

Unclassified

EDU/IMHE/AHELO/GNE(2011)21/ANN2/FINAL

Organisation de Coopération et de Développement Économiques  
Organisation for Economic Co-operation and Development

02-May-2012

English - Or. English

DIRECTORATE FOR EDUCATION  
INSTITUTIONAL MANAGEMENT IN HIGHER EDUCATION GOVERNING BOARD

## Group of National Experts on the AHELO Feasibility Study

### INTERNATIONAL SCORING MANUAL

#### 8th Meeting of the AHELO GNE

Paris, 28-29 November 2011

*This document was prepared by the ACER Consortium.*

*It is a finalised version of the document presented at the 8th meeting of the AHELO GNE and is declassified with the GNE and Consortium's approval.*

*It is only available in PDF format.*

Contact:

Consortium: [ahelo@acer.edu.au](mailto:ahelo@acer.edu.au)

OECD Directorate for Education: [Diane.Lalancette@oecd.org](mailto:Diane.Lalancette@oecd.org)

**JT03320864**

Complete document available on OLIS in its original format

*This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.*

EDU/IMHE/AHELO/GNE(2011)21/ANN2/FINAL  
Unclassified

English - Or. English





## **Contents**

Contents .....	2
Introduction.....	3
The Lead Scorer.....	4
Recruiting the national scoring team .....	6
Training the national scoring team .....	7
Scoring operations .....	10
Quality monitoring.....	11
Appendix A: AHELO confidentiality form .....	14
Annex A: Generic Skills Scoring Guide.....	15
Annex B: Economics Scoring Guide .....	15
Annex C: Engineering Scoring Guide .....	15

## Introduction

1. This International Scoring Manual forms part of a suite of scoring materials used for the AHELO Feasibility Study. Detailed operational and technical information about AHELO can be obtained from National Project Managers (NPMs) in participating countries. Public information is available from [www.oecd.edu.au/edu/ahelo](http://www.oecd.edu.au/edu/ahelo).
2. This document provides general procedures for scoring the constructed response tasks in all three testing strands of the AHELO Feasibility Study: Generic Skills, Economics and Engineering. Multiple choice questions are scored automatically and do not require human scoring.
3. Specific information on scoring for the Generic Skills, Economics and Engineering strands can be found in the relevant Scoring Guide, included as annexes to this document. These Scoring Guides provide detailed strand-specific information about test design, software systems, scoring rubrics and sample responses.
4. This International Scoring Manual is a public document. The strand-specific Scoring Guides contain secure test material, and are confidential.
5. Further details on test management can be found in AHELO's National Management Manual, Institution Coordination Manual, and Test Administration Manual.
6. It is essential that all countries participating in AHELO follow the scoring procedures in this document. Any anticipated or desired variations need to be discussed and approved by the AHELO Consortium in advance (email: [ahelo@acer.edu.au](mailto:ahelo@acer.edu.au)).
7. Broadly, this International Scoring Manual provides information on these processes:
  - NPM recruitment of the Lead Scorer;
  - NPM and Lead Scorer participation in international training;
  - NPM and Lead Scorer recruitment of national scoring teams;
  - training of the scoring teams by Lead Scorers;
  - scoring of student responses, coordinated by Lead Scorers; and
  - management of quality control by the NPM and Lead Scorer.
8. Table 1 shows the scoring schedule for the AHELO Feasibility Study.

**Table 1: AHELO Feasibility Study scoring schedule**

<b>Activity</b>	<b>Timing</b>
Develop international scoring documents	August 2011 to November 2011
Recruit Lead Scorer	By November 2011
NPM and Lead Scorer training	23 to 27 November 2011
Lead Scorer recruits scoring team	By February 2012
Develop national scoring manual and guides	August 2011 to November 2011
Train scoring team	Days/week before scoring
NPM meeting	April 2012
Testing	February to May 2012
Scoring	Days/week after testing
Reporting	June 2012

**The Lead Scorer**

9. In each country a Lead Scorer is responsible for the scoring of all student responses. In countries which are participating in more than one strand of AHELO, one Lead Scorer is required for each strand.
  
10. The AHELO Consortium will work with NPMs to identify the person who will serve as the Lead Scorer in each participating country. The Lead Scorer is responsible for:
  - working with the NPM and the AHELO Consortium to complete the scoring of the constructed response tasks according to the requirements provided and within the timeline allotted;
  - participating in an international training session in late November 2011 to learn about scoring rubrics and procedures;
  - working with NPMs to adapt and translate relevant information in the AHELO Scoring Guides (only the scoring rubrics will be translated and adapted by the AHELO Consortium);
  - working with NPMs to ready relevant software systems for scoring operations;
  - identifying and recruiting from three to ten individuals (depending on test volume) who will score student responses;
  - working with NPMs to select and assemble student sample responses as scorer training materials;
  - pre-assigning scores to student sample responses included in the scorer training material;
  - training selected scorers using international materials, scoring rubrics and guidelines;
  - monitoring scoring to ensure that scorers are completing scoring activities in an accurate, consistent, and timely fashion; and

- responding to scorer questions about general requirements or specific student responses.
11. As this list suggests, Lead Scorers play a very important role in AHELO. They are responsible for working with NPMs to interpret international contexts and requirements, prepare for and manage scoring activities, and monitor the quality of scoring processes and outcomes. Lead Scorers play a vital role in ensuring that student responses are scored accurately and consistently in-country to international standards. Unless this is done correctly it is not possible to compare student scores.
  12. If NPMs have the appropriate professional background it is possible for them to act as the Lead Scorer. Otherwise, a Lead Scorer will need to be recruited.
  13. Lead Scorers, ideally, should have appropriate qualifications and experience. They should have:
    - a PhD in a relevant academic field;
    - national or international high standing in their field;
    - extensive experience assessing university students;
    - strong language skills in both English and the language of testing;
    - recent experience teaching in the relevant AHELO discipline (if applicable) at the university level;
    - ability to travel to Paris in late November 2011 for international training;
    - ability to recruit, train and manage multiple scorers;
    - ability to score responses accurately and explain the reasoning behind the allocation of scores; and
    - ability to monitor scoring and manage the performance of scorers.
  14. The remainder of this AHELO International Scoring Manual assumes that a Lead Scorer for each country has been recruited and briefed on the study. The NPM should contact the AHELO Consortium if it would be helpful to have further information about the study or requirements of this role.
  15. The remainder of the document also assumes that the NPMs and Lead Scorers will have participated in the November 2011 international training session.

## **Recruiting the national scoring team**

16. Scorers need to be recruited to score constructed response tasks in the AHELO assessment instruments. In some countries, there may be experienced scorers from related national or international assessment programs who are available to score for AHELO. In other cases, suitable scorers need to be recruited for AHELO.
17. All scorers must undergo scorer training specific to AHELO regardless of whether or not they have previous scoring experience. If helpful, the AHELO Consortium can provide materials to assist countries recruit scorers.
18. Scorers, in general, should have appropriate qualifications and experience. They should have:
  - a graduate qualification, or be currently enrolled in a relevant graduate qualification;
  - extensive experience assessing university students;
  - strong language skills in English and the language of testing; and
  - ability to score accurately and explain reasoning behind scores allocated to student responses.
19. In addition to qualifications and experience above, scorers in the Engineering and Economics strands must possess at minimum a graduate qualification in that discipline. It is anticipated that the scoring team may comprise of several graduate or doctoral students. This is sufficient, as long as the students possess at minimum graduate training. If NPMs are uncertain about the suitability of a scorer, please discuss the situation with the AHELO Consortium.
20. Scorers should have experience in scoring assessment tasks and be available to work in the estimated time period in which scoring will take place.
21. AHELO's test instruments are secure, and before being shown test materials all scorers must sign a confidentiality form, legally obliging them not to disclose the content of the AHELO assessment instruments beyond the scoring team environment. A confidentiality form is provided in Appendix A. All confidentiality forms should be sent to the AHELO Consortium ([ahelo@acer.edu.au](mailto:ahelo@acer.edu.au)) and kept on file by NPMs at the AHELO National Centre.
22. It is important that scorers who are recruited commit necessary time for training (about one day) and for the duration of the scoring. For each strand in which a country

participates, around 10 higher education institutions will participate in AHELO. At each institution, up to 150 students may complete the test. This means that in each strand up to 1,500 student responses could be scored in each country. This number does not take double scoring into account. The average time required to score constructed response tasks depends on the strand in question. Expected scoring times for each strand and recommended scoring time allocations are specified in the relevant Scoring Guide.

### **Training the national scoring team**

23. It takes about one day to train a group of scorers to the point that they generally agree on the score that should be assigned.
24. Scorer training should take place close to actual scoring. This ensures that scorers have been recently trained in the material to be scored and thus results in more accurate and consistent scoring.
25. Lead Scorers need to be thoroughly familiar with both the assessment instrument and the corresponding Scoring Guide prior to scorer training.
26. Lead Scorers also need to be familiar with the use of the AHELO test system to view student responses and enter scores.
27. Lead Scorers need to work with NPMs to prepare the AHELO test system for national scoring operations.
28. The scoring rubrics in the Scoring Guides will go through a thorough translation and adaptation process managed by the AHELO Consortium. Lead Scorers should work with NPMs, however, to adapt and translate other relevant information in the AHELO Scoring Guides into the language used for testing. The translation of these materials will not be verified internationally.
29. The general training materials provided to NPMs and Lead Scorers at the November international training is the first resource for national training sessions. These materials demonstrate how to use the Scoring Guides and provide a number of straightforward and more complex examples.
30. NPMs, in consultation with the Lead Scorer, should select sample responses from their country to use as part of the Scorer training and quality monitoring. It is recommended

that at least two actual student examples be selected for each score point available for every item in the instrument. There is no need to search for a long time for responses that occur rarely. Fewer examples for straightforward responses will suffice.

31. If no responses are available from the country, the NPM should contact the AHELO Consortium and ask for sample responses to use in training. Sample responses can be provided in English only.
32. The sample responses should be organised so that they correspond to each item in the constructed response tasks. Sample responses should be photocopied so that each Scorer has one complete set for use during training.
33. Scorers may not remove the sample responses, Scoring Guide, rubrics or any other material from the training session.
34. The basic steps involved in the training the national scoring team are:
  - the Lead Scorer gives scorers an overview of the AHELO Feasibility Study;
  - the Lead Scorer provides an overview of the relevant testing strand;
  - the Lead Scorer explains the approach to scoring used in AHELO;
  - the NPM or Lead Scorer explains the role of scorers;
  - the NPM or Lead Scorer provides scorers with an introduction to the relevant AHELO test and explains its structure;
  - the NPM or Lead Scorer provides scorers with the constructed response test items and scorers work individually to provide their own responses to each item;
  - scorers read through the Scoring Guide and rubrics;
  - the Lead Scorer discusses the Scoring Guide and sample responses with scorers and explains how to score student responses;
  - scorers read and score additional sample responses to all test items;
  - the Lead Scorer leads a discussion of the scores awarded to each sample response;
  - the Lead Scorer advises scorers on how to accurately score student responses; and
  - the Lead Scorer reviews the item and Scoring Guide with scorers, and answers final questions.
35. Scorers are asked to complete the assessment items themselves because this helps to familiarise them with the tasks and different ways in which students may respond.

36. Scorers read the relevant Scoring Guide to become familiar with the test content. This also helps scorers to understand the kind of problems students might have in responding to items.
37. Lead Scorers should emphasise to scorers that they are expected to follow the Scoring Guides rigorously and with a high level of reliability. The role of Lead Scorers should be explained and it should be made clear to scorers that reliability checks, such as double marking, are made throughout the scoring process.
38. Scorers should be encouraged to ask questions and engage in discussion during the training. The objective is that the group of scorers will reach a consensus on the score that should be assigned to each sample response. These discussions need to be carefully informed by reference to international specifications and standards in the International Scoring Manual and the relevant strand-specific Scoring Guide.
39. After the initial training is complete, trainee scorers should undertake additional practice with the sets of locally assembled materials. These materials will need to be assembled by the Lead Scorer. The trainer should ensure that this process is conducted as a formal exercise, using testing software as appropriate.
40. Trainee scorers should undertake this exercise without asking questions of the trainer or consulting other scorers. They should do this scoring in the same way as they would score actual student responses, using testing software as appropriate.
41. Once this task is completed, the trainer should review the scores assigned and conduct further discussion to ensure that the scorers understand the rationale for the scores as assigned in the Scoring Guide. The scores assigned by the trainee scorers should not be altered at this point. It is recommended that this process is done for each item one by one. This process should be completed before actual scoring begins.
42. The Lead Scorer should aggregate and tabulate the scores assigned as soon as possible. The scores assigned by the trainee scorers should be compared to determine whether the group are achieving a high level of reliability.
43. Scoring reliability should be calculated in the following way: If a score matches the pre-assigned scores, this should be regarded as a 'hit'. Each score which does not match the pre-assigned scores should be regarded as a 'miss'. The percentage of scorer 'hits' to the total of 'hits plus misses' for an exercise should total at least 85%—preferably higher.

44. If the group of scorers are not achieving 85% reliability, additional practice with local examples should be undertaken before actual scoring begins. Any scoring undertaken prior to results of this reliability test being known should be reviewed.
45. If the Lead Scorer identifies that one or two scorers are having more troubles than others, they will need to spend time with them to identify where the problems lie and to help them achieve a better understanding of the Scoring Guide.
46. If, after additional practice exercises, the same one or two scorers are consistently unable to use the Scoring Guide with sufficient reliability, then these scorers should not be used for any further scoring. Any scoring of actual student result that they may have already done should be reviewed.

### **Scoring operations**

47. Scorers should only proceed to scoring actual student responses once Lead Scorers are satisfied with the quality and consistency of scoring as covered in the training sessions.
48. It is very important that scoring is completed both thoroughly and methodically, with strict adherence to Scoring Guides. The Lead Scorer is responsible for ensuring the quality of the scoring process and for monitoring the work of scorers.
49. Once the actual scoring starts, Lead Scorers should not do any actual scoring. They should only engage in ‘back reading’—systematically monitoring scoring performance. Lead Scorers check scores from other members of the scoring team and are available to answer questions from scorers during the scoring process.
50. It is preferable for scorers to work in silence and to refer queries to the Lead Scorer, rather than to other scorers. If the need arises, Lead Scorers should advise all scorers individually or as a group if a particular query justifies this.
51. Scoring constructed response tasks can be mentally demanding and requires high levels of continued concentration. Thus, it is recommended that scorers do not undertake more than six hours of scoring per day.

52. It is vital that scorers take a short break (of approximately 10 minutes) every 90 minutes. This is to ensure that concentration is maintained. After two 90-minute sessions, scorers should take a break of at least 30 minutes.
53. All scorers need to be given a unique identification number labelled as ‘SCORERID’. This number will comprise of an A, B or C depending on the strand (Generic Skills = A; Economics = B; Engineering = C), followed by a 2 digit number. This SCORERID will allow Lead Scorers to better monitor the quality and reliability of scoring by being able to identify which responses scorers are scoring student responses.
54. The order in which item scoring is to be conducted is prescribed by each strand’s Scoring Guide.
55. Scored student responses are recorded using strand-specific testing software. Further details are supplied in the relevant Scoring Guide.
56. Scoring activities should be conducted in less than two weeks. Lead Scorers should refer to the Scoring Guide for the relevant strand to ascertain the amount of time which scoring will require and, hence, the number of scorers required. In general, for quality control and efficiency it is recommended that at least three scorers be recruited per strand. In addition, it is recommended that at least one reserve scorer be trained in each participating country in case of any unforeseen absence.

### **Quality monitoring**

57. It is essential that the AHELO Group of National Experts (GNE), OECD, AHELO Consortium, NPMs, Lead Scorers and scorers have confidence in the integrity of scoring operations and outcomes.
58. The steps described in this International Scoring Manual set minimum standards for scoring activities. If NPMs or Lead Scorers wish to implement more extensive forms of training or management during scoring they are encouraged to do so. Strand-specific Scoring Guides provide additional information.
59. Robust scoring operations play an essential role in AHELO. To ensure high-quality outcomes it is essential to ensure that there is:

- international consistency—that there is comparability between scoring operations in participating countries;
- national consistency—that there is consistency between individual scorers within a national team; and
- individual consistency—that each individual scorer has a sufficiently consistent understanding and application of rubrics, and is consistent across time.

60. The following materials and procedures are important for helping to ensure international consistency:

- provision of an International Scoring Manual and Scoring Guides;
- international training of NPMs and Lead Scorers;
- ongoing international training and support conference calls;
- a secure online international Lead Scorer query service;
- discussion and review of sample student scripts across countries;
- use of particular item materials and psychometric control processes; and
- undertaking studies of international, national and individual scorer reliability.

61. NPMs and Lead Scorers play an important role in ensuring the reliability of national scoring operations by:

- providing the International Scoring Manual and Scoring Guides to the national scoring team;
- training the national scoring team;
- ongoing international communication, training and support;
- providing summary evaluative feedback on the scoring process;
- using the testing system to double-score a minimum of 20% of student responses; and
- monitoring ‘scorer drift’—any unwanted systematic application of the rubric.

62. Individual scorers help to manage reliability by:

- being aware of key facets of the AHELO Feasibility Study;
- having a good working knowledge of scoring materials and procedures;
- discussing possible and actual responses with fellow scorers;
- spot-checking scoring performance at the start and end of each scoring session; and
- comparing scoring performance with the scripts selected for training.

63. Lead Scorers have a key role to play regarding quality monitoring. As well as participating in the scoring themselves, and being available to field queries throughout the scoring process, Lead Scorers need to undertake spot-checking during scoring.
64. During each scoring session, and at the end of a scoring session, Lead Scorers should perform a random audit of student responses that have already been scored. The aim of this spot-check is to identify any potential problems to discuss with individual scorers or with the group as a whole. If there are indications that a specific scorer is having difficulty in using the Scoring Guide consistently, then more of the student responses scored by that scorer should be checked.
65. At the beginning of the next session, the Lead Scorer should review the outcomes of the spot checking with the scorers. This activity should be seen as a constructive form of feedback for the scoring team. It is important that scorers who are not meeting the standards of reliability are carefully monitored. If these standards cannot be consistently met, then it is of best interest to the participating country, and to the AHELO project as a whole, to remove such a scorer from the scoring team.
66. If during the scoring, the Lead Scorer is unable to deal with a specific issue relating to a particular student response, the Lead Scorer should direct the query to their NPM. NPMs should direct queries to the AHELO Consortium.
67. Lead Scorers are supplied with support materials to help them monitor the quality of scoring.

## Appendix A: AHELO confidentiality form



Improving Learning



Australian Council for Educational Research

### CONFIDENTIALITY AGREEMENT

#### OECD AHELO PROJECT

I understand that the signing of this form is an acknowledgment of my professional responsibility to maintain complete integrity of security for this project. I declare that I will not divulge any project information, trial materials, test materials, processes, contents or results, or any other materials, documents or information pertaining to the project, or its clients or suppliers, to any person or organisation as directed under the terms of the project unless I have prior written approval from ACER. I understand that the above does not apply to information that is in the public domain.

I have read and accept the conditions as outlined above.

Signed: \_\_\_\_\_ Dated: \_\_\_\_\_

Australian Council for Educational Research Limited  
19 Prospect Hill Road (Private Bag 55) Camberwell VIC 3124 Australia  
t +61 3 9277 5555 f +61 3 9277 5500 w www.acer.edu.au  
ACN 004 398 145 ABN 19 004 398 145

**Annex A: Generic Skills Scoring Guide**

**Annex B: Economics Scoring Guide**

**Annex C: Engineering Scoring Guide**