**Measuring the generic skills of higher education students and graduates:
Implementation of CLA+ International**

Doris Zahner[1]
Dirk Van Damme[2]
Roger Benjamin[1]
Jonathan Lehrfeld[1]

Introduction: why we need to know and understand the learning outcomes of higher education students

In many parts of the world higher education systems are going through a process of transformation. Technological changes and associated developments in the economy and labor markets have pushed the demand for high-skilled workers and professionals to unprecedented levels. Higher education has become the most important route for the human capital development of a country. Higher education is the part of the learning trajectory where young people acquire the higher levels of generic and specific skills that are needed in the knowledge economy. At the individual level a higher education qualification still offers the prospect of significant benefits in employability and earnings, despite the fact that in most countries the enrolment and graduation rates have increased massively. The higher education system also plays an important role in developing the social and emotional skills for becoming effective citizens that are able to participate in the social and political processes of developed economies. Higher education attainment rates thus also correlate highly with indicators of social capital and social cohesion such as interpersonal trust and volunteering.

But how do we know that higher education systems and individual institutions affectively fulfill this role of developing the skills that matter? The OECD Survey of Adult Skills (PIAAC), assessing foundation skills such as literacy, numeracy and problem solving in digital environments has demonstrated that higher education qualifications, the most commonly used measure of human capital, are a poor indicator of the actual skills level of the population. There is growing evidence that qualification do not match skills. In contrast with secondary school systems, where OECD's PISA has become the global benchmark of learning outcomes of 15 year-old students and hence of the quality of school systems, there is no valid and reliable measure of the learning outcomes of higher education students and graduates.

This puts a severe strain on the credibility of higher education systems and institutions to effectively develop the skills that matter for today's and tomorrow's knowledge economy and society. Massification and grade inflation further underscore doubts on the value of degrees and qualifications. There are signs that global employers start to distrust university qualifications and develop their own assessment tools and procedures to test students for the skills that they think are important. Also governments start to worry in a context where not only the overall cost but also the per student cost is rising. They confront universities with concerns about efficiency and "value-for-money". And they increasingly shift the balance in the funding mix of higher education from public to private sources, thereby increasing the cost for students and families. When students are asked to pay more for the degree they hope to earn, they also become stakeholders in the value-for-money debate.

In essence, these developments point to a transparency problem in higher education, very similar to what economists identify as an "information asymmetry" problem (Van Damme, 2015). The system provides very little information to the user (the student), the stakeholders (tax-payers, employers, social

---

[1] Council for Aid to Education (CAE)
[2] Organisation for Economic Co-operation and Development (OECD)

partners) or the government. Data-driven transparency systems in higher education overly rely on metrics related to the research function of universities. And the rankings which are built on such systems define the power balances in the global higher education arena and drive the reputation race without saying anything about the teaching and learning function of universities, nor of the vast bulk of institutions serving the large majority of students below the absolute global top.

In this context the interest in direct assessment of learning outcomes is rapidly growing. In recent years various approaches and experiments have been developing, among them the CLA+ International. Endorsed by the OECD, the Council for Aid to Education (CAE) has enrolled a significant number of postsecondary institutions in several countries to participate in the launch of CLA+ International, an assessment of generic skills. This chapter aims to illustrate the relevance of and outline the process for developing and administering a generic skills assessment in higher education in an international context.

The Case for Generic Skills Assessment
Institutions of higher education globally are being challenged to improve instruction so that tomorrow's workforce will have the knowledge and skills necessary to meet the demands of modern careers and contribute to the global economy. Indeed, a higher education degree has never been more necessary for productive participation in society. Employers now seek individuals able to think critically and communicate effectively in order to meet the requirements of the new Knowledge Economy (AHELO, 2012a; Hart Research Associates, 2006; Levy & Murname, 2004). Therefore, the skills taught in higher education are changing with more emphasis being placed on so called "generic skills"(Clanchy & Ballard, 1995; Crebert, Bates, Bell, Patrick, & Cragnolini, 2004; Kearns, 2001) such as analytic reasoning and evaluation, problem solving, and written communication.

Because generic skills are so critical to workforce productivity, one of the best alternatives to solve the skills mismatch issue (Montt, 2015) which has been identified as globally problematic, is to emphasize the development of generic skills in education and training of all students which will equip them to "…learn field or job-specific skills on the job."

This alternative is aligned with and supports the most basic rationale for focusing on generic skills. In the knowledge economy in which the service sector is dominant, definitions of knowledge and learning have shifted from an emphasis on content to the ability to apply what one knows to new situations. Human capital, the most important asset nations have, includes the knowledge, education, experience, and skills a nation's citizens possess. In today's knowledge economy this privileges the ability to access, structure and use information not merely recall facts. This places the focus squarely on the importance of generic skills in every occupation.

Measuring Generic Skills
Increasing recognition of the essential role of generic skills in the knowledge economy portends significant changes in teaching and learning as reflected in the educational reform movement now underway and assisted by education technology. Although this reform is present in elementary and secondary education, most advances have occurred in higher education in Europe and the United States. The reform movement can be characterized along three dimensions: the shift from the long-standing lecture format to a student-centered approach emphasizing students' active class participation; the change in the balance of curricular and textbook focus from its current emphasis on content to case and problem-based materials requiring students to apply what they know to novel situations; and the innovation in assessment instruments from multiple-choice tests that are best used for measuring the

level of content absorbed by students to open-ended assessments that are aligned with several goals of the reform initiative.

Although significant advances have been made on the first two dimensions of this education reform movement, assessment has lagged behind. As universities focus increasingly on developing generic skills in their students, assessments need to evolve to measure how well students are learning—and institutions are teaching—such skills. The recall, recognition, and regurgitation paradigm is no longer sufficient.

Multiple-choice and short-answer assessments remain the dominant testing regime, not only for facts, but also for generic skills. As a result, in higher education and elsewhere, the testing regime is not assessing the most critical skills required of students in the workplace and—just as importantly—is not supporting the other two dimensions of reform. We believe the promise of educational reform developing in today's knowledge economy cannot be achieved without employing open-ended, performance-based assessments, not only in higher education, but in primary and secondary education as well as other points along the education-to-work continuum. In the workplace and contemporary human resources management approaches to recruitment, selection and upskilling performance-based assessments have become standard practice, as in so-called assessment centers.

Another important advantage of performance assessments is that they are seen as tests worth teaching to (Benjamin & Klein, 2006). The practice of "teaching to the test" is generally frowned upon when referring to traditional multiple-choice and short-answer assessments (Lazear, 2006; Moloney, 2006; Popham, 2001; Volante, 2004), and there is ample evidence that this practice occurs, especially when educators are held accountable for their students' test performance. However, "teaching to the test" for performance assessments should be encouraged. That is, class time spent preparing students to apply knowledge and analysis and problem solving skills to complex, real-world problems is time well spent. If performance assessments are integrated into accountability systems, this has the potential to positively impact classroom practice by encouraging teachers to foster the development of competencies in generic skills. This effect has yet to be established, so it would be worthwhile to investigate whether the introduction of performance assessment for accountability purposes has the desired effect on teaching and learning. One potential barrier to investigate is the perceived level of effort required to use performance assessments regularly in the classroom.

In addition to negative effects on pedagogy, a critical shortcoming of today's principal educational assessment regime is that it pays little attention to how much an institution contributes to developing the competencies students will need after graduation. For instance, the outcomes that are typically looked at by higher education accreditation teams, such as an institution's retention and graduation rates and the percentage of its faculty in tenured positions, say nothing about how well the school fosters the development of its students' analytic reasoning, problem-solving, and communication skills. This situation is unfortunate because the ways in which institutions are evaluated significantly affects institutional priorities. If institutions were held accountable for student learning gains and student achievement, they would likely direct greater institutional resources and effort toward improving teaching and learning.

All these conditions point to the need to support advances in performance assessment, particularly in the field of education. If the human capital school demonstrates the importance of education, the implications of the knowledge economy and recent theories of learning place the focus on improving the generic skills of the next generation of students. These developments create an urgent need to

generate and implement a testing paradigm that measures and simulates these skills. Because of its broad reach, there is an excellent opportunity to demonstrate the effective use of performance assessment to measure generic skills through the use of CLA+ International.

CLA+: Performance-based assessments of generic skills

The Collegiate Learning Assessment (CLA/CLA+) is a performance-based assessment of critical thinking and written communication. Traditionally, the CLA was an institutional level assessment that measured student learning gains (Klein & Benjamin, 2008; Klein, Benjamin, Shavelson, & Bolus, 2007), specifically using a value-added model (Steedle, 2009, 2012) within a university. The CLA employed a matrix sampling approach, under which students were randomly distributed either a Performance Task (PT) or an Analytic Writing Task, for which students were allotted 90 minutes and 75 minutes, respectively. The CLA Performance Tasks presented real-world situations in which an issue, problem, or conflict was identified, and students were asked to assume a relevant role to address the issue, suggest a solution, or recommend a course of action based on the information provided in a document library. Analytic Writing Tasks consisted of two components— one in which students were presented with a statement around which they must construct an argument (Make an Argument), and another in which students were given a logically flawed argument that they must then critique (Critique an Argument).

In its original form, the utility of the CLA was limited. Because the assessment consisted of just one or two responses from each student, reliable results were only available at the institutional level, and students' results were not directly comparable. Likewise, reporting for the CLA was restricted to the purposes of its value-added measure, and institutions were not eligible for summary results unless they had tested specified class levels in the appropriate testing windows.

Thus, the creation of the CLA+, which has a PT similar to the original CLA PT as the anchor of the assessment. There is an additional set of 25 selected response questions (SRQs) to increase the reliability of the instrument (Zahner, 2013) for reporting individual student results. The SRQ section is aligned to the same construct as the PT and is intended to assess higher-order cognitive skills rather than the recall of factual knowledge. Similar to the PT, students are presented with a set of questions as well as one or two documents to refer to when answering each question. The supporting documents include a range of information sources, such as letters, memos, photographs, charts, and/or newspaper articles. Each student receives both components (PT and SRQ) of the assessment.

Subscores

The CLA+ has six separate subscores. The open-ended student responses from the PT are scored on three sub-score which have a range from 1 – 6: Analysis and Problem Solving, Writing Effectiveness, and Writing Mechanics[3]. The SRQs consist of three sub-sections: Scientific and Quantitative Reasoning, Critical Reading and Evaluation, and Critiquing an Argument. Students have 60 minutes to complete the PT and 30 minutes to complete the SRQs. There is a short demographic survey following the assessment which should be completed within 15 minutes.

Additionally, CLA+ includes a new metric in the form of mastery levels. The mastery levels are qualitative categorizations of total CLA+ scores, with cut scores that were derived from a standard-setting study (Zahner, 2014). The five mastery level categories are: Below Basic, Basic, Proficient, Accomplished, and Advanced.

---

[3] https://cae.org/images/uploads/pdf/CLA_Plus_Scoring_Rubric.pdf

CLA+ International is the translated and adapted version of the domestic (to the United States) CLA+.

Distribution of mastery levels by race/ethnicity
One of the first studies using CLA+ was a longitudinal survey which followed graduating students who had taken the CLA+ in spring 2014 for one year after graduation. CAE also surveyed the employers of students who volunteer their managers and asked about the importance of the skills measured by CLA+. Results from the first major study of the longitudinal data (Zahner & James, 2016) revealed unsurprising results.

Arum and Roksa (2014) found better post-university outcomes for engineering and computer science major, but only by comparison to the rather bleak outcomes of their fellow student graduates. In comparison, CLA+ was found to be a positive predictor of post-university outcomes as measured by employment, full time employment, salary, or enrolment in a graduate school program (Zahner & James, 2016). However, the research also indicated that there are racial biases with respect to hiring, salary, and enrollment in continuing education. White, male, business majors had the best post-university outcomes when compared to others. These biases, though, may be conflated with whether students from underrepresented or minority groups attend selective or non-selective colleges.

Race/ethnicity by generic skills by institution competitiveness (Figure 1)
Race/ethnicity was self-reported by students in the demographic survey from the CLA+. Four categories were selected for analysis: Asian, African-American/Black non-Hispanic, Hispanic or Latino, and White, non-Hispanic. Students were also categorized into two groups based upon their mastery of the skills measured on CLA+: those who were proficient in critical thinking and written communication and those who only possessed basic or below basic skills. The final variable was whether the student attended a competitive or non-competitive institution (Barron's, 2014).
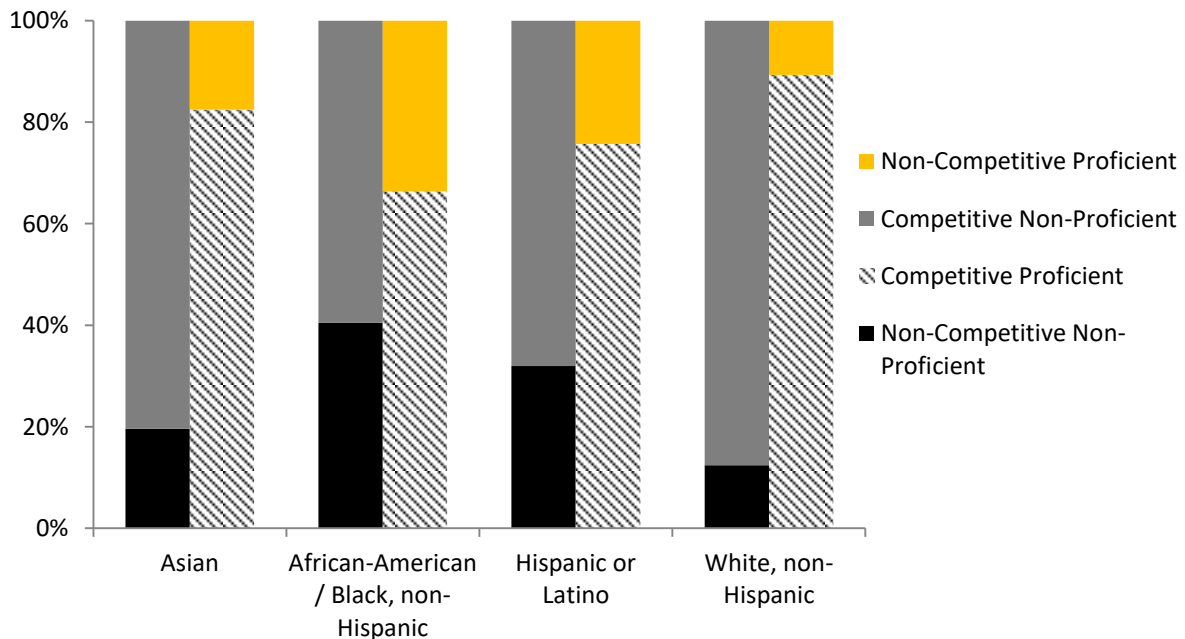

Figure 1: Distribution of CLA+ proficiency and institution competitiveness by race/ethnicity

In general, the percentage of minority students in less- or non-competitive institutions from

CAE's sample is almost double (50.6%) than from competitive institutions (26.8%). Yet, about 40% of students with proficient scores, as measured by CLA+, at the less- and non-competitive institutions are minorities. This means that there is a significantly large group of qualified university graduates from underrepresented or minority groups who may be overlooked as viable candidates due to the school they attended.

Benchmarking students from universities
As an extension of the above analysis on race/ethnicity and university competitiveness, there are approximately 144 competitive higher education institutions with 950,000 places for undergraduate students (Barron's, 2014). The remaining 13 to 14 million four-year students attend less or non-competitive institutions that are largely public, or not renowned. About 30% of the students in the competitive universities fall into the high ability category as measured by the CLA+ whereas 9 % of the students in the less competitive colleges are in this same category (Benjamin, 2015). However, simple arithmetic indicates that there are approximately 1.2 million high-ability students in the less competitive institutions versus 300,000 in the selective ones (Figure 2). The 1.2 million high ability students in the less competitive institutions that have generic skills similar to their counterparts in the competitive colleges do not get the opportunity to compete for the higher value added jobs their skills qualify them for because the employers do not know how to find them.
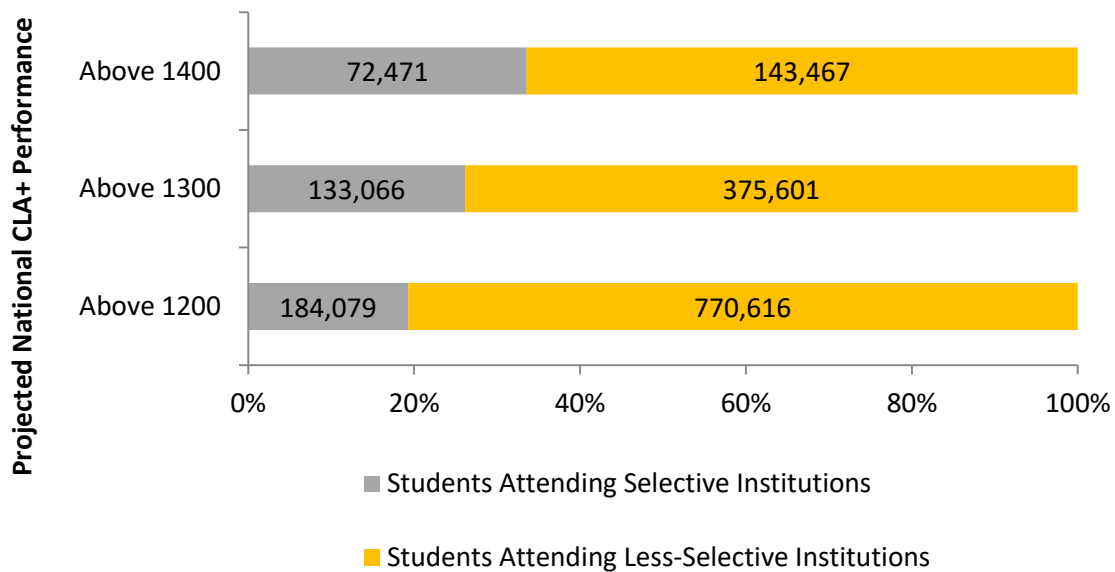

Figure 2: Projected national student attendance and CLA+ total scores

Grade inflation has made the undergraduate degree less useful in discriminating students' absolute skills so employers settle on for students from the selective (branded) colleges. The OECD finds similar problems in their research grade inflation globally (OECD, 2016). This example illustrates just how many students, who are proficient in critical thinking, are potentially being overlooked because the playing field for hiring recent university graduates is not level. Many of these students, as shown in Figure 1, are from underrepresented groups, typically first-generation university students or ethnic minorities. These individuals would be great candidates to employ and also to help organizations fulfill their goal to increase diversity in the workplace. Oftentimes, organizations are not able to identify these individuals due to their unawareness that these talented and able candidates are attending these less-selective institutions. They go to the same set of institutions from which they have previously recruited, but this

doesn't help with the goal of increasing workforce diversity. CLA+ offers an opportunity to close this gap.

There are potentially millions of students graduating from these institutions (Benjamin, 2015) who are proficient in the skills that employers say they desire (Hart Research Associates, 2013, 2015). Given that there is increasing enrollment at these non- and less-selective institutions, which have higher proportions of minority students (Benjamin, 2015), employers should expand their recruitment searches beyond the elite colleges and universities in order to have a representative and diverse workforce.

Employers could also consider hiring students from varied fields of study for entry-level positions. It appears from that business majors are more likely to obtain full-time positions within three months of graduation (Zahner & James, 2016). However, these students have the lowest CLA scores and GPAs (Jeffrey T. Steedle & Michael Bradley, 2012). If employers want to hire candidates with generic skills (Hart Research Associates, 2013, 2015), they should seek candidates with degrees in the social sciences, humanities and languages and science and engineering fields of study. This will also increase diversity in the workplace since students coming from varied fields of study will bring different perspectives to a team, which has been shown to be true when applied to diversifying based on gender (Badal, 2014).

Findings from this study offer support for the conclusion that critical-thinking and written-communication skills are important in predicting career placement and workplace success (Arum & Roksa, 2014). Additionally, assessments like the CLA+, a well-established standardized assessment of critical-thinking and written-communication skills serve as both an effective instrument for identifying high-achieving students from less and non-competitive institutions and making their skills more visible to perspective employees. The high-performing students who attended non- and less-competitive institutions (Hoxby & Avery, 2012) do in fact have the same critical-thinking skills that can potentially lead to positive post-college outcomes as do their peers at competitive institutions. Future studies will include continued longitudinal tracking of this cohort of students as well as surveys of employers who have hired college graduates with verified critical-thinking and written-communication skills. This will corroborate evidence of the findings from this study, furthering the validity of these skills as predictors of post-college outcomes.

What do the employers think?
In a follow-up study to Zahner & James (2016), CAE asked the 2014 longitudinal survey participants for the contact information of their manager or graduate advisor. The eighty nine managers and advisors that agreed to participate in the follow-up were asked how important they felt analysis and problem solving, writing effectiveness, and writing mechanics were. They were also asked to rate the students on these same skills and rank the student on how well they do compared to other recent college graduates within their organization. Although only a small subset of participants provided this information, they were representative of the original cohort of students with the exception of having slightly higher average GPA and CLA+ scores (Table 1).

Table 1: Demographic descriptive statistics

	Employer Survey Students	All Participants
N	89	21,513
% Female	66.3	60.0
% White	66.3	59.2
% English primary language spoken at home	89.5	84.5
% Parent with at least bachelor's degree	66.2	51.9
Mean (St. Dev) cumulative GPA (out of 4.0)	3.37 (.45)	3.24 (.48)
Mean (St. Dev) SAT (or converted ACT)	1114 (153)	1066 (172)

The survey consisted of a series of questions (Table 2), regarding how important critical thinking and written communication skills are to successful performance by the student, how proficient the students are as measured by these skills, and how the students ranked in comparison to their peers in the workplace or graduate program.

Table 2: Employer survey questions

How important are the following skills to successful performance in the participant's position:	1 = Unimportant	2 = Of little importance	3 = Moderately important	4 = Important	5 = Very important
Analysis and Problem Solving					
Writing Effectiveness					
Writing Mechanics					
How would you rate the participant on the following skills:	1 = Unsatisfactory	2 = Needs Improvement	3 = Satisfactory	4 = Good	5 = Outstanding
Analysis and Problem Solving					
Writing Effectiveness					
Writing Mechanics					
Overall, where does the participant's performance rank compared to other recent college graduates in your workplace?	1 = Well below other employees	2 = Below other employees	3 = About the same as other employees	4 = Above other employees	5 = Well above other employees

Results indicate that employers and graduate advisors indeed find critical thinking and written communication skills, as measured by analysis and problem solving, writing effectiveness, and writing mechanics, important (Figure 4).
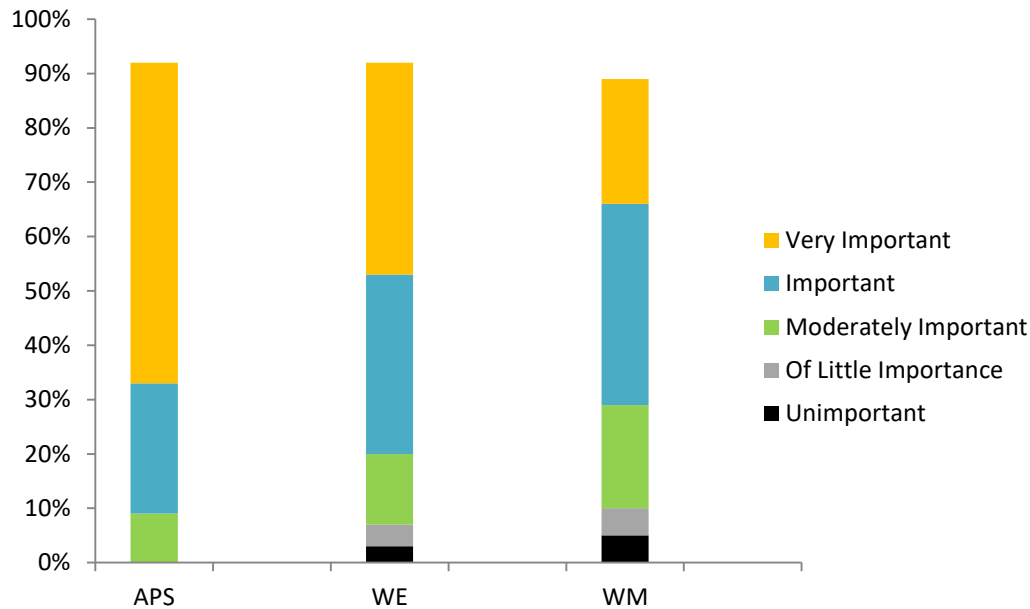
Figure 4: Distribution of responses to "Importance" questions (from Zahner & Lehrfeld, under review)

Additionally, Table 3 shows the ordinal logistic regression coefficients, their standard errors, 95% confidence intervals, and the *t*-statistics ($p < .001$ for all analyses). The regression coefficients can be interpreted as the log-odds of being rated higher given a 1-point increase in CLA+ total score. For instance, in the analysis and problem solving model, the estimated coefficient is given as .0033. Thus, for a 1-point increase in CLA+ total score, the log-odds of "jumping" to a higher rating category ("Good" instead of "Satisfactory or worse," or "Outstanding" instead of "Good") increases by .0033. The regression coefficients are small because CLA+ total scores are on a large scale (400-1600), so one extra point is not expected to make much of a difference. Two factors would increase the interpretability of the results: (1) using a more meaningful score increase, such as 50 points, and (2) converting the log-odds to odds by exponentiating the coefficient. Thus, if one student scores 50 points higher than a second student, the log-odds of being rated one category higher than the second student is 50*.0033 = .165, and the odds are exp(.165) = 1.18. This first student is 18% more likely than the second student to be rated one category higher ("Good" rather than "Satisfactory or worse," or "Outstanding" rather than "Good") due to the higher CLA+ total score.

The students with higher scores from their managers and advisors tended to have higher CLA+ scores. This is important to note because despite approximately 1.8 million individuals graduating within the United States each year (Hussar & Bailey, 2014), employers are still finding a skills gap (Hart Research Associates, 2015).

Table 3: Ordinal logistic regression models for predicting participants' post-college performance as measured by their managers or graduate advisors

Covariate	Est. Coefficient	Std. Error	t-statistic	95% CI	
				Lower	Upper
Analysis and Problem Solving					
CLA+ Score	.0033	.0002	14.33	.0029	.0038
Writing Effectiveness					
CLA+ Score	.0043	.0002	18.36	.0039	.0048
Writing Mechanics					
CLA+ Score	.0046	.0002	19.33	.0041	.0051
Rank Comparison of Participant					
CLA+ Score	.0049	.0002	22.18	.0045	.0053

Note: Estimated coefficients are log-odds of being rated one category higher given a one-point increase in CLA+ total score.

Findings from this study offer support for the conclusion that critical-thinking and written-communication skills are seen by hiring managers and employers as important for success in the workplace(Hart Research Associates, 2006, 2013). Additionally, it offers additional evidence that these skills are important in predicting career placement and workplace success (Arum & Roksa, 2014). CLA+ can serve as both an effective instrument for not only identifying those high-achieving students but also for making their critical thinking and written communication skills more visible to prospective employers, allowing them to identify skilled individuals who may be overlooked in a talent search if based solely on GPA and the brand name of their institution.

Implementation of CLA+ in an international context
Due to CAE's abundant research and development (Benjamin, 2008, 2014, 2015; Benjamin, Chun, & Jackson, 2009; Klein & Benjamin, 2008; Klein et al., 2007; Klein, Liu, et al., 2009; Klein, Steedle, & Kuglemass, 2009; Klein, Zahner, Benjamin, Bolus, & Steedle, 2013; Steedle, 2009, 2010a, 2010b; Jeffrey T. Steedle & M Bradley, 2012; Steedle & Elliot, 2012; Steedle, Zahner, & Kuglemass, 2014; Zahner, 2013, 2014; Zahner & James, 2016; Zahner, Kornhauser, Benjamin, Wolf, & Steedle, 2015; Zahner, Ramsaran, & Steedle, 2012; Zahner & Steedle, 2014, 2015) and the quality of the CLA+ instrument for measuring higher education students' generic skills, in March 2016, the OECD and CAE entered into an agreement to collaborate on CLA+ International. The goal was to attract a sufficient number of international participants for the first administration to support a strong research design. The design of the CLA+ International was intended to address some of the major challenges encountered in the AHELO Feasibility Study (OECD, 2011, 2012) (e.g., motivation, sampling, and scoring equivalency(Zahner & Steedle, 2014)).

In the international context, students, institutions, consortia, and ministries are facing the same challenges that their counterparts encounter domestically in the United States. More students are entering university than ever before (OECD, 2013). Grade inflation has been reported internationally as well (Bachan, 2017; OECD, 2016; Wissenschaftsrat, 2012). And graduating students are faced with the task of finding employment once they graduate. Anecdotally, countries have reported that their graduation rates for students is an issue due to the fact that students want to remain in school so they can continue to receive support. Students who are unable to obtain employment post-graduation would not necessarily be motivated to graduate. Perhaps employers globally are also finding it difficult to identify qualified students. An assessment like CLA+ International would be able to close the gap

between students seeking employment and employers seeking personnel by identifying those who are proficient in the requisite generic skills that are essential for success in the workplace.

CAE and the OECD completed implementation of CLA+ International for the 2017 – 2018 academic year. The process of implementing CLA+ International was developed following best practice recommendations (AERA, APA, & NCME, 2014) as well as improvements and modifications to address some of the limitation of the AHELO feasibility study (AHELO, 2012a, 2012b).

Sampling

Since the goal of CLA+ International is not as a feasibility study, not all participating countries and institutions need to follow the same exact model of assessing their students' generic skills. CLA+ International allows for multiple sampling models to accomplish the specific goals of the institutions and ministries.

If participants are interested in using the cross-sectional model, then CAE recommends sampling and testing 200 entering students and 400 exiting students. This will yield an institutional-level measure of student learning gains within one academic year.

If participants are interested in using the longitudinal model, then CAE recommends sampling and testing 600 entering students tracked longitudinally. This will yield both student and institutional measures of student learning gains. However, this will take up to three or four academic years to yield results.

A hybrid sampling plan is also recommended which requires a minimum of two years. During the first year, institutions would sample 500 incoming (longitudinal cohort) and 100 exiting (cross-sectional cohort) students. The subsequent years would require sampling a minimum of 100 entering and 100 exiting students to maintain the cross-sectional as well as the persisting 500 students from the original longitudinal cohort. This model will yield both student and institutional measures of student learning gains, but institutional measures will be available within one academic year.

Translation & Adaptation

Translation and adaptation of a performance assessment is a process that is more complex than a simple word-for-word replacement from one language to another. Translation and adaptation experts must ensure that the assessments are not only consistent with the original version in the source language, but also that the tasks will be interpreted by students in their native language as the developers intended. These experts will confirm that the assessment topics maintain their authenticity and meaning for the target student population as they do for the original student population for which the tasks were initially developed.

CAE uses an internationally accepted five step translation process that is in compliance with International Translation Committee (ITC) guidelines (Gregoire, 2018). CAE follows the guidelines used for the localization process of major international studies such as PISA, TIMSS, PIRLS, PIAAC, and AHELO. The process includes translatability review, double translation and reconciliation, client review, focused verification, and cognitive labs.

During the translatability review, source material is reviewed to confirm that the text will adapt well to the native language and culture. Particular attention is paid to disambiguation of source, respecting key correspondences between stimuli and questions, and deciding what should or should not be adapted to

local context. Two independent translators then review the text and provide translations. The translations are reconciled and sent to the Lead Project Manager for review and an opportunity to provide minor suggestions. The translated CLA+ items are then sent for a focused verification. Cognitive labs are then carried out with the assistance of participating institutions and the institutional teams to ensure the translation and adaptation process was effective.

Administration
CLA+ International is administered through an Internet-based test platform. Students enter the exam through a secure browser that locks down computer functions and distributes a 60-minute PT and a 30-minute, 25-item SRQ section to each student. All testing sessions require a proctor to authorize students into the interface and manage the testing environment. The assessment is designed to be completed in approximately 90 minutes. The assessment will include an optional tutorial that students can scroll through. The assessment must be administered under standardized, controlled testing conditions. CAE will provide training materials for Institutional Administrators and Proctors.

Scoring
For CLA+ International, all responses are double-scored by human scorers. CAE staff direct the training for the scoring process. This starts with an in-person group training for lead scorers from all participating institutions. This training will be conducted in English with American student exemplary responses. The lead scorer then undergoes rigorous training in order to become a CLA+ International scorer. Training includes an orientation to the prompts and scoring rubrics/guides, repeated practice grading a wide range of student responses, and extensive feedback and discussion after scoring each response.

Following the training, CAE serves as a resource for the lead scorer, who is responsible for recruiting and training the in-country team of scorers. This ensures quality and consistency both within and across institutions. The scorers should be recruited from participating institutions and will need to be able to judge university student generic skills. Institutions will often appoint professors, institutional research fellows, post-doctoral associates, or doctoral students to score the student responses. Once trained, the scorers will receive a randomized selection of anonymized student responses and will enter the score results directly into CAE's Internet-based scoring platform. The scorers will not know the institution to which each student belongs. CAE's system automatically monitors human scorer calibration and inter-rater reliability and notifies the lead scorers of any problems through the E-Verification system.

The E-Verification system was developed to improve and streamline scoring. Calibration of scorers through the E-Verification system requires scorers to score previously-scored results, or "verification papers," when they first start scoring, as well as throughout the scoring window. The system periodically presents verification papers to scorers in lieu of unscored student responses, though they are not flagged to the scorers as such. The system does not indicate when a scorer has successfully scored a verification paper, but if the scorer fails to accurately score a series of verification papers, he or she will be removed from scoring and must participate in a remediation process. At this point, scorers are either further coached or removed from scoring.

Subscores are assigned on a scale of 1 (lowest) to 6 (highest). Blank responses or responses that are entirely unrelated to the task are flagged for removal from results.

Reporting
Each participating institution receives its own set of reports and data files. If agreed upon by all participating institutions, a collective data file for all students from all participating institutions within

the country, along with comparative information of CAE's domestic national data, will be provided. In addition, individual institutional reports and data files for each participating institution, as well as comparative data across all institutions within the country, will be prepared. Finally, there are individual student reports for all participating students.

In addition to these reports, for any new items or forms that are piloted, CAE provides a detailed item analysis of the native language SRQ. CAE also provides scaling and equating analyses for the PT and SRQs of the national forms to the English language forms. No data will be shared without the explicit permission of a particular institution.

Standard Setting
CAE will conduct a standard-setting validation study for one form of CLA+ (one PT and 25 Selected-Response Questions—one CRE set, one SQR set, one CA set) based upon the already established standards for CLA+ (Zahner, 2014). The design and execution of the standard setting study for CLA+ International will be consistent with procedures adopted in the Standards for Educational and Psychological Testing (AERA et al., 2014). This study will require a meeting, facilitated by CAE, with subject matter experts (SMEs) selected collectively by the participating institutions, who will review student responses and determine various cut scores for the tests using the Bookmark (Lewis, Mitzel, Green, & Patz, 1999) standard setting procedure. The SMEs will consist of professionals who have experience working with graduating university students who are entering the workforce or continuing their studies in graduate school. The goal of the standard setting study is to obtain a consensus on the international standards of mastery of generic skills that students need. CAE recognizes that this may be an extremely challenging task and is therefore open to having country-specific standard setting studies. Although, this approach also has limitations since it will not be possible to easily compare these results cross-nationally.

International Benchmarking & Cross-Country Comparisons
By participating in CLA+ International, participating institutions have the opportunity to obtain international benchmarking information (for comparisons beyond their native country and the United States). Participating countries who are interested in international benchmarking and cross-country comparisons will have the opportunity to share their results and collaborate on research projects. CAE has a rich research history of investigating international comparative studies (Benjamin et al., 2012; Klein et al., 2013; Wolf & Zahner, 2015; Wolf, Zahner, & Benjamin, 2015; Wolf, Zahner, Kostoris, & Benjamin, 2014; Zahner & Kostoris, 2016; Zahner & Steedle, 2014) and hopes to continue this line of research with global partners. No data will be shared without the explicit permission of a particular institution or country.

Certificates & Badging
Using the results from the standard setting study, students earning Proficient, Accomplished, or Advanced scores will be eligible to receive digital badges (Figure 5). These badges will allow students to showcase their generic skills achievement levels to potential employers. They will also be able to share their CLA+ scores directly on employment boards, through social media, and on their resumes.

Figure 5 : CLA+ Mastery Level Badges

Research & Development Efforts
In addition to the administration of CLA+ International, all participating institutions will have the opportunity to contribute to the ongoing research and development initiative at CAE. This will include identifying the topics for future PTs and SRQs and innovating new assessments for additional generic skills, such as creativity and collaboration. There will also be opportunities for participants to jointly author research articles and chapters on their collaborations (e.g., Zahner & Ciolfi, 2018).

Participation
The CLA+ International initiative functions as a loose network of institutions, systems and countries, supported jointly by the OECD and CAE. No formal process of registration or accession is in place. By the time of drafting this chapter the CLA+ International initiative connects the following institutions, systems and countries. Finland has decided to do a system-wide implementation of the CLA+ in 2019. Two language versions (Finnish and Swedish) are being developed and tested. In Italy, the national quality assurance and accreditation agency ANVUR has implemented the CLA+ in two rounds of the TECO project and is continuing using a modified version of CLA+ in its assessment tools. Hungary is developing a large-scale study of the learning outcomes of its students, using a short version of the CLA+. A pilot will be conducted in the Spring of 2019. In Mexico, the University of Guadalajara is testing several cohorts of students with CLA+. A group of universities in Chile, Peru, Paraguay and Colombia is also starting testing. In England, a group of post-1992 universities is using the CLA+ within the Teaching Excellence Framework. There is strong interest in Croatia and Ukraine to implement the CLA+ International. Discussions have been ongoing with countries such as the Russian Federation, China, Singapore, the Slovak Republic and Korea.

Summary and Conclusion
The collaboration between the OECD and CAE in the development and implementation of the CLA+ International certainly is not the only promising initiative in the field of assessing higher education students' and graduates' learning outcomes. At an international level the situation probably can be best described as one in which institutions, systems and countries are exploring various conceptual and methodological approaches. It is clear that the interest is growing, not at least with governments and stakeholders in higher education. At the same time there continues to be resistance and outright opposition among higher education institutions.

Compared with alternative and competing approaches and initiatives the CLA+ International stands out in several aspects. It is based on a clear argument about generic skills assessment. It focuses on assessing those skills where higher education is supposed to make a real difference, while at the same time being critically important in the context of skills demand in the knowledge economy. The assessment connects very well to core academic values and to what the higher education community

itself considers to be its added value. It has a long history of development and fine-tuning, as well as a long and successful history of implementation in the US higher education system. The assessment methodology, as well as the translation, adaptation, implementation and scoring are based on state-of-the-art approaches. Institutions, systems and countries participating can easily benchmark their results against a rich and growing database of assessment data for the US and other countries.

At the same time there still are hurdles to overcome. In several countries motivating students to participate to the best of their abilities can be a challenge, and certificates and badges are not yet familiar enough to students and employers, but might become so in the future. The support of employers in recognizing the value of the assessment and its outcomes both at system level and individual level will be critically important.

Still, the benefits for institutions and systems are clear. Assessing students' and graduates' learning outcomes offers an invaluable point of comparison with the internal selection, assessment, examination and certification processes and can add tremendous value to evaluating the quality and equity of these processes. To institutions it offers a unique external point of reference for internal quality improvement. To students it offers an additional attestation of skills that are valued by employers, in addition to the certification acquired from university. Finally, to governments and external stakeholders it offers an exceptional rich dataset to assess the quality, equity and value-added of higher education institutions and systems.

References

AERA, APA, & NCME. (2014). *Standards of Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

AHELO. (2012a). *AHELO feasibility study interim report*. Retrieved from Paris:

AHELO. (2012b). Testing students and university performance globally: OECD's AHELO. Retrieved from http://www.oecd.org/edu/skills-beyond-school/testingstudentanduniversityperformancegloballyoecdsahelo.htm

Arum, R., & Roksa, J. (2014). *Aspiring Adults Adrift*. Chicago, IL: University of Chicago Press.

Bachan, R. (2017). Grade inflation in UK higher education. *Studies in Higher Education, 42*(8), 1580-1600.

Badal, S. B. (2014). The business benefits of gender diversity. Retrieved from http://www.gallup.com/businessjournal/166220/business-benefits-gender-diversity.aspx

Barron's. (2014). *Barron's profiles of American colleges* (Vol. 31): Barron's Educational Series.

Benjamin, R. (2008). *The Contribution of the Collegiate Learning Assessment To Teaching and Learning*. Council for Aid to Education. New York, NY.

Benjamin, R. (2014). Two questions about critical thinking tests in higher education. *Change: The Magazine of Higher Learning, 46*(2), 32-39.

Benjamin, R. (2015). *Leveling the Playing Field From College to Career*. Retrieved from New York:

Benjamin, R., Chun, M., & Jackson, C. (2009). *The Collegiate Learning Assessment's place in the new assessment and accountability space*. Retrieved from New York:

Benjamin, R., & Klein, S. (2006). *Assessment Versus Accountability in Higher Education: Notes for Reconciliation*. Council for Aid to Education. New York, NY.

Benjamin, R., Klein, S., Steedle, J. T., Zahner, D., Elliot, S., & Patterson, J. A. (2012). The case for generic skills and performance assessment in the United States and international settings. *CAE - Occasional Paper*. Retrieved from http://www.collegiatelearningassessment.org/files/The_Case_for_Generic_Skills_and_Performance_Assessment_in_the_United_States_and_International_Settings.pdf

Clanchy, J., & Ballard, B. (1995). Generic skills in the context of higher education. *Higher Education Research and Development, 14*(2), 155-166.

Crebert, G., Bates, M., Bell, B., Patrick, C. J., & Cragnolini, V. (2004). Developing generic skills at university, during work placement, and employment: graduates' perceptions. *Higher Education Research and Development, 23*(2), 147-165.

Gregoire, J. (2018). ITC guidelines for translating and adapting tests. *International Journal of Testing, 18*(2), 101-134.

Hart Research Associates. (2006). *How Should Colleges Prepare Students to Succeed in Today's Global Economy? - Based on Surveys Among Employers and Recent College Graduates*. Retrieved from Washington, DC:

Hart Research Associates. (2013). *It takes more than a major: Employer priorities for college learning and student success*. Retrieved from Washington, DC: https://www.aacu.org/leap/documents/2013_EmployerSurvey.pdf

Hart Research Associates. (2015). *Falling short? College learning and career success*. Retrieved from Washington, DC: http://www.aacu.org/sites/default/files/files/LEAP/2015employerstudentsurvey.pdf

Hoxby, C. M., & Avery, C. (2012). *The Missing" One-Offs": The Hidden Supply of High-Achieving, Low Income Students*. Retrieved from

Hussar, W. J., & Bailey, T. M. (2014). *Projections of education statistics to 2022*. Retrieved from Washington, DC: https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2014051

Kearns, P. (2001). *Generic Skills for the New Economy. Review of Research*. 252 Kensington Road, Leabrook, South Australia 5068, Australia: National Centre for Vocational Education Research.

Klein, S., & Benjamin, R. (2008). *The Collegiate Learning Assessment (CLA)*. Council for Aid to Education. New York, NY.

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., Bolus, R., Bridgeman, B., Kugelmass, H., . . . Steedle, J. T. (2009). Test Validity Study (TVS) Report. Supported by the Fund for the Improvement of Postsecondary Education.   Retrieved from http://www.cae.org/content/pdf/TVS_Report.pdf

Klein, S., Steedle, J. T., & Kugelmass, H. (2009). *CLA Lumina Longitudinal Study Summary Findings*. Retrieved from New York:

Klein, S., Zahner, D., Benjamin, R., Bolus, R., & Steedle, J. T. (2013). *Observations on AHELO's Generic Skills Strand Methodology and Findings*. Retrieved from

Lazear, E. P. (2006). Speeding, terrorism, and teaching to the test. *The Quarterley Journal of Economics, 121*(3), 1029-1061.

Levy, F., & Murname, R. J. (2004). Education and the Changing Job Market: An Education Centered on Complex Thinking and Communicating is a Graduate's Passport to Prosperity. *Educational Leadership, 62*(2), 80-83.

Lewis, D. M., Mitzel, H. C., Green, D. R., & Patz, R. J. (1999). *The Bookmark standard setting procedure*. Monterey: McGraw-Hill.

Moloney, K. (2006). Teaching to the test. *International Journal of Learning, 13*(6).

Montt, G. (2015). *The causes and consequences of field-of-study mismatch*. Retrieved from Paris, France:

OECD. (2011). *AHELO feasibility study progress report* Retrieved from Paris, France:

OECD. (2012). *AHELO Feasibility Study Report*. Retrieved from Paris, France:

OECD. (2013). *How are univerversiyt students changing?* Retrieved from Paris:

OECD. (2016). *Getting Skills Right: Assessing and Anticipating Changing Skill Needs*: OECD Publishing.

Popham, W. J. (2001). Teaching to the test? *Educational Leadership, 58*(6), 16-21.

Steedle, J. T. (2009). *Advancing Institutional Value-Added Score Estimation (draft)*. Retrieved from New York, NY:

Steedle, J. T. (2010a). *Improving the Reliability and Interpretability of Value-Added Scores for Post-Secondary Institutional Assessment Programs.* Paper presented at the Annual Meeting of the American Educational Research Association, Denver, CO.

Steedle, J. T. (2010b). *Incentives, Motivation, and Performance on a Low-Stakes Test of College Learning.* Paper presented at the Annual Meeting of the American Educational Research Association, Denver, CO.

Steedle, J. T. (2012). Selecting value-added models for postsecondary institutional assessment. *Assessment & Evaluation in Higher Education, 37*(6), 637-652.

Steedle, J. T., & Bradley, M. (2012). *Majors matter: Differential performance on a test of general college outcomes.* Paper presented at the Annual Meeting of the American Educational Research Association, Vancouver, Canada.

Steedle, J. T., & Bradley, M. (2012). *Majors Matter: Differential Performance on a Test of General College Outcomes.* Paper presented at the Annual Meeting of the American Educational Research Association, Vancouver, Canada.

Steedle, J. T., & Elliot, S. (2012). *The Efficacy of Automated Essay Scoring for Evaluating Student Responses to Complex Critical Thinking Performance Tasks*. Retrieved from New York, NY: http://www.collegiatelearningassessment.org/files/Steedle__2012__The_Efficacy_of_Automated_Essay_Scoring.pdf

Steedle, J. T., Zahner, D., & Kuglemass, H. (2014). *Test administration procedures and their relationship with effort and performance on a college outcome test.* Paper presented at the American Educational Research Associatoin, Philadelphia, PA.

Van Damme, D. (2015). Global higher education in need of more and better learning metrics. Why OECD's AHELO project might help to fill the gap. *European Journal of Higher Education, 5*(4), 425-436.

Volante, L. (2004). Teaching to the test: What every educator and policy-maker should know. *Canadian Journal of Educational Administration and Policy*.

Wissenschaftsrat. (2012). *Prüfungsnoten an Hochschulen im Prüfungsjahr 2010. Arbeitsbericht mit einem Wissenschaftspolitischen Kommentar.* Retrieved from Germany: https://www.wissenschaftsrat.de/download/archiv/2627-12.pdf

Wolf, R., & Zahner, D. (2015). Mitigation of test bias in international, cross-cultural assessments of higher-order thinking skills. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of research on technology tools for real-world skill development*. Hershey, PA: IGI Global.

Wolf, R., Zahner, D., & Benjamin, R. (2015). Methodological challenges in international comparative post-secondary assessment programs: lessons learned and the road ahead. *Studies in Higher Education*, 1-11. doi:10.1080/03075079.2015.1004239

Wolf, R., Zahner, D., Kostoris, F., & Benjamin, R. (2014). *A Case Study of an International Performance-Based Assessment of Critical Thinking Skills*. Paper presented at the American Educational Research Association, Philadelphia, PA.

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

Zahner, D. (2014). *CLA+ standard setting study final report*. Retrieved from New York, NY: http://cae.org/images/uploads/pdf/cla_ss.pdf

Zahner, D., & Ciolfi, A. (2018). International Comparison of a Performance-Based Assessment in Higher Education. In O. Zlatkin-Troitschanskaia, M. Toepper, H. A. Pant, C. Lautenbach, & C. Kuhn (Eds.), *Assessment of Learning Outcomes in Higher Education* (pp. 215-244). Cham: Springer.

Zahner, D., & James, J. K. (2016). *Predictive validity of a critical thinking assessment of post-college outcomes.* Paper presented at the American Educational Research Association, Washington, DC.

Zahner, D., Kornhauser, Z. G. C., Benjamin, R., Wolf, R., & Steedle, J. T. (2015). Using the Collegiate Learning Assessment to address the college to career space. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), Handbook of research on computational tools for real-world skill development. Hershey, PA: IGI Global.

Zahner, D., & Kostoris, F. (2016). *International testing of a performance-based assessment of higher education.* Paper presented at the American Educational Research Association, Washington, DC.

Zahner, D., Ramsaran, L. M., & Steedle, J. T. (2012). *Comparing alternatives in the prediction of college success.* Paper presented at the Annual Meeting of the American Educational Research Association, Vancouver, Canada.

Zahner, D., & Steedle, J. T. (2014). *Evaluating Performance Task Scoring Comparability in an International Testing Program*. Paper presented at the American Educational Research Association, Philadelphia, PA.

Zahner, D., & Steedle, J. T. (2015). *Comparing longitudinal and cross-sectional school effect estimates in postsecondary education.* Paper presented at the National Council on Measurement in Education, Chicago, IL.