; Graham & Harris, 1997), reflecting a lack of goal setting and self-regulation. In contrast, expert writers and domain experts are likely to use a *knowledge-transforming* approach, which involves viewing the writing task as a problem-solving process. Writers who adopt this approach do not only deal with knowledge and beliefs related to the topic, but also consider the rhetorical goals of the composition; experts make decisions about how to represent this knowledge best in terms of the appropriate language for the intended audience, which is directly reflected in the structure of the text (Bereiter & Scardamalia, 1987).

Reading and Writing From Sources

Reading and comprehending source texts gives writers content knowledge about which they can write (e.g., Hayes, 1996; see also Hillocks, 1987, 2005). Expert researchers across multiple domains rely on synthesis of multiple sources to situate their ideas within a particular literature and to build support for their knowledge claims (e.g., Bazerman, 1985; Goldman, 2004; Goldman et al., 2010; Latour & Woolgar, 1986). Writers of arguments, reports, and other research-based genres of writing must interpret sources, determine what information is relevant to their task and purpose, and decide what quotations or paraphrases to embed in the text to support their ideas. These reading-writing connections, including the importance of critically analyzing and using source texts to support one's writing, are emphasized in the *Framework for Success in Postsecondary Writing,* as described previously. However, according to reports from the National Adult Literacy Survey (Kutner, Greenberg, & Baer, 2006), fewer than one third of college graduates surveyed were proficient in comprehending prose (extended texts, such as newspaper articles) and other documents (practical directions, such as a prescription medicine label), suggesting that many college students' writing difficulties may be due to failures of reading comprehension.

Even students who read proficiently may have difficulty writing syntheses or arguments because they fail (and perhaps do not know how) to evaluate or to cite sources appropriately. Empirical research demonstrates that attention to sources (i.e., author expertise, publication venue, possible biases) supports understanding and integration of information from multiple documents (e.g., Bråten, Strømsø, & Britt, 2009; Britt & Aglinskas, 2002; Sparks, 2013; Wineburg, 1991), suggesting that students who are more attentive to the characteristics of source documents are better equipped to write essays or reports based on those sources. Unfortunately, empirical research generally suggests that undergraduates fail to attend to source information unless given specific instructions or tasks to consider it critically (e.g., Britt & Aglinskas, 2002; Rouet, Britt, Mason, & Perfetti, 1996; Sparks & Rapp, 2011; Wiley et al., 2009). These difficulties with sourcing likely contribute to several common issues observed in undergraduates' source-based essays, including plagiarism, inclusion of quotations without source attribution, excessive use of quotations (i.e., *quote pastiche*), little use of explicit citations (e.g., "according to Carnegie, … "), and little evidence of synthesis across sources (Britt, Wiemer-Hastings, Larson, & Perfetti, 2004). In one study, Britt et al. (2004) asked 108 undergraduates to write opinion essays from a set of seven sources on a history topic, finding that "only 28% of the essays included at least one explicit reference. Considering that no participants made more than two explicit references, it appears that undergraduates are not fully proficient at sourcing" (p. 2). Students tended to cite one to two key sources rather than incorporating content and ideas from across a variety of documents. However, findings from experimental tasks that require students to write and cite sources from memory may not fully generalize to situations where students write essays with source texts and notes available to them, such as in classrooms or assessment situations. It remains an open question whether these contexts might encourage additional attention to critical analysis and incorporation of sources.

Given the preceding theoretical discussion, it is worth considering the extent to which these research perspectives on written communication correspond to the instructional goals and outcomes observed in educational settings, both within higher education and in K–12, where college readiness is a particular concern. The following section outlines the writing skills that are important for success in college writing.

Writing Instruction and Learning to Write in College

Writing for College Readiness: Connections to Common Core State Standards

In developing a framework for written communication at the college level, it is critical to have expectations concerning incoming students' knowledge and skills. To understand what writing skills are expected of someone who is ready to take on college-level work, one can consider the upper levels of the Common Core State Standards (CCSS) for ELA/literacy (National Governors Association & Council of Chief State School Officers, 2010), with a particular focus on the writing standards for Grades 11–12. As these standards define the highest levels of K–12 performance, they are equivalent to the incoming skills expected of a first-year undergraduate who demonstrates readiness for college-level writing literacy.

As seen in Table 3, students who meet the expectations of the CCSS college and career readiness standards can comprehend and evaluate a variety of different texts and documents; construct effective arguments and explications of complex or multifaceted information; build and share their knowledge with others through writing; tailor communications to particular audiences, tasks, purposes, genres, and disciplines; select and use evidence that is appropriate for the discipline (e.g., history, science); conduct research and inquiry from multiple sources, evaluating their reliability and credibility; evaluate sources for their use of evidence; and cite specific textual evidence to support claims and explanations in one's writing. While it is certainly the case that many students will enter into higher education settings with these skills being less than fully developed, it is important to note that the standards correspond to many of the major dimensions of writing that emerged from the review of frameworks above, including attention to one's task, purpose, and audience; writing in the genres of argument and explanation; developing and organizing one's ideas coherently; proficiency with the writing and revision process; conducting research; and engaging in close reading and synthesis of sources.

Writing Instruction in the College Classroom

Undergraduates' experience with writing instruction varies with, historically, the bulk of this instruction occurring in first-year composition courses with little continuing writing instruction when students move on from general education courses to more specialized work within their major discipline. Since the 1980s, however, Writing Across the Curriculum (WAC) programs have emerged, emphasizing "active student engagement with the material and with the genres of the discipline through writing, not just in English classes, but in all classes across the university" (McLeod, 2012, p. 54). WAC views writing as a skill that must be continuously integrated into curricula, so that students can learn to communicate effectively within the constraints and values of their discipline through exposure to and practice of the conventions and genres that are valued for success in that discipline (i.e., *writing in the disciplines*). Importantly, writing is viewed not just as a way to demonstrate learning, but also as a method of learning and of refining one's thinking (i.e., *writing to learn*; WAC Clearinghouse, 2014). Writing to learn emphasizes the reflective and sense-making functions of writing, which can help the writer to organize and represent his or her thoughts coherently. Writing in the disciplines asks students to produce genres and forms of products that are used routinely by working professionals within the field (e.g., lab reports, position papers, literature reviews, journal articles, and project or grant proposals), consistent with sociocognitive perspectives described previously.

Despite its popularity, WAC can pose significant challenges for students. As described by Haswell (2008), "academic fields differ in the way they regulate every aspect of writing, from usage as minute as the function of the colon in titles to usage as pervasive as the way evidence is respected, gathered, and presented" (p. 416). For undergraduates, learning to write in the disciplines exposes them to "unfamiliar composing processes, novel genres and tasks, shifting standards and expectations" (p. 416), often resulting in discrepant feedback across courses (e.g., Anson, Schwiebert, & Williamson, 1993).

Table 3 Common Core Standards for Writing, Grades 11–12

CCSS standard	Description
W.11-12.1	Write arguments to support claims in an analysis of substantive topics or texts, using valid reasoning and relevant and sufficient evidence.
W.11-12.2	Write informative/explanatory texts to examine and convey complex ideas, concepts, and information clearly and accurately through the effective selection, organization, and analysis of content.
W.11-12.3	Write narratives to develop real or imagined experiences or events using effective technique, well-chosen details, and well-structured event sequences.
W.11-12.4	Produce clear and coherent writing in which the development, organization, and style are appropriate to task, purpose, and audience. (Follow standards 1-3)
W.11-12.5	Develop and strengthen writing as needed by planning, revising, editing, rewriting, or trying a new approach, focusing on addressing what is most significant for a specific purpose and audience. (Editing for conventions should demonstrate command of language standards 1-3 through grades 11-12).
W.11-12.6	Use technology, including the Internet, to produce, publish, and update individual or shared writing products in response to ongoing feedback, including new arguments or information.
W.11-12.7	Conduct short as well as more sustained research projects to answer a question (including a self-generated question) or solve a problem; narrow or broaden the inquiry when appropriate; synthesize multiple sources on the subject, demonstrating understanding of the subject under investigation.
W.11-12.8	Gather relevant information from multiple authoritative print and digital sources, using advanced searches effectively; assess the strengths and limitations of each source in terms of the task, purpose, and audience; integrate information into the text selectively to maintain the flow of ideas, avoiding plagiarism and overreliance on any one source and following a standard format for citation.
W.11-12.9	Draw evidence from literary or informational texts to support analysis, reflection, and research.
W.11-12.10	Write routinely over extended time frames (time for research, reflection, and revision) and shorter time frames (a single sitting or a day or two) for a range of tasks, purposes, and audiences.

The ability to understand and adapt one's writing to the current social and situational context (including disciplinary considerations) is an important part of the development of skilled writing (Carroll, 2002). After having mastered fluency with text production skills, college writers learn to produce different text structures, for different audiences, with various goals or purposes. However, development of these skills is uneven and dependent on students' experiences with various instructional strategies. The most common strategies for teaching college writing, as observed across a sample of more than 2,300 teacher intervention studies, include audience awareness, coauthoring and peer discussion, journaling, planning and prewriting (e.g., outlining, concept mapping), editing and proofreading, detecting and correcting errors, drafting or revising, and grammar instruction (Haswell, 2008). As these strategies are most commonly taught, one might predict that writing skills associated with those strategies would be among the most likely candidates for improvement during college and, therefore, could be considered potential targets for assessment.

What Skills Can Be Expected to Develop in College Writers?

Evidence from cross-sectional comparisons of first-year and senior students' writing reveals that advanced undergraduates show the largest gains with respect to vocabulary development, organization, reasoning and argumentation, use of composition strategies, and use of longer sentences with more complex syntactic structures (e.g., Flowers, Osterlind, Pascarella, & Pierson, 2001; Haswell, 1991; Hunt, 1970). However, there are clear limitations to drawing inferences about student improvement from cross-sectional data. Oppenheimer, Zaromb, Pomerantz, Williams, and Park (2014) conducted both cross-sectional and longitudinal analyses of growth in undergraduates' writing performance in response to persuasive (e.g., convince an audience about the importance of an issue; 20 minutes) and/or expository (e.g., explain a game or hobby so that someone could read the instructions and participate in the activity; 15 minutes) writing prompts. Writing samples were scored by trained raters on a 4-point scale and submitted to cross-sectional and longitudinal analysis. Cross-sectional results revealed average gains from first-year to fourth-year students of 0.33 points for persuasive and 0.25 points for expository writing. Longitudinal comparisons of first to third year, or second to fourth-year growth showed similar patterns, with essay scores improving by an average of 0.30 points. Evidence of growth was stronger for higher performing students, with approximately half of these students showing some improvement, but this study did not indicate what specific aspects of writing might improve over time.

Haswell (2000) reported results of a longitudinal study of the specific features of writing that can be observed to develop from the first to third year of college, using a random sample of 64 students' responses to an impromptu persuasive writing prompt. Haswell found significant improvement in several areas, including mean holistic rating (8-point scale, human scored, consistent with Oppenheimer et al., 2014); mean sentence, clause, and overall essay length (related to fluency and content development); proportion of words in introductory paragraphs, proportion of words in free modifiers (i.e., independent phrases or clauses, which can be moved to sentence-initial, mid-sentence, or sentence-final positions); and vocabulary (use of words with nine or more letters; p. 331). In sum, students can improve the cohesion, elaboration, logic, and overall quality of their persuasive and expository writing through instruction, with these improvements in the direction of the skills expected in professional practice (Haswell, 1986, 2000). The extent to which skill in the writing process and use of sources can be expected to develop during college is yet unclear, but because these dimensions differentiate expert from more novice writers, one could predict that college may help students develop these skills, to the extent that students receive appropriate instructional support. In the next section, we review existing assessments of written communication and the design challenges and considerations associated with developing such an assessment.

Review of Existing Assessments and Design Challenges

Existing Assessments of Written Communication

In support of the goal of developing an operational definition of written communication, we reviewed a variety of existing writing assessments designed to be administered to students approaching the entry or exit point of their college education. Specifically, we reviewed the assessments with a goal of understanding the advantages and disadvantages of existing approaches to assessing writing, and to inform both our notions of the construct and our recommendations for negotiating particular challenges in designing such an assessment. Key features of the assessments reviewed are described in terms of the assessment purpose, format, construct coverage, and reliability and validity evidence. Detailed information on each assessment appears in Table 4. Table 5 shows the correspondence between the targeted skills (i.e., as described in rubric statements and definitions) and the dimensions of the writing construct outlined in the section of this report.

Purpose of the Assessments

Assessments are created for various purposes, and these purposes affect how the assessment is designed, used, and interpreted. Because our goal is to support the development of SLO assessment of written communication in higher education, we examined assessments designed for a range of purposes and use cases, including placement into developmental or college-level English courses (e.g., ASSET, COMPASS, ACCUPLACER, English Placement Test, AP and CLEP tests), admission to graduate or professional programs (e.g., GRE Analytical Writing, GMAT Analytical Writing), assessment of student learning outcomes (e.g., Collegiate Assessment of Academic Progress, Collegiate Learning Assessment, ETS Proficiency Profile; ETS, 2010a; Liu, 2008), or multiple purposes (e.g., College BASE is used for both placement and SLO assessment).

Assessment Format and Construct Coverage

Large-scale writing assessment in the United States typically takes the form of selected-response (SR) tests, extended constructed-response (CR) tests (i.e., composing an essay response to a prompt), or writing portfolios, which consist of multiple examples of student writing across contexts, genres, and purposes, collected over time (Yancey, 1999). Portfolio assessment could be an effective complement for higher education institutions wishing to get a more detailed view of students' writing performance, particularly across genres, disciplines, and modes of expression, supported by interactions with and feedback from instructors (cf. Behizadeh, 2014; see Hamp-Lyons & Condon, 2000, for an in-depth review of portfolio assessment). Here, however, we focus on tests with SR and CR item formats.

Table 4 Existing Assessments of Written Communication

Test	Vendor	Purpose	Length	Delivery	Format	Forms & items	Scoring	Task type
ACT Writing Test	ACT	Placement	30 m	Paper	Essay (1, optional)		Essays scored by 2-3 human raters; holistic score (1-6), with scores summed across 2 raters (2-12)	Prompts describe an issue relevant to high school students; examinees are asked to explain their perspective on the issue.
ASSET Writing Skills Test	ACT	Placement	25 m	Paper	MC	36 items, with 12 items for each of 3 passages		Revision in passage context; assesses usage/mechanics, rhetorical skills (strategy, organization, style).
Collegiate Assessment of Academic Proficiency (CAAP)	ACT	SLO assessment	40 m	Paper	Essays (2)		Essays scored by 2 human raters; holistic score (1-6), with scores averaged across 2 essays	Prompt specifies an issue under discussion and a specific audience; examinees write an argument explaining their position on the issue.
CAAP	ACT	SLO assessment	40 m	Paper	MC	72 items, with 12 items for each of 6 passages	Students receive subscores for usage/mechanics and rhetorical skills	Revision in passage context; assesses usage/mechanics, rhetorical skills (strategy, organization, style).
COMPASS Writing Essay Test (e-Write)	ACT	Placement	60 m (varies)	Computer	Essay (1)		Automated scoring; holistic score from 2-8 (or 2-12), with subscales from 1-4 (1-6) on focus, content, organization, style	Students respond to an issue or problem (e.g., "take a position and offer a solution supported with specific example or evidence regarding the position taken").
COMPASS Writing Skills Placement Test	ACT	Placement	Untimed	Computer	MC	22 to 24 MC questions per essay prompt; 20 forms		Revision in passage context; assesses usage/mechanics, rhetorical skills (strategy, organization, style).
Medical College Admissions Test Writing Section	American Association of Medical Colleges (AAMC)	Admissions	60 m (30 m/ prompt)	Computer	Essays (2)		Essays scored by 2 human raters	Each prompt is a 3-part expository task: (a) explain the prompt: explain what you think the statement means; (b) explain a view opposite the prompt; (c) bring new meaning: explain a view that resolves the sides explained in (a) and (b).

Table 4 Continued

Test	Vendor	Purpose	Length	Delivery	Format	Forms & items	Scoring	Task type
English Placement Test (EPT)	California State University	Placement	45 m	Paper	Essay (1)	New prompt per each admin.; all concurrent examinees take same prompt	Human scored by 1 faculty member, 1-6 scale (0 for off topic or unscorable; 3 or below suggests that examinee lacks college readiness)	Prompts take a position on or present an argument about a situation or issue; examinees are instructed to evaluate the argument's reasoning and may invoke reasons and examples from their own experience, observations, and reading.
EPT	California State University, ETS	Placement	30 m	Paper	MC	45 items	40 items count toward total score; 5 are for field test purposes	4 item types: construction-shift items (select best continuation of sentence given a new introduction), sentence correction items (select best revision of underlined portion of a sentence), missing sentence items (select best sentence for paragraph initial, middle, or final-position blanks), and supporting sentence items (select sentence that best supports or explains preceding sentence).
CUNY Assessment Test in Writing (CATW)	City University of New York (CUNY)	Placement	90 m	Paper	Essay (1)		Human scored by 2 faculty members, with analytic (trait) scoring (1-6) in 5 categories	Prompts consist of a 300-350 word passage on topic of general interest. Examinees must respond to the passage as follows: (a) summarize the passage in their own words; (b) identify and explain a significant idea from the passage; and (c) support claims with evidence or examples (from the reading, prior experience, or other sources).
ACCUPLACER® Sentence Skills	College Board®	Placement	Untimed	Computer	SR	20 items		Revision in sentence context and construction-shift items.

Table 4 Continued

Test	Vendor	Purpose	Length	Delivery	Format	Forms & items	Scoring	Task type
ACCUPLACER *WritePlacer*®	College Board	Placement	Untimed	Computer	Essay (1)		Automated scoring; holistic score from 1-8, with 0 for off topic or unscorable; examinees may receive feedback (1-3 level proficiency statements) for each of 6 dimensions	Prompts present contrasting views on a topic of general interest; test takers are asked to write an essay that presents and explains their position on the issue to a specified audience; students are asked to "plan, write, review, and edit" the essay in the allotted time.
Advanced Placement® English Language and Composition (Section ii)	College Board	*AP*® course outcomes, credit	2 h (40 m/ prompt)	Computer	Essays (3)	Synthesis, passage analysis, and argument tasks	Human scored by 2-3 raters per essay; holistic scoring (0-9)	4 tasks. *Synthesis*: test takers respond to a general topic by synthesizing at least 3 of the accompanying 7 sources for support; *passage analysis*: test takers read a passage and write an essay that analyzes writing techniques (e.g., rhetorical devices, organization strategies, argumentative techniques, etc.); *argument*: the prompt presents a general issue and instructs test takers to write an essay presenting their own view of the issue, drawing from their own readings, experiences, observations, etc.
College Level Examination Program (*CLEP*®): College Composition Modular	College Board	Placement, credit	70 m for prompts provided by CLEP, or institution can set own test time	Computer (online), or determined by institution	Essays (2) provided by CLEP, or institution can substitute other direct writing assessment		Essays scored by institution	Prompts provided by CLEP include two tasks. *Argument*: write a response to a prompt using the student's own observations, experiences, or readings as support. *Synthesis*: write a response involving the synthesis and citation of 2 provided source texts.

Table 4 Continued

Test	Vendor	Purpose	Length	Delivery	Format	Forms & items	Scoring	Task type
CLEP: College Composition Modular	College Board	Placement, credit	90 m	Computer (online)	SR	90 items		Items assess revision skills, rhetorical analysis, ability to use source materials, and conventions of standard English. Item types include identifying errors in sentence contexts, revision in passage context, and items pertaining to the passage's use of rhetoric or sources.
CLEP: College Composition	College Board	Placement, credit	70 m	Computer (online)	Essays (2)		Essays scored by 2 human raters; holistic score (1-6), with scores summed across 2 raters (2-12)	2 tasks. *Argument*: write a response to a prompt using the student's own observations, experiences, or readings as support. *Synthesis*: write a response involving the synthesis and citation of 2 provided source texts.
CLEP: College Composition	College Board	Placement, credit	50 m	Computer (online)	MC	50 items	MC and essays weighted approximately equally; total score 20-80	Item types include identifying errors in sentence contexts, revision in passage context, and items pertaining to the passage's use of rhetoric or sources.
Collegiate Learning Assessment (CLA+)	Council for Aid to Education (CAE)	SLO assessment	60 m	Computer (online)	Essay (1)	Performance task	Essays scored by human raters, with automated scoring as supplement (Pearson's *Intelligent Essay Assessor*); examinees receive subscores (1-6) in 3 dimensions	Performance task requires students to read a scenario, analyze and take a position or draw a conclusion about the issue described, using sources from a document library to support one's claims, including citing those sources.
Criterion® Online Writing Evaluation Service	ETS	Placement, formative assessment, various	Varies	Computer (web-based)	Essays (varies)	Topic pool includes over 120 prompts appropriate for college writing courses	Automated scoring (*e-rater*®); examinees receive a holistic score (1-6 or 1-4), and diagnostic feedback on 5 dimensions	*Criterion* is an integrated assessment and instructional system that collects writing samples (i.e., essays in response to prompts) and provides instant scores and annotated feedback, to identify weaknesses in writing and to support students in the revision process. Instructors can create their own prompts or choose from a library of prompts.

Table 4 Continued

Test	Vendor	Purpose	Length	Delivery	Format	Forms & items	Scoring	Task type
ETS Proficiency Profile (EPP), formerly Measure of Academic Proficiency and Progress (MAPP) and the Academic Profile	ETS	SLO assessment	30 m	Computer (online)	Essay (1, optional)		Automated scoring (e-rater); examinees receive a holistic score (1-6)	Prompts present a claim about a topic of general interest that can be discussed from various perspectives; examinees are asked to think critically about the claim and to construct a well-organized, clear, and effective response that takes a position on the issue and supports that position with reasons and evidence.
EPP, formerly MAPP and the Academic Profile	ETS	SLO assessment	30 m/10 m abbreviated	Paper	MC	27 items (9 for abbreviated form)		Revision in sentence context and construction-shift items.
GRE® revised General Test: Analytical Writing	ETS	Admissions	1 h(30 m/ prompt)	Internet-based testing (iBT)	Essays (2)		Essays scored by 1-2 human rater (holistic scoring, 1-6), with automated scoring (e-rater) used as a check score; scores on 2 tasks are averaged	2 tasks. *Issue*: prompts provide generalizations, positions, reasons, etc. about a particular topic, along with a specific analytical task that candidates should enact in their responses (e.g., consideration of counterarguments to their own positions). *Argument*: prompts present arguments that include claims, examples, and flawed reasoning, sometimes from a particular source, and examinees must analyze the argument (e.g., discussing alternative explanations, identifying assumptions, or describing specific evidence needed as support).
TOEFL iBT® Writing Assessment	ETS	Placement, admissions	55 m (30 m for independent task; 20 m for integrated task; 5 m break)	Internet-based testing (iBT)	Essays (2)		Responses scored by 1-2 human raters who focus on content and meaning (holistic scoring, 0–5), with automated scoring (e-rater) of linguistic features. Scores on 2 tasks are summed and converted to a scale score of 0–30	2 tasks. *Independent task*: examinees are asked whether they agree or disagree with an opinion on a topic of general concern and to support their positions with reasons and examples (similar to GRE issue task). *Integrated task*: examinees read a passage and then listen to a lecture on the same topic (i.e., the lecture either supports or opposes the reading); they are instructed to summarize the information in the lecture and describe the relationship between that information and the information provided in the passage.

Table 4 Continued

Test	Vendor	Purpose	Length	Delivery	Format	Forms & items	Scoring	Task type
Graduate Management Admission Test (GMAT): Analytical Writing Assessment	Graduate Management Admission Council (GMAC)	Admissions	30 m	Computer	Essay (1)		Essays scored by 1–2 human raters (holistic scoring, 0–6), with automated scoring (Vantage Learning's *Intellimetric*) used as a check score; scores from 2 raters are averaged	*Analysis of an argument task*. Prompts present arguments that include claims, examples, and flawed reasoning in a fictional context, sometimes from a particular source, and examinees must analyze the argument in terms of specifics of the prompt (e.g., discussing alternative explanations, or describe the specific evidence needed to evaluate the argument).
Law School Admissions Test (LSAT) writing sample	Law School Admissions Council (LSAC)	Admissions	35 m	Paper	Essay (1)		Essays are not scored but are sent with LSAT scores to institutions identified by examinees	The prompt presents a fictional scenario describing 2 "equally defensible" decisions or positions; candidates must choose which of the 2 to support and provide reasons supporting the decision with reference to both alternatives presented.
Analytical Writing Placement Examination (AWPE; formerly Subject A)	Pearson/State of California	Placement	2 h	Computer	Essay (1)		Essays scored by 2–3 human raters (holistic scoring, 1–6)	Prompts present a prose passage (700–1,000 words) on an accessible general topic; examinees are asked to write an essay analyzing the passage and developing their own thoughts about issues described in the passage, using specific examples from experience, observations, or readings.
Georgia Regents' Essay Test	Regents' Testing Program Office	SLO assessment	60 m	Paper	Essay (1)	Topic list available (http://www2.gsu.edu/~wwwrtp/topics.htm)	Essays scored by 3 raters (holistic scoring, 1–3), with scores of 2–3 passing	Examinees are asked to write an essay responding to their choice of 1 from a set of 4 possible prompts on topics of general interest, relevant to society, politics, history, literature, or other subjects (e.g., "What could be done to make students more interested in learning about science? Discuss.").

Table 4 Continued

Test	Vendor	Purpose	Length	Delivery	Format	Forms & items	Scoring	Task type
College Basic Academic Subjects Examination (College BASE)	University of Missouri	Admission to educator preparatory programs, SLO assessment	20 m	Paper	MC	16–18 items	Students receive an overall writing cluster score and skill scores for understanding the writing process, use of language conventions, and the essay section	Items assess student skill in understanding the writing process, including (a) identifying and applying appropriate prewriting strategies, organizational methods, and research techniques (e.g., select the best source for a particular purpose) and (b) improving the clarity, coherence, organization, and style through revision (e.g., select the best revision of a sentence) and use of language conventions (e.g., recognize or repair flaws in mechanics and punctuation).
College BASE	University of Missouri	Admission to educator preparatory programs, SLO assessment	40 m	Paper	Essay (1, optional)		Essays scored by 2–3 human raters with holistic (1–6) scale	Prompts describe a scenario of general interest to college students, such as a proposal; examinees must express an attitude toward or take a position on the issue and defend that position using supporting examples and direct their response to a specified audience.
Writing Placement Exam	Washington State University	Placement	2 h	Online or paper	Essays (2)		Essays scored by human raters	2 tasks. *Argument*: examinees construct an argument about a short reading passage. *Reflection*: examinees reflect on the writing sample they just completed.

Note. m = minutes; h = hours; admin = administration; AP = advanced placement; MC = multiple choice; SR = selected response; SLO = student learning outcome.

Table 5 Correspondence Among Existing Assessments and Dimensions of Written Communication

Assessment	Task, context and purpose	Audience aware-ness	Genre conven-tions	Modes and forms	Deve-lopment and organization	Use of sources	Disciplinary conven-tions	Style, word choice, tone	Language use and conven-tions	Writing process
Selected-response assessments[a]										
ACCUPLACER	X	—	—	—	—	—	—	X	X	~
ASSET Writing Skills Test	X	—	—	—	X	—	—	X	X	~
CAAP Writing Skills	X	—	—	—	X	~	—	X	X	~
College BASE	X	X	X	—	X	X	—	X	X	X
COMPASS Writing Skills	X	—	—	—	X	—	—	X	X	—
EPT-CSU	—	—	—	—	X	—	—	~	X	—
EPP	—	—	—		—	—	—	X	X	~
Constructed-response assessments[a]										
ACT Writing Test (Essay)	~	—	~	—	X	—	—	—	X	—
ACCUPLACER WritePlacer (Essay)	~	—	~	—	X	—	—	X	X	—
AWPE (Essay)	~	—	~	—	X	~	—	~	X	—
CAAP Essay	~	X	X	—	X	—	—	—	X	—
CATW (Essay)	X	—	X	—	X	X	—	—	X	—
CLA+ (Essay)	X	—	X	—	X	X	—	X	X	—
CLEP: College Composition (Essay)	X	—	X	—	X	X	—	—	X	—
College BASE (Essay)	X	X	~	—	X	—	—	X	X	—
COMPASS E-Write (Essay)	X	X	~	—	X	—	—	X	X	—
EPP (Essay)	X	—	~	—	X	—	—	—	X	—
EPT-CSU (Essay)	X	—	X	—	X	X	—	—	X	—
Georgia Regents Essay Test	X	—	~	—	X	—	—	—	X	—
GMAT Writing	X	—	X	—	X	~	—	X	X	—
GRE-R Analytical Writing	X	—	~	—	X	~	—	X	X	—
MCAT Writing	X	—	X	—	X	—	—	—	X	—
TOEFL Writing	X	—	X	—	X	~	—		X	—
WSU Writing Placement Exam	X	—	X	—	X	~	—	—	X	—

Note. X = assessment provides evidence of this aspect; ~ = assessment provides partial evidence of this aspect; — = assessment does not provide evidence of this aspect.
[a]See Table 4 for full names of tests.

Selected-Response Format (Indirect Writing Assessment)

Eight assessments we reviewed included an SR section; of these, assessments designed for placement purposes were most prevalent (e.g., ASSET, ACCUPLACER, and EPT), but SLO assessments such as the EPP and CAAP include SR items as well.[5] Most common SR item types include (a) *revision-in-context* items, in which a section of a sentence or passage is underlined, and examinees can either select the most appropriate revision to correct an error in grammar, usage, or syntax or indicate that no revision is needed and (b) *construction-shift* items, which present an alternate beginning to a stimulus sentence and require examinees to select the best continuation of that stem from the options provided. ACCUPLACER, EPT, and EPP include both of these item types; CAAP includes revision-in-context items presented within a passage. Other SR item types ask examinees to select the best sentence to fill a blank in a paragraph (in initial, middle, or final-sentence position; e.g., EPT) or to answer questions about the writer's rhetorical or stylistic goals (e.g., CAAP).

From a measurement perspective, SR assessments have some advantages, as they tend to be more cost-effective in terms of administration and scoring than CR items, and they are considered more objective (i.e., with specific and distinct correct responses versus open-ended items that may not have a single correct answer). SR items are often faster to complete, meaning that examinees can respond to more of these items in the allotted time compared to the number of CR items (e.g., for the CAAP, test takers are asked to respond to 72 SR items, as compared to two CR essays, in 40-minute sessions) and, therefore, an SR test typically has a higher reliability than a CR test taking the same amount of time.[6] The SR items may also demonstrate better prediction of criterion scores (i.e., scores on a series of short essay tasks) than a single holistically scored CR essay (e.g., Godshalk, Swineford, & Coffman, 1966).

However, with respect to construct representation, use of SR items has some clear limitations. SR items have been said to "fail to address the cognitive and reflective processes involved in creating a text—such as making plans for writing, generating and developing ideas, and making claims and providing evidence" (Murphy & Yancey, 2008, p. 450; see also Odell, 1981), suggesting that SR items underrepresent the writing construct. Consistent with this notion, the SR items we reviewed overwhelmingly assessed lower level writing abilities, such as language conventions (grammar, usage, mechanics), style (i.e., word choice, sentence variety, and register), and organization (text structure and sequence of ideas), at the expense of higher order writing skills. SR item types such as revision-in-sentence-context and construction-shift items are typically used to assess students' knowledge of local organization, style, and language conventions while only indirectly assessing skill in the revision process. CAAP Writing, for example, provides students with subscores for usage/mechanics and rhetorical skills (i.e., strategy, purpose, organization, and style). Revision-in-passage-context items assess skill in usage, mechanics, or style, targeting a specific text section, while strategy and organization items might ask about the passage as a whole; a typical rhetorical strategy question might ask examinees to evaluate the appropriateness of a quotation in the passage, given a particular communicative goal. This example item indirectly addresses use of sources but is considered mainly in terms of attention to rhetorical purpose. The College BASE SR writing test also includes items assessing skill in selecting appropriate prewriting strategies, text structure and organization, choosing sources for a particular purpose, and revision; these items contribute to a subscore for understanding the writing process. While the College BASE had the widest construct definition of any SR assessment we reviewed, given the limited number of items on the test (i.e., 16–18), it is unlikely that the intended construct can be adequately covered such that it provides useful information about rhetorical or conceptual skills.

Thus, while SR items can be used to evaluate linguistic as well as more rhetorical dimensions, the emphasis on rhetorical skills—and the extent to which they can be measured reliably—may vary with particular test designs. For example, students who demonstrate proficiency with SR revision items can be said to possess the abilities to manipulate sentences, to correct errors in diction and syntax, and to recognize inappropriate relations among clauses (e.g., ACCUPLACER). Given the review presented in the first section of this article, it is clear that the writing skills that can be effectively assessed by asking students to correct errors within single sentences are largely limited to those related to language conventions (i.e., grammar, usage, syntax, and mechanics). This tendency of SR items to focus on low-level mechanics and usage in lieu of higher order cognitive skills is a primary objection to the use of SR tests to measure writing proficiency (cf. Murphy & Yancey, 2008). Arguably, revision-in-passage-context items can assess discourse-level, rather than sentence-level, processing, which is more consistent with the kinds of literacy practices that are expected of college writers who deal more often in extended text and discourse than with discrete sentences presented in an isolated fashion. However, in general, SR assessments still only estimate students' *probable* writing ability, by testing discrete knowledge and skills that are *associated* with writing, rather than evaluating students' ability to produce coherent, error-free writing, as in CR assessments.

Constructed-Response Format (Direct Writing Assessment)

The majority of writing assessments we reviewed included CR items, which directly assess students' writing skills. Typically, CR assessments require examinees to compose an essay in response to a prompt or stimulus under controlled conditions; the texts produced are then evaluated, whether by human raters, automated writing evaluation systems, or some combination of the two. The CAAP essay, COMPASS e-Write, CLA+ Performance Task, CUNY CATW, and the GRE Analytical Writing are all examples of CR tests. Many such assessments ask examinees to take a position and present a well-developed argument using supporting evidence from one's own readings and experiences (e.g., CAAP essay, GRE issue task) or to critically analyze arguments or information presented in a text (e.g., CLA+, GRE argument task).

Use of CR format items is consistent with the widely held perspective that the most valid measures of writing ability are those that actually require students to write extended text (cf. Fowles, 2012; Yancey, 1999). In contrast to SR tests, CR items are more authentic, in that they treat writing as an active, social, communicative process (Murphy & Yancey, 2008). That is, CR tasks require examinees to deploy and demonstrate proficiency with the social, cognitive, and linguistic processes that are necessary to solve the rhetorical problem posed by the prompt (cf. Bereiter & Scardamalia, 1987). Under timed conditions, the writer's fluency with these processes becomes particularly important, as he or she needs to be able to produce clear and effective text with a logical and coherent organization and structure, despite limited time for planning and revision (Hayes & Flower, 1980). The greater one's fluency with low-level language processes, the more one can use available cognitive and conceptual resources to develop and organize ideas, to engage with the intended audience, and to address the rhetorical goals of the piece. Fluent writers' essays are also less likely to be marked by errors in syntax, mechanics, grammar, and word choice.

However, the extent to which examinees are expected to engage in higher level social and rhetorical problem solving in a given CR assessment depends on the assigned prompt. Most assessments we reviewed assessed students' response to the assigned task and genre of writing requested (i.e., argument); organization and content development; word choice and style; and adherence to conventions and control of grammar, usage, and mechanics. For example, CR prompts that ask examinees to take a position on an issue, to support that position with reasons and examples, and to anticipate counterarguments, while addressing the response to a specific audience (e.g., CAAP, COMPASS e-Write), provide evidence of students' skill in several aspects of writing: adapting writing to purpose and audience, adherence to genre conventions for argument structure and quality, development and organization of ideas, and facility with both stylistic and grammatical language conventions. It is notable that while any CR assessment could potentially evaluate audience awareness by asking writers to address a specified audience, only three assessments we reviewed included this task requirement (i.e., CAAP, COMPASS e-write, and College BASE). In the majority of CR assessments, then, aside from the raters that score the essay, examinees are writing arguments to no one in particular, which does not truly count as an instance of written communication (Condon, 2013); this lack of authentic social features has led some researchers to claim that CR tests largely ignore social and cultural elements, using "one narrow version of literacy to represent a broad construct" (Behizadeh, 2014, p. 128). Given the importance of audience awareness in advanced writing proficiency, it is important to assess this aspect of writing; yet, many current CR assessments fail to do so.

CR assessments can also be used to evaluate students' use of sources in writing. For example, the CLA+ performance task presents examinees with a document library, which they can consult and use as evidentiary support in addressing key questions and making an argument about an issue described in the prompt. Others include a more limited text stimulus, yet they still require students to critically evaluate or summarize information from sources. For example, CATW asks examinees to respond to a reading passage of 300–350 words by summarizing the most important ideas of the author and explaining the significance of one key idea, using supporting evidence and examples from prior learning or experience (CUNY, 2012). Students are assessed in terms of understanding and responding to the main ideas in the passage and the use of supporting details and examples, including specific references to the passage. Other assessments partially deal with use of sources, by either asking examinees to summarize or explain the ideas in a passage (e.g., TOEFL integrated task) or to critique those ideas, without requiring examinees to quote or cite information from those sources as support for their ideas. For example, studies of expert raters indicate that although the GRE Analytical Writing tasks provide much information that is relevant to important writing skills at both the undergraduate and graduate levels (e.g., organizing ideas and information coherently, following conventions of standard written English), they do *not* provide information about students' ability to credit sources appropriately or to integrate quoted or referenced material into their own text (Rosenfeld, Courtney, & Fowles, 2004). Such assessments provide better measurement of students' critical reading and analytic skills, rather than their skill in writing from multiple sources, per se.

Further, no CR assessment provided information about students' writing process, other than drafting. The nature of most on-demand writing assessments precludes assessment of planning or revision, because examinees respond to a single prompt in a limited amount of time. While the final written product is saved and evaluated, the composition process is not captured. However, with technology-enhanced delivery, the writer's process can be captured for subsequent analysis. Evidence from analyses of keystroke logs suggests that process-level features can predict students' writing proficiency (e.g., Deane, 2014), though efforts to use keystroke-logging techniques on the fly to evaluate and score the efficiency and effectiveness of students' processes or to deliver just-in-time feedback are still in early stages. Systems like ETS's *Criterion* Online Writing evaluation service can help support students and teachers in understanding and engaging in the writing process by providing planning tools and a collection of prompts to which writers can compose responses and receive instant feedback (provided by the e-rater engine) about aspects of the text that could be improved through revision. When students successfully address the feedback provided by the system, their scores may improve if they resubmit the revised essay to *Criterion* (though evidence suggests that some implementations of *Criterion* may not take full advantage of the planning and revision tools; Warschauer & Grimes, 2008). However, even this system does not provide assessment of the writing process per se. New assessment designs that incorporate distinct planning and revision tasks, or traditional revision-in-context SR items, may be required beyond typical CR tasks, if assessing the writing process is considered a priority. Similarly, assessments intended to provide information about students' proficiency with composing in multiple modes and formats (i.e., using technology-enhanced composition tools); disciplinary conventions; or genres other than argument, critique, or explanation (such as a research report) will necessitate different assessment design strategies than those currently observed in the market.

In sum, relative to SR items, CR items demonstrate better coverage of the written communication construct. However, CR items have other notable constraints, such as the extended testing time required for essay writing tasks and the increased costs associated with scoring the responses, particularly if human raters are to be used (Williamson, Xi, & Breyer, 2012). We return to these issues when discussing assessment design challenges; but, first, we examine reliability and validity evidence for the assessments reviewed.

Reliability and Validity Evidence

Table 6 presents a summary of reliability and validity evidence available for each assessment reviewed. Reliability and validity evidence has been examined throughout the literature, particularly for three popular, widely used assessments: CAAP, CLA/CLA+, and EPP. Substantial validity evidence has also been gathered for the GRE Analytical Writing assessment. Importantly, for many of the assessments we reviewed, written communication represents only a part of a larger assessment; accordingly, for the purposes of the current review, only reliability and validity evidence pertaining to the writing sections will be examined here.

Reliability Evidence

Often, written communication assessments represent a subtest of a larger suite of assessments; therefore, it is important to demonstrate evidence of adequate reliability (i.e., internal consistency) for those subtest scores. The reliability of a particular test score is highly related to the number of items within that test, so test length is an important consideration with respect to reliability (Sinharay, Puhan, & Haberman, 2011). In part due to this, CR assessments often demonstrate low reliability compared to SR assessments, where a larger number of items can be administered within the same testing time. Still, sufficient numbers of items must be administered to achieve adequate reliability. For example, the EPP only reports individual subtest scores (i.e., a separate score for EPP Writing, Reading, Critical Thinking, and Mathematics) if individuals take the standard form, with 27 items per section, but not the abbreviated form, with only nine. EPP Writing has demonstrated alpha reliability coefficients of .81 (ETS, 2010a) and school-level reliability of .91 (Klein et al., 2009); estimates for the SR CAAP writing section are similarly high, with school-level reliability of .88 (Klein et al., 2009) and KR-20 of .92 (CAAP Program Management, 2012). CAAP also reports sufficient reliability of the Rhetorical Skills and Usage/Mechanics subscales, with KR-20 ranging from .84 to .86 across forms (CAAP Program Management, 2012).

In contrast to SR tests, CR assessments are typically less reliable. For example, school-level reliabilities for CAAP Essay (.75), CLA Make an Argument (.84), and the CLA Performance Task (.75) are lower than estimates observed for SR-format assessments, but all reliability estimates exceeded .70 except for the school-level reliability of the CAAP Essay for first-year students, which was .68 (Klein et al., 2009). Reliability for the GRE Analytical Writing section is estimated at .82 (ETS, 2013b), similar to the figures for the CAAP and CLA MA tasks, but slightly higher than the estimated reliability (.77) of the analytical writing section in the version of the GRE used prior to August 1, 2011[7] (ETS, 2010b). The CLA+, the most recent version of the CLA, also yields relatively low individual-level reliability estimates for the Performance Task; specifically, CAE reports coefficient alphas for the Performance Task of .43 and .57 for test forms A and B, respectively (Zahner, 2013). The CLA+ Performance Task provides measures of students' writing mechanics and writing effectiveness, in addition to analytic reasoning and problem solving (i.e., a critical thinking measure), by using trait scoring, rather than holistic scoring. The total CLA+ test achieves a higher reliability (i.e., alpha between .85–.87) by combining the CR Performance Task with highly reliable SR items assessing other skills. However, the low reliability estimates observed suggest that a subscore for writing should not be reported.

Interrater Reliability

Interrater reliability measures the degree of agreement among raters for CR assessments. Many studies measure interrater reliability by estimating the consistency between raters using correlation methods or percent agreement; for these consistency estimates, values exceeding .70 are considered acceptable, yet thresholds for interrater agreement may vary,

Table 6 Reliability and Validity Evidence for Existing Written Communication Assessments

Assessment	Author/Year	Subjects	Sample	Reliability	Validity evidence
ACT Writing Test	ACT (2009)	Students	20	Interrater reliability = 0.94	
		Students	6,346	Alternate forms reliability = 0.67	
		Colleges	10		ACT Writing alone accurately placed 65% into college classes, an estimated 7% increase over placing all students into a college- level course. 66% of students accurately placed earned a grade of B or better.
		Students from 147 writing courses from 36 colleges	4,598		Grades in writing courses increase with increases in ACT Writing scores. Regression analysis predicting course grades is improved by added ACT Writing vs. using high school English grades and ACT English alone ($R = 0.448$).
ASSET Writing Skills Test	Hughes and Nelson (1991)	Entry-level college students	578		57.61% of students were correctly classified.
	Moss and Yeaton (2006)	Students who successfully passed college-level English	1,473		In grades for college-level English, 89.3% of students who were sorted into developmental English first scored C or better, and 87% of students sorted into college-level English scored C or better. In grades for college-level English, 53.8% of students who were sorted into developmental English first scored B or better, and 68.1% of students sorted into college-level English scored B or better.
Collegiate Assessment of Academic Proficiency (CAAP)	Council of Presidents and State Board for Community College Education (1989)	Baccalaureate students who earned between 75 and 105 credit hours and community college students who earned at least 70 credit hours	1,302		The test received a favorable rating by no more than 30% of the faculty on any dimension.
	Klein et al. (2009)	Freshman and seniors of 11 colleges	1,051	$\alpha^{b} = .75$	At the student level, the essay section correlated with MAPP at $r = 0.33$, and the CLA MA at $r = 0.37$. It had a precision-weighted average observed effect size of 0.37, with a standard error of 0.092. At the institution level, the essay section correlated with the MAPP at $r = 0.70$, the CLA MA at $r = 0.67$, and the CAAP itself at $r = 0.74$.

Table 6 Continued

Assessment	Author/Year	Subjects	Sample	Reliability	Validity evidence
	CAAP Program Management (2012)	Students participating in national administrations from 1998 to 2001	80,010	KR-20[a] = .92 for Writing Skills	
COMPASS Writing Skills Placement Test	ACT (2006)	Students in participating colleges	68 colleges	.90 for writing	Accurately placed 66% of students expected to earn a B or higher based on test cutoff scores.
	Belfield and Crosta (2012)	College freshmen	3,425		Accurately placed 60.5% of students expected to earn a B or higher based on test cutoff scores.
	Davey et al. (1997)	Students from postsecondary institutions	13,106		Average estimated accuracy rate of 60% for students to earn a B or higher, a 13% improvement over placing all students in an entry-level course.
	Matzen and Hoyt (2004)	Incoming freshmen	431		Accurately placed 62% of students.
	Scott-Clayton (2012)	Students enrolled in a large urban community college system (LUCCS)	36,917		Accurately placed 61% of students expected to earn a B or higher based on test cutoff scores.
COMPASS Writing Essay Test (e-Write)	ACT (2006)	Responses to 6 prompts	300		The correlation between human raters and the e-Write program varied between $r = 0.67$ to $r = 0.83$ across prompts.
	ACT (2006)	Responses to 6 prompts	900		The correlation between human raters and the e-Write program varied between $r = 0.55$ to $r = 0.60$ across prompts when evaluating for the analytic subscales of the e-Write program.
	Matzen and Sorensen (2006)	College freshmen	~300		Essays scored by e-Write correlated at $r = 0.267$ with the ACT COMPASS test, $r = 0.290$ with the ACT English test, $r = 0.192$ with the ACT Reading test, and $r = 0.209$ with the ACT Composite score.
English Placement Test (EPT)	Michael and Shaffer (1978)	Incoming freshmen	637		EPT-Essay correlates with grade in English class at $r = 0.35$ and fall semester GPA $r = 0.21$. EPT total score correlates with grade in English class at $r = 0.47$ and fall semester GPA $r = 0.30$. EPT-Essay correlates with *SAT*® Verbal score $r = 0.42$. EPT total score correlates with SAT Verbal score $r = 0.74$
	White (1995)	Students enrolled fall 1978	3771		Of percent enrolled in spring 1981, those who did not take EPT had fallen to 37.8%, while those placed in remedial English had fallen to 51.8%, and those placed in regular English had only fallen to 57.8%.

Table 6 Continued

Assessment	Author/Year	Subjects	Sample	Reliability	Validity evidence
ACCUPLACER and WritePlacer	Belfield and Crosta (2012)	College freshmen	1,983		Accurately placed 59% of students expected to earn a B or higher based on test cutoff scores.
	James (2006)	Students	60		Correlation between WritePlacer plus and human scorers $r = 0.63$. Correctly predicted the results of students 70% of the time, 90% in an introductory writing course and 66% for a literature and composition course.
	Mattern and Packman (2009)	Students at 17 different colleges	3,408		For B or better adjusted $r = 0.35$, with 59% of students correctly placed, for C or better adjusted $r = 0.29$, with 75% of students correctly placed.
Collegiate Learning Assessment (CLA)	Benjamin and Chun (2003)	Students at 14 different colleges	1,300		At the institution level, the entire CLA has a correlation of $r = 0.90$ with the SAT. At the student level, the scores on CLA have a correlation of $r = 0.50$ with GPA, and $r = 0.65$ when adjusted for reliability.
	Klein et al. (2007)	Freshmen and seniors			At the student level, the essay section correlates with the SAT at $r = 0.44$ for freshmen and $r = 0.46$ for seniors.
	Klein et al. (2009)	Freshmen and seniors at 11 colleges	544	$\alpha^b = .84$	At the student level, the CLA MA has a correlation of $r = 0.37$ with the CAAP essay, and $r = 0.44$ with the MAPP writing section. It had a precision-weighted average observed effect size of 0.28, with a standard error of 0.089. At the institution level, the CLA MA has a correlation of $r = 0.67$ with the CAAP essay and $r = 0.86$ with the MAPP writing section.
Collegiate Learning Assessment PLUS (CLA+)	Council for Aid to Education (2013)	Students at four colleges		$\alpha^b = .43 - .57$ between 2 forms for Performance Tasks	78.5% of students believed that the test was at least a moderately good test of writing (face validity).
Criterion Online Writing Evaluation Service	Klobucar et al. (2012)	First-year students	1,482		Criterion scores and SAT Writing correlated at $r = 0.43$ for 2009 and $r = 0.41$ in 2010. In 2010, Criterion predicted course grades at statistically significant levels for all groups with sufficient sample size.

Table 6 Continued

Assessment	Author/Year	Subjects	Sample	Reliability	Validity evidence
ETS Proficiency Profile (EPP), formerly Measure of Academic Proficiency and Progress (MAPP) and the Academic Profile	Banta and Pike (1989)	College seniors	1,228	KR-20[a] = .79 – .84	Multiple choice writing subscore correlated at r = 0.58 with ACT scores, total score correlated at r = 0.72 with ACT scores.
	Council of Presidents and State Board for Community College Education (1989)	BA students who earned between 75 and 105 credit hours and community college students who earned at least 70 credit hours	1,302		Academic Profile received a favorable rating by no more than 30% of the faculty on any dimension.
	Klein et al. (2009)	Freshmen and seniors at 11 colleges	1,051	α[b] =.91	At the student level, the MC section correlated with CAAP at 0.72, the CAAP essay at $r = 0.33$, and the CLA MA at $r = 0.44$. It had a precision-weighted average observed effect size of 0.34, with a standard error of 0.063.
ETS Proficiency Profile (EPP) Optional Essay	Liu et al. (2012)	Students from 3 colleges	757		Essay correlated with SAT scores at $r = 0.34$ for the research university, at $r = 0.27$ for the master's university, and with a college placement exam at $r = 0.51$ for the community college.
Graduate Record Examinations (GRE): Analytical Writing (AW)	Klieger et al. (2014)	Doctorate students from a statewide university system	4,229		GRE-AW correlated with doctoral level graduate GPA at $r = .16$ for all fields of study. Psychology was the field with the highest correlation $r = .24$.
	Klieger et al. (2014)	Master's students from a statewide university system	21,127		Correlation of GRE-AW tasks with master's level graduate GPA for any field is $r = .16$. Math students had the lowest correlation at $r = .11$, and the highest correlation was for English language and literature students, $r = .28$.
	Ramineni et al. (2012)	Essays drawn from test takers between Sep. 2006 and Sep. 2007	750,000		Correlation of GRE-AW issue task rated by humans with undergraduate GPA is $r = 0.13$, the correlation of GRE-AW argument task as rated by humans with undergraduate GPA is $r = 0.20$. While the correlation of GRE-AW issue task as rated by humans with GRE Verbal score is $r = 0.51$, and the argument task is $r = 0.55$

Table 6 Continued

Assessment	Author/Year	Subjects	Sample	Reliability	Validity evidence
	Schaeffer, Briel, and Fowles (2001)	Students from 26 colleges	2,326	$r_{xx}^{c} = .70$	Correlation between the issue task and the argument task varied between $r = 0.56$ and $r = 0.46$, depending on order of task given.
TOEFL iBT Writing	ETS (2011)	Operational data from 2007		0.74 for writing	
	Weigle (2011)	Nonnative speakers of English from 8 colleges	386		Correlation of TOEFL iBT Writing task and instructor rating of writing ability was $r = 0.32$ when scored by human raters, but rose to $r = 0.37$ when writing and assessing in the student's subject area.
Georgia Regents' Essay Test	The Regents' Testing Program (n.d.)	Students at the 45-credit hour mark from 2002–2009	30,417		By the 45-credit hour mark, 89.4% pass the essay portion of the test, 88.4% pass the reading portion of the test, resulting in 84.6% passing both tests.
College Basic Academic Subjects Examination (College BASE)	Cole and Osterlind (2008)	Students from more than 51 institutions	1,318	$\alpha^{b} = .771$ for English	
	Gao (2003)	Students who took College BASE in 2000	8,009	$\alpha^{b} = .79$ for English	English portion correlated at $r = 0.49$ with SAT Verbal scores, at $r = 0.57$ with ACT composite scores, and at $r = 0.36$ with college GPA.
	Osterlind, Robinson, and Nickens (1997)	Students from 56 institutions	74,535	KR-20[a] = .67 for Writing	
	Pike (1992)	Students	1,037		Both total score and subscore scoring models are supported by the empirical structure of the test. Writing loaded 0.50 on the total score of the test and 0.82 on English/Writing subscore of the test.
Washington Writing Placement Exam	Haswell (1998)	Entering students	2,800	KR-20[a] = .83	78% of students placed into regular English pass, 9% pass at a questionable pass, and 13% fail. 63% of students placed into regular English plus a tutorial pass, 15% questionably pass, and 22% fail. Only 58% of students who did not take the exam passed.
	Haswell (1998)	Juniors	2,500	KR-20[a] = .83	81% of native writers, transfer and nontransfer, and nontransfer nonnative writers pass.

Table 6 Continued

Assessment	Author/Year	Subjects	Sample	Reliability	Validity evidence
Analytical Writing Placement Examination (AWPE; aka Subject A)	University of California (2014)	New freshmen	32,947		5,529 students (17% of total enrolled) passed the writing requirement; 8,792 (27% of total enrolled) did not pass the writing requirement.
Advanced Placement (AP) English Language and Composition (Section ii)	Dodd, Fitzpatrick, De Ayala, and Jennings (2002)	AP students	959		AP scores above 3 scored significantly higher grades on a subsequent English course than those who scored 2-, $F = 3.86$, $p < .0002$
Graduate Management Admission Test (GMAT): Analytical Writing Assessment (AWA)	Kass, Grandzol, and Bommer (2012)	MBA students	69		Adding the analytical writing section to the regression model did not significantly improve the prediction in MBA GPA (value-added $R = .019$) over undergraduate GPA, GMAT-Verbal, and GMAT-Quantitative. However, AWA had significant and positive correlations with four of five managerial competencies (leadership, decision making, communication, and teamwork, but not organization).

Note. CLA MA = Collegiate Learning Assessment Make an Argument; MAPP = Measure of Academic Proficiency and Progress; GPA = grade point average; MBA = master of business administration.

[a]KR-20 (Kuder-Richardson Formula 20) is an index of the internal consistency reliability of a measurement instrument, such as a test, questionnaire, or inventory. [b]α Cronbach's alpha is a measure of internal consistency, that is, how closely related a set of items are as a group. [c]r_{xx} is the proportion of variance that is true variance in score and is a measure of reliability.

depending on the stakes of the assessment. For the CLA+, interrater correlations of .67 to .75 have been observed across several forms of the performance task, which is scored using a 3-trait rating system (Zahner, 2013). For the CAAP Essay, interrater reliability estimates range from .68 to .74 across seven prompts, with percentage of perfect agreement ranging from 70 to 78% on a 1–6 holistic rating scale (CAAP Program Management, 2012). Thus, these assessments appear to achieve acceptable interrater reliability as measured by consistency estimates (Stemler, 2004).

For CR items scored by both human raters and automated scoring systems, correlations between those two ratings are reported as a measure of the extent to which the human and automated system agree. Correlations observed between human and automated scores were somewhat lower than correlations among human raters. As one example, the WritePlacer online assessment reports Pearson correlations of $r = .63$ between holistic scores assigned by humans and those assigned by the IntelliMetric automated essay scoring system (James, 2006); operationally, this assessment uses automated scoring as the sole method of scoring student essays. For COMPASS e-Write, observed correlations between human raters and automated scores ranged from $r = .67$ to .83 across prompts for holistic scores; correlations between human and automated scores on analytic (trait scored) subscales ranged from $r = .55$ to .60 across prompts (ACT, 2006), suggesting that automated scoring methods may not be able to provide sufficiently reliable trait scores.

Further research on the validity of ETS's e-rater has reported correlations between human and e-rater scores that are comparable to those observed between two human raters. For example, Burstein, Kukich, Wolff, Lu, and Chodorow (1998) examined scores on a sample of 500 GMAT analytical writing essays, across a sample of eight argument and five issue prompts. They reported correlations of .82 to .89 between two human raters, compared to .79 to .87 between e-rater and each of the human raters. Reported human/e-rater correlations for the GRE are somewhat lower (.73 to .74) compared to human/human correlations (.83 for argument and .85 for issue prompts; Powers, Burstein, Chodorow, Fowles, & Kukich, 2002a, 2002b). Research on the IntelliMetric system has reported average correlations between human and automated scores of .83 across six different GMAT analytical writing prompts (Rudner, Garcia, & Welch, 2006); this correlation is the same as the average observed correlation among two human raters ($r = .83$), indicating comparable interrater reliability across automated essay scoring (AES) and human scoring methods. Correlations ranged from .80 to .84 across forms of the argument task and from .83 to .87 for the issue task, indicating that both task types achieved good reliability. However, because this agreement is imperfect, Powers et al. (2002a) suggested only using automated scores to supplement human ratings, particularly under high-stakes testing conditions.

Convergent Validity Evidence

Convergent validity evidence concerns the relationship between scores across tests measuring similar constructs (AERA, APA, & NCME, 1999). Klein et al. (2009) reported correlations between the two SR EPP and CAAP Writing tests of .72 at the student level and .97 for the school level, representing a very strong relationship in the aggregate. Overall, SR assessments of writing skill as an SLO appear to be better correlated with one another than comparable CR format tests. The lowest student-level correlations among the writing measures administered by Klein et al. (2009) were observed between the CAAP essay and the EPP ($r = .33$), the CLA Performance Task ($r = .32$), and the CLA MA task ($r = .37$). Again, at the institution level, these correlations were somewhat higher ($r_{EPP} = .70$, $r_{CLA\text{-}PT} = .58$, $r_{CLA\ MA} = .67$), indicating that, to some extent, both SR and CR assessments measure a comparable construct, but that this relationship is far from perfect. Klein et al. (2009) attributed low correlations with open-ended measures of written communication as due, in part, to the low reliability of CR assessments with few items, noting that multiple essays would enhance test reliability. Others have suggested that good estimates of students' writing ability can be obtained by combining SR and CR formats (e.g., Breland, Camp, Jones, Morris, & Rock, 1987). This logic is evident in the designs of the CLA+ and CAAP, which each combine SR with extended CR item formats.

For the GRE, moderate correlations have been observed among GRE Analytical Writing tasks and the GRE Verbal section, with estimates ranging from $r = .51$ for the issue task and $r = .55$ for argument (Ramineni, Trapani, Williamson, Davey, & Bridgeman, 2012) to .66 overall (ETS, 2013b), suggesting that the CR Analytical Writing section measures skills that are related to, but somewhat distinct from, verbal reasoning skills. Further, scores from the Criterion Online Writing evaluation system showed moderate correlations with SAT writing in 2009 ($r = .43$) and 2010 ($r = .41$; Klobucar, Elliot, Deess, Rudniy, & Joshi, 2012). Thus, CR assessments achieve moderate evidence of convergent validity.

Concurrent Validity Evidence

Concurrent validity refers to the relationship between an outcome and a criterion measured at the same time (AERA et al., 1999). Evidence of concurrent validity has been evaluated for several assessments we reviewed, particularly by computing correlations among the assessment scores and other measures, such as ACT scores, SAT scores, or GPA. For example, EPP writing correlates with ACT scores ($r = 0.58$; Banta & Pike, 1989), while data sampled over a 10-year period shows that students with a higher college GPA consistently achieved higher EPP writing scores (Liu & Roohr, 2013).

For CR tests, the optional EPP essay correlates with both community college placement exams ($r = .51$) and SAT scores ($r = 0.27-0.37$; Liu, Bridgeman, & Adler, 2012). Compared to EPP, the CLA Performance Task demonstrates higher correlations with SAT (.56 and .54 for first-year and senior students, respectively; Klein, Benjamin, Shavelson, & Bolus, 2007). Total CLA scores have a school-level correlation with SAT of .90, while at the student level, CLA total has a moderate correlation with college GPA ($r = .50$, increasing to .65 when adjusted for reliability; Benjamin & Chun, 2003). Correlations between GRE Analytical Writing scores and college GPA are relatively low, ranging from $r = .13-.20$ (Powers, Fowles, & Welsh, 2001; Ramineni et al., 2012), with the highest correlations observed with GPA in writing-intensive courses ($r = .34$; Powers et al., 2001). In terms of placement tests, the COMPASS e-Write essay, which is scored using automated scoring techniques, correlates $r = .27$ with the total COMPASS test, $r = .29$ with ACT English, and $r = .21$ with ACT composite (Matzen & Sorensen, 2006).

Predictive Validity Evidence

Predictive validity concerns the extent to which outcomes such as college GPA can be predicted from the assessment scores. The predictive validity of the SR CAAP writing skills test was evaluated by examining the relationship between sophomore CAAP writing scores and junior-level GPA. Across seven institutions ($n = 1{,}514$), junior English GPA had a median correlation of .25 with sophomore CAAP writing skills scores (ACT, 2010). Further, the median cross-institutional correlation between sophomore CAAP writing skills and cumulative English GPA was .37, with a range of .26–.57 across a sample of eight postsecondary institutions (ACT, 2010). Thus, sophomore CAAP scores have modest predictive ability for junior-level GPA. For the EPP, Lakin, Elliott, and Liu (2012) observed a significant relationship between college credit hours and EPP Writing score ($r = .31$). Marr (1995) also reported significant Spearman rank correlations between EPP Writing and percent of total core college courses completed ($r = .19$) and core courses completed in humanities ($r = .07$), social science ($r = .06$), natural science ($r = .12$), and mathematics ($r = .12$).

While predictive validity evidence for CR format tests is more limited, the English Placement Test essay has been demonstrated to correlate $r = .35$ with English grades and $r = .21$ with fall semester GPA (Michael & Shaffer, 1978). For the GRE, Klieger, Cline, Holzman, Minsky, and Lorenz (in press) reported small but significant correlations ($r = .16$) between GRE Analytical Writing scores and graduate-school GPA for samples of more than 24,000 graduate students. The highest correlation of GRE-AW with GPA at the master's level was observed for English language and literature students ($r = .28$), suggesting that this test was most predictive for fields requiring considerable reading, critical analysis, and writing of texts.

For placement tests, predictive validity is evaluated in terms of placement accuracy, or the extent to which students are placed in a course of study in which they are likely to be successful. ACT Writing accurately placed 65% of students, with 66% of students earning a B or better (ACT, 2009). The ASSET Writing Skills test performs similarly, with Moss and Yeaton (2006) reporting that in college-level English classes 68% of students correctly placed in college English earn a B or better, compared to 54% of students who were initially sorted into developmental English courses. Placement accuracy rates for the COMPASS Writing Skills test range from 60% (Davey, Godwin, & Mittelholtz, 1997) to 66% earning a B or higher (ACT, 2006; Belfield & Crosta, 2012; Scott-Clayton, 2012). Placement accuracy rates are lower for the ACCUPLACER (59% earning a B or higher; Belfield & Crosta, 2012; Mattern & Packman, 2009).

Challenges in Designing Written Communication Assessment

Designing educational innovations involves negotiating a series of tradeoffs, which requires considering and making decisions to prioritize certain design aspects over others, which may be in tension with one another (cf. Collins, 1996). Designing assessments of written communication presents a number of specific challenges, which we describe below.

Balancing Authenticity and Psychometric Quality

Authenticity of writing assessment is considered a critical component of the validity of writing assessments, concerning both face and construct validity (Murphy & Yancey, 2008). Authenticity can be defined as the extent to which the features of an assessment task correspond to the features of the situations in which the skills being assessed will be used and applied in the real world (Bachman & Palmer, 1996). For higher education, the notion of authentic writing assessment suggests that the tasks included in the assessment design should correspond to the types of writing assignments required of students in their undergraduate coursework, such as writing arguments or research articles (Burstein et al., 2014). The notion of authentic writing assessment is consistent with the position statement on writing assessment released by the Conference on College Composition and Communication (CCCC, 2009), which asserts that best assessment practices ask students to produce writing within a meaningful context:

> The assessment of writing must strive to set up writing tasks and situations that identify purposes appropriate to and appealing to the particular students being tested. . . . What is easiest to measure—often by means of a multiple choice test—may correspond least to good writing; choosing a correct response from a set of possible answers is not composing. (Principles 2A and 2B)

Generally, CR format assessments are considered more authentic relative to tests consisting solely of discrete SR items, because they require students to compose extended text. However, some scholars argue that CR tasks are still not particularly authentic because they ask students to write about unfamiliar topics under highly constrained conditions. For example, Weigle (2002) argued that on-demand CR assessments "[do] not accurately reflect the conditions under which most writing is done in nontesting situations or writing as it is taught and practiced in the classroom" (p. 197).

A balance of authenticity and psychometric quality could be achieved through a combination of direct and indirect writing assessment (i.e., use of both SR and CR item formats; Breland et al., 1987). Providing students with a meaningful and realistic task context for writing an essay (e.g., to persuade the Board of Trustees to adopt a particular policy; to identify and explain to key stakeholders the critical flaws in a business proposal) offers a more authentic assessment task with a specific purpose, audience, and context for the writing task, consistent with the notion that all writing is fundamentally social (CCCC, 2009). The authenticity of SR items assessing skill in identifying and revising errors could be enhanced by presenting items in the context of an extended passage (versus discrete sentences) and a realistic task (e.g., attending to a peer's feedback on a passage; Haswell, 2008). Research on scenario-based assessments (Sabatini et al., 2013; Sheehan & O'Reilly, 2012) can inform the design of literacy assessments that have a balance of authentic purposes and desirable measurement properties.

Assessment Purposes: Supporting Institutional or Individual Goals

Members of the higher education writing community have suggested that assessment should primarily function to support evidence-based decision making intended to improve the teaching and learning of writing (CCCC, 2009; NCTE-WPA, 2010). Further, the intended purpose of a writing assessment should influence its design (CCCC, 2009). The desire for assessment results to provide actionable information to the institution in the service of improving teaching and learning suggests a need for alignment between the constructs measured in the assessment and the competencies that are relevant to the local curricular and instructional context. Alignment between instruction and assessment is also important for the measurement of student growth attributable to a curriculum or course of study; as Haswell (2008) noted, "The gain [from an intervention strategy] most often occurs when the classroom intervention is clear and concrete and when the measurement of writing accomplishment focuses analytically on traits associated with the teaching method" (p. 410). Thus, to have instructional value, the assessment results should inform institutions about the aspects of writing that pose challenges for their students, which could be addressed through instruction.

Some assessment formats may be more appropriate for supporting some institutional goals. For example, portfolio assessment represents an approach to evaluating student writing that is highly tailored to the local context (Behizadeh, 2014; Yancey, 1999), which may be quite useful for informing local curricular and instructional improvements. Institutions may also wish to make comparative evaluations of writing proficiency for groups of students across institutions, for purposes of benchmarking or accountability; this requires assessments that are not so locally defined that the test will fail to yield meaningful comparisons when administered to a different population of students, at schools using different curricula or instructional approaches. This goal suggests a relatively domain- and discipline-general approach to designing writing assessment such that assessment tasks should measure aspects of the construct that are practiced across a range of student majors and fields of study, so as not to advantage students from a specific curriculum or course of study. This logic is evident in SR assessments that measure the surface-level, linguistic aspects of writing skill, which may yield reliable comparisons yet have more limited instructional relevance, particularly for low-level editing skills that students are presumed to have mastered prior to enrolling in college.[8] Despite this presumption, usage and mechanics may still be considered instructionally relevant, given that college student populations are increasingly diverse, with many students enrolling in college courses, unprepared for the writing assignments that are required of them. Because institutions need assessments that provide actionable information about the strengths and weaknesses of highly diverse student groups, evaluating the linguistic aspects of students' writing remains an important goal. Therefore, to support the goals of comparability and instructional improvement across populations and institutions, writing assessments should provide evidence of students' proficiency with linguistic, as well as conceptual and rhetorical, aspects of the writing construct.

Beyond institutional goals, the extent to which a writing assessment is intended to support individual-level goals dictates the extent to which scores that are reliable at the student level are required. If the assessment is designed for institutional use only, the scores provided by the assessment need only be reliable at the group level (i.e., at the level of the institution), rather than reliable at the level of the individual student. For example, some SLO assessments are primarily designed to provide information at the aggregate group level and may not require highly reliable individual scores. Alternatively, placement tests, which have stakes in terms of the course of study an individual may pursue, must be reliable at the individual level, due to the potential consequences for the student's educational trajectory. Similarly, if the results will be used for credentialing, such as a certificate or badge, it is important that such certifications be reliable at the individual level—particularly if those credentials have consequences for educational attainment or employment.

Reporting Overall Scores Versus Subscale Scores

A related issue is the extent to which a single score can be used to represent students' writing proficiency and whether meaningful subscale scores can be reported to institutions or to examinees. From an institutional perspective, subscales can yield valuable information about the relative strengths and weaknesses among students' proficiency with particular aspects of written communication and whether proficiency varies as a function of students' major, years of college experience, and so on. Such information can be used to make improvements to curriculum and instruction. For the individual, subscale scores can provide useful feedback about the aspects of writing in which additional practice is needed. Thus, to support learning and instruction, provision of subscale scores may provide greater diagnostic information beyond overall scores. However, from a measurement perspective, it is only defensible to offer examinees subscale scores if these scores are reliable and valid. Haberman (2008) described methods for determining the added value of subscores relative to total test scores; these methods should be applied when determining whether or not subscores should be reported to examinees. In some cases, subscores do not add useful information to examinees; therefore, these subscores should not be reported (Sinharay & Haberman, 2008).

Subscale scores can be computed from SR assessments by having sufficient numbers of items assessing each skill of interest, such as rhetorical skills or mechanics and usage, such that reliable subscores can be reported for each skill (e.g., CAAP provides scores for these two dimensions). For CR assessments, subscale scores can be obtained by applying analytic or trait scoring. In contrast to holistic scoring, in which raters assign a single numerical score to the examinee based on an overall evaluation of the work, trait scoring requires raters to assign a numerical score for each of the qualities (or traits) that are important in the assessment, considered separately. For example, the CLA+ performance task scoring rubric asks raters to evaluate students' responses for three traits: analytic reasoning and problem solving, writing effectiveness, and writing conventions. For a given essay, a rater must provide three separate scores. Accordingly, trait scoring can provide more detailed, diagnostic information to examinees about their writing compared to a single, holistic score, which may not provide detailed information with respect to the writer's specific weaknesses but rather descriptions of the types of weaknesses commonly exhibited by responses receiving the same score (as in GRE Analytical Writing).

In a study comparing the use of trait and holistic scoring in the CUNY CATW, Faggen (2001) found that holistic scoring was somewhat more efficient than trait scoring, with raters divided as to which method they preferred. While some believed that trait scoring would provide more diagnostic information, evidence of high correlations among the traits suggested that trait scoring might not provide more detailed information than a holistic score, provided that the scoring criteria were comparable. Thus, holistic scoring is often preferred. In assessments with both SR and CR components, these test sections may either be reported separately, combined into a single, weighted proficiency score, or both.

Human Scoring Versus Automated Scoring

Beyond the issue of whether to report an overall proficiency score or multiple subscores is the issue of whether to employ automated scoring engines to support—or supplant—the use of human raters to score CRs. Recruiting and providing training and calibration to human raters is a time-consuming and often costly process, particularly in the case of large-scale assessments administered to thousands of students. Automated scoring engines offer two distinct advantages relative to human raters, in terms of reliability and cost. The scores provided by automated essay scoring systems are highly reliable (i.e., internally consistent), in that they apply an identical scoring algorithm each time an essay is scored; further, they demonstrate high correlations with human ratings (i.e., interrater reliability; Burstein & Chodorow, 2003; Chodorow & Burstein, 2004), often comparable to the agreement among two human raters. With respect to cost, automated essay scoring systems require little time per essay to apply the scoring model after model development has occurred, making the average cost to score one essay minimal compared to that of a human rater. These advantages have led automated scoring systems such as the ETS e-rater engine (Burstein & Marcu, 2003) to be used as a check score or second score in operational scoring of CR assessments (cf. Deane, 2013). For example, each GRE essay is scored by e-rater and at least one human rater, using a holistic scoring rubric with a 1–6 scale. If e-rater and the human rater agree within a certain threshold, the human rater's score is accepted as the final score; however, if the discrepancy between human and e-rater scores exceeds that threshold, a second human rater will score the essay, with the final score being the mean of the two human scores, rounded to the nearest half-point (ETS, 2013b). In contrast, systems like Criterion (Burstein, Chodorow, & Leacock, 2004), as well as WritePlacer and the optional EPP essay, use automated scores as the primary method of evaluating students' writing.

Although automated scoring has clear advantages, the decision to use such methods should take into consideration the validity of the test scores for a particular intended use. Critiques of the use of automated scoring methods alone (or in general) hinge on the notion that the features of text that can be feasibly evaluated using automated methods are not necessarily coincident with the features that correspond to good writing, including logical and accurate content (Condon, 2013; CCCC, 2009; Perelman, 2012). As noted by Deane (2013), writing is a complex skill, some aspects of which can be better captured by automated writing evaluation methods than others. Automated scoring methods rely on natural language processing techniques to detect and compute features of the text that are associated with higher quality writing. Many of these features are low level, such as nonstandard grammar, spelling, and punctuation, which are relatively easy for automated methods to detect, but automated scoring models also attempt to evaluate higher level features of writing quality. For example, e-rater is designed to measure both lower and higher level features, categorized under dimensions such as organization and development, vocabulary (i.e., word choice), grammar, usage, mechanics, and style (i.e., sentence variety). However, as Deane (2013) emphasized, concepts like *organization* and *development* as they are instantiated in an e-rater model are not interpreted in ways that humans would understand and apply these terms; development, for example, is largely a measure of length, rather than the quality of supporting ideas or examples. Research by Attali and Powers (2008) suggested that the text features measured by e-rater can be collapsed into three factors: fluency (including organization and development), accuracy of text production (i.e., skill in producing error-free text), and vocabulary sophistication (i.e., use of low-frequency vocabulary words). These factors do not correspond to the social and rhetorical elements of writing that are emphasized in the frameworks reviewed in the first section of this article.

Overall, automated writing evaluation systems seem to measure a restricted version of the construct, which excludes some critical communicative elements. Deane (2013) concluded that e-rater and other state-of-the-art scoring engines provide a measure of *text quality* based on surface linguistic features, rather than a measure of writing skill per se. With respect to using information from sources, previous research projects have developed effective methods for detecting use of explicit citations, plagiarism from sources, and other sourcing related issues (e.g., Britt et al., 2004; Deane, 2014; Hastings, Hughes, Magliano, Goldman, & Lawless, 2012), but this development often requires hand-coding of sources and training of prompt- and/or task-specific models in order to detect certain anticipated strings in student responses (e.g., according to Blake, as Carnegie writes; Britt et al., 2004). These model building efforts would be required for each prompt or test form, making them costly to develop. Further, these efforts have not yet effectively developed automated methods for detecting critical analysis and synthesis of sources, and other higher-order skills, such as argumentation and the accuracy of content, pose significant challenges as well (e.g., Powers et al., 2002a). Research in these areas is ongoing, but existing automated scoring models do not yet provide reliable, valid assessment of these aspects of writing sufficient for operational use.

To assess the features of students' writing that are important at the higher education level, it is likely that humans will be required to read, evaluate, and provide ratings of students' work with respect to a holistic or analytic rubric that takes into account these social and conceptual aspects, until automated scoring methods advance significantly. The CCCC (2009) asserted that best practice of writing assessment is to use direct assessment with human raters, particularly in the case of high-stakes assessment. While use of automated methods alone may be sufficient for a low-stakes assessment, we concur with the CCCC that the greater the assessment stakes, the more important it is to use human scoring. Because writing is fundamentally a social act, done to communicate meaning to an audience, it is important that a human reader evaluate the extent to which that communication successfully achieved the task goals.

An Operational Framework for Next-Generation Written Communication Assessment

Below, we outline a proposed operational framework to support the design of next-generation written communication assessments. We present our framework and construct definition, followed by a description of the structural features and task types that such an assessment might include. We then describe how the current framework compares with existing frameworks and assessments.

Proposed Framework and Definition

Informed by the preceding review and synthesis presented in the first and second parts of this article, the proposed operational framework and construct definition for written communication appears in Table 7. We have organized the construct definition for written communication into four major dimensions:

- *Knowledge of social and rhetorical situations*, which concerns the purpose-driven, social nature of all written communication, includes the ability to adapt one's writing to the demands of the specific context, audience, and purpose for writing; adherence to genre conventions, such as those for writing arguments or explanations; and skill in creating multimodal or multimedia products, using traditional and digital methods of production.
- *Domain knowledge and conceptual strategies*, which concerns the use of relevant content knowledge and development strategies, includes the ability to develop one's ideas using sufficient and effective reasons, evidence, and examples; presenting those ideas in an organized, logical, and coherent sequence; use of information drawn from sources to support one's ideas without distorting the author's original meaning; and adherence to disciplinary conventions, such as evidentiary or organizational standards.
- *Knowledge of language use and conventions*, which concerns the linguistic elements of writing, includes the ability to convey meaning clearly by using appropriate word choice, tone, and style, given the purpose of the writing, as well as the ability to produce relatively error-free text without substantial flaws in usage, syntax, and mechanics.
- *Knowledge of the writing process*, which cuts across the preceding social, conceptual, and linguistic dimensions, concerns the various strategies used to support prewriting or planning, drafting, and revision of text, as well as reading and appropriately responding to others' feedback.

Taken together, these dimensions represent a rather comprehensive view of written communication, spanning social and rhetorical, conceptual, and linguistic aspects of producing quality writing, including knowledge of the writing process (planning, drafting, and revision) as a major aspect of the framework. Importantly, the purpose-driven social and conceptual aspects of writing should be the primary focus of the assessment, in contrast to lower level language elements; further, information about students' proficiency with the writing process could provide useful feedback to both instructors and students in the service of supporting teaching and learning. However, these framework dimensions and corresponding definitions alone reveal little about how these aspects of writing will be assessed. Below, we propose a set of structural features and task types that may be used to evaluate these various aspects of written communication.

Table 7 Proposed Operational Definition and Construct for Written Communication

Construct dimension	Definition	Example assessments	Example rubric statements
Knowledge of social and rhetorical situations			
Task, context, and purpose	The ability to effectively consider and adapt one's writing to particular purposes (to inform, to argue, to persuade), contexts (academic, professional, social), and task instructions.	Any CR that asks test takers to respond to a specific task or prompt (e.g., "develop a position on the issue described, supported by reasons and examples"); essays that do not respond to the prompt are "off topic" and receive a 0. Explicitly, CUNY CATW (critical response to task and text is one trait dimension). Rhetorical skills CAAP items dealing with "strategy" assess purpose (e.g., "is X appropriate, given a particular purpose?").	CLEP College Composition: degree of focus on the assigned task; CATW: critical response to the writing task and the text; COMPASS Writing Skills: rhetorical skills (strategy: appropriateness of expression for audience and purpose, supporting material to strengthen writing, effective choice of theme or purpose statements); TOEFL iBT: appropriateness of the essay for topic and task (independent); GRE-AW: degree to which the writer addresses the specific task directions.
Audience awareness	The ability to effectively consider and adapt one's writing to particular audiences (e.g., experts, nonexperts, specialist, general).	Writing process items of College BASE (ask about whether X would be appropriate, given Y audience); CAAP, COMPASS E-Write, and College BASE each specify an audience for the essay task(s), for example, "Start your letter, Dear School Board:".	CAAP: supporting assertions appropriately for a given audience; CLA+: effectiveness of essay in persuading audience; COMPASS Writing Skills: rhetorical skills (strategy: appropriateness of expression for audience and purpose).
Genre conventions (text types/forms)	The ability to compose texts that adhere to conventions (formal and informal guidelines as to what is appropriate for a piece of writing) specific to the genre or type of writing (e.g., argument, exposition, essay, critique, summary). For higher education, writing arguments and research reports are common and valued genres.	Most CR tasks require argument writing or a critique, typically in the format of a 5-paragraph essay.CLA + Performance Task (argument), CAAP Essays, CLEP College Composition (argument), GRE issue (argument); MCAT writing (expository), TOEFL integrated (expository), CSU English Placement Test (source-based expository).	APELC: development of a position on the topic; CLA+: effectiveness of essay in persuading audience, identifying flaws in a specific argument.
Composing in multiple modes and forms	The ability to use a variety of technologies (pen and paper, digital software, online environments) to create written products, which may include multimedia elements, particularly when communicating complex information and ideas.	No assessments we reviewed tested this. Portfolio assessments;innovative item types (e.g., select an image that best illustrates your point).	n/a

Table 7 Continued

Construct dimension	Definition	Example assessments	Example rubric statements
Domain knowledge and conceptual strategies			
Disciplinary conventions (major/field)	The ability to compose texts that adhere to conventions (formal and informal guidelines as to what is appropriate for a piece of writing) specific to one's discipline or field of study; related to genre conventions. Includes conventions related to source attribution, content, tone, style, organization, and use of evidence, as appropriate given the discipline.	No assessments we reviewed tested this. Portfolio assessment	n/a
Content development and organization	The development and logical expression of ideas in writing. The ability to fully develop one's ideas with supporting information and examples from one's prior knowledge, reading, and experiences, and to present information and ideas in a logical, organized, and coherent way.	Any direct writing assessment; e-rater scores a limited version of this construct (i.e., development as sentence length, rather than quality or appropriateness of examples chosen to support a point, as a human might interpret it; organization can look for discourse elements such as thesis statement).	*Content development*. ACCUPLACER: elaboration of ideas and presentation of supporting details; AWPE: quality of reasons and examples; development and elaboration; APELC: quality of evidence and explanations in support of the position; CAAP: supporting assertions with appropriate evidence; CLA+: developing relevant support for a position; COMPASS Writing Essay: extent to which the topic is addressed by the development of ideas and the specificity of details and examples; CATW: development of ideas; CLEP College Composition: development of ideas and support; TOEFL iBT: development and support of ideas using examples or evidence; GRE-AW: the development of reasons and/or examples to support a position/analysis.

Table 7 Continued

Construct dimension	Definition	Example assessments	Example rubric statements
			Organization. ACCUPLACER: overall structure of response and sequence of ideas; AWPE: organizational structure; ASSET Writing Skills: organization; CAAP Essay: organizing and connecting ideas; CAAP SR: organization of ideas, relevance of statements (order, coherence, unity); CLA+: organizing arguments, organizing an essay and using transitions; COMPASS Writing Skills: rhetorical skills (organization: organization of ideas, relevance of statements [order, coherence, unity]); COMPASS Writing Essay: unity and coherence achieved through logical sequence of ideas; CATW: structure of the response (thesis and connection of ideas); CLEP College Composition: organization; TOEFL iBT: organization, unity, progression, and coherence; GRE-AW: organization (for both task types); MCAT: organization.
Use of sources and textual evidence	The ability to comprehend and critically analyze a source text (i.e., text, document, data table, image, etc.) and to effectively incorporate information drawn from source texts to develop and support one's ideas, using appropriate attribution.	CUNY CATW; CLA+ Performance Task; CLEP College Composition; CSU English Placement Test.	APELC: quality use of sources; CLA+: use and analysis of specific sources, use of sources to support decisions; CLEP College Composition: degree to which candidates synthesize two sources; CATW: critical response to the writing task and the text; TOEFL iBT: task-appropriate use of sources (integrated).

Table 7 Continued

Construct dimension	Definition	Example assessments	Example rubric statements
Knowledge of language use and conventions			
Language use: word choice, tone, voice, and style of language	The ability to compose text that conveys meaning clearly by using appropriate word choice, sentence variety, tone, voice, and style; what is appropriate will be determined by the context, purpose, and genre of writing.	Any direct writing assessment; e-rater scores this construct in terms of word choice (sophistication of vocabulary) and use of varied syntax; tone/voice cannot quite be assessed in the way that a human rater might interpret them.	ACCUPLACER: effectiveness of sentence constructions; AWPE: word choice, variety and complexity of sentence structure; APELC: control of the elements of effective writing; CAAP: expressing ideas clearly and effectively; CLA+: effectiveness of word choice; COMPASS Writing Skills: rhetorical skills (style: precision and appropriateness of word choice, effective management of sentence elements, avoidance of ambiguous pronoun references, economy in writing); COMPASS Writing Essay: effectiveness of style; CATW: sentence structure and word choice; GRE-AW: sentence structure; MCAT writing: control of vocabulary and sentence structure; EPP: organize units of language for coherence and rhetorical effect.
Language use: grammar, usage, syntax, and mechanics	The ability to compose text that is relatively free of errors in grammar, usage, mechanics, syntax, and spelling. Command of the fundamental skills needed to produce fluent text.	Any direct writing assessment; e-rater scores this construct, with most differentiation among low scorers. Revision-in-sentence-context items assess errors at the sentence level.	ACCUPLACER: control of language usage and mechanics; AWPE: control of the conventions of standard written English; CLA+ facility with the conventions of standard written English, control of grammar; COMPASS Writing Essay: control of the conventions of standard written English; CLEP College Composition: control of the conventions of standard written English; CATW: grammar, usage, and mechanics; TOEFL iBT: overall language facility, including grammar, usage, mechanics; GRE-AW: control of the conventions of standard written English; EPP: organize elements into larger units of meaning.

Table 7 Continued

Construct dimension	Definition	Example assessments	Example rubric statements
Knowledge of the writing process			
Writing processes (planning, drafting, and revision)	Strategic knowledge of the writing process, including prewriting strategies (idea generation, research), drafting, reviewing, revising, editing, and responding to others' feedback.	Revision-in-context items (revising, editing); rhetorical skills items (research, prewriting); keystroke logging (measures actual writing process, but not operational).	EPP: organize units of language for coherence and rhetorical effect, organize elements of writing into larger units of meaning. *Note that all assessments requiring revision-in-context type items, where examinees detect and correct errors in texts, would fall here.*

Note. CR = constructed response; CUNY CATW = City University of New York/CUNY Assessment Test in Writing; CAAP = Collegiate Assessment of Academic Proficiency; CLEP = College Level Examination Program; GRE-AW = Graduate Record Examinations Analytical Writing; College BASE = College Basic Academic Subjects Examination; CLA+ = Collegiate Learning Assessment; MCAT = Medical College Admission Test; CSU = California State University; APELC = Advanced Placement English Language and Composition; AWPE = Analytical Writing Placement Examination; SR = selected response; EPP = ETS Proficiency Profile.

Item and Task Types

As described in the second section of this article, existing assessments of written communication typically use CR items, though several couple these with SR items to compensate for the relatively low reliability of CR items. While many of the SR items included in these assessments are traditional discrete items (i.e., single-selection multiple choice [MC]; EPP), other assessments use groups of SR items associated with a particular passage (i.e., a set leader and set members; CAAP). Single-selection SR items presented in a passage context could also be administered as drop-down menus, drag-and-drop, or other more innovative technology-enhanced item types. While more complex technology-enhanced item types can sometimes require more effortful and time-consuming processing to read and make a response compared to basic MC response types (see Graf, 2009), they may provide evidence about an examinee's higher order reasoning skills as opposed to the passive recognition skills often elicited by traditional MC. Because some item types and tasks are more appropriate for measuring some aspects of the writing construct than others, we recommend using various item types to provide a more complete view of students' proficiency with writing, spanning the social and rhetorical, conceptual, and linguistic dimensions of this skill. For example, single-selection SR items may be better suited to measure the lower level linguistic aspects of writing, while drag-and-drop formats may be useful for assessing students' use of sources (e.g., by dragging an in-text citation to an appropriate location in a passage).

Structural Features of Items

Table 8 presents a proposed taxonomy of structural features for items assessing written communication, based on the framework and operational definition described above. Consistent with other widely used writing assessments, we propose the use of both CR items (i.e., direct writing assessment) and SR items (i.e., indirect writing assessment) to achieve a balance of authenticity and psychometric quality (i.e., reliability and validity). Beyond the typical single-selection SR items, we propose the use of more interactive structural features (e.g., drop-down menu, select in passage, drag-and-drop) where appropriate for measuring the intended construct (e.g., using drag-and-drop to add appropriate supporting evidence or citations to a stimulus passage). The use of technology-enhanced item types affords different kinds of measurement opportunities compared to traditional MC assessment. For example, such item types could be used to assess multimodal composition skills that cannot easily be assessed with typical CR item types, such as selecting an image or graph that best supports one's arguments and inserting it into a particular location in the text. Further, composing text on a computer can provide information about the writing process that cannot be captured with traditional CR items (i.e., pencil and paper). The use of technology-enhanced items also makes the assessment experience more dynamic and potentially more engaging to students, which can provide more robust, valid information about their abilities.

Task Types

The specific nature of the assessment task(s) is also an important consideration for assessment design. The structural features described above could be used to support several task types that we consider promising for measuring the aspects of written communication defined in the current framework. Table 9 presents descriptions for several CR and SR assessment task types, with their correspondence to the operational framework, and examples of similar assessments.

As consistent with best practices in writing assessment (CCCC, 2009), CR should be preferred when possible, because these item types permit direct assessment of multiple aspects of students' writing simultaneously and directly, while SR items tend to target a specific aspect of the construct, such as organization or syntactical errors, and then only indirectly. Importantly, with respect to use of SR items, we do not advocate the use of discrete, sentence-level traditional MC items for an assessment of written communication at the higher education level. Low-level items such as these do not represent the skills and competencies that are required of real-world writers, who work with ideas in the context of extended discourse rather than discrete and isolated sentences. Therefore, if these linguistic-level skills are to be assessed, they should be done so in the context of extended written discourse, which examinees must either read, respond to, and make revisions to—or which they produce themselves in a CR task. In addition, to the extent possible, the tasks should be introduced in such a way that they represent an authentic context and purpose for writing, with a specified audience to be addressed. For example, a set of revision-in-passage-context items could be framed as a peer-editing task or as responding to feedback from an instructor rather than an abstract task done solely for the purpose of taking the

Table 8 Descriptions of Structural Features of Written Communication Items

Item type	Description
CR prompt	Writing-based CR item, in which examinees compose an open-ended response to a prompt, which may or may not include source texts.
Set leader	Stimulus (i.e., passage) for which there are one or more items (set members) that are based on the stimulus content.
Set member, single selection MC	Stem with multiple answer choices, of which one is the correct response; displayed along with set leader for reference. Revision-in-passage context items follow this format (e.g., CAAP).
Single selection MC	A stem with multiple answer choices, of which one could be a correct response.
Multiple selection MC	A stem with multiple answer choices, of which two or more could be a correct response.
Drop-down menu	A variation of a traditional MC item, where one answer choice is selected via a drop-down menu.
Select in passage (single selection/multiple selection)	Item where the answer choices are predefined set of words, phrases, sentences, or paragraphs within a set leader. When test taker clicks on selection, the word/sentence is highlighted in the passage. If only one answer, use *Select in Passage SS*. If two or more answers, use *Select in Passage MS*.
Drag-and-drop	An examinee selects objects (i.e., text segments, citations, or images) and places them in a specific location or order within a text.

Note. CR = constructed response: MC = multiple choice (MC); CAAP = Collegiate Assessment of Academic Proficiency.

assessment. Such tasks may be more reflective of the real-world settings in which college-level writers engage in the practice of writing.

Unique Features of this Framework

The framework and construct definition presented in this article are informed by current research on writing and writing instruction, which views learning to write as a process of socialization into a particular set of practices for achieving particular social and rhetorical goals (e.g., presenting a scientific argument or advancing a particular historical or literary interpretation), and by current higher education frameworks, which recognize that the construct of writing must be updated to reflect the place of written communication in a contemporary social and technological context. The abilities to produce multimedia compositions, to synthesize information from a wide variety of information sources, and to convey complex information effectively and succinctly are increasingly important for success in both academic and workforce domains in the 21st century. Consistent with the developmental competency model of literacy that underlies the design of CBAL assessments in K–12 (Deane, 2011; Deane et al., 2011; Sabatini et al., 2013), we conceptualize written communication as involving the coordinated recruitment of social, conceptual, and linguistic (i.e., discourse, verbal, and print) representations, on which the writer's cognitive processes operate. Fluency with lower level linguistic processes frees up cognitive resources for engaging in the conceptual and social aspects of the writing. We include rhetorical aspects of writing in the social dimension, as rhetorical considerations are a part of the social and communicative goals of writing. By addressing social, conceptual, linguistic, and process-level dimensions of writing, we present a comprehensive operational framework that can be used to evaluate existing assessments and to support the development of new assessments.

We have included knowledge of composing in multiple modes and forms (including use of technological tools to compose text) under the social and rhetorical dimension of the writing construct, and this represents a unique feature of the current assessment framework. It is important to note that while the vast majority of frameworks reviewed mentioned this skill as important for higher education, particularly in the 21st century, none of the assessments of written communication we reviewed made any attempt to provide evidence of students' proficiency with creating multimedia compositions or using technology-enhanced composition methods. Skill in composing multiple different types or forms of text (including multimedia, PowerPoint presentations, etc.) as a writing outcome is typically assessed through portfolio assessment methods, if at all. Most assessments of writing skill do not evaluate this dimension of student writing explicitly. Specific information about student proficiency with multimedia composition skills might also be provided by an assessment of another, related construct, such as digital or information and communication technology (ICT) literacy, which concerns

Table 9 Descriptions of Task Types of Written Communication Items

Task type	Description	Example task	Dimensions assessed
Make an argument	CR prompt that presents an issue, opinion, or hypothetical situation and requires the examinee to take a position on the issue, use reasons and evidence to support the position, and organize the response in a logical and coherent manner with clear and effective language use.	GRE issue; CAAP Essay; EPP Essay	Task/Context/Purpose/Audience,[a] Genre (Argument), Development/Organization,Word Choice & Style, Grammar/Usage/ Mechanics, Process (Drafting)
Critique an argument	CR prompt that presents a flawed and problematic line of reasoning about a fictional situation and requires the examinee to analyze the argument with respect to task instructions, writing a response that presents a high-quality analysis, with sufficient development and support for one's ideas, and organizes the response in a logical and coherent manner with effective language use.	GRE argument; CLA Critique an argument	Task/Context/Purpose/Audience,[a] Genre (Argument), Use of Sources (Evaluation only)[b], Development/Organization,Word Choice & Style, Grammar/Usage/ Mechanics, Process (Drafting)
Source-based synthesis	CR prompt that presents examinees with one or more stimulus passages that serve as the basis for one's response; examinees should summarize or explain information from the passage, incorporate paraphrases or quotes from sources to support one's ideas, and include appropriate citations and references to the source texts. Examinees may also be asked to critically evaluate the ideas in the source text(s).	CLA + Performance Task; CLEP College Composition; CUNY CATW	Task/Context/Purpose/Audience, Genre (Argument or Expository), Use of Sources, Development/Organization, Word Choice & Style, Grammar/Usage/ Mechanics, Process (Drafting)
Revision in passage context	SR item types that present examinees with a stimulus passage containing errors at the level of organization, word choice and style, or usage and mechanics, and ask examinees to identify the most appropriate revision to the error or indicate that no change is needed. These errors should often extend beyond a single-sentence context.	CAAP Writing; COMPASS Writing Skills	Development/Organization,Word Choice & Style, Grammar/Usage/ Mechanics, Process (Revision)
Selected-response in passage context	SR item types that present examinees with a stimulus passage accompanied with questions that span longer segments of text (i.e., paragraph to whole passage) but do not require the examinee to detect and correct an error. This task type could be used to evaluate aspects of writing that are not effectively assessed by revision-in-context items (e.g., use of supporting evidence, audience awareness).	CAAP Writing; COMPASS Writing Skills (i.e., rhetorical strategy items spanning the entire passage)	Task/Context/Purpose/Audience, Genre (Argument or Expository), Use of Sources, Development/Organization

Note. CR = constructed response; SR = selected response; GRE = Graduate Record Examinations; CAAP = Collegiate Assessment of Academic Proficiency; EPP = ETS Proficiency Profile; CLA = Collegiate Learning Assessment; CLEP = College Level Examination Program; CUNY CATW = City University of New York/CUNY Assessment Test in Writing.
[a]The current GRE Analytical Writing measure does not require examinees to address their response to a specific audience, so the audience awareness dimension is not evaluated by this assessment. [b]Argument critique tasks do not typically require examinees to summarize, paraphrase, or quote from the stimulus prompt, nor do they require examinees to cite sources. Therefore, this task does not fully correspond to the definition of use of sources and textual evidence as defined in the current framework.

the extent to which students can use technological tools to compose multimodal communication, such as writing an e-mail to a colleague explaining data displayed in a graph (e.g., Katz & Macklin, 2007).

We also include use of sources and adherence to the conventions of argument and expository genres, which are particularly critical skills at the higher education level, yet are not explicit components of many assessment frameworks. Similar to use of sources, disciplinary considerations are considered a part of the conceptual aspects of writing, because they directly affect the writer's interaction with the content; however, our goal is to design a writing assessment that can inform the evaluation of general student learning outcomes across curricular or disciplinary boundaries, and, thus, our proposed operational definition and task types do not directly address this aspect of the framework. Rather, adherence to disciplinary conventions could be assessed locally, within a particular school or department, through some form of portfolio assessment, if disciplinary writing assessment is sought beyond typical classroom assessment practices.

In sum, the proposed framework offers several advantages, which support its use for developing written communication assessments at the higher education level. The framework captures multiple dimensions of writing, informed by a review of extant frameworks and literature from the learning sciences and the higher education writing community. It affords the use of multiple assessment formats, including extended CRs, traditional SR items, and more innovative item types. The use of technology-enhanced item types as proposed here has the potential to provide more robust measurement of student proficiency by obtaining evidence of skills that are difficult to measure with traditional methods and by potentially enhancing student engagement in the assessment experience. Such item types have been developed and administered in the context of assessing the language skills of English learners; these designs could be adapted for use in measuring undergraduates' proficiency with college-level writing tasks. Further, combining a direct writing assessment with multiple indirect items designed to assess aspects of the construct that are not covered by the specific CR prompt can provide a balance of authenticity and technical quality. This framework can also support the design of assessments that are reliable at the group or at the student level, depending on the intended purpose of the assessment, though the specific degree of reliability obtained is an empirical question, to be revealed through pilot testing.

Conclusions

Written communication has been identified as one of the most important learning outcomes among higher education institutions, as well as employers. Frameworks from higher education, educational institutions, national associations, the workforce, K–12 standards, and the research literature have each offered definitions of proficiency with written communication. At the higher education level, in particular, writing should involve critical and reflective engagement with others' ideas, development and support of one's own ideas, skill in producing compelling arguments directed to an audience, and fluency with producing coherent and logical written text that is free of errors. The operational definition proposed in the current article emphasizes the intersection of social, conceptual, and linguistic processes in the writing process, providing a comprehensive view of what skilled written communication involves, which can be used to obtain more complete evidence of students' proficiency with various aspects of writing. This framework aligns with current writing assessments but extends beyond current offerings by emphasizing the authentic social contexts and tasks in which real-world written communication skills will be deployed.

Acknowledgments

The first author's contribution to this work was supported in part by an American Educational Research Association Postdoctoral Fellowship in Educational Measurement at Educational Testing Service. The authors are grateful to Doug Baldwin, Paul Deane, and Donald Powers for their thoughtful comments on the manuscript; to Jennifer Bochenek for research assistance; and to the respective members of the ETS R&D and Automated Scoring Technical Advisory Committees for their feedback.

Notes

1 Note that not all of the frameworks provided an explicit definition of writing or written communication; therefore, in some cases, a definition of the targeted construct was inferred from the statements describing the desired student learning outcomes (e.g., rubric statements) relevant for a particular aspect of writing skill.

2 Related competencies appear in other frameworks under the heading of information literacy or critical thinking, which deal with evaluating the relevance, reliability, and credibility of various information sources and using those sources to make and defend arguments, develop solutions to problems, and so forth.

3 Second-language communication was mentioned across several frameworks (ATC21S, LEAP, BOLOGNA, DQP), but we do not deal with this issue in detail as second-language learning is outside the scope of the written communication per se.

4 The linguistic aspects of literacy can be further decomposed into discourse, verbal, and print levels of representation. The discourse representation includes information about text structure, organization, and the situation being described in the text (i.e., a situation model of the text). The verbal level of representation includes information about the meaning and usage of words (i.e., vocabulary knowledge). The print level includes representations of print conventions (i.e., knowledge of spelling, morphology, and phonology). Facility with print, verbal, and discourse-level representations is required for skillful command of the linguistic aspects of writing.

5 The CLA+ now includes an SR section, but these items assess students' skill in scientific literacy, critical analysis and evaluation of sources, and critiquing arguments, rather than writing skill.

6 Of course, compared to SR tests, CR format assessments often have lower reliability due to other reasons, such as failure to achieve high reliability in CR scoring, which are separate from concerns about test length per se.

7 The GRE revised General Test was implemented after August 1, 2011.

8 The presence of such items on college placement tests such as the ACCUPLACER and COMPASS which are used to determine whether students demonstrate readiness for college-level writing instruction or require remediation through developmental coursework, is consistent with this notion.

References

ACT, Inc. (2006). *COMPASS/ESL reference manual*. Iowa City, IA: Author.

ACT, Inc. (2009). *ACT writing test technical report*. Iowa City, IA: Author.

ACT, Inc. (2010). *CAAP technical handbook: 2011–2012*. Iowa City, IA: Author.

Adelman, C., Ewell, P., Gaston, P., & Schneider, C. G. (2011). *The degree qualifications profile*. Indianapolis, IN: Lumina Foundation.

Alvermann, D. E. (2002). Effective literacy instruction for adolescents. *Journal of Literacy Research*, *34*(2), 189–208.

American Educational Research Association, American Psychological Association, & National Council for Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.

Anson, C. M., Schwiebert, J. E., & Williamson, M. M. (1993). *Writing across the curriculum: An annotated bibliography*. Westport, CT: Greenwood.

Association of American Colleges and Universities. (2011). *The LEAP vision for learning: Outcomes, practices, impact, and employers' view*. Washington, DC: Author.

Attali, Y., & Powers, D. (2008). *A developmental writing scale* (Research Report No. RR-08-19). Princeton, NJ: Educational Testing Service. doi: 10.1002/j.2333-8504.2008.tb02105.x

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, England: Oxford University Press.

Banta, T. W., & Pike, G. R. (1989). Methods for comparing outcomes assessment instruments. *Research in Higher Education*, *30*, 455–470.

Barab, S. A., & Duffy, T. M. (2000). From practice fields to communities of practice. In D. H. Jonassen & S. M. Land (Eds.), *Theoretical foundations of learning environments* (pp. 25–55). Mahwah, NJ: Lawrence Erlbaum.

Bazerman, C. (1985). Physicists reading physics: Schema-laden purposes and purpose-laden schema. *Written Communication*, *2*, 3–23.

Bazerman, C. (2004). Speech acts, genres, and activity systems: How texts organize activity and people. In C. Bazerman & P. Prior (Eds.), *What writing does and how it does it* (pp. 309–339). Mahwah, NJ: Lawrence Erlbaum.

Behizadeh, N. (2014). Mitigating the dangers of a single story: Creating large-scale writing assessments aligned with socio-cultural theory. *Educational Researcher*, *43*, 125–136.

Behizadeh, N., & Engelhard, G. (2011). Historical view of the influences of measurement and writing theories on the practice of writing assessment in the United States. *Assessing Writing*, *16*, 189–211.

Belfield, C., & Crosta, P. M. (2012). *Predicting success in college: The importance of placement tests and high school transcripts* (Community College Research Center Working Paper 42). New York, NY: Teachers College, Columbia University.

Benjamin, R., & Chun, M. (2003). A new field of dreams: The Collegiate Learning Assessment project. *Peer Review*, *5*(4), 26–29.

Bennett, R. E. (2010). Cognitively Based Assessment of, for, and as Learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research & Perspective*, *8*(2-3), 70–91.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., & Rumble, M. (2010). Defining 21st century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17–66). New York, NY: Springer Science and Business Media B.V.

Bråten, I., Strømsø, H. I., & Britt, M. A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly*, *44*(1), 6–28.

Breland, H., Camp, R., Jones, R., Morris, M., & Rock, D. (1987). *Assessing writing skill*. New York, NY: College Entrance Examination Board.

Britt, M. A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction*, *20*, 485–522.

Britt, M. A., Wiemer-Hastings, P., Larson, A. A., & Perfetti, C. A. (2004). Using intelligent feedback to improve sourcing and integration in students' essays. *International Journal of Artificial Intelligence in Education*, *14*(3), 359–374.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32–42.

Burstein, J., & Chodorow, M. (2003). Directions in automated essay scoring. In R. Kaplan (Ed.), *Handbook of applied linguistics*. New York, NY: Oxford University Press.

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The Criterion online service. *AI Magazine*, *25*(3), 27–36.

Burstein, J., Elliott, N., Holtzman, S., Lentini, J., Molloy, H., Shore, J. Steinberg, J., & Vezzu, M. (2014). *Genre research and automated writing evaluation: Using the lens of genre to understand exposure and readiness in teaching and assessing school and workplace writing.* Paper presented at the annual meeting of the National Council on Educational Measurement, Philadelphia, PA.

Burstein, J. C., Kukich, K., Wolff, S., Lu, C., & Chodorow, M. (1998, April). *Computer analysis of essays.* Paper presented at the annual meeting of the National Council of Measurement in Education, San Diego, CA.

Burstein, J., & Marcu, D. (2003). Automated evaluation of discourse structure in student essays. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 209–229). Mahwah, NJ: Routledge.

CAAP Program Management. (2012). *ACT CAAP technical handbook 2011–2012*. Retrieved from http://www.act.org/caap/pdf/CAAP-TechnicalHandbook.pdf

Carroll, L. A. (2002). *Rehearsing new roles: How college students develop as writers (Studies in writing and rhetoric)*. Carbondale: Southern Illinois University Press.

Casner-Lotto, J., & Barrington, L. (2006). *Are they really ready to work?* Washington, DC: Partnership for 21st Century Skills.

Chodorow, M., & Burstein, J. (2004). *Beyond essay length: Evaluating e-rater's performance on TOEFL essays* (TOEFL Research Report No. RR-73). Princeton, NJ: Educational Testing Service.

City University of New York. (2012). *CUNY assessment test in writing (CATW): Information for students*. Retrieved from http://www.cuny.edu/academics/testing/cuny-assessment-tests/faqs/CATWInformationforStudentsandpracticeweb.pdf

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. *Language and Comprehension*, *9*, 287–299.

Cole, J. S., & Osterlind, S. J. (2008). Investigating differences between low- and high-stakes test performance on a general education exam. *The Journal of General Education*, *57*, 119–130.

Collins, A. (1996). Design issues for learning environments. In S. Vosniadou (Ed.), *International perspectives on the design of technology-supported learning environments* (pp. 347–361). Mahwah, NJ: Lawrence Erlbaum.

Condon, W. (2013). Large-scale assessment, locally developed measures, and automated scoring of essays: Fishing for red herrings? *Assessing Writing*, *18*(1), 100–108.

Conference on College Composition and Communication Committee on Assessment. (2009). *Writing assessment: A position statement.* Retrieved from http://www.ncte.org/cccc/resources/positions/writingassessment

Cope, B., & Kalantzis, M. (Eds.) (2000). *Multiliteracies: Literacy learning and the design of social futures*. New York, NY: Routledge.

Council for the Advancement of Standards in Education. (2006). *CAS professional standards for higher education* (6th ed.). Washington, DC: Author.

Council for the Advancement of Standards in Education. (2009). CAS learning and development outcomes. In Council for the Advancement of Higher Education (Ed.), *CAS professional standards for higher education* (7th ed.). Washington, DC: Author. Retrieved from http://standards.cas.edu/getpdf.cfm?PDF=D87A29DC-D1D6-D014-83AA8667902C480B

Council for Aid to Education. (2013). *CLA+ overview*. Retrieved from http://cae.org/performance-assessment/category/cla-overview/

Council of Presidents and State Board for Community College Education. (1989). *The validity and usefulness of three nationally standardized tests for Washington college sophomores: General report*. Bellingham, WA: Western Washington University Office of Publications.

Council of Writing Program Administrators, National Council of Teachers of English, & National Writing Project. (2011). *Framework for success in postsecondary writing.* Retrieved from http://wpacouncil.org/files/framework-for-success-postsecondary-writing.pdf

Cumming, A., Kantor, R., Powers, D., Santos, T., & Taylor, C. (2000). *TOEFL 2000 writing framework: A working paper* (Research Memorandum No. RM-00-05). Princeton, NJ: Educational Testing Service.

Davey, T., Godwin, J., & Mittelholtz, D. (1997). Developing and scoring an innovative computerized writing assessment. *Journal of Educational Measurement*, *34*(1), 21–41.

Deane, P. (2011). *Writing assessment and cognition* (Research Report No. RR-11-14). Princeton, NJ: Educational Testing Service. doi: 10.1002/j.2333-8504.2011.tb02250.x

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, *18*(1), 7–24.

Deane, P. (2014). *Using writing process and product features to assess writing quality and explore how those features relate to other literacy tasks* (Research Report No. RR-14-03). Princeton, NJ: Educational Testing Service. doi: 10.1002/ets2.12002

Deane, P., Sabatini, J., Feng, G., Sparks, J. R., Song, Y., Fowles, M. E., . . . Foley, C. (in press). *Key practices in the English language arts: Linking learning theory, assessment, and instruction.* (Research Report Series). Princeton, NJ: Educational Testing Service.

Deane, P., Sabatini, J., & O'Reilly, T. (2011). *English language arts literacy framework*. Princeton, NJ: Educational Testing Service.

Dewey, J. (1938/1997). *Experience and education*. New York, NY: Simon & Schuster.

Dodd, B. G., Fitzpatrick, S. J., De Ayala, R. J., & Jennings, J. A. (2002). *An investigation of the validity of AP® grades of 3 and a comparison of AP and non-AP student groups* (Research Report No. 2002-9). New York, NY: College Entrance Examination Board.

Educational Testing Service. (2010a). *ETS proficiency profile user's guide*. Princeton, NJ: Author.

Educational Testing Service. (2010b). *Graduate Record Examinations guide to the use of scores 2010–2011*. Retrieved from http://www.ets.org/s/gre/pdf/2010-11_gre_guide.pdf

Educational Testing Service. (2011). Reliability and comparability of TOEFL iBT scores. *TOEFL iBT Research Insight*, *1*(3), 1–8.

Educational Testing Service. (2013a). *Quantitative market research* [PowerPoint slides]. Princeton, NJ: Author.

Educational Testing Service. (2013b). *GRE guide to the use of scores 2013-2014.* Retrieved from http://www.ets.org/s/gre/pdf/gre_guide.pdf

European Higher Education Area. (2005). *The framework of qualifications for the European Higher Education Area*. Retrieved from http://www.ehea.info/Uploads/QF/050520_Framework_qualifications.pdf

Faggen, J. (2001). *Three research studies based on Task I of the CUNY Proficiency Examination.* Unpublished manuscript, Educational Testing Service, Princeton, NJ.

Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication*, *32*(4), 400–414.

Fitzgerald, J. (1987). Research on revision in writing. *Review of Educational Research*, *57*, 481–506.

Fitzgerald, J., & Markman, L. (1987). Teaching children about revision in writing. *Cognition and Instruction*, *4*, 3–24.

Flowers, L., Osterlind, S., Pascarella, E., & Pierson, C. T. (2001). How much do students learn in college? Cross-sectional estimates using the College BASE. *Journal of Higher Education*, *72*(5), 565–583.

Fowles, M. (2012). Writing assessment for graduate and professional programs: Lessons learned and a note for the future. In N. Elliot, & L. Perelman (Eds.), *Writing Assessment in the 21st century* (pp. 135–146). New York, NY: Hampton Press.

Gao, R. (2003). *Construct validity of College Basic Academic Subject Examination*. Dissertation Abstracts International Section A, 64.

Godshalk, F. I., Swineford, F., & Coffman, W. E. (1966). *The measurement of writing ability*. New York, NY: College Entrance Examination Board.

Goldman, S. R. (2004). Cognitive aspects of constructing meaning through and across multiple texts. In N. Shuart-Ferris, & D. M. Bloome (Eds.), *Uses of intertextuality in classroom and educational research* (pp. 313–347). Greenwich, CT: Information Age.

Goldman, S. R., Lawless, K. A., Gomez, K. W., Braasch, J. L. G., Macleod, S., & Manning, F. (2010). Literacy in the digital world: Comprehending and learning from multiple sources. In M. G. McKeown & L. Kucan (Eds.), *Bringing reading research to life* (pp. 257–284). New York, NY: Guilford Press.

González, J., & Wagenaar, R. (Eds.). (2003). *Tuning educational structures in Europe: Final report phase one*. Bilbao, Spain: University of Deusto.

Gordon Commission. (2013). *To assess, to teach, to learn: A vision for the future of assessment.* Retrieved from http://www.gordoncommission.org/rsc/pdfs/gordon_commission_technical_report.pdf

Graf, E. A. (2009). *Defining mathematics competency in the service of cognitively based assessment for grades 3 through 8* (Research Report No. RR-09-42). doi: 10.1002/j.2333-8504.2009.tb02199.x

Graham, S., & Harris, K. R. (1997). Self-regulation and writing: Where do we go from here? *Contemporary Educational Psychology*, *22*(1), 102–114.

Graham, S., & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools*. New York, NY: Carnegie Corporation.

Graves, D. H., & Murray, D. M. (1980). Revision: In the writer's workshop and in the classroom. *Journal of Education*, *162*, 38–56.

Haberman, S. J. (2008). *Subscores and validity* (Research Report No. RR-08-64). Princeton, NJ: Educational Testing Service. doi: 10.1002/j.2333-8504.2008.tb02150.x

Hamp-Lyons, L., & Condon, W. (2000). *Assessing the portfolio: Principles for practice, theory, and research*. Cresskill, NJ: Hampton Press.

Hastings, P., Hughes, S., Magliano, J. P., Goldman, S. R., & Lawless, K. (2012). Assessing the use of multiple sources in student essays. *Behavior Research Methods*, *44*(3), 622–633.

Haswell, R. H. (1986). Change in undergraduate and post-graduate writing performance [part 1]: Quantified findings. *ERIC Clearinghouse on Reading and Communication Skills* (ED 269 780). Retrieved from http://files.eric.ed.gov/fulltext/ED269780.pdf

Haswell, R. H. (1991). *Gaining ground in college writing: Tales of development and interpretation*. Dallas, TX: Southern Methodist University Press.

Haswell, R. H. (1998). Multiple inquiry in the validation of writing tests. *Assessing Writing*, *5*(1), 89–109.

Haswell, R. H. (2000). Documenting improvement in college writing: A longitudinal approach. *Written Communication*, *17*(3), 307–352.

Haswell, R. H. (2008). Teaching of writing in higher education. In C. Bazerman (Ed.), *Handbook of research on writing* (pp. 405–424). New York, NY: Lawrence Erlbaum.

Hayes, J. R. (1996). A new model of cognition and affect in writing. In M. Levy & S. Ransdell (Eds.), *The science of writing*. Hillsdale, NJ: Erlbaum.

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Hillsdale, NJ: Lawrence Erlbaum.

Hayes, J. R., & Flower, L. S. (1986). Writing research and the writer. *American Psychologist*, *41*, 1106–1113.

Hillocks, G. (1987). Synthesis of research on teaching writing. *Educational Leadership*, *44*(8), 71–82.

Hillocks, G. (2005). At last: The focus on form vs. content in teaching writing. *Research in the Teaching of English*, *40*(2), 238–248.

Holland, V. M., Rose, A. M., Dean, R. A., & Dory, S. L. (1985). *Processes involved in writing effective procedural instructions* (Technical Report No. 1). Washington, DC: American Institutes for Research.

Hughes, R. E., & Nelson, C. H. (1991). Placement scores and placement practices: An empirical analysis. *Community College Review*, *19*(1), 42–48.

Hunt, K. W. (1970). Syntactic maturity in school children and adults. *Monographs of the Society for Research in Child Development*, *35*(1), 1–67.

James, C. L. (2006). Validating a computerized scoring system for assessing writing and placing students in composition courses. *Assessing Writing*, *11*(3), 167–178.

Kass, D., Grandzol, C., & Bommer, W. (2012). The GMAT as a predictor of MBA performance: Less success than meets the eye. *Journal of Education for Business*, *87*(5), 290–295.

Katz, I. R., & Macklin, A. S. (2007). Information and communication technology (ICT) literacy: Integration and assessment in higher education. *Journal of Systemics, Cybernetics and Informatics*, *5*(4), 50–55.

Kinsler, K. (1990). Structured peer collaboration: Teaching essay revision to college students needing writing remediation. *Cognition and Instruction*, *7*(4), 303–321.

Klein, S., Benjamin, R., Shavelson, R., & Bolus, R. (2007). The Collegiate Learning Assessment facts and fantasies. *Evaluation Review*, *31*(5), 415–439.

Klein, S., Liu, O. L., Sconing, J. A., Bolus, R., Bridgeman, B., Kugelmass, H., . . . Steedle, J. (2009). *Test validity study (TVS) report* [Technical report]. Princeton, NJ: Educational Testing Service.

Klieger, D. M., Cline, F. A., Holzman, S. L., Minsky, J. L., & Lorenz, F. (2014). *New perspectives on the validity of the GRE General Test for predicting graduate school grades* (Research Report No. GRE-14-03). Princeton, NJ: Educational Testing Service.

Klobucar, A., Elliot, N., Deess, P., Rudniy, O., & Joshi, K. (2012). Automated scoring in context: Rapid assessment for placed students. *Assessing Writing*, *18*, 62–84.

Krahn, H., & Silzer, B. J. (1995). A study of exit surveys: The Graduand Survey at the University of Alberta. *College and University*, *71*(1), 12–23.

Kuh, G. D., Jankowski, N., Ikenberry, S. O., & Kinzie, J. (2014). *Knowing what students know and can do: The current state of student learning outcomes assessment in U.S. colleges and universities*. Champaign, IL: National Institute for Learning Outcomes Assessment.

Kutner, M., Greenberg, E., & Baer, J. (2006). *A first look at the literacy of America's adults in the 21st century* (Report No. NCES 2006-470). Washington, DC: National Center for Education Statistics.

Lakin, J. M., Elliott, D. C., & Liu, O. L. (2012). Investigating ESL students' performance on outcomes assessments in higher education. *Educational and Psychological Measurement*, *72*(5), 734–753.

Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. New York, NY: Princeton University Press.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York, NY: Cambridge University Press.

Liu, O. L. (2008). *Measuring learning outcomes in higher education using the Measure of Academic Proficiency and Progress (MAPP)* (Research Report No. RR-08-47). Princeton, NJ: Educational Testing Service. doi: 10.1002/j.2333-8504.2008.tb02133.x

Liu, O. L., Bridgeman, B., & Adler, R. M. (2012). Measuring learning outcomes in higher education: Motivation matters. *Education Researcher*, *41*(9), 352–362.

Liu, O. L., & Roohr, K. C. (2013). *Investigating 10-year trends of learning outcomes at community colleges* (Research Report No. RR-13-34). Princeton, NJ: Educational Testing Service. doi: 10.1002/j.2333-8504.2013.tb02341.x

MacArthur, C. A., Schwartz, S., & Graham, S. (1991). Effects of a reciprocal peer revision strategy in special education classrooms. *Learning Disabilities Research and Practice*, *6*, 201–210.

Markle, R., Brenneman, M., Jackson, T., Burrus, J., & Robbins, S. (2013). *Synthesizing frameworks of higher education student learning outcomes* (Research Report No. RR-13-22). Princeton, NJ: Educational Testing Service. doi: 10.1002/j.2333-8504.2013.tb02329.x

Marr, D. (1995). *Validity of the academic profile*. Princeton, NJ: Educational Testing Service.

Mattern, K. D., & Packman, S. (2009). *Predictive validity of ACCUPLACER scores for course placement: A meta-analysis* (Research Report No. 2009-2). New York, NY: College Board.

Matzen, R. N., & Hoyt, J. E. (2004). Basic writing placement with holistically scored essays: Research evidence. *Journal of Developmental Education*, *28*, 2–13.

Matzen, R. N. Jr., & Sorensen, C. (2006). E-write as a means for placement into three composition courses: A pilot study. In P. F. Ericsson & R. H. Haswell (Eds.), *Machine scoring of student essays: Truth and consequences* (pp. 130–137). Logan: Utah State University Press.

McLeod, S. (2012). The pedagogy of writing across the curriculum. In T. Zawacki & P. Rogers (Eds.), *Writing across the curriculum: A critical sourcebook* (pp. 53–68). Boston, MA: Bedford-St. Martin's.

Michael, W. B., & Shaffer, P. (1978). The comparative validity of the California State University and Colleges English Placement Test (CSUC-EPT) in the prediction of fall semester grade point average and English course grades of first-semester entering freshmen. *Educational and Psychological Measurement*, *38*(4), 985–1001.

Moss, B. G., & Yeaton, W. H. (2006). Shaping policies related to developmental education: An evaluation using the regression-discontinuity design. *Educational Evaluation and Policy Analysis*, *28*(3), 215–229.

Murphy, S., & Yancey, K. B. (2008). Construct and consequence: Validity in writing assessment. In C. Bazerman (Ed.), *Handbook of research on writing* (pp. 448–474). New York, NY: Lawrence Erlbaum.

National Council of Teachers of Engish, & Council of Writing Program Administrators. (2010). *NCTE-WPA white paper on writing assessment in colleges and universities*. Retrieved from http://wpacouncil.org/whitepaper

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common core state standards for English language arts*. Washington, DC: Authors.

Odell, L. (1981). Defining and assessing competence in writing. In C. Cooper (Ed.), *The nature and measurement of competency in English* (pp. 95–138). Urbana, IL: National Council of Teachers of English.

Office of Human Resources at the National Institutes of Health. (2014). *Competencies dictionary: Communications.* Retrieved from http://hr.od.nih.gov/workingatnih/competencies/core/communication.htm

Oppenheimer, D., Zaromb, F., Pomerantz, J. R., Williams, J. C., & Park, Y. S. (2014). *Statistically reliable improvement of writing skills during college: A multi-year cross-sectional and longitudinal study of undergraduate writing performance.* Manuscript submitted for publication.

Organisation for Economic Co-operation and Development. (2012). *Education at a glance 2012: OECD indicators*. Paris, France: OECD.

Osterlind, S. J., Robinson, R. D., & Nickens, N. M. (1997). Relationship between collegians' perceived knowledge and congeneric tested achievement in general education. *Journal of College Student Development*, *38*, 255–265.

Perelman, L. (2012). Construct validity, length, score, and time in holistically graded writing assessments: The case against automated essay scoring (AES). In C. Bazerman, C. Dean, J. Early, K. Lunsford, S. Null, P. Rogers, & A. Stansell (Eds.), *International advances in writing research: Cultures, places, measures* (pp. 121–131). Fort Collins, CO: WAC Clearinghouse/Anderson, SC: Parlor Press.

Perry, K. (2012). What is literacy? A critical overview of sociocultural perspectives. *Journal of Language and Literacy Education*, *8*(1), 50–71. Retrieved from http://jolle.coe.uga.edu/wp-content/uploads/2012/06/What-is-Literacy_KPerry.pdf

Pike, G. (1992). The components of construct validity: A comparison of two measures of general education outcomes. *The Journal of General Education*, *41*, 130–160.

Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M. E., & Kukich, K. (2002a). Stumping e-rater: Challenging the validity of automated essay scoring. *Computers in Human Behavior*, *18*, 103–134.

Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M. E., & Kukich, K. (2002b). Comparing the validity of automated and human scoring of essays. *Journal of Educational Computing Research*, *26*, 407–425.

Powers, D. E., Fowles, M. E., & Welsh, C. K. (2001). Relating performance on a standardized writing assessment to performance on selected academic writing activities. *Educational Assessment*, *7*, 227–253.

Quality Assurance Agency (QAA). (2008). *The framework for higher education qualifications in England, Wales and Northern Ireland: August 2008*. Mansfield, UK: Author.

Ramineni, C., Trapani, C., Williamson, D., Davey, T., & Bridgeman, B. (2012). *Evaluation of e-rater® for the GRE® issue and argument prompts* (Research Report No. RR-12-02). Princeton, NJ: Educational Testing Service. doi: 10.1002/j.2333-8504.2012.tb02288.x

Regents' Testing Program. (n.d.). *Percentage of students completing the Regents' Skills requirement by the semester during which they reached the 45 credit hour mark 2002–03 through 2008–2009*. Retrieved from http://www2.gsu.edu/~wwwrtp/pass_sophomore.htm.

Rhodes, T. L. (Ed.). (2010). *Assessing outcomes and improving achievement: Tips and tools for using rubrics*. Washington, DC: Association of American Colleges and Universities.

Rosenfeld, M., Courtney, R., & Fowles, M. (2004). *Identifying the writing tasks important for academic success at the undergraduate and graduate levels* (Research Report No. RR-04-12). Princeton, NJ: Educational Testing Service. doi: 10.1002/j.2333-8504.2004.tb01969.x

Rouet, J.-F., Britt, M. A., Mason, R. A., & Perfetti, C. A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology*, *88*, 478–493.

Rudner, L. M., Garcia, V., & Welch, C. (2006). An evaluation of IntelliMetric™ essay scoring system. *The Journal of Technology, Learning and Assessment*, *4*(4), 1–21.

Sabatini, J. P., O'Reilly, T., & Deane, P. (2013). *Preliminary reading literacy assessment framework: Foundation and rationale for assessment and system design* (Research Report No. RR-13-30). Princeton, NJ: Educational Testing Service. doi: 10.1002/j.2333-8504.2013.tb02337.x

Schaeffer, G. A., Briel, J. B., & Fowles, M. E. (2001). *Psychometric evaluation of the new GRE writing assessment* (GRE Board Professional Report No. 96-11P). Princeton, NJ: Educational Testing Service.

Scott-Clayton, J. (2012). *Do high-stakes placement exams predict college success?* (CCRC Working Paper No. 41). New York, NY: Community College Research Center.

Sheehan, K., & O'Reilly, T. (2012). The case for scenario-based assessments of reading competency. In J. P. Sabatini, E. R. Albro, & T. O'Reilly (Eds.), *Reaching an understanding: Innovations in how we view reading assessment* (pp. 19–33). Lanham, MD: Rowman & Littlefield.

Sinharay, S., & Haberman, S. J. (2008). *Reporting subscores: A survey* (Research Memorandum No. RM-08-18). Princeton, NJ: Educational Testing Service.

Sinharay, S., Puhan, G., & Haberman, S. J. (2011). An NCME instructional module on subscores. *Educational Measurement: Issues and Practice*, *30*(3), 29–40.

Sommers, N. (1980). Revision strategies of student writers and experienced adult writers. *College Composition & Communication*, *31*(4), 378–388.

Sparks, J. R. (2013). *Critical evaluation of information credibility: Contextual dimensions and implications for memory and learning from text* (Unpublished doctoral dissertation). Northwestern University, Evanston, IL.

Sparks, J. R., & Deane, P. (2014). *Cognitively based assessment of research and inquiry skills: Defining a key practice in the English language arts*. Manuscript submitted for publication.

Sparks, J. R., & Rapp, D. N. (2011). Readers' reliance on source credibility in the service of comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *37*(1), 230–247.

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research, & Evaluation, 9*(4). Retrieved from http://pareonline.net/getvn.asp?v=9&n=4&

U.S. Department of Labor Employment and Training Administration. (2014). *Competency model clearinghouse: Writing.* Retrieved from http://www.careeronestop.org/competencymodel/blockModel.aspx?tier_id=2&block_id=8

University of California. (2014). *Entry level writing requirement legislative report*. Retrieved from http://ucop.edu/operating-budget/_files/legreports/1314/ELWRlegrpt-3-4-14.pdf

WAC Clearinghouse. (2014). *A fuller definition of writing to learn*. Fort Collins, CO: WAC Clearinghouse.

Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, *3*(1), 22–36.

Weigle, S. C. (2002). *Assessing writing*. Cambridge, England: Cambridge University Press.

Weigle, S. C. (2011). *Validation of automated scores of TOEFL iBT® tasks against nontest indicators of writing ability* (Research Report TOEFL iBT-15). Princeton, NJ: Educational Testing Service.

White, E. M. (1995). The importance of placement and basic studies: Helping students succeed under the new elitism. *Journal of Basic Writing*, *14*, 75–84.

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal*, *46*(4), 1060–1106.

Williamson, D. M., Xi, X., & Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practice*, *31*(1), 2–13.

Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, *83*, 73–87.

Yancey, K. B. (1999). Looking back as we look forward: Historicizing writing assessment. *College Composition and Communication*, *50*(3), 483–503.

Zahner. (2013). *Reliability and validity of CLA+*. Retrieved from http://cae.org/images/uploads/pdf/Reliability_and_Validity_of_CLA_Plus.pdf

Zimmerman, B. J., & Risemberg, R. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology*, *22*, 73–101.

Suggested citation:

Sparks, J. R., Song, Y., Brantley, W., & Liu, O. L. (2014). *Assessing written communication in higher education: Review and recommendations for next-generation assessment* (ETS Research Report No. RR-14-37). Princeton, NJ: Educational Testing Service. doi:10.1002/ets2.12035

Action Editor: Donald Powers

Reviewers: Doug Baldwin and Paul Deane