

**DIRECTORATE FOR EDUCATION  
INSTITUTIONAL MANAGEMENT IN HIGHER EDUCATION GOVERNING BOARD**

**Group of National Experts on the AHELO Feasibility Study**

**PROJECT PLAN FOR THE ADAPTATION OF THE CLA INSTRUMENT FOR INTERNATIONAL IMPLEMENTATION**

**27-28 April 2009  
OECD Conference Centre, 2 rue André-Pascal, 75016 Paris**

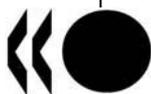
*The AHELO GNE is invited to:*

- *TAKE NOTE, COMMENT and AGREE on proposed project plan; and*
- *AGREE to contract the Council for Aid to Education (CAE) as the sole provider to develop instruments towards an international pilot test of the Collegiate Learning Assessment (CLA), being aware that this is a derogation from the standard OECD call for tender procedure.*

*This document is available in pdf format only.*

Roger Benjamin (Collegiate Learning Assessment, E-mail: [roger@cae.org](mailto:roger@cae.org)) or Karine Tremblay,  
Directorate for Education, Education Management and Infrastructure division  
[Tel: +33 1 45 24 91 82; E-mail: [Karine.Tremblay@oecd.org](mailto:Karine.Tremblay@oecd.org)]

**JT03263364**



## **PROJECT PLAN FOR THE ADAPTATION OF THE CLA INSTRUMENT FOR INTERNATIONAL IMPLEMENTATION**

1. At its 17-18 December meeting, the AHELO GNE agreed to use the Collegiate Learning Assessment (CLA) instrument for the generic skills strand of the AHELO feasibility study, subject to a number of conditions set out in the summary record of the meeting [EDU/IMHE/AHELO/GNE/M(2008)1/REV2].
2. Accordingly, the OECD Secretariat asked the Council of Aid to Education – which developed the CLA instrument – to prepare a project plan and budget for the adaptation of the CLA instrument for international implementation.
3. The first project proposal was reviewed internally and submitted to independent experts for technical feedback<sup>1</sup>. This document presents the revised project plan based on the independent experts' recommendations and advice.
4. The AHELO GNE is invited to:
  - TAKE NOTE, COMMENT and AGREE on proposed project plan; and
  - AGREE to contract the Council for Aid to Education (CAE) as the sole provider to develop instruments towards an international pilot test of the Collegiate Learning Assessment (CLA), being aware that this is a derogation from the standard OECD call for tender procedure.

---

<sup>1</sup> The experts consulted are Diane Pennock and Pierre Brochu – who have both been extensively involved in PISA implementation in Canada.



**CLA Project Plan and Budget  
Assessment of Higher Education Learning Outcomes (AHELO):  
Generic Skills Strand**

**April 20, 2009**

**Table of Contents**

<b>Overview</b>	<b>2</b>
<b>Summary of Work Plan</b>	<b>4-8</b>
<b>Assumptions</b>	<b>4-7</b>
<b>Project Organization</b>	<b>7-8</b>
<b>Phases</b>	
<b>Pre-Start Up</b>	<b>8</b>
<b>Phase 1 (3 months, July 1-October 31, 2009)</b>	<b>9-10</b>
<b>Milestones</b>	<b>10</b>
<b>Phase 2 Assessment Task Adaptation (13 months, October 1, 2009—October 31, 2010)</b>	<b>11-15</b>
<b>2.1 Introduction and Task Selection and Adaptation (1 month, October 2009)</b>	<b>1-12</b>
<b>Milestones</b>	<b>12</b>
<b>2.2 Task Adaptation (3 months, November 1, 2009-January 31, 2010)</b>	<b>12-13</b>

<b>2.3 Task Adaptation and Translation (9 months)</b>	<b>13</b>
<b>2.4 Feasibility Study Design (9 months, co-terminus with phases 2.2/2.3)</b>	<b>13-14</b>
<b>Milestones</b>	<b>14</b>
<b>2.5 Test Translation and Preparation for Testing Phase (3 months, July 1, 2010-October 31, 2010)</b>	<b>14-15</b>
<b>Milestones</b>	<b>15</b>
<b>Roles and Responsibilities</b>	<b>15-16</b>
<b>Publications and Projects Related To the Collegiate Learning Assessment</b>	<b>16-18</b>
<b>Appendix: Translation and Translation Review</b>	<b>18-21</b>

### **Overview**

The proposed work plan and accompanying proposed budget presented here are designed to accomplish the goals of the generic skills strand of the OECD's AHELO feasibility study adapting the Collegiate Learning Assessment (CLA) for the international arena. The central question to be answered is whether performance assessments, developed in one national cultural context and in one language, can be adapted for use in other nations with distinctly different cultures and languages. The study also seeks to shed light on how closely related is the CLA to another, reliable and valid (multiple-choice) measure of critical thinking, in terms of testing time, student performance (scores) *and* student cognition (thinking).

Performance assessments are realistic work samples, case studies, or real world scenarios presented to students to analyze. They differ from multiple choice or short answer tests

in that the student is provided documents, figures, tables, graphs, poems and pictures to answer a set of questions. There are no “right” or “wrong” answers; rather the student is encouraged to develop a well reasoned response. The CLA performance tasks are intended to measure critical thinking, problem solving, analytical reasoning, and written communication, often called higher order skills in the United States. Consequently, they meet the requirements of the concept of generic skills as defined by the OECD. The strategy to be employed is to see whether the pool of performance tasks developed for the CLA can be adapted for reliable and valid use in very different languages and cultural contexts.

The CLA has been offered widely in the United States since 2003 with hundreds of colleges and universities and many thousands of students participating in the CLA testing program. As with any major national testing program the assessment instruments used must meet the highest standards of reliability and validity, and, in the case of the CLA, be engaging to the students and understood to be authentic by faculty. Readers interested in scientific documentation of the characteristics of the CLA—correlations between the measures, reliability of the scoring system, validity and the value added results achieved—are invited to access the methodological publications listed in the bibliography.

In this feasibility study we need to collect evidence that the CLA tasks as adapted for use in Finland, Korea, Mexico, and Norway are working cross-nationally as intended. More specifically, the study seeks to determine if the CLA tasks, adapted and translated for use in all four countries (with a formal translation monitoring process), measure in each language and culture the same critical thinking, analytic reasoning, problem solving and communication skills (indexed by students’ scores and patterns of thinking), and that the CLA scores correlate positively and moderately highly with a multiple-choice measure of critical thinking. Moreover, the study needs to collect evidence about sampling and administration feasibility, technical qualities, and interpretation of findings from this diverse set of nations (e.g. see OECD AHELO/GNE working papers 12 and 13). The study also needs to assess whether colleagues in each of these four countries can

successfully adapt candidate CLA performance assessments so as to work cross-nationally.<sup>1</sup> Benchmarks of success for stages of this study are noted below.

The generic skills to be focused on are one important component of the AHELO feasibility study. In an age in which one can “google” for facts, it is critical to improve citizens’ ability to access, structure and use information. This is a central goal common to all member countries of the OECD whose leaders recognize that human capital is a most vital resource. Thus, the goal to benchmark the successful development of generic skills by colleges and universities across all member countries is worth pursuing.

### **Summary of Work Plan**

The purpose of this document is to sketch a plan and time line for the Council for Aid to Education’ (CAE) carrying out the Generic Skills Strand of the Assessment of Higher Education Learning Outcomes (AHELO) Feasibility Study. The proposal highlights major tasks as currently conceived and major outcomes to be achieved in each phase of the study.

### **Benchmarks of Success for the Feasibility Study**

The feasibility study will be deemed a success if we have a clear decision as to whether:

- a) It is feasible technically, linguistically, logistically, economically, and culturally to field a reliable, valid and useful cross-national assessment with the CLA;
- b) National assessment teams can be trained to adapt CLA performance tasks demonstrating the possibility of a large-scale study with multiple performance tasks, and whether we have a firm grasp on how to train other countries to develop tasks on a large scale basis economically.<sup>2</sup>

---

<sup>1</sup> Problems in test adaptation, test translation, and international comparisons loom large in a study of this nature. In addition to the expertise offered by Dr. Richard Shavelson, the leader of the research team, this is an area of expertise of Dr. Guillermo Solano-Flores, one of the principal members of the team. See his resume attached.

<sup>2</sup> Definitions of the key terms are as follows. A) Technically refers to producing reliable and valid scores and score interpretation. We will examine reliability using generalizability theory. We will examine validity in a number of ways including correlations among performance tests and the multiple-choice test of critical thinking. (Our current construct validation study suggests correlations of about .60.) and verbal think-aloud protocols supporting the interpretation higher order skills are implicated in performing the performance tasks across the four participating countries. B) Linguistically refers to the challenge presented by translating and preserving the same construct across four languages. In part, the validity

## Assumptions

In developing the plan and time line, several assumptions were made and are enumerated here at the outset.

1. The duration of the study has been shortened as requested by OECD and assumes that the work will begin no later than July 1, 2009. Consequently we will modify and use only currently existing CLA tasks for assessing student learning in the feasibility study, as agreed upon at the generic skills strand project's first meeting with the countries. This will provide for ample experience in learning how to modify such performance tasks (PTs) to meet the possibly varying considerations brought by each of the participating countries.
2. The approach taken in (1) above will not, however, permit us to see how feasible it is to have participating countries contribute original PTs for consideration by participating countries for inclusion in the final assessment. To address this limitation, and as called for in the Pennock and Brochu review, we could implement a separate assessment-task development feasibility study at the conclusion of this study.
3. Participating countries agree to deliver the agreed upon version of the CLA via the Internet in a proctored environment. This is the only possible CLA delivery method especially given the constraints of the budget and timeline. We will work with each of the four participating countries to find a way to make this work. Based on our assessment each country has the potential capacity to deliver the CLA on an internet platform. By the time of any widespread use of the AHELO by the 30 member OECD countries internet infrastructure will likely be widely available in all OECD countries.
4. Participating countries agree to collaborate with CAE in putting together the assessment team for each country and carrying out agreed upon tasks in a timely fashion to meet the strict time line.<sup>3</sup> Each Country Team will ideally consist of the country representative to OECD along with one or two specialists in assessment of

---

information in A will bear n success. And agreement among experts from the participating countries (assessment adaptation group, test translation committee, test review committee) as to the adequacy of translation will provide evidence. C Logistically refers to the delivery, data collection, and scorer training and scoring. Etc. D Economically is self-explanatory. We can provide cost estimates and AHELO can decide what is feasible. E Culturally refers to the fact that the countries differ from one another on facts other than language and within a country where there is typically regional variation.

<sup>3</sup> From October 1,2009 to October 31, 2010, the Country Team time commitment will be as follows: About 12 days of meetings in New York including travel time plus work at home: Country AHELO Rep: 12 days meetings + 10 days = 25 days; Experts: 12 days meetings each + 35 days each working on tasks at home= 50 days each total

learning from universities or research institutes following agreed upon criteria for selecting these specialists. The representatives will sign confidentiality agreements with CAE to protect CAE's proprietary methods and other intellectual property. The participating countries will be responsible for any costs of travel, other expense reimbursement, and compensation (if any) of the country teams.

5. Participating countries agree, in collaboration with CAE, to assist in identifying translators following agreed upon OECD/CAE-developed translation protocol. (OECD protocol plus CAE standards.)
6. Participating countries agree, in collaboration with CAE, to assist in identifying scorers who will be trained and monitored by CAE to evaluate student performance on the CLA/AHELO in accordance to procedures established and disseminated by CAE. Note that it would be desirable to have translators bilingual in two of the four languages represented by the feasibility study countries. However, this may not be practical or possible.
7. While CAE's value-added approach to learning assessment is an important feature of the CLA, we assume that value-added estimation will not be implemented in the feasibility study. (We understand that another part of the AHELO feasibility study will address value added generally.)
8. While it is technically possible, the students' PT responses in the feasibility study will not be machine scored. In the future, these responses might be used to test machine scoring in four languages for a large scale implementation of AHELO, but that is beyond the scope of this plan.
9. Only one language per country will be chosen for the field test.
10. We assume that all countries will participate in the additional multiple choice test to be delivered via the Internet to some or all students taking the PT. The question of whether all students will take all tests or whether a matrix sampling design will be used will be decided in the early stages of Phase I. All four countries should use the same multiple-choice test. Otherwise, it will not be possible to compare one country to the others. Advice should and will be sought from the country teams about which test they prefer.
11. It is understood that other contractor(s) will be dealing with sampling and translations/adaptations for the other strands of the feasibility study. CAE will share its work and experience in these issues in order to build synergies across strands and generate economies of scale. It is understood that countries may opt to participate in

the contextual strand and that in such cases, students participating in the generic skills strand will also have to respond to a 15 minutes background questionnaire.

### Project Organization

The generic skills strand feasibility study will be conducted by CAE and representatives of the four participating countries. An initial organization chart for the study is provided in Figure 1. Roger Benjamin, President of CAE, will serve as the principal investigator and have overall responsibility for the project. Richard Shavelson, a consultant to CAE, will serve as Project Director. He will have overall day-to-day and technical responsibility for running the project (assisted by Amy Kurpius, Program Associate). A Technical Advisory Group (TAG) will advise the Project Director. Dr. Ron Hambleton will serve on the TAG. An additional two members for the TAG will be drawn from the larger group of advisors to the emerging CAE Performance Assessment Institute: Dr. Richard Duran, Professor of Education, University of California, Santa Barbara; Dr. George Engelhard, professor of educational measurement, Emory University; Dr. Edward Haertel, Professor of Psychological Studies in Education, Stanford University; Dr. Larry Hedges, Professor of Statistics and Social Policy, Institute for Policy Research, Northwestern University; Dr. Paul Holland, chief scientist emeritus, ETS; Dr. Robert Linn, Professor Emeritus, University of Colorado at Boulder (willing to consult but does not travel); Dr. William Mehrens, Prof of Ed Psych emeritus Michigan State University; Dr. Paul Sackett, Professor of Psychology, University of Minnesota; Dr. Matt Sera, Director of Institutional Research, Duke University; Dr. Laurie Wise, Human Resources Research Organization. In addition, the AHELO staff is asked to suggest an additional member of the technical advisory group from non U.S. members of OECD.

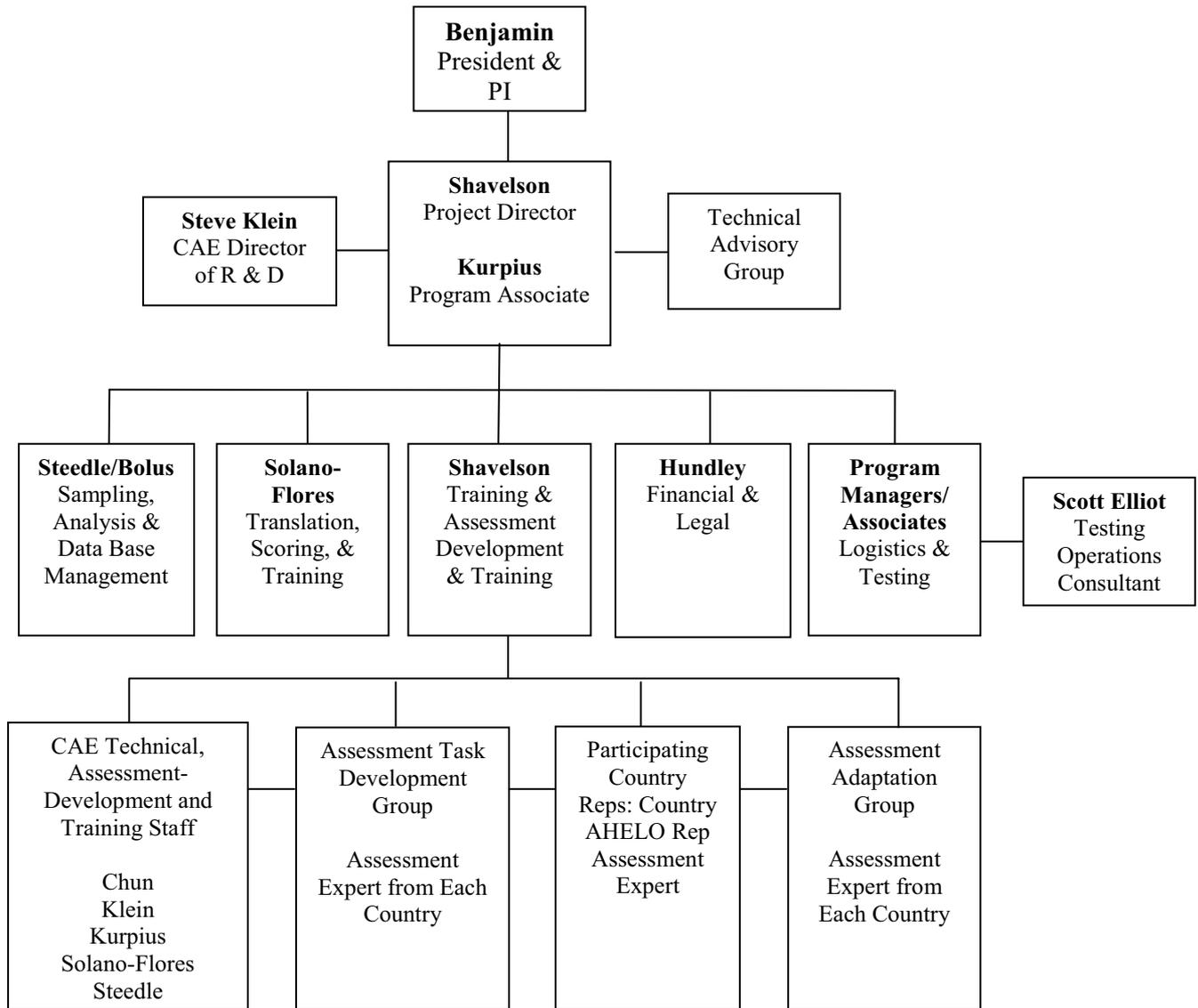


Figure 1. Project organization chart

**Phase 0. Pre-Start-up Planning and Organization (5 months) February 1, 2009 – June 30, 2009**

- Complete preliminary project plan, time line, and budget.
- Identify CAE and consultant staff and conceive project organization.
- Prepare for start-up.
- **Milestones:**
  - Plan, time line, and budget submitted to and approved by OECD
  - Richard Shavelson, Project Director, and CAE staff and consultants ready for start up

### **Phase 1. Start-up: Three Months July 1, 2009 – September 30, 2009**

- Operational planning for all phases
- Establish criteria for country representative selection (see assumptions above).

We assume two or three representatives will be available from each country. One representative will be the country's AHELO representative who will function on all matters of policy, including recruiting colleges and universities and their students (200 per institution) and assuring that country's human subjects and ethical policies are followed throughout the study. The other representatives will have expertise in test development and evaluation (sampling, translation, cultural and language adaptation, reliability, validity). We recognize that a single individual may not have all the expertise envisioned and that is, in part, why two experts have been recommended. The assessment experts will be front-end loaded as development of tasks begins early; the country representative will be involved at a low level throughout the project but will be quite active in recruitment.
- Request countries (Finland, Korea, Mexico, Norway), in collaboration with CAE and following established criteria, to select representatives for project.
- Contact and discuss with AHELO participating country representatives (by phone) special needs or concerns that need to be given consideration in the first meeting of the country representatives.
- Send package of material to participating countries:
  - Project overview and time line if different from the one submitted to OECD on April 17, 2009.
  - Criteria for selecting country representatives (see above).
  - CLA crime performance task with structured discussion questions.
  - CLA research background papers.
  - Rough organizational chart (Figure 1, with contact information).
  - Rough activity timeline.
  - Legal agreements between CAE and country representatives (to protect confidentiality of CAE's Intellectual Property).
- Set preliminary dates and locations for meetings over the course of the project.

- Develop conceptual framework and procedures for training country experts (Assessment Adaptation Group or AAG) in adapting CLA PT contents to reflect their contexts.
- Research critical-thinking multiple-choice assessment options. The GNE requested that an objective, critical thinking test be administered in addition to the adapted CLA for time, cost, and validity reasons. A review of tests as to their applicability to the international context and their technical characteristics (construct definition, reliability and validity) will be carried out with the goal of recommending to the participating countries one or two critical thinking tests for their selection. The chosen test will be used to concurrently validate CLA test scores. The Technical Advisory Group will review the report and its recommendations before being circulated to the countries.
- Assemble materials for initial meeting (see Phase 2): (a) Logistical materials such as agenda, contact information of attendees, list of participants, (b) CLA in the classroom “workshop” materials, (c) short list and examples of critical thinking tests, (d) rationale for having countries active in modifying CLA tasks, (e) document that discusses PTs for cross-cultural/cross-language use, (f) all CLA PTs.
- **Milestones:**
  - Country assessment teams in place.
  - Meetings of the teams scheduled.
  - Countries contacted and special concerns addressed or noted for first meeting’s discussion among country representatives.
  - Short list of critical thinking tests assembled for review and selection.
  - Framework and criteria for modifying PTs ready for review.
  - Materials ready for first joint team meeting
  - Brief progress report.<sup>4</sup>

---

<sup>4</sup> Brief progress reports will be sent approximately three times a year (or as a phase or sub-phase is completed). They will be approximately two pages long and be comprised of a checklist as to milestones accomplished and a brief explanation as to why a milestone, if any, might have not been met as expected.

**Phase 2. Assessment Task Adaptation: (13 months) October 1, 2009 – October 31, 2010**

*2.1 Introduction and Task Selection and Adaptation (1 month) October, 2009*

- Meeting 1 (3 days in New York). Goals are to:
  - Introduce country teams (3 representatives from each country: AHELO country representative and two assessment specialists).
  - Familiarize teams with the CLA.
  - Review and select a set of 3 CLA PTs for modification and use in the AHELO feasibility study.<sup>5</sup>
  - Train the AAG in task modification.
  - Train AAG in issues of translation and translation review, including guidelines for the recruitment of test translation and test translation review teams.<sup>6</sup>
  - Agree upon an appropriate multiple choice critical thinking test.
  - Establish objectives and process for recruiting participating colleges and universities.

In the meeting, then, the following tasks will be carried out:

- Introductions of country team and CAE staff members including a special dinner meeting as a means of getting to know one another.
- Familiarize and train teams on the CLA (CLA in the Classroom exercise).
- Review and evaluate 9 CLA PTs for possible adaptation to the international context.
- Select three CLA tasks for adaptation and field testing. These three tasks will be used in the field testing and, consequently, administered in all participating countries.
- Agree upon procedures for adapting/translating PT contents to country context.

---

<sup>5</sup> Our rationale for selecting three CLA tasks is based on several assumptions: (1) not all CLA tasks are appropriate for the international context (we will find out in the feasibility study); (2) the cost in time and Euros to adapt, translate, and score more tasks would be prohibitive; and (3) to have adequate sample sizes in a matrix sampled assessment, dividing 200 students by 3 is about as small a number of students per task as desirable.

<sup>6</sup> This would include an exercise where the teams go through a hands-on translation and review activity.

- Train team members in task adaptation as well as on translation and review process procedures.
- Review and select critical-thinking multiple-choice test.
- AAG team representatives return to their respective countries after meeting 1 with the task of reviewing the contents of the selected and agreed upon three PTs and modifying each to fit within their country's context following agreed upon procedures, as well as beginning to recruit participating colleges and universities.
- **Milestones:**
  - Three CLA PTs selected for adaptation into the CLA/AHELO assessment.
  - Critical thinking test selected.
  - Procedures for adapting/translating tasks agreed upon.
  - Procedures for developing new tasks agreed upon.
  - AAG representatives are assigned the task of modifying the three chosen and agreed upon CLA PTs
  - Begin the task of recruiting test translation teams and test translation review teams according to CLA (see Appendix) and OECD guidelines.
  - Begin translation of the agreed upon multiple choice critical thinking test (see 2.2 below).
  - Begin recruiting participating colleges and universities
  - Progress report.

## 2.2 *Task Adaptation (AAG)* (3 months) November 1, 2009 – January 31, 2010

This phase of the project focuses on the adaptation of CLA tasks for the international context, the development of new PTs as a test of the feasibility of so doing for a large-scale implementation of AHELO, and the translation of the critical thinking test into the respective languages for review.

- Countries review and adapt contents of selected PTs to fit context (AAG member takes lead in each country).
- Countries review translated versions of the multiple-choice critical thinking test.
- Also continue recruitment of colleges and universities
- Meeting 2 (- February, 2010—3 days in New York). Goals are to:
  - Present each county's recommended modifications of selected CLA PTs for discussion/review (AAG member takes lead).

- Gain consensus on the three PTs based on CAE and country's evaluations and recommendations for final development.
- Review and recommend new PTs to be developed, one by each participating country.
- Review translations of critical thinking test.
- Review progress recruiting colleges and universities

### 2.3 Task Adaptation and Translation (9 months)

This phase focuses on the full development, pilot testing, and translation of the (1) adapted CLA PTs (CAE and AAG), (2) critical thinking test (CAE).

- CLA Task Adaptation and Translation
  - CAE fully develops three PTs (tasks, response formats, scoring rubrics and IT administration procedures) following agreed upon modifications at Meeting 2 and sends draft PTs for review, comment, and revision as they are completed.
  - Test translation team (see Appendix for a description) roughly translate PTs as they are finalized so that small pilots (~10 students) can be conducted in home countries following CAE protocol (including cognitive workshops).<sup>7</sup>
  - CAE conducts small pilot in the U.S.A. (~20-30 students) for each modified PT following CAE protocol.
- Critical thinking multiple choice test translation is piloted (CAE and translators).

### 2.4 Feasibility Study Design (9 months co-terminus with Phase 2.3)

At the same time tasks are being adapted, developed, and translated, CAE will fully develop the feasibility design. CAE will circulate the design for review and revisions will be one of the major items at the next meeting of the countries.

- CAE develops feasibility study design and protocols including sampling, data collection and proctoring, and scoring (following CAE protocol and AHELO/GNE Report 13).
- Following review by the Technical Advisory Group, CAE circulates among countries, and revises:

---

<sup>7</sup> We anticipate a translation cycle as follows: translation team creates first version in home language, circulates it to the country review team, gets feedback and modifies translation. The same cycle is carried out for the revised version. Once these two cycles are completed, the tasks are pilot tested.

- Feasibility study design
- Sampling frame
- Data collection plan including proctoring
- Scoring plan<sup>8</sup>
- Internet interface
- Data base
- Training for (a) assessment administration proctors and (b) assessment scorers in each country
- Meeting 3 (June, 2010—3 days in New York). Goals are to:
  - Report, evaluate, and discuss findings on mini-pilots.
  - Discuss challenges of rough translations.<sup>9</sup>
  - Finalize the three performance tasks.
  - Review, discuss and agree upon feasibility study design and protocols.
  - Confirm that all colleges and universities have been recruited
- **Milestones:**
  - Three CLA PTs modified and pilot-tested (in very small studies) in the U.S.A. and each country.
  - Multiple choice critical thinking test translated and pilot tested.
  - Feasibility study design and data collection protocols agreed upon.
  - All participating colleges and universities recruited
  - Progress report.

## 2.5 Test Translation and Preparation for Testing Phase (Months 9-13) July 1, 2010 – October 31, 2010

Translation teams translate modified (see footnote under Meeting 3) tasks following OECD/CAE guidelines.

---

<sup>8</sup> One possibility for PT scoring: (1) Recruit ~2 scoring trainers for each testing language (2 for Spanish, 2 for Korean, etc.); (2) train them in English using a common set of training papers and scoring criteria; (3) the scoring trainers then train monolingual scorers in their respective native languages. For example, the Korean scoring trainers would train Korean-speaking scorers to score Korean responses. For training, they would use the scoring criteria (translated into Korean) and a set of Korean training papers. This would also work if bilingual scorers were used.

<sup>9</sup> Specifically, the four countries would compare their respective language translations with each other and the English version. Then subsequent versions of the translation-review cycle is carried out producing a third and fourth version in each language.

- CAE representatives either (a) visit with countries during this process, or (b) video conference with countries to support translation process.
- CAE carries out proctor and scorer training as PT translations are finalized.
- Meeting 4 (October, 2010—3 days in New York) Goals are to review and finalize translations and finalize design and data collection process for feasibility study.
- The final version of the translated performance tasks are produced.
- **Milestones:**
  - Final translated PTs for each country.
  - Proctors and scorers trained or in training process.
  - Progress Report. The analyses in the progress reports, including the final one will include quantitative estimates of reliability and validity and qualitative information regarding validity and other aspects of the feasibility study (see footnote 2 for more specific description of what will be included to measure success of the feasibility study.)

## **Roles and Responsibilities**

Roles and responsibilities are summarized as follows:

Richard J. Shavelson, Ph. D. will be project director. He is a professor of education at Stanford University in California and one of the founders of the CLA. He will be assisted by Guillermo Solano-Flores, Ph. D. and associate professor of education at the University of Colorado. Dr. Solano-Flores will be responsible for ensuring construct consistency across cultures. A part-time program manager in California, Amy Kurpius, will also assist Dr. Shavelson. In addition to the research design and analysis of results, Dr. Shavelson and these associates will be overseeing adaptation and translation of the measures working with CLA staff. As such, they will coordinate the advisory work of country teams.

CAE program managers and associates will coordinate project operations including organizing meetings in New York. They have extensive expertise implementing the CLA in the United States. Their work will also include preparing for testing operations with CAE's exclusive test delivery partner, Internet Testing Systems Inc., whose proprietary

web testing application was designed for online delivery of the CLA tests. It will also involve helping the testing sites implement testing in the testing phase, working through a primary contact at each college. The program managers and associates will be advised by Dr. Scott Elliot, CAE's testing operations consultant.

Other CAE staff involved in the project will include Dr. Stephen Klein, Ph. D., CAE consulting director of research and development who will advise the project director; Marc Chun, Ph. D. and director of academic programs, who will lead the workshops on CLA test development and assist in test adaptation; Jeffrey Steedle, Ph. D., CAE measurement scientist, who will work with Dr. Shavelson on test adaptation, research design and results analysis; and, Roger Bolus, Ph. D., CAE's consulting CLA data management coordinator who will supervise data management.

Roger Benjamin, Ph. D., CAE's president, will provide leadership for CAE programs and projects and will coordinate liaison with OECD headquarters staff. James Hundley, CAE's chief operating and finance officer, has management responsibility for budgeting, cost control, financial reporting, compliance and contracting.

### **Publications and Projects related to the Collegiate Learning Assessment.**

*An asterisks (\*) indicates particular attention to methodological issues, including reliability and validity.*

Benjamin, Roger (2009) and Marc Chun, et al. (2009). *Returning To Learning in an Age of Assessment*. New York: CAE.

\*Benjamin, Roger (2008). "The Case for Comparative Institutional Assessment of Higher Order Thinking Skills." *Change*, November/December: 51-55.

Benjamin, Roger. (2007) "Recreating A Faculty Role in University Governance." *Fixing the Fragmented University*. Ed. J. Burke. Bolton, MA: Anker Publishing, 2007: 70-98.

Benjamin, Roger and Stephen Klein (2007). "Assessment Versus Accountability in Higher Education: Notes on Reconciliation." United Nations Educational, Scientific and

Cultural Organization (UNESCO) Commissioned Paper Series: 1-26.

Benjamin, Roger, Marc Chun and Richard Shavelson (2007). *Holistic Tests in A sub-Score World: The Diagnostic Logic of the CLA.* New York: CAE.

Benjamin, Roger. (2003) "The Environment of American Higher Education: A Constellation of Changes," *The ANNALS of the American Academy of Political and Social Science* 585: 8-30.

\*Carini, Robert, George Kuh, and Stephen Klein. "Student Engagement and Student Learning." *Research in Higher Education* 47.1 (2006): 47-68.

Chun, Marc. "Looking Where the Light is Better: A Review of the Literature on Assessing Higher Quality Education." *Peer Review* 4.2/3 (2002): 16-25.

\*Klein, Stephen, David Freedman, Richard Shavelson, Roger Bolus (2008). *Assessing School Effectiveness.* *Evaluation Review*, 32, 6: 510-525.

\*Klein, Stephen, Roger Benjamin, Roger Bolus, and Richard Shavelson. "The Collegiate Learning Assessment: Facts and Fantasies." *Evaluation Review* 31.5 (2007): 415-439.

\*Klein, Stephen. "Characteristics of Hand and Machine-Assigned Scores to College Students' Answers to Open-Ended Tasks." *Probability and Statistics: Essays in Honor of David A. Freedman.* Eds. Deborah Nolan and Terry Speed. Beachwood, Ohio: Institute of Mathematical Statistics, 2005. 76-89.

\*Klein, Stephen. "Direct Assessment of Cumulative Student Learning." *Peer Review* 4.2/3 (2002): 26-28.

\*Klein, Stephen, Marc Chun, Laura Hamilton, George Kuh, and Richard Shavelson. "An Approach to Measuring Cognitive Outcomes Across Higher Education." *Research in Higher Education* 46.1 (2005): 251-276.

Klein, Stephen. "The Costs and Benefits of Performance Testing on the Bar Examination." *The Bar Examiner* 65.3 (1996): 13-20.

Shavelson, Richard. "Assessing Student Learning Responsibly: From History to an Audacious Proposal." *Change* 39.1 (2007): 26-33.

Shavelson, Richard. *A Brief History of Student Learning: How We Got Where We Are and a Proposal for Where to Go Next*. Washington, DC: Association of American Colleges and Universities, 2007.

Shavelson, Richard and Leta Huang. "Responding Responsibly to the Frenzy to Assess Learning In Higher Education." *Change* 35.1 (2003): 10-17.

### **Forthcoming**

\*Hardison, Chaitra and Anna Marie Vilamovska. "Critical Thinking Performance Tasks: Setting and Applying Standards for College-Level Performance." Forthcoming RAND Corporation.

Shavelson, Richard, Amy Kurpius and Matt Bundick with Richard Hersh, Daniel Silverman, Corey Keyes, and Lynn Swaner (2009). "On Assessing Learning Broadly and Responsibly", forthcoming.

\*Shavelson, Richard. "The Collegiate Learning Assessment," in *The Quest to Assess Learning and Hold Higher Education Accountable*. Stanford, CA: Stanford University Press.

### **Appendix: Translation and Translation Review**

Our test translation and test translation review procedures address the goals of translation guidelines used by OECD projects in international test comparisons (e.g. OECD, 2007; 2009). All test materials (e.g., tasks and scoring rubrics) and surveys will be translated according to those documents. However, the uniqueness of this project makes it necessary to make some modifications in these procedures:

1. Our procedures need to address the fact that important aspects of the cultural differences that motivate a feasibility project like this are the countries tremendous differences regarding assessment capacities and information technology infrastructures. Test translation issues will be discussed and

addressed—from the very beginning of the project and through the entire process of task adaptation and test translation.

2. While double translation from two source languages is currently accepted practice in international test comparisons (e.g., Grisay, 2003), in this project we will use double translation from one language source. For each country, two independent translators will translate the original, English version into the corresponding target language. A reconciler will merge the two independent translations into one that best captures the intended meaning of the two translations. Since the CLA tasks were developed originally in English, we believe that the approach of double translation from English is more consistent with the main goal of this project—examining whether performance assessments developed in one national cultural context and in one language can be adapted for use in other nations with distinctly different cultures and languages.
3. Our actions will be guided by the theory of test translation error (Solano-Flores, Backhoff, & Contreras-Niño, 2009). This theory, and the review and coding procedures based on it will allow operationalization of the test translation guidelines used by OECD projects in international comparisons such as Hambleton’s (2005) guidelines<sup>10</sup> and OECD’s (2007) guidelines. The theory addresses test translation error on several error dimensions (including semantics, register, culture) with an unprecedented level of precision. The theory and the translation review procedures based on it allow identification of potential sources of invalidity of existing translated tasks or tasks that are in the process of translation.
4. In this project, the verification of the test translations will not be made by a third party, as it is customary in international comparisons. Instead, the verification will be performed at two review stages:
  - a. In the first stage, translation review will be performed within each country by independent translation review panels, which will provide their countries’ translation teams with feedback on the errors identified. With project staff supervision, each country will assemble a translation review team which will work independently from the translation team.

---

<sup>10</sup> Professor Ron Hambleton is a member of this project’s Technical Advisory Committee.

This team will be multidisciplinary and will be composed of university professors, content area specialists, certified professional translators, assessment specialists (psychometricians), and linguists. This multidisciplinary team will review draft versions of the translated items according to a coding system and coding procedures based on the theory of test translation, and will provide the translation team with detailed feedback on potential sources of invalidity and ways in which the translation of the tasks can be improved.

- b. In the second stage, translation review will consist of comparing the different language versions with each other and with the original version in English. We will convene a meeting with representatives of the four participating countries. With the facilitation of project staff, these representatives will discuss in length how their language versions of each task are consistent with or differ from each other and how they are consistent or differ from the original, English version. Also with the facilitation of project staff and based on the theory of test translation error, the country representatives will discuss and resolve any discrepancies across language versions encountered. This process will allow us to identify and document the kinds of issues that are relevant to this feasibility project.

### **Translation and Review Teams**

For each country, there will be a test translation team, and test translation review team (see table below). These teams will be independent—no individual will be a member of more than one team. The CAE Project staff will approve the participation of each individual based on their documented experience and the professional background and on the language background/personal life profiles that the project staff will develop. The AAG from each country will oversee the translation process and represent the process at project meetings at OECD.

Test Translation Team	Test Translation Review Team
<ul style="list-style-type: none"> <li>• One translation leader (linguist or professional translator)</li> <li>• One specialist in the content area</li> <li>• Two certified professional translators English-Target language</li> <li>• One reconciler</li> </ul>	<ul style="list-style-type: none"> <li>• One translation review leader (linguist or assessment specialist)</li> <li>• One specialist in the content area</li> <li>• Two certified professional translators English-Target language</li> <li>• One linguist (depending on leader's background)</li> <li>• Two university professors</li> </ul>

### References

- Grisay, A. (2002). Translation and cultural appropriateness of the test and survey material. In R. Adams & M. Wu (Eds.), *PISA 2000 Technical Report* (pp. 57-70). Paris: Organization for Economic Co-operation and Development.
- Hambleton, R.K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R.K. Hambleton, P.F. Merenda, & C.D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3-38). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- OECD (2007). PISA 2009 translation and adaptation guidelines. Doc: NPM(0709)1, National Project Managers' Meeting, Dubrovnik, Croatia, 24-28 September.
- OECD (2009). PISA 2006 technical report.
- Solano-Flores, G., Backhoff, E., & Contreras-Niño, L.A. (In press). Theory of test translation error. *International Journal of Testing*.