

Report No: ACS19285

Republic of India

Measuring Learning in Higher Education in India

March 25, 2016

GED06

SOUTH ASIA



Standard Disclaimer:

This volume is a product of the staff of the International Bank for Reconstruction and Development/ The World Bank. The findings, interpretations, and conclusions expressed in this paper do not necessarily reflect the views of the Executive Directors of The World Bank or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

Copyright Statement:

The material in this publication is copyrighted. Copying and/or transmitting portions or all of this work without permission may be a violation of applicable law. The International Bank for Reconstruction and Development/ The World Bank encourages dissemination of its work and will normally grant permission to reproduce portions of the work promptly.

For permission to photocopy or reprint any part of this work, please send a request with complete information to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA, telephone 978-750-8400, fax 978-750-4470, <http://www.copyright.com/>.

All other queries on rights and licenses, including subsidiary rights, should be addressed to the Office of the Publisher, The World Bank, 1818 H Street NW, Washington, DC 20433, USA, fax 202-522-2422, e-mail pubrights@worldbank.org.

Background on the Role of Assessments in Higher Education for the Technical/Engineering Quality Improvement Project in India

Kimberly Parekh

I. The Context

Higher education institutions (HEI) have three major functions: **education**, **research**, and **service**. Each HEI differs in the priority of these functions based on their mission and consists of multiple stakeholders who are all invested in performing these functions. These stakeholders may include **policy makers** (from the political and legislative sphere), **employers**, **university administrators**, **faculty members**, **parents**, and **students**.

HEIs use multiple forms of data to measure each of the three major functions. However, the majority of information collected prioritizes inputs, processes, and outputs at the student, program, and institution level. The gap in information pertains to outcomes. Most stakeholders will agree that the desired outcomes of higher education include “the development of appropriate levels of knowledge and skills; the ability to integrate and apply knowledge to a variety of problems; and the acquisition of intellectual and social habits and dispositions in preparation for productive, responsible citizenship” (New Leadership Alliance for Student Learning and Accountability, 2012). Data to measure these outcomes is often scarce.

In terms of inputs, HEIs collect information at the student level through the number of student applications, number of students admitted, and number of students accepted who enroll all of which provide basic ‘quantity’ measures. They also look at average performance on secondary school leaving exam (i.e. the SAT and the ACT in the case of the US) and average performance in secondary school (i.e. average high school grade point average (GPA) in the case of the US) which provides basic ‘quality’ measures (Dwyer, Millet, & Payne, 2006). HEIs collect more information at the institutional level on the side of the student body, academic reputation through ranking, faculty academic credentials, size of library, and size of endowment (Dwyer et al., 2006). Typical analyses of the quality of institutions assume that HEIs with better inputs lead to better educational outputs and outcomes. For example, students with higher admissions test scores or faculty with qualifications result in higher student learning outcomes (Klein, Kuh, Chun, Hamilton, & Shavelson, 2005) (p. 254).

In terms of outputs, HEIs collect information at the student level on individual GPA, standardized tests, and persistence and graduation rates as basic ‘quantity’ measures (Cunha & Miller, 2014; Dwyer et al., 2006). Individual GPAs are typically norm-referenced making it difficult to compare students across universities and even disciplines. Standardized tests offer one way to understand student’s generic and domain-specific knowledge, skills, and competencies. Graduation rates signal that students have acquired knowledge, skills and competences in their field but they do not provide the extent to which students have acquired such knowledge, skills, and competencies while in university. The rates are also a very long term indicator making any shifts in policy difficult to understand. In this case, persistence rates provide alternative shorter term understanding of how policy changes can impact students (Cunha & Miller, 2014). HEIs can also collect information on wages and earnings after students leave the HEI. However, this assumes that students with greater knowledge and skills have higher wages, but higher learning outcomes do not always relate to higher income (Cunha & Miller, 2014). Institutions also collect information on average performance of graduate and/or professional school admission tests (GRE, GMAT, LSAT, MCAT in the case of the US), performance on licensure exams in professional and technical fields (in the case of the US, Singapore, Japan, etc.), and percentage of students who are employed after graduation as basic ‘quality’ measures. The ‘quality’ measures have limitations as they reflect the scores of students in particular disciplines and not all students (Dwyer et al., 2006). At best, this ‘quality’ data represents proxy measures on student learning outcomes.

In recent times, the pressure to go beyond typical input and output measures as proxies for quality has encouraged the need to assess student learning outcomes. In addition, international agendas such as the Spellings Commission in the US (2006) and Bologna process in the EU heightened the need to measure student learning outcomes (U.S. Department of Education, 2006).

II. Problem Statement

Like all HEIs around the world, India’s HEIs have systems to measure inputs, processes, and outputs. Currently, Indian engineering HEIs gather input information of students at the beginning of their HEI experience. Indian engineering HEIs first assess student achievement during the admissions process. Students take the JEE in two stages known as the JEE Mains and JEE Advanced. The Central Board of Secondary Education (CBSE) conducts the JEE Mains, a discipline-specific achievement test in three major subject matters namely physics, chemistry, and math for admission to 32 National Institutes of Technology (NITs), Indian Institutes of Information Technology (IIITs) and Government Funded Technical Institutes (GFTIs). Students with high scores are eligible to take the JEE Advanced. The Indian Institute of Technologies (IITs) then conducts the JEE Advanced, also a discipline-specific achievement test, in three major subject matters namely physics, chemistry, and math for admission into 16 Indian Institutes of Technologies (IITs) (Speed News Desk, 2016).

Last year, MHRD announced a decision to reform the admissions process with an emphasis on the entrance exam. According to MHRD, the push for reform is driven partly because students are able to successfully pass entrance exams due

to extensive preparation through coaching institutions but ultimately do not have the ability to be successful while at their HEI (Kumar, 2015). In addition to entrance exam scores, current admission processes also consider secondary school leaving exams known as school board exams. The MHRD identifies several problems in using secondary school board exam scores particularly that they measure learning at one point in time and do not provide additional opportunities to improve scores (Government of India, 2015).

In response, MHRD, through a Committee of Eminent Persons (CEP) headed by Ashok Mishra, suggested a compulsory standardized aptitude exam (Speed News Desk, 2016). Similar to the US Educational Testing Service (ETS), the newly created National Authority for Testing (NAT) will design and administer the test. MHRD will initially fund the NAT but is intended to be financially independent in the future (Mukul, 2016). The MHRD plans to register the NAT under their ministry as an independent agency which will require Cabinet approval (Speed News Desk, 2016). The NAT will design a mandatory high stakes aptitude test similar to the US SAT for Grades 11 and 12 students. The test is intended to measure generic competencies in both cognitive and non-cognitive skills including 'innovative abilities, logical thinking, problem solving, comprehension, decision making, critical thinking and general IQ' (Government of India, 2015). The new exam is intended to better identify students oriented towards scientific inquiry. The MHRD will then shortlist 40,000 students with the highest scores to take the JEE exam for admission to the IITs (Mukul, 2016). The admissions process will also discontinue the use of school board exam scores as a criteria (Speed News Desk, 2016). Policy makers believe that changing the admissions process for admission into HEIs, and that too primarily focusing on the exam, will allow for a student body that is better able to meet the demands of typical engineering HEIs. Like many universities around the world, Indian engineering HEIs consistently gather information on student achievement over the course of the university experience, but this information is not standardized making it difficult to understand what the student has actually gained throughout their university experience. The National Board of Accreditation (NBA) conducts accreditation processes at the program level and the National Assessment and Accreditation Council (NAAC) conducts accreditation processes at the institution level. But neither the NBA nor the NAAC processes take student learning into consideration.

III. Solution

The MHRD shift in assessments during the admissions process and the growing need to measure learning outcomes must allow for discussion of the purpose and use of assessments. This paper also provides a framework of understanding assessments and illustrative examples of such assessments.

For the purposes of this paper, **assessments** are "any systematic method of obtaining information, used to draw inferences about characteristics of people, objects, or programs; a systematic process to measure or evaluate the characteristics or performance of individuals, programs, or other entities, for purposes of drawing inferences" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014)(p. 216). Assessments, including exams or tests, measure knowledge, skills, and competencies. **Knowledge** is defined as "a body of information (such as facts, concepts, theories, and principles) students are expected to learn in a given content area", **skills** are defined as "the ability to apply knowledge to complete tasks and solve problems", and **competencies** are defined as "ability to use knowledge or skills in professional or academic situations" (From European Qualifications Framework <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=URISERV:c11104>)

All assessments designed to measure student learning test the student. However, HEIs can analyze student learning through multiple units of analysis, namely the student, program, and institution. Regardless of the unit, the assessment process can have greater or lesser consequences. High stakes tests have direct and important consequences for the participating students, programs, or institutions and low stake tests have indirect or unimportant consequences for participating students, programs, or institutions (American Educational Research Association et al., 2014). Almost all HEIs participate in high stakes exams for students during the admissions process. Increasingly, many HEIs participate in low stakes testing of students as a means to improve course curricula and instructional processes and create accountability at the institutional level.

HEIs measure student learning through assessments at various levels, namely HEIs use assessments to **certify student aptitude and/or achievement and verify student advancement at the student level**, **improve course curricula and instructional processes at the program level**, and **create accountability at the institutional level**.

A. Student Level Assessments

Assessments to certify student achievement and verify student advancement in higher education may include local achievement assessments and standardized tests such as generic aptitude and/or achievement tests; discipline-specific tests; and licensing exams (Martin, 2014).

Local achievement assessments offer an opportunity to examine learning over a longer period of time while the remaining generic aptitude and/or achievement tests, discipline-specific tests, and licensing exams typically offer an opportunity to measure learning at one single point in time (Shavelson, 2007). Local achievement assessments provide the most direct understanding of faculty instruction and student learning, but can lack the standardization required for comparability of student performance across programs and institutions.

Generic aptitude and/or achievement tests are often developed externally and provide comparability. For entry into higher education, both generic aptitude and/or achievement tests can measure general skills and competencies common to all disciplines. In particular, *generic aptitude tests* measure students' ability to learn new skills and competencies (Linn &

Gronlund, 2000). These tests may be suitable for students who have not been trained in any particular skills or competencies. They may be better for students who come from diverse backgrounds and may have not been exposed to adequate schooling or coaching. The use of aptitude tests are often debated and discussed as a predictor of success in higher education. In the U.S., the best predictor for success in higher education is the combined SAT (including critical reading, mathematics, and writing) and high school GPB but only for the first year of university (Korbin, Patterson, Shaw, Mattern, & Barbuti, 2008). Aptitude tests tend to have fewer questions and therefore require a shorter time than achievement tests. In the US, examples used for admission into higher education include the SAT and GRE and examples used to measure learning throughout multiple points in higher education include CAPP, MAPP, and CLA (Martin, 2014).

Generic achievement tests measure skills and competencies that have already been learned (Linn & Gronlund, 2000). Distinctions between aptitude and achievement are often confused and used interchangeably. Achievement tests may be conducted for students trained in specific skills and competencies and may be better suited for students from less diverse backgrounds and who have been trained in specific instruction either through schooling or coaching institutions. Achievement tests tend to have more questions and therefore require a longer time than aptitude tests.

Discipline-specific tests have a higher validity than generic competency tests, but discipline-specific tests make it difficult to compare students amongst the disciplines (McGrath, Guerin, Harte, Frearson, & Manville, 2015). However, discipline-specific achievement tests are often the best predictor of future achievement in the *same* discipline. For example, a math achievement test has good predictive value for future study in math particularly if the content of the assessment is aligned with the content of future learning.

Licensing exams test competency levels in professional and technical areas and provide signals to demonstrate a readiness to begin practice in the field. However, they are not very common around the world. Rather, countries rely on program-level accreditation (Coates, 2014).

Types of assessments that **certify student achievement and verify student advancement:**

Local Achievement Assessments: Administrators require course creators and faculty members administer formative assessments such as research papers, projects, tests, etc. or course-specific rubrics. These assessments are typically domain-specific skills and result in course grades.

Generic Aptitude and/or Achievement Tests: Policy makers and supporting organizations administer standardized tests to measure generic cognitive skills upon entry and throughout multiple time points exit from HEIs.

Discipline-Specific Achievement Tests: Policy makers and supporting agencies administer standardized tests to measure domain-specific skills upon entry from undergraduate HEIs.

Licensing Exams: Supporting agencies administer licensing exams to certify that students have mastered domain-specific skills to practice in their field.

Assessments measure a range

BOX 1: Faculty Involvement in Assessment

Tennessee Tech University's Center for Assessment and Improvement of Learning provides an example of how a university used assessments to improve course curricula and the instructional process. It conducted a 'generic competency test' known as the Critical Thinking Assessment Test (CAT). Test designers actively engaged faculty from a variety of institutions and disciplines and provide them guidance on how to design the assessment. They asked faculty to identify a set of core skills common to all disciplines. They identified four major skills: (1) evaluating information; (2) creative thinking (3) learning and problem solving; (4) and communication (Stein & Haynes, 2011). They then asked the faculty to create test items to assess these skills. Each HEI's faculty also scored student responses. In a survey, the faculty noted that their involvement in creating test questions and scoring test items allowed them to better understand student gaps and modify teaching methods to address student weaknesses.

such as critical thinking and problem solving, creativity and innovation, and communication **non-cognitive life and career** skills include "individual 's personality, temperament, and adaptability, initiative and self-direction, social and cross-cultural skills, productivity and accountability; and (4) **generic cross-cutting** skills include information, media, and technology (21st Century Learning, 2015).

Traditionally, HEI assessments typically focus on domain specific skills. But there has been a shift towards measuring generic skills even in the technical and professional disciplines. Researchers in general have found that measuring generic skills while understanding broader student learning outcomes. In fact, the lack of generic skills has risen to the forefront as a prevailing problem for graduates. The eye-opening *Academic Achievement Study* (2011) revealed that more than 45% of more than 2,000 students in multiple disciplines showed significant gains in reasoning and writing skills from the beginning of their first year to the end of their first year at university (Arum & Roksa, 2011). Specific disciplines are also recognizing the need for generic skills. The Accreditation Board of Engineering and Technology (ABET) has included higher order thinking skills as requisite skills in the real world. As such, they were included while measuring student learning outcomes for accreditation standards (Stein et al., 2009).

B. Program/Instructional Level Assessments

Assessments at the program level serve to improve course curricula and the instructional process. These assessments may often be in conflict with assessments for accountability at the institutional level (Erickson, 2009). Different stakeholders at have different goals for the assessment. Faculty members use local or discipline-specific tests, essays, research papers, research projects, etc. and standardized tests to measure how students are learning course material and make adjustments in instruction or curriculum to improve student better achievement gains. While adjustments can be made, faculty members often believe that the lack of resources related to student level of preparation given the resources available (time, facilities, technology) is a major barrier (Martin, 2014)(p. 203).

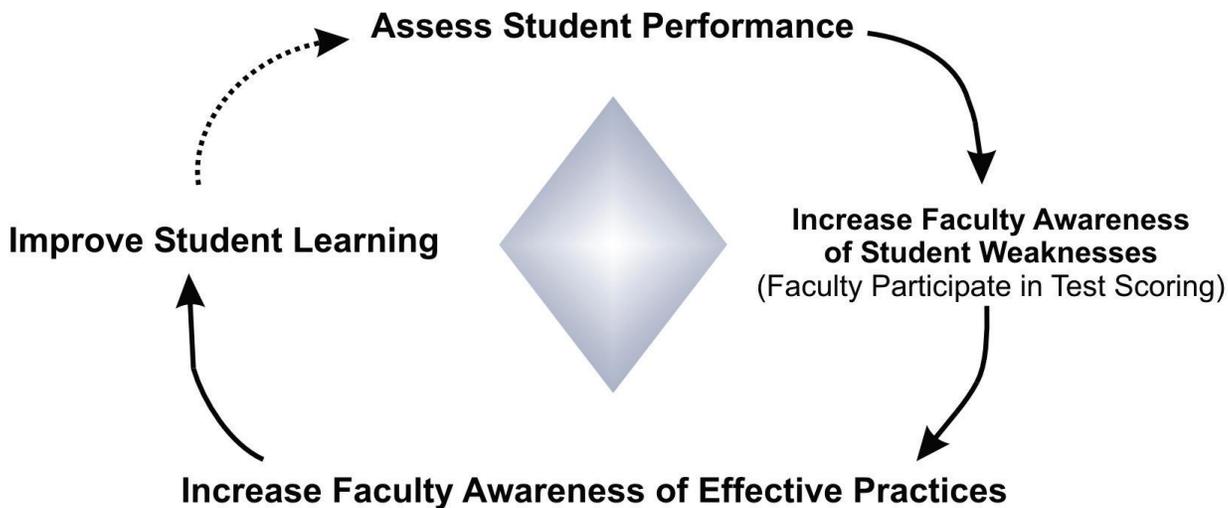
Often, faculty members and sometimes administrators view standardized tests as less useful because the goals for each program and university are so different. Some believe any attempt to measure learning outcomes will lead to over

simplification and result in a poor comparison of programs and universities (Benjamin, 2012; Miller, 2006). In particular, faculty note that the content of these tests do not necessarily align with the content of highly individualized and specific course content even within disciplines and can only be useful when the assessment instruments align with course curricula (Benjamin, 2012). Furthermore, there is a disconnect as instructional processes in higher education are often student-centered, case-study/problem-centered, or open ended while typical assessments use select response such as multiple choice (Benjamin, 2012). Some faculty members and even administrators are highly suspicious because of the commercial aspects of testing organizations and because of the costs associated with external testing (Jankowski et al., 2012; Miller, 2006). They are reluctant to allocate additional resources for testing. Instead, they prefer use of in-house assessments to keep financial resources within the institutions (Jankowski et al., 2012). Others are cautious about the dissemination and public reporting of results and fear punitive actions by policy makers and administrators for low test scores (Benjamin, 2012; Miller, 2006).

On the other hand, some stakeholders such as some administrators, policy makers, and supporting organizations use standardized tests such as 'generic aptitude and/or achievement tests', 'discipline-specific achievement tests', and 'licensing exams' to better understand how students perform on a broader segment of learning outcomes, benchmark progress, and compare programs and institutions with each other (Benjamin, 2012). They may believe that while local achievement tests are a necessary part of understanding the instructional process, they are not reliable in the test taking sense. They believe that is therefore necessary to use standardized testing which have greater reliability and validity (Benjamin, 2012). Administrators often believe assessments measuring student learning at the program level are a reflection of program strength and faculty member's ability to teach effectively (Borden & Peters, 2014). But while test scores can be a useful signal to administrators and faculty members about students' knowledge, skills, and competency levels, additional understanding needs to take place on why students are performing the way that they are (Benjamin, 2012).

This conflict can be reconciled if policy makers, university administrators, assessment professionals, and faculty members work together to take ownership of the assessment process and align HEI curricular and instructional goals with assessment (Kuh et al., 2015). Stein & Hayes (2011) offer a model on how faculty participation in assessment enables them to be more aware of the teaching and learning process in order to improve student learning outcomes. Box 1 of Tennessee Tech University's Critical Assessment Test (CAT) provides a relevant example of how faculty use assessments to improve teaching and learning.

Closing the Loop in Assessment and Quality Improvement



(Stein & Haynes, 2011)

In addition to these direct measurements of learning, **it is also possible to understand indirect or proxy measures of learning.** For example, some US HEIs administer the National Survey of Student Engagement (NSSE) which provides information on student engagement in the learning process (Dwyer et al., 2006). NSSE asks students to self-assess their understanding of academic based knowledge and skills prioritizing 'higher-order learning', 'reflective and integrative learning', 'learning strategies', and 'quantitative reasoning', their assessment of peer-based learning (with a particular focus on 'collaborative learning' and diversity), their experience with faculty (including student faculty engagement and effective faculty pedagogy), and their experience with the enabling campus learning environment (including quality and support systems) (National Survey of Student Engagement (NSSE), 2015). These self-assessed areas represent an indirect measure of quality. Like standardized tests, HEIs aggregate and analyze this data at the institution level and create benchmarks for comparison. This data is also intended to provide institutional accountability.

C. Institutional Level Assessments

Assessments for accountability can be made through **(1) quality assurance through institution and/or program audits; (2) accreditation of programs; and (3) national assessments of students.** Quality assurance audits are conducted by external quality assurance agencies. These program or institutional level assessments collect multiple pieces of data on management and administration.

Accreditation processes are conducted by external accreditation agencies and enable programs or institutions to be certified as meeting certain standards (Mahat & Goedegebuure, 2014). Over the years, accreditation processes in countries like the US and Australia have increasingly included learning outcomes as a part of their processes by aggregating student level assessment data by program or institution. Before the 1990s, accreditation processes in many institutions emphasized input, process, and output oriented indicators such as faculty credentials, facility characteristics, curriculum requirements, library capacity, and fiscal soundness (Kuh et al., 2015)(p.5). After the 1990s, accreditation processes moved towards the inclusion of outcome oriented indicators such as test scores, number of years until graduation, graduation rates, and licensure and certification rates (U.S. Department of Education, 2006). More importantly, accreditation processes have not only focused on whether HEIs have assessment processes of learning outcomes, but whether the HEIs are actually using the results to improve curricula and pedagogy (Ewell, 2009). This is important because accreditation processes which consider learning view graduates in aggregate and therefore do not consider individual students. In other words, all students who graduate from the program are deemed to be qualified and competent in their field (Coates, 2014).

In engineering, the Washington Accord, an international agreement, is one of the leading frameworks for accrediting HEI engineering programs. Signatory countries such as the UK, Australia, New Zealand, etc. accredit HEI engineering programs and assume students who successfully graduate from these programs have the knowledge, skills, and competencies to successfully perform in their field. In 2014, India, as represented by the National Board of Accreditation (NBA), became a signatory of the Washington Accord for Tier 1 HEIs. Country-specific accreditation processes such as ABET from the U.S. also consider learning outcomes. For example, ABET's eight 'Criteria for Accrediting Engineering Programs' for the 2016-2017 accreditation cycle emphasize student performance and specify 11 learning outcomes including domain-specific, cognitive, non-cognitive, and cross-cutting skills and competencies.¹

Beyond quality assurance and accreditation, institutions can also assess their performance by administering **national level student assessments**. These types of assessments allow HEIs to understand student performance against a fixed standard and compare students amongst institutions (Klein et al., 2005) (p. 4).

For example, students take the Collegiate Learning Assessment (CLA) at the end of higher education. It assesses generic cognitive skills and provides an understanding of the learning gained during the HEI experience. It is "intended to send a signal to administrators, faculty, and students about some of the competencies that need to be developed, the level of performance attained by the students at their institution, and most importantly, whether that level is better, worse, or about the same as what should be expected given the ability of its incoming students" (Klein et al., 2005) (p. 4). Like other standardized tests, the CLA itself does not identify the reasons why students do better or worse than expected or what measures the school should take either in terms of instruction or curriculum to improve student performance. In this way, the HEI must look at a compilation of other data including local achievement tests or even standardized tests through accountability reviews to determine how to raise student scores.

Regardless of the type of accountability measure, it is important to draw comparisons by comparing likeminded programs or institutions. HEIs can benchmark performance relative to peer institutions or can aspirationally benchmark performance to the next level of institutions (Dwyer et al., 2006). At the institution level, peer groups can be created by geographic location or similar goals. For example, in the case of Indian Engineering HEIs, groupings can be made with Tier 1 institutions. At the student level, it is important to understand performance based on student characteristics (Dwyer et al., 2006).

¹ ABET's student learning outcomes include: (a) an ability to apply knowledge of mathematics, science, and engineering; (b) an ability to design and conduct experiments, as well as to analyze and interpret data; (c) an ability to design a system, component, or process to meet desired needs within realistic constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability; (d) an ability to function on multidisciplinary teams; (e) an ability to identify, formulate, and solve engineering problems; (f) an ability to communicate effectively; (g) an ability to communicate effectively; (h) the broad education necessary to understand the impact of engineering solutions in a global, economic, environmental, and societal context; (i) a recognition of the need for, and an ability to engage in life-long learning; (j) a knowledge of contemporary issues; and (k) an ability to use the techniques, skills, and modern engineering tools necessary for engineering practice (ABET, 2015).

IV. A Framework for Creating Assessments

The following framework identifies brief *key steps* for creating and implementing relevant and appropriate student-level, program-level, and institutional-level assessments²:

A. Choose assessment methods appropriate for educational decisions

Set goals about learning outcomes upon completion of instructional period:

Start with the central claim the policy makers and/or institutions are willing to make about what students should know, do, and apply upon completion of the instructional period.

Identify coordinating body and key persons responsible for creating the assessment:

Select key stakeholders in the assessment process. Key stakeholders can include policy makers (from the political and legislative sphere), employers, university administrators, and faculty members. Students can also be used in piloting the assessment.

Define use of assessment including any intended and unintended consequences:

Determine the use of the assessment. HEIs use assessments to certify student aptitude and/or achievement and verify student advancement at the student level, improve course curricula and instructional processes at the program level, and create accountability at the institutional level.

B. Develop assessment methods appropriate for educational decisions

Create framework including a detailed table of specifications of the domain-specific, cognitive, and non-cognitive skills to be measured:

Create a table of specifications including the exact domain-specific, cognitive, and non-cognitive skills to be measured, including the weight of skills to be measured. For professional and technical fields, there has been much debate and discussion on whether to separate or combine or separate domain-specific skills and cognitive and non-cognitive skills. Some psychometricians believe that it is not possible to measure both domain-specific and cognitive skills at the same time; others believe that especially for professional and technical fields understanding of domain-specific skills are only meaningful if they are understood and applied in real-life settings. Test designers must discuss what is relevant and appropriate for their context.

For example, in the Assessment of Higher Education Learning Outcomes (AHELO) civil engineering assessment, test developers combined domain-specific skills with both generic cognitive and non-cognitive skills. The following constructed response test item includes a multiple part question on domain-specific skills that includes multiple cognitive levels from understanding to application (Assessment of Higher Education Learning Outcomes (AHELO), 2012). In addition, correct responses include generic cognitive and non-cognitive skills like communications.

(Question 1a³): Imagine that a new dam is being planned today in a different location. **Briefly explain two environmental effects of the dam (which could also be upstream or downstream) than an engineer would need to consider in an environmental impact statement).**

(Answer 1a): The right answer assumes that the student identifies habitats; soil and/or siltation/erosion; ground stability around storage itself; CO₂ emissions/greenhouse gases; aesthetics; effluent impact; and community impact. [Note (author italics): *The correct answer assesses domain-specific knowledge, including lower-order cognitive skills like identification.*]

(Question 1b): A taller arch dam, built in 1959 and now disused, is the Vajont Dam in Italy. In 1963, a massive landslide occurred upstream of the dam, and some 260 million cubic meters of earth and rock fell into the reservoir. Over 2000 lives were lost as water spilled over the dam and caused massive flooding in the valley below. The question then includes pictures of the damn, before/after the landslide. **Briefly explain two geo-technical assessments you would expect to have done before the dam was constructed.**

(Answer 1b): The right answer assumes that the student refers to material assessment; stability assessment; geological assessment; and hydro-geological assessment. [Note (author italics): *The correct answer assesses domain-specific knowledge, including higher-order cognitive skills like assessment and evaluation.*]

(Question 1c): Consider the following scenario: Results of the geotechnical analyses performed before the dam is constructed indicate that there is the potential for a serious slop failure risk, but with very low probability of occurrence. The engineering team is presented with two options: (i) Proceed with current design, on the basis that the serious problem is unlikely to occur; and (ii) Redesign the damn, which will incur greater cost, but reduce the risk of failure. **Discuss two aspects of engineering practice that favor option (ii).**

(Answer 1c): The right answer assumes that the student refers to ethics; safety/risk; action; and communicating risk. [Note (author italics): *The correct answer assesses domain-specific knowledge, including higher-order cognitive*

² Based on: American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME)'s *Standards for Educational and Psychological Testing*

³ Based on: AHELO's sample illustrative items from AHELO, Volume 1 Design and Implementation

skills like problem solving. Also, the correct answer considers generic cognitive skills like communications.]

(Question 1d): **After construction of the Vajont Dam, and recognizing the possibility of hillside failure, outline two planning measures an engineer could have suggested to minimize potential harm to people.**

(Answer 1d): The right answer assumes that the student refers to evacuation procedures; town planning; town protection; monitoring; operation; communication plan regarding the risks; and strengthening/reinforcement of the dam wall and/or the hillside. [*Note (author italics): The correct answer assesses domain-specific knowledge, including higher-order cognitive skills like problem solving. Also, the correct answer considers generic cognitive skills like communications and non-cognitive skills like leadership and accountability.*]

Align competencies for both generic aptitude and/or achievement test with educational standards, course curricula, and instructional goals:

Ensure that policy makers and HEI administrators and faculty align assessment with stated and taught educational standards, course curricula, and instructional goals. In addition, the select competencies measured should reflect intended learning outcomes. For aptitude tests, the type of competencies measured should be the type of competencies expected for success during HEIs. For achievement tests, the competencies measured must be in line with and agreed upon as what is stated and taught and is reflected in the educational standards, course curricula, and instruction.

Determine format of the test including length of exam, number of test items, type of test questions (selected response or constructed response), etc.:

Determine the format of the assessment. Aptitude tests require less number of items because they generally test verbal and mathematical reasoning, while achievement tests require a larger number of items to demonstrate proficiency of competencies.

Tests can either have select or constructed response test items. Select response test items include multiple-choice, true-false, matching, etc. The multiple choice item has been the most widely used format in standardized testing in the United States for more than half a century. However, select response test items tend to emphasize discrete facts or procedural skills, which assess lower-order cognitive skills. These test items can come at the expense of higher-order cognitive skills like critical thinking and problem solving skills which involve constructing and applying knowledge (Linn & Gronlund, 2000)(p. 39).

Constructed response test items require students to create their own responses but are guided by specific rubrics to standardize scoring. In recent years, assessments entirely guided by constructed responses have become more common. Performance assessments tend to test higher order cognitive skills like problem solving and authentic assessments situate performance assessments in real world contexts (Linn & Gronlund, 2000) (p. 40). Critics say that performance assessments are time-consuming to administer and are problematic in scoring than fixed response tests because human judgment is involved in scoring and scorers require a high degree of expertise and training (Linn & Gronlund, 2000)(p. 39).

Guarantee that the testing of identified competences is fair and accessible for all students regardless of individual characteristics such as gender, ethnicity, religious affiliation, disability, language, etc.:

Ensure test items do not have any undue bias based on student characteristics. The best way to ensure test items do not have bias is to pilot items with a wide variety of students.

Pilot assessment, including for multiple uses, to meet basic psychometric principles including validity and reliability:

Ensure test items are both valid and reliable. "Validity refers to the degree to which evidence support the interpretation of test scores" (Millet, Stickler, Payne, & Dwyer, 2007) (pg. 4). ETS identifies several key issues around validity including the degree to which what you want measured is actually being measured; the degree to which what you did not intend to be measured is actually being measured; and the degree of intended and unintended consequences of the assessment altogether (Dwyer et al., 2006) (pg. 11). "Reliability refers to the consistency with which an assessment measures the construct(s) that it purports to measure." Test items that are valid are not necessarily reliable and test items that are reliable are not necessarily valid (Millet et al., 2007) (pg. 4).

Create appropriate manuals and guidelines for test administrators and scorers:

Create manuals and guidelines for test administrators and scorers. This is particularly important if the test is to be administered across a wide context. Also, guidelines for scorers are imperative especially for constructed response test items. In this case especially, consistent training should be provided to scorers and multiple scorers for the same student can ensure reliability of scores.

C. Administer, score, and interpret the results of assessment

Determine test administering agency, location of tests, delivery of test (paper or online), translation of test, grading and scoring (answer sheets and rubrics), and interpretation of results:

Determine the administering and scoring of the assessment. Test administering agencies vary in their autonomy from the government and HEIs all over the world. In Colombia, ICFES is an independent test making organization but has very close ties with the government. On other hand, in Mexico, CENEVAL is an independent test administering agency.

Tests can be administered in the higher education institutes themselves or independent third party locations affiliated with the test administering agency.

The delivery of the test can perhaps be one of the most important aspects in the assessment process. For example, Stanford University faced many challenges delivering their electrical engineering assessments across various countries because students had various experiences in taking paper and online versions of the test. Test developers have to ensure that the delivery of the test does not result in score variance, i.e. competencies measured are not impacted by paper versus online delivery.

Test developers interpret student scores on a test in two different ways: norm-referenced interpretation and criterion-referenced interpretation. **Norm-referenced score interpretation** is “a score interpretation based on a comparison of a test taker’s performance with the distribution of performance in a specified reference population” (American Educational Research Association et al., 2014) (p. 221). Norm-referenced interpretation describes a student’s performance relative to others in the same group. **Criterion-referenced score interpretation** is “the meaning of a test score for an individual or an average score for a defined group, indicating the individual’s or group’s level of performance in relationship to some defined criterion domain” (American Educational Research Association et al., 2014) (p. 218). Criterion-reference interpretation describes a student’s performance against fixed criteria (Linn & Gronlund, 2000) (p. 42).

Strictly speaking, norm-referenced and criterion-referenced are terms specifically used for the interpretation of scores. However, tests are designed taking into consideration norm-referenced and criterion-referenced interpretation of scores. Norm-referenced tests tend to select items of average difficulty so that there can be a widespread distribution of scores and thereby allow discrimination amongst student achievement at various levels. This is useful for student selection, grouping, and grading. Criterion-referenced tests tend to select items that represent specific learning outcomes, without regard to comparison amongst students. These tests do not typically eliminate easy or difficult items, and therefore the test can range in ease or difficulty. These tests are useful if students’ understanding and application of knowledge and skills are intended to cover specific learning (Linn & Gronlund, 2000) (p. 43).

Identify whether the test will be voluntary or compulsory for students:

Determine which students will take the tests. If the tests are high stakes, then it is likely the tests will be compulsory and all students will take the tests. Admissions tests into HEIs are often compulsory. If the tests are low stakes, the tests may be voluntary or compulsory for students. It is important to note that individual student test takers must be motivated to make a sufficient effort in taking the test in order for the test scores to accurately represent student’s understanding of knowledge, skills and competencies gained during the period of performance. Test design must therefore eliminate any threats to validity by offering appropriate and relevant incentives so that students can meaningfully participate in the test. This issue is particularly important for students who take the test as a part of program improvement or institutional accountability (Dwyer et al., 2006) (p.11).

For example, in the case of Colombia, ICFES first created a voluntary HEI leaving exam called SABER PRO. The purpose of the test was to assess the strength of HEI programs by testing students in domain-specific skills. HEIs had the responsibility of encouraging students to take the voluntary test but did not invoke penalties if students did not take the test. The voluntary nature of the test resulted in many challenges including difficulties in getting students to take the test. As a result, each discipline had only a small sample size making it difficult to compare. After 2009, ICFES changed the test to assess generic cognitive skills, semi-specific domain-specific skills, and specific domain-specific skills. They also made taking the test a compulsory requirement for graduation but left it to the HEIs to determine specific incentives. Some HEIs required minimum results for graduation, while other HEIs created financial incentives for top performers. Students could also use test results for scholarships. The increased number of test takers for multiple skills made it much easier to compare test scores within and across disciplines.

While the Colombia experience made SABER PRO compulsory, some studies have showed that there has been a negative association in requiring students to take standardized learning assessments. Specifically, mandatory testing has been known to depress testing effort, thereby negatively impacting validity. For example, a research study showed a negative association between requiring students to take the Collegiate Learning Assessment (CLA) after controlling for other factors. Some studies have shown positive correlations related to faculty encouragement and involvement in the administration of such assessments; other studies show no significance while controlling for other factors. Other studies have shown positive correlations related to the use of incentives such as payment for completion of assessments or payment for performance (e.g. top scorers receive payment) (Steedle, Zahner, & Kugelmass, 2014) (p. 10). While further inquiry on the use of financial incentives in particular contexts needs to take place, research shows incentives may be a good strategy for students taking low-stakes tests. The Council on Academic Education recommends that students volunteer for testing but be offered financial incentives for completion. In any case, test administration agencies administering voluntary test taking must ensure that the testing sample is representative of the entire student body (CAE, 2014, pg. 10).

D. Use assessment results when making decisions about individual students, planning teaching, developing curriculum, and improving university/program

Document assessment results and take the appropriate steps to verify those who are using the results are qualified to make decisions within the HEI:

Ensure assessments created for one purpose are not used for another purpose. It is important to document which stakeholders are using the results and for what purpose. These results should be incorporated into a continuous cycle of understanding student learning.

E. Communicate assessment results to key stakeholders

Contextualize reporting to external and internal key stakeholders, including institutional and program information and the disaggregation of data based on student characteristics in order to provide a better understanding of results:

Communicate results with external and internal key stakeholders. While doing so, communications of results must (1) meet audience specific needs; (2) provide appropriate context; and (3) share evidence in multiple formats and forums.

For external stakeholders, disclosure and comparability of information are the two most important considerations. Disclosure of information can be controversial, especially if HEIs are asked to publicly report data. Careful consideration must be made in considering the HEI contexts and the implications of results when disclosing data particularly if it will lead to key decisions made (Kuh et al., 2015) (p.204). Academics note that more information may not necessarily lead to better-informed consumers.

What is needed is not just the disclosure of more information but more targeted, purposeful communication of relevant evidence of student learning to specific audiences that need and can use it. By focusing on sharing evidence-based stories of institutional improvement, for example, institutions are able to both meet external accountability demands and show they are assuring quality through improving results (Kuh et al., 2015) (p.204).

Comparability is also an important tool. But providing comparisons does not always lead to change. For example, the National Institute for Learning Outcomes Assessment's (NIOLA) Voluntary System of Accountability (VSA) presents HEI data on multiple dimensions, including student learning. Careful understanding of website traffic revealed that very few visitors of the website accessed the student learning outcomes page. However, it is important to note that those accessing the website have also commented that the student learning outcomes page does not present data in a user-friendly way (Jankowski et al., 2012).

If the assessment is intended for program improvement, there must be a focus on internal communication. Factors that impede sharing internally include: "organizational silos and a lack of buy-in or training among administrators around transparent communication, faculty cultures that might dissuade from engagement in assessment work due to skepticism or distrust, and the highly decentralized nature of many colleges and universities" (Jankowski et al., 2012). NIOLA notes that the single most useful way of communicating results internally are during larger faculty retreats and faculty professional development opportunities.

F. Recognize unethical, illegal, and otherwise inappropriate methods and uses of assessment information

All stages of the assessment process from selecting, developing, administering, using, and communicating results must be fair and must be conducted in professional and ethical manner. All stakeholders should be familiar with their roles and responsibilities in conducting assessments. Stakeholders should also ensure results are not used inappropriately such as embarrassing students, violating student's right to confidentiality, and using students' scores to measure teaching and learning effectiveness (American Federation of Teachers, National Council on Measurement in Education, & Association, 1990).

V. Appendices

A. International Civil Engineering Assessment – AHELO Case Study

The OECD's Assessment of Higher Education Learning Outcomes (AHELO) was the first major international assessment designed to measure learning outcomes in higher education. It was designed as a voluntary international comparative assessment of learning outcomes at the higher education level similar to international assessment programs at the school level such as the Program for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS). Specifically, it aims to provide higher education institutions in both developing and developed countries a tool to measure overall performance of their students. The pilot assessment was designed to measure both generic and domain-specific skills and prioritizes a higher understanding of skills. The generic skills were modeled after the US Collegiate Learning Assessment. The domain-specific skills measured skills from two specific disciplines: economics and civil engineering.

The civil engineering assessment was carefully created through a collaborative and extensive design process. The Australian Council for Educational Research (ACER), Japan's National Institute for Educational Policy Research (NIER), and the University of Florence created the engineering assessment. These groups created an international Engineering Expert Group which guided the development of an Engineering Assessment Framework. The Framework was built upon a long and extensive process drawing upon on the AHELO-tuning document, the AHELO Engineering Assessment workshop held at ACER in Melbourne in January 2010, the Tertiary Engineering Capability Assessment (TECA), and broader AHELO technical materials (Assessment of Higher Education Learning Outcomes (AHELO), 2012)(pg. 121-122). The Framework identified specific learning outcomes to be measured including the specific domain to be tested and their definitions. "The framework also provided details of how much the domain was to be operationalized for measurement purposes with discussions of issues such as time, language level, item type, scoring, assessment delivery and administration, and reporting" (Assessment of Higher Education Learning Outcomes (AHELO), 2012) (pg. 122).

The Civil Engineering Assessment aimed to test the 'proficiency' of final-year Bachelor's degree of civil engineering students which they define as "demonstrated capacity to solve problems by applying basic engineering and scientific principles, engineering processes and generic skills. It includes the willingness to engage with such problems in order improve the quality of life, address social needs, and improve the competitiveness and commercial success of society" OECD 2012a in (Pearce, 2015)(pg. 5). In summary, the Assessment measured: (1) Engineering Generic Skills (two competencies); (2) Basic and Engineering Sciences (seven competencies); (3) Engineering Analysis (six competencies); (4) Engineering Design (two competencies); and (5) Engineering Practice (six competencies). The Framework raised several issues such as creating a common definition of what the assessment is intended to measure; the nature of basic engineering generic skills and how this differs from the generic skills assessment; the proportion and extent to which basic engineering skills and discipline-specific skills (in this case civil engineering) and should be measured.

The civil engineering assessment consists of both select-response (constructed-response questions) and fixed-response (multiple choice questions) and is designed to be completed in 90 minutes. The constructed-response questions were created by ACER. The constructed-response questions were designed to assess: (1) engineering generic skills; (3) engineering analysis; (4) engineering design; and (5) engineering practice. Constructed-response items were designed to assess higher-order understanding of concepts. These test items also offered authentic engineering contexts and identified problems within this context. It offered multiple pieces of information including photographs, diagrams, tables, charts, etc. Students were required to use both short and long answers to analyze or evaluate and/or provide solutions and/or recommendations to the problems. The Engineering Expert group created scoring rubrics and took cultural and linguistic differences into consideration. After a round of piloting, each student was to complete one constructed response item in 30 mins (though three constructed response items in total were administered through a rotated test design).

The multiple choice questions drew from existing instruments designed to assess fundamentals of recently trained engineers such as ABET & etc. The multiple choice questions were designed to assess: (1) basic engineering science. These questions allowed a quicker understanding of knowledge, skills and competencies and verified student understanding of the constructed response question. After a round of piloting, 25 multiple choice were to be administered (though 30 multiple choice questions in total were administered through a rotated test design.)

In consultation with the Engineering Expert Group, all test items were mapped against the established framework. The pilot resulted in a few findings (Pearce, 2015):

Validity, Reliability & Bias

- Basic psychometric analysis removed between 2 to 11 test items from the total 30 multiple choice questions and 3 constructed response questions, with the majority being multiple choice items because of bias.
- It is unclear whether or not the breadth of competencies adequately represented the construct (but it is important to note that feasibility not construct was the focus of pilot).
- Content validity of instrument through expert consensus of the development of the framework, instrument, and scoring guide. Positive student feedback also confirmed though based on acceptance rather scientific data validating construct.
- Inter-rater score reliability was around 80% which is considered good.

Item Difficulty & Targeting

- Constructed response items proved to be difficult. Scoring rubric includes three measures: correct interpretation of competency (score of 1), incorrect interpretation of competency (score of 0), and no interpretation of competency. Most students (between 20-70%) responded with 0.

Motivation & Efforts

- Students self-reported that CRT were relevant to their program. Furthermore, questions were interesting and clear. But student-self reported effort was only moderate (on a four point scale).

Phase I first included five countries and was later expanded to Phase II which included 9 'systems' (since there was variability within countries). These countries include Australia, Canada (Ontario), Japan, Colombia, Slovak Republic, United Arab Emirates (Abu Dhabi), Egypt, Mexico and Russia. These countries solicited participation from 92 higher education institutions and 6,078 students (Pearce, 2015) (pg. 15).

B. National Engineering Assessment – Mexico Case Study

The National Center for Evaluation of Higher Education (CENEVAL) in Mexico was founded by the National Association of Universities and Higher Education Institutions (ANUIES) in 1994. CENEVAL's mission is to design and administer assessments for academic purposes. They broadly administer four major assessments: (1) entrance tests (EXANI); (2) student learning outcomes for specific disciplines (EGEL); student learning outcomes for generic skills (EXDIAL); and ad-hoc tests upon requests (Uribe, 2013).

The *Examen General Para el Egreso de la Licenciatura* (EGEL) is a discipline-based criterion assessment for students completing their higher education from a four-year degree program. The test is designed to assess whether students have minimum knowledge, skills, and competencies in their field. Test items emphasize real-life situations for young professionals. EGEL exams can be grouped into three categories: (1) life sciences and behavioral sciences which include 11 disciplines; (2) social sciences and humanities which includes 10 disciplines; and (3) engineering and technology which includes 12 disciplines. They cover the majority of disciplines studied in Mexico.

The Generic Skills tests are divided into four tests: Basic Science for Engineering Programs Test (EXIL-CBI); Statistics Test (EXTRA-ES); Communication and Critical Thinking Test (ECCyPEC); and Written Expression Test (EEE-II). The EXIL-CBI exam tests basic science skills in three subjects: math (algebra, calculus, differential equations, probability and statistics), physics (mechanics, thermodynamics, electromagnetism), and chemistry (pure substances and mixtures and chemical reactions). The ECCyPEC tests reading comprehension, knowledge of written expression, and critical thinking. The EEE-II test is an essay-based exam scored by two professors who use a rubric. It tests conventions of language, syntactic knowledge, lexical variety, thematic progression of text, global consistency, planned speech, information sources, and creativity.

In terms of test design, CENEVAL follows the AERA, APA, and NCME's *Standards for Educational and Psychological Testing* standards in terms of "defining the construct domain, producing the test specification, building the item bank, and creating and administering the actual tests, scoring the results and delivering reports (Uribe, 2013)(p. 140). A Technical Expert Group (TEG) comprised of faculty, employers, and professionals coordinates the minimum levels to be achieved in each discipline-based exam. Trained specialists write each of the items, which are always piloted and tested in the field. All items follow a multiple-choice format. After the administration of every test, TEG reviews the tests for validity and reliability.

The exam is administered four times a year in over 60 locations. They are offered in two four-hour sessions. The exam is not compulsory for students, unless specified by the higher education institution. It is used to help higher education administrators and faculty understand whether their students are achieving a benchmark. While the test does test minimum standards, it categorizes performance on a three point-scale: outstanding, satisfactory, and not yet satisfactory. Students can show results to potential employers.

CENEVAL provides individual reports to each student, including an overall and subscale results. It also provides institutional reports to each institution. The reports are confidential and not provided to anyone else. However, once a year, CENEVAL aggregates all students by program in a specific university and compares the results with other universities.

C. National Engineering Assessment – Colombia Case Study

Colombia's ICFES administers compulsory standardized tests for secondary school graduates called the SABER 11 and compulsory standardized tests for final year higher education students called the SABER PRO. The SABER 11 was created in 1968 and is taken by 600,000 students per year. Students provide SABER 11 test scores are part of their admission process into HEIs. Test scores can also be used for financial scholarships and awards for higher education. Secondary schools also use SABER 11 test scores as 'unofficial ranking of high schools'. In this regard, SABER 11 is considered 'high stakes' for students and schools. Since 2000, the SABER 11 is a subject-matter specific competency test but also recently moved towards generic and semi-specific subject matter competencies as well. The change in SABER 11 will allow the SABER PRO to make better comparisons (Julian P. Marino von Hildebrand, 2014).

The SABER PRO was created in 2003 and is taken by 300,000 students per year. Students use SABER PRO test scores for admission process into graduate studies and for financial scholarships and awards in higher education. Institutions use SABER PRO to better understand how students are performing, inform accreditation processes, and maintain university rankings (Julian P Marino von Hildebrand & Molina Mantilla, 2016). In this regard, SABER PRO is considered 'high stakes' for institutions and 'low stakes' for students. From 2003-2007, SABER PRO was designed to measure domain-specific skills in 55 specific disciplines and was voluntary for students. In some cases, challenges included too few students taking tests to

make valid comparisons as well as capacity and costs (Julian P Marino von Hildebrand & Molina Mantilla, 2016). Because the tests were voluntary, HEIs had difficulty in getting students to take the test and perform to the best of their ability (Julian P Marino von Hildebrand & Molina Mantilla, 2016). From 2009, SABER PRO was modified to measure domain-specific disciplines and became compulsory for students. But it also included domain semi-specific skills such as scientific inquiry or engineering project management depending on the discipline and generic skills such as critical reading, quantitative reasoning, written communication, citizenship, and English (Julian P. Marino von Hildebrand, 2014). ICFES provides SABER PRO scores to individual students and aggregate scores to individual programs and institutions.

D. Generic Skills Assessment – Collegiate Learning Assessment (CLA) Case Study

The US' Collegiate Learning Assessment+ (CLA+) is a generic achievement test and measures generic cognitive skills in critical thinking and written communications. The test is formatted to offer both select response and constructed response questions. The select response section consists of 25 multiple choice questions which students are expected to take 30 minutes to complete. The section assesses scientific and quantitative reasoning (10 questions); critical reading and evaluation (10 questions); and the critiquing of an argument (5 questions). These questions are supported by real life documents such as 'information sources including letters, memos, photographs, charts, and newspaper articles'(Council for Aid to Education, 2014). The constructed response section assesses "analysis and problem-solving, writing effectiveness, and writing mechanics" and consists of one performance task which is expected to take students 60 minutes to complete (Council for Aid to Education, 2014).

E. National Entrance Exam – Russia Case Study

Russia's Unified State Examination (USE) was created in 2009 as an entrance exam for the purposes of certifying secondary school students and selecting students for higher education (Tyumeneva, 2013)(p.V, XI). This high stakes exam is compulsory and is taken every year by nearly a million students in 83 subject areas (Tyumeneva, 2013)(p.1). The exam's dual purpose was developed because Russia did not have a large scale assessment program at the secondary school level. The USE therefore uses the exam partly for that purpose even though it was not designed or intended for such use (Tyumeneva, 2013) (p.I). As such, the exam also informs pedagogy, monitors education quality, and ensures accountability at the secondary level (Tyumeneva, 2013)(p. V). Specifically, the exam has informed new national learning standards, curricula, and textbooks (Tyumeneva, 2013) p. XI). The unfortunate consequence is that the exam has also resulted in the training of students on how to take the exam (Tyumeneva, 2013) (p.2).

VI. Illustrative Examples of HEI Engineering Assessments

Name of Test	Use of Assessment (Student, Program, Institutional)	Administered by	Type	Competencies	Format	Length
Assessment of Higher Education Learning Outcomes (AHELO)	Institutional Assessment	OECD (International Organization)	Generic Competency Test Discipline-Specific Achievement Test	<i>General:</i> Generic Competency Skills <i>Specific:</i> Civil Engineering (Engineering Generic Skills, Basic and Engineering Sciences, Engineering Analysis, Engineering Design, Engineering Practice)	Short Answer Multiple Choice (25 items) and Performance Task (1 Constructed Response item)	90 mins (60 mins for Multiple Choice and 30 mins for Performance Task)
Critical Thinking Assessment Test (CAT)	Program and Institutional Assessment	Tennessee Technological University (TTU)	Generic Competency Test	Critical Thinking: Evaluating Information, Creative Thinking, Learning and Problem Solving, Communication	Short Answer Essay (15 test items)	1 hour
Collegiate Learning Assessment + (CLA+)	Program and Institutional Assessment	Council for Aid to Education (Supporting Organization)	Generic Competency Test	<i>General:</i> Combined Modules in Critical Thinking, Analytic Reasoning, Problem Solving, and Written Communication	Performance Task (1 Constructed Response Item) and Multiple Choice (25 items)	90 mins (60 mins for Performance Task and 30 mins for Multiple Choice)

				s		
Examen General Para el Egreso de la Licenciatura (EGEL)	Student, Program, and Institutional Assessment	Mexico's National Center for Evaluation of Higher Education (CENEVAL) (Supporting Organization)	Generic Competency Test Discipline-Specific Achievement Test	<p><i>General:</i> Separate Modules in Communication & Critical Thinking (ECCyPEC); Written Communications (EEE-II); etc.</p> <p><i>Semi-Specific:</i> Basic Science for Engineering Programs Test (EXIL-CBI);</p> <p><i>Specific:</i> 12 Engineering and Technology discipline based EGEL tests</p>	Mostly Multiple Choice & Essay (for Written Communications) Multiple Choice (200-250 items)	4 hours
HEIghten	Program and Institutional Assessment	Educational Testing Services (Supporting Organization)	Generic Competency Test	<p><i>General:</i> Separate Modules in Critical Thinking; Quantitative Literacy; & Written Communication</p>	Multiple Choice (24-26 items for each module) & Essay (for Written Communications)	45 mins each module
Colombian State Examination of the Quality of Higher Education (SABRE PRO)	Student, Program, and Institutional Assessment	Colombia's Instituto Colombiano para la Evaluación de la Educación (ICFES) (Supporting Organization)	Generic Competency Test Discipline-Specific Achievement Test	<p><i>General:</i> Separate Modules in Critical Reading, Quantitative Reasoning, Written Communication, Citizenship, English</p> <p><i>Semi-Specific:</i> Scientific Inquiry; Engineer Program Management,</p>	Multiple Choice (35 items for each module) and Essay (for Written Communications) Multiple Choice (30-45 items) for each module Multiple Choice (30-45 items) for each module	4.5 hours (to complete all five modules) 4 hours (to complete up to 4 semi-specific and specific modules)

				etc. <i>Specific:</i> Multiple Engineering disciplines	
Stanford	Program and Institutional Assessment	Stanford University (research association)	Generic Competency Test Discipline-Specific Achievement Test	<i>General</i> Separate Assessment Modules: Critical Thinking and Quantitative Literacy, Creativity, Ability to Learn and Make Choices <i>Semi-Specific:</i> Math, Physics, and Informatics <i>Specific:</i> Electrical Engineering	Multiple Choice for the Critical Thinking and Quantitative Literacy (24-26 items); Free response for Creativity; "Choice Data" for Ability to Learn and Make Choices Multiple Choice (45 items) for each module Multiple Choice (30-45 items)

ABET. (2015). Criteria for Accrediting Engineering Programs *Effective for Reviews During 2016-2017 Accreditation Cycle*. Baltimore, MD.

ACT. (2014). Cognitive and Noncognitive Skills *WorkKeys*.

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

American Federation of Teachers, National Council on Measurement in Education, & Association, N. E. (1990). Standards for Teacher Competence in Educational Assessment of Students from <http://buoros.org/standards-teacher-competence-educational-assessment-students>

Arum, R., & Roksa, J. (2011). *Academically Adrift*. Chicago: The University of Chicago Press.

Assessment of Higher Education Learning Outcomes (AHELO). (2012). Volume 1 Design and Implementation. In OECD (Ed.), *Feasibility Study Report*.

Benjamin, R. (2012). The Seven Red Herrings About Standardized Assessments in Higher Education *Occasional Paper #15*.

Borden, V., & Peters, S. (2014). Faculty Engagement in Learning Outcomes Assessment. In H. Coates (Ed.), *Higher Education Learning Outcomes Assessment* (Vol. Part III Review). Frankfurt: Peter Lang

Coates, H. E. (2014). *Higher Education Learning Outcomes Assessment* (Vol. 6): Peter Lang Edition.

Council for Aid to Education. (2014). CLA+ Practice Performance Task

Cunha, J. M., & Miller, T. (2014). Measuring Value-Added in Higher Education: Possibilities and Limitations in the Use of Administrative Data. *Economics of Education Review, 42*.

Dwyer, C. A., Millet, C. M., & Payne, D. G. (2006). Postsecondary Assessment and Learning Outcomes - Recommendations to Policymakers and the Higher Education Community *A Culture of Evidence*.

Ewell, P. T. (2009). Assessment, Accountability, and Improvement *Occasional Paper*

- Government of India. (2015). National Testing Scheme 61st CABE Meeting.
- Jankowski, N. A., Ikenberry, S. O., Kinzie, J., Kuh, G. D., Shenoy, G. F., & Baker, G. R. (2012). Transparency & Accountability: An Evaluation of the VSA College Portrait Pilot *A Special Report from the National Institute for Learning Outcomes Assessment for the Voluntary System of Accountability*.
- Klein, S. P., Kuh, G. D., Chun, M., Hamilton, L., & Shavelson, R. (2005). An Approach to Measuring Cognitive Outcomes Across Higher Education Institutions. *Research in Higher Education, 46*(3).
- Korbin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). Validity of the SAT for Predicting First-Year College Grade Point Average (Vol. 5). New York.
- Kuh, G. D., Ikenberry, S. O., Jankowski, N. A., Cain, T. R., Ewell, P. T., Hutchings, P., & Kinzie, J. (2015). *Using Evidence of Student Learning to Improve Higher Education* Jossey-Bass.
- Kumar, B. (2015, December 28,). HRD Considering Aptitude Test to Help Students in Their Career Choice. *Hindustan Times*.
- Linn, R. L., & Gronlund, N. E. (2000). *Measurement and Assessment in Teaching*. Upper Saddle River, New Jersey: Prentice-Hall, Inc.
- Mahat, M., & Goedegebuure, L. (2014). Transparent Reporting of Learning Outcomes. In H. Coates (Ed.), *Higher Education Learning Outcomes Assessment* (Vol. Part IV: Improvement, pp. 263-278). Frankfurt: Peter Lang.
- Marino von Hildebrand, J. P. (2014). *Colombia's Assessment System*, Philadelphia
- Marino von Hildebrand, J. P., & Molina Mantilla, A. (2016). In K. Parekh (Ed.).
- Martin, L. (2014). Assessing Student Learning Outcomes: Research Trajectories. In H. Coates (Ed.), *Higher Education Learning Outcomes Assessment* (pp. 49-68). Frankfurt: Peter Lang.
- McGrath, C. H., Guerin, B., Harte, E., Frearson, M., & Manville, C. (2015). Learning Gain in Higher Education
- Miller, M. A. (2006). *The Legitimacy of Assessment*.
- Millet, C. M., Stickler, L. M., Payne, D. G., & Dwyer, C. A. (2007). A Culture of Evidence: Critical Features of Assessments for Postsecondary Student Learning.
- Mukul, A. (2016). Testing Authority Planned On Lines of SAT, GRE.
- National Survey of Student Engagement (NSSE). (2015). Engagement Indicators and High-Impact Practices Retrieved March 1, , 2016, from http://nsse.indiana.edu/pdf/Els_and_HIPs_2015.pdf
- New Leadership Alliance for Student Learning and Accountability. (2012). Guidelines for Assessment and Accountability in Higher Education Washington, DC.
- Partnership for 21st Century Learning. (2015). Framework for 21st Century Learning. Washington, DC.
- Pearce, J. (2015). Assessing Vocational Competencies in Civil Engineering: Lessons from AHELO for Future Practice. *Empirical Research in Vocational Education and Training, 7*(1).
- Shavelson, R. J. (2007). A Brief History of Student Learning Assessment: How We Got Where We Are and a Proposal for Where to Go Next.
- Speed News Desk. (2016, January 15,). JEE Exam to be Scrapped; Get Ready for NAT 2017.
- Steedle, J. T., Zahner, D., & Kugelmass, H. (2014). *Test Administration Procedures and Their Relationships with Effort and Performance on a College Outcomes Test*. Paper presented at the Annual Meeting of the American Educational Research Association, Philadelphia, Pennsylvania
- Stein, B., & Haynes, A. (2011). Engaging Faculty in the Assessment and Improvement of Students' Critical Thinking Using the Critical Thinking Assessment Test. *The Magazine of Higher Learning, 43*(2), 44-49.
- Stein, B., Haynes, A., Redding, M., Harris, K., Tylka, M., & Lisc, E. (2009). *Faculty Drive Assessment of Critical Thinking: National Dissemination of the CAT Instrument*. Paper presented at the In Proceedings of the 2009 International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering, 2010.
- Tyumeneva, Y. (2013). Disseminating and Using Student Assessment Information in Russia *Systems Approach for Better Education Results (SABER) Student Assessment*
- U.S. Department of Education. (2006). A Test of Leadership: Charting the Future of U.S. Higher Education *A Report of the Commission Appointed by Secretary of Education Margaret Spellings*.
- Uribe, R. V. (2013). Measurement of Learning Outcomes in Higher Education: The Case of Ceneval in Mexico. In S. Blomke, O. Zlatkin-Troitschanskaia, C. Kuhn & J. Fege (Eds.), *Modeling and Measuring Comptencies in Higher Education*. The Netherlands: Sense Publishers.
- ABET. (2015). Criteria for Accrediting Engineering Programs *Effective for Reviews During 2016-2017 Accreditation Cycle*. Baltimore, MD.
- ACT. (2014). Cognitive and Noncognitive Skills *WorkKeys*.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

- Arum, R., & Roksa, J. (2011). *Academically Adrift*. Chicago: The University of Chicago Press.
- Assessment of Higher Education Learning Outcomes (AHELO). (2012). Volume 1 Design and Implementation. In OECD (Ed.), *Feasibility Study Report*.
- Benjamin, R. (2012). The Seven Red Herrings About Standardized Assessments in Higher Education *Occasional Paper #15*.
- Borden, V., & Peters, S. (2014). Faculty Engagement in Learning Outcomes Assessment. In H. Coates (Ed.), *Higher Education Learning Outcomes Assessment* (Vol. Part III Review). Frankfurt: Peter Lang
- Coates, H. E. (2014). *Higher Education Learning Outcomes Assessment* (Vol. 6): Peter Lang Edition.
- Council for Aid to Education. (2014). CLA+ Practice Performance Task
- Cunha, J. M., & Miller, T. (2014). Measuring Value-Added in Higher Education: Possibilities and Limitations in the Use of Administrative Data. *Economics of Education Review, 42*.
- Dwyer, C. A., Millet, C. M., & Payne, D. G. (2006). Postsecondary Assessment and Learning Outcomes - Recommendations to Policymakers and the Higher Education Community *A Culture of Evidence*.
- Ewell, P. T. (2009). Assessment, Accountability, and Improvement *Ocasional Paper*
- Government of India. (2015). National Testing Scheme *61st CABE Meeting*.
- Jankowski, N. A., Ikenberry, S. O., Kinzie, J., Kuh, G. D., Shenoy, G. F., & Baker, G. R. (2012). Transparency & Accountability: An Evaluation of the VSA College Portrait Pilot *A Special Report from the National Institute for Learning Outcomes Assessment for the Voluntary System of Accountability*.
- Klein, S. P., Kuh, G. D., Chun, M., Hamilton, L., & Shavelson, R. (2005). An Approach to Measuring Cognitive Outcomes Across Higher Education Institutions. *Research in Higher Education, 46*(3).
- Korbin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). Validity of the SAT for Predicting First-Year College Grade Point Average (Vol. 5). New York.
- Kuh, G. D., Ikenberry, S. O., Jankowski, N. A., Cain, T. R., Ewell, P. T., Hutchings, P., & Kinzie, J. (2015). *Using Evidence of Student Learning to Improve Higher Education* Jossey-Bass.
- Kumar, B. (2015, December 28,). HRD Considering Aptitude Test to Help Students in Their Career Choice. *Hindustan Times*.
- Linn, R. L., & Gronlund, N. E. (2000). *Measurement and Assessment in Teaching*. Upper Saddle River, New Jersey: Prentice-Hall, Inc.
- Mahat, M., & Goedegebuure, L. (2014). Transparent Reporting of Learning Outcomes. In H. Coates (Ed.), *Higher Education Learning Outcomes Assessment* (Vol. Part IV: Improvement, pp. 263-278). Frankfurt: Peter Lang.
- Marino von Hildebrand, J. P. (2014). *Colombia's Assessment System*, Philadelphia
- Marino von Hildebrand, J. P., & Molina Mantilla, A. (2016). In K. Parekh (Ed.).
- Martin, L. (2014). Assessing Student Learning Outcomes: Research Trajectories. In H. Coates (Ed.), *Higher Education Learning Outcomes Assessment* (pp. 49-68). Frankfurt: Peter Lang.
- McGrath, C. H., Guerin, B., Harte, E., Frearson, M., & Manville, C. (2015). Learning Gain in Higher Education
- Miller, M. A. (2006). *The Legitimacy of Assessment*.
- Millet, C. M., Stickler, L. M., Payne, D. G., & Dwyer, C. A. (2007). A Culture of Evidence: Critical Features of Assessments for Postsecondary Student Learning.
- Mukul, A. (2016). Testing Authority Planned On Lines of SAT, GRE.
- National Survey of Student Engagement (NSSE). (2015). Engagement Indicators and High-Impact Practices Retrieved March 1, , 2016, from http://nsse.indiana.edu/pdf/EIs_and_HIPs_2015.pdf
- New Leadership Alliance for Student Learning and Accountability. (2012). Guidelines for Assessment and Accountability in Higher Education Washington, DC.
- Partnership for 21st Century Learning. (2015). Framework for 21st Century Learning. Washington, DC.
- Pearce, J. (2015). Assessing Vocational Competencies in Civil Engineering: Lessons from AHELO for Future Practice. *Empirical Research in Vocational Education and Training, 7*(1).
- Shavelson, R. J. (2007). A Brief History of Student Learning Assessment: How We Got Where We Are and a Proposal for Where to Go Next.
- Speed News Desk. (2016, January 15,). JEE Exam to be Scrapped; Get Ready for NAT 2017.
- Steedle, J. T., Zahner, D., & Kugelmass, H. (2014). *Test Administration Procedures and Their Relationships with Effort and Performance on a College Outcomes Test*. Paper presented at the Annual Meeting of the American Educational Research Association, Philadelphia, Pennsylvania
- Stein, B., & Haynes, A. (2011). Engaging Faculty in the Assessment and Improvement of Students' Critical Thinking Using the Critical Thinking Assessment Test. *The Magazine of Higher Learning, 43*(2), 44-49.
- Stein, B., Haynes, A., Redding, M., Harris, K., Tylka, M., & Lisic, E. (2009). *Faculty Drive Assessment of Critical Thinking: National Dissemination of the CAT Instrument*. Paper presented at the In Proceedings of the 2009 International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering, 2010.

- Tyumeneva, Y. (2013). Disseminating and Using Student Assessment Information in Russia *Systems Approach for Better Education Results (SABER) Student Assessment*
- U.S. Department of Education. (2006). A Test of Leadership: Charting the Future of U.S. Higher Education *A Report of the Commission Appointed by Secretary of Education Margaret Spellings*.
- Uribe, R. V. (2013). Measurement of Learning Outcomes in Higher Education: The Case of Ceneval in Mexico. In S. Blomke, O. Zlatkin-Troitschanskaia, C. Kuhn & J. Fege (Eds.), *Modeling and Measuring Competencies in Higher Education*. The Netherlands: Sense Publishers.