

Revising the Evaluation Procedure of the European Academic Software Award

Sophie David*, Rachel Panckhurst†, Lisa Whistlecroft††

* CNRS UMR 7114, MoDyCo, Université Paris 10 – Nanterre, France
sophie.david@u-paris10.fr

† Praxiling ICAR, UMR 5191 CNRS Université Paul-Valéry, Montpellier 3, France
rachel.panckhurst@univ-montp3.fr

†† PALATINE, Lancaster University, UK
L.Whistlecroft@lancaster.ac.uk

Abstract

The European Academic Software Award was initiated in 1994. It is a biennial competition which is organised by the European Knowledge Media Association. Academics and students are able to submit software they have developed which is then evaluated by a team of European expert-jurors. In this paper, we describe how the current EASA evaluation process runs and we then comment on evaluation in general, and specific aspects of the procedure which need to be reorganised, modified, abolished or developed for the next competition. A methodological revision of the EASA evaluation grid and a synthesis of the results are proposed.

Keywords: Evaluation, Software, e-learning.

1 Introduction

The European Academic Software Award (EASA) was initiated in 1994. It is a biennial competition, organised by a European body called the European Knowledge Media Association — EKMA (for more information on EKMA and EASA, *cf.* [1] and the website: www.ekma.net). Academics and students are able to submit software they have developed which is then evaluated by a team of European expert-jurors. Approximately 30 entries are short-listed for an award, and 10 awards are made after a final judging process.

We originally became interested in researching the EASA evaluation procedure as a result of our own experience in the EASA process, but also because we noticed, during discussions at the finals' jurors' meetings at each EASA competition, that certain aspects were discussed time and time again, without being properly resolved. A number of these were recently presented (*cf.* [2]) and an *Evaluation Working Group* emerged from the Montpellier workshop (8 people including the authors), in order to rethink the current evaluation process in time for the 2006 EASA competition. We initiated a questionnaire in which suggestions to conduct extensive revision of the evaluation procedure were made, and we also reflected upon the EASA com-

petition as a whole. This questionnaire and two working papers (*cf.* [2], [9]) were sent out by the authors to the Evaluation Working Group. Remarks by contributors were included in a final report, along with a number of recommendations (*cf.* [4]). The full results of our conclusions appear in a written report (*cf.* [4]) which was initially validated by the Evaluation Working Group, then presented to the EKMA board at Oxford Brookes University on 11 April 2005.

In this presentation, we briefly describe how the current EASA evaluation process runs, from the moment authors submit their software or materials up until the final evaluation. We then comment on the notion of evaluation in general, and the specific aspects of this procedure which need to be reorganised, modified, abolished or developed for the next competition in 2006. We have undertaken a methodological revision of the EASA evaluation grid and we present a synthesis of the results here.

2 Current procedure

There are currently 14 official disciplines (plus one discipline designated 'other') within the European Academic Software Award. Different categories of product may be submitted to EASA: educational software, research software and production tools.

The whole process takes place in three stages. In the first stage the authors complete a web form giving information on several aspects. The organising team checks the validity of the entries and inappropriate or incomplete submissions are eliminated; the authors then submit their software or materials and give more detailed information for the second stage. This is when the major evaluation and judging process takes place. Between 30-35 submissions (out of 150-200) are then accepted to proceed to stage 3 (the finals). During stage 3, the finalists' submissions are evaluated once more and 10 prizes are then allocated to the winners.

In stage 1, the following aspects are checked:

- Authors/software (higher education or higher education and commercial)
- Country of origin (a European state)
- Completion of the mandatory sections of the form.

In stages 2 and 3, the following 5 evaluation criteria are applied (*cf.* Appendices for the full evaluation grids):

- Innovation
- Design and ease of use
- European portability
- Educational materials and approach
- Evaluation of use

In stage 2, each criterion is marked and a general appreciation is also given in order to indicate if a specific entry should proceed to the finals or not ('accepted', 'borderline', 'rejected'). In stage 3, the five criteria are marked and then an overall average is calculated according to two mathematical methods, which are weighted in different ways according to the importance the jurors accord the different criteria; the winners are designated at the end of the final jurors' meeting, where numerical scores may be moderated or amended after open discussion amongst the jurors. (*cf.* [7]).

3 EASA & evaluation

The process described here stems from our own juror experience, but also from initial reflection on forms of evaluation which are traditionally used in domains such as Linguistics, Natural Language Processing & Information Retrieval (*cf.* [5], [6], [3]). This research enabled us to situate EASA and to outline a methodological framework: (i) the procedure evaluates the performance of technological devices; it is not linked to real users and the devices are evaluated out of their context of use; (ii) the devices are judged one by one; (iii) the evaluation procedure uses a black box method; (iv) experts and non-experts of the discipline are involved at different stages; (v) the evaluation procedure does not use referentials; (vi) the evaluation is measured quantitatively. Over and above the importance of evaluating these aspects, having been able to identify them and use a precise methodological framework helped guide us in measuring the coherence of our proposals.

As stated in the introduction, several aspects of the current EASA evaluation process needed to be revised. In this paper, we focus on several important aspects¹.

It seems that the problems concerning the evaluation grid which have arisen during the evaluation procedure will only be solved if EKMA first takes a firm position on the following points:

1. How does EASA position itself compared to other national, European and international competitions?

¹Due to space restrictions, we cannot go into depth in this paper, but we have nevertheless highlighted here the major points for both short-term and long-term changes to the competition.

2. What values are promoted by EASA?
3. As a consequence: What types of 'objects' are allowed to enter?

The final report answers these three questions in the following way:

- EASA must be a unique European competition (§ 3.1);
- EASA promotes the following values: differing European cultures and languages, innovation, best practice, standardisation, accessibility (§ 3.2);
- EASA evaluates products emanating from academia (§ 3.3).

3.1 EASA: a unique European competition

We have examined in detail two other European competitions, (Medida Prix – www.medidaprix.org and EURELEA – www.eurelea.de): EURELEA is not about software, and both EURELEA and Medida Prix concentrate on e-learning. There are also other differences but this initial distinction is sufficient for our needs. Our position is that EASA must be unique. This does not mean that cooperation is impossible but, as a competition, it must remain specific, in order, for instance, to obtain funding. We also focused on a definition of Europe, and EKMA decided to retain a political/educational definition of Europe as those countries which are signatories to the Bologna process.

3.2 EASA's values

EASA promotes the following values: cultural and linguistic European diversity, innovation, best practice, up-to-date standards, accessibility. We do not go into detail of cultural and linguistic aspects in this paper; the current European context and the importance of justifying the European nature of the EASA competition in a positive manner have led us to consider our differing cultural and linguistic heritages as essential resources. The types of products which can enter the competition must incorporate some sort of knowledge of languages and cultural diversity (*cf.* Appendix 2). Additionally, we retained the principle of making a significant number of awards in each competition. EASA is not the Olympic games where many compete and an overall winner emerges, with two runners-up being acknowledged. EASA recognises and rewards outstanding excellence in several disciplines across several academic contexts and the ten Awards recognise both the excellence and the diversity.

3.3 EASA evaluates academic products

Over the years, EASA has implicitly become a competition which may include not only software as such but also virtual

learning environments (VLEs) and pedagogical innovations using VLEs. In this instance, submissions to EASA could include one or more of the following: stand-alone software; virtual learning environments (VLE); coursework built within a VLE. This aspect needs to be clarified since there is a difference between evaluating a piece of software, and a course designed *within* an existing VLE. Furthermore, we observe the inapplicability of the current grid for certain types of tools: research tools cannot compete properly, given that they are ‘non-applicable’ for the ‘educational materials and approach’ criterion; all types of submissions are currently non-applicable for the ‘evaluation of use’ criterion (see § 3.4 below); different criteria are not necessarily equal for different disciplines.

We finally decided to concentrate on software and/or programmed content emanating from academia. Thus, programming (of some sort) must be a significant part of a valid entry to EASA. EASA should therefore accept the following sorts of software from eligible higher educational institutions:

- stand-alone (possibly single task) software (e.g. for teaching, learning, research, learning support, educational management, etc.);
- generic VLEs (with example content);
- subject-specific VLEs (i.e. with content);
- programmed content presented via a commercial VLE.

3.4 Revised evaluation grid

After this initial work, we redesigned the evaluation grid, with specific aims, as follows: (i) include values promoted by EASA; (ii) make sure the grid is coherent in relation to the evaluation framework and the types of objects/products which may enter the competition; (iii) check the trans-disciplinary and trans-tools nature of the grid; (iv) design a grid incorporating a set of themes. These various reformulations are presented below:

- Innovation
The main problem with this criterion was that there was a confusion between programming innovation vs. content innovation. As EASA is a ‘black-box’ competition, there is no way in which technical aspects can be verified precisely. This criterion is now explicitly conceptual and not at all technical.
- Design and ease of use
The ‘Design and ease of use’ criterion maintained the ambiguity between the architecture of the system itself and the user interface. The new criterion only looks into installation/access and the interface.
- European portability
Several different interpretations were also possible for portability: computer portability or portability from one

natural language to another. We have removed this ambiguity by requiring entries to specify, in some way, their precise European nature, from a linguistic and cultural perspective.

- Educational materials and approach
This criterion posed a problem for research software and generic tools. By reformulating it to read: ‘Users, approach and content’, target users and needs, as well as approach and content, via particular activities, are now clearly defined.
- Evaluation of use
At the moment, it is extremely difficult to check that a piece of software has been evaluated by *real users in real-life situations*, because the current grid does not address this problem clearly. In addition, more importantly, as we mentioned above, EASA is not a real-user oriented procedure (*cf.* [3]):

- evaluation of use should not be reduced to a mere estimate of the number of users of a software entry, or to the (positive) appreciations indicated by the authors themselves;
- evaluation of use should not rely solely on jurors’ intuitions, even if, as experts in their discipline and/or in e-learning, they may have valid intuitions about the real use of the entry. If real evaluation of use is to be conducted, we need to answer the following questions: who is the entry used by? at what moment? in what context? what kind of other resources does it compete with? etc.

To summarise, *evaluation of use* versus *evaluation of systems* require radically different methodologies (*cf.* [3] and [8]). We have proposed to suppress this criterion entirely².

Hence, in the new grid, we have decided to use 4 criteria instead of 5 (*cf.* Appendix 2):

- Innovation;
- Installation/access & interface;
- Europe: language & culture;
- Users, approach and content.

3.5 New evaluation grid: case study

In a solely distance education online course which we ran for the first time this year (5th-year French Mastère students specialising in e-learning approaches, learning management – ‘Gestion des apprentissages et formation ouverte et a distance’, at Montpellier 3 University), we gave the students a group task to complete³: evaluate a piece of software or online program

²Some aspects are treated in the previous criterion ‘Users, approach and content’. A ‘usability’ criterion could also be used, in which information is collected at stage 1. Whatever the reformulation is, this theme would remain dependent on information submitted by the authors. As a minimum, one could recommend that the competitor provides evidence of use, without implying that this constitutes evaluation of use.

³This task was allocated once the students had had sufficient electronic discussions and chat sessions with their supervisor (R. Panckhurst) as well as reading material to synthesise.

as if they were student jurors for the EASA 2004 competition. Once they had completed the task, they were then given the real evaluations that were provided by the EASA 2004 jurors (in anonymous format) in order to compare both series of evaluations. A second task was then requested: re-evaluate the initial piece of software with the new evaluation grid we had devised, and make comments about the revised criteria.

We found several interesting aspects in their evaluations, concerning both the old and new grids: the students came across problems that we had highlighted in previous competitions, for instance: what is research software? How can it be evaluated? How is e-learning linked with EASA? How does one differentiate between technical and pedagogical innovation? How does one evaluate a piece of software if it is integrated within a VLE? Can one differentiate between ease of use for the teacher and the learner? How are European standards incorporated etc.

Concerning the new grid, students made many interesting remarks (three of which appear below):

‘La nouvelle grille est plus précise, plus détaillée et permet de mieux comprendre ce qui est demandé.’

[The new grid is more precise, more detailed and allows [the juror] to understand what is required in a better manner.]

‘Le critère 1 spécifie bien mieux les questions. Dans l’ancienne grille, on avait tendance à mélanger l’aspect technique et pédagogique.’

[Criterion 1 is more specific in relation to the questions it addresses. In the old grid, technical and pedagogical aspects were confused.]

‘Le critère 5 étant supprimé, cela évite la confusion entre évaluation: tests pour les apprenants et évaluation du produit sur le marché.’

[Since criterion 5 has now been suppressed, there is no longer a confusion between tests for learners [users] and market evaluation of the product.]

4 Conclusion

Over and above a full revision of the EASA evaluation grid by the Evaluation Working Group set up for this purpose, we feel that the organising body of the competition, EKMA, has a crucial role to play concerning the key points which emerged in

the recommendations presented by the authors to the EKMA board⁴ and highlighted in our paper:

- EASA must be unique, distinct from all other European competitions.
- languages and cultures within the European area: all software products entered into the competition need to be more precisely identified as truly European, both from a linguistic and cultural perspective, i.e., some sort of language and/or cultural adaptation needs to be made mandatory (*cf.* Criterion 3 of new grid).
- EASA must remain a competition for academic products in any discipline, produced in higher(/further) educational institutions and designed for use at higher(/further) educational levels.
- programming (of some sort) must be a significant part of a valid entry to EASA. EASA should therefore accept the following sorts of software from eligible higher educational institutions:
 - stand-alone (possibly single task) software (e.g. for teaching, learning, research, learning support, educational management, etc.);
 - generic VLEs (with example content);
 - subject-specific VLEs (i.e. with content);
 - programmed content presented via a commercial VLE.

EASA should not accept ‘online’ content in a commercial VLE where there has been no programming input by the authors.

Acknowledgements

We would like to thank the other 5 members of the Evaluation Working Group – Peter Baumgartner (De), Jonathan Darby (UK), Per-Gotthard Lundquist (Se), Debra Marsh (Fr), Sabine Payr (Aut) – who made many crucial remarks and suggestions about the revised evaluation procedures.

The following institutions have sponsored this work: Centre National de la Recherche Scientifique (C.N.R.S.), The European Knowledge Media Association (EKMA), The Higher Education Academy, The Joint Information Systems Committee (JISC), Lancaster University, Université Paris 10 Nanterre, Université Paul-Valéry Montpellier 3.

Gilles Pérez-Lambert deserves our special thanks for his precious help with the L^AT_EX layout of this paper.

⁴All of the recommendations presented to the EKMA board by the authors were accepted for future incorporation in the forthcoming evaluation procedure.

References

- [1] Panckhurst R., David S., Whistlecroft L. (eds), 2004, *Evaluation in e-learning: the European Academic Software Award*. Montpellier: Presses universitaires de Montpellier.
- [2] David S., Panckhurst R. 2004, 'Comments on the current EASA evaluation process', Paper presented at a European workshop held in Montpellier, France, in November 2004, 'Evaluation in e-learning: review & future directions'.
- [3] David S., Panckhurst R. 2005 (submitted paper), 'Des approches évaluatives: un classement méthodologique', 'Evaluative approaches: a methodological classification', 15 p.
- [4] David S., Panckhurst R., Whistlecroft L. (2005), 'Many Forms of the Future. A report on future options for the organisation of EASA', Report submitted to the EKMA board and presented in Oxford, April 11, 2005, 38p + vii.
- [5] Cori M., David S., Léon J. 2002. 'Pourquoi un travail épistémologique sur le TAL', TAL 'Problèmes épistémologiques', Cori M., David S., Léon J. (eds), 43 (3), 7-22.
- [6] Chaudiron S. (ed.). 2004, *Evaluation des systèmes de traitement de l'information*, Paris: Hermès.
- [7] Hammond N. 2004, 'Evaluating academic software: can comparing chalk and cheese be valid, reliable or accountable?', in Panckhurst R., David S., Whistlecroft L. (eds), 2004, *Evaluation in e-learning: the European Academic Software Award*. Montpellier: Presses universitaires de Montpellier, 61-74.
- [8] Le Marec J. 2004, 'Les études d'usage', in Chaudiron S. (ed.), *Evaluation des systèmes de traitement de l'information*, Paris: Hermès, 353-372.
- [9] Payr S. 2004, 'EASA 2004. A Discipline Coordinator's Report', Paper presented to the EKMA board.

Appendices

Appendix 1. Individual juror evaluation grid (used in recent EASA competitions)



Individual juror evaluation form

Use this form to keep a record of your evaluation

Name of juror:

Id no of entry:

Short name of entry:

Criterion 1: Innovation

Is the project novel in approach or in terms of the activities it supports?

Added value: enables activities difficult to carry out by other means; *Distinctiveness:* supports activities not supported by other products; *Effectiveness:* supports novel activities more effectively than other products; *General:* advances the use of technology for education or research within the discipline area

Overall judgement for *innovation*: 4 3 2 1

Circle one number: 4 is best, 1 is worst; write your comments overleaf

Criterion 2: Design, and Ease of Use

Is the product or approach well-designed and easy to use or apply?

Installation/access: product is easy to install or access; *User interface:* user interface is easy to understand and use, and follows appropriate up-to-date standards; *Support:* documentation, on-line help etc is provided, and is appropriate and of high quality; *Screen design:* presentation and interactions are attractive, effective and appropriate for the target users and tasks; *Transferability:* product can be used on a range of machines available to intended users

Overall judgement for *design and ease of use*: 4 3 2 1

Circle one number: 4 is best, 1 is worst; write your comments overleaf

Criterion 3: European portability

Can the software or approach be used (or adapted for use) across Europe?

Language for use: product is available in different European languages; *Language for support:* installation procedures and support materials are appropriate for the languages of intended users; *Language adaptability:* the product can easily be adapted for different European languages; *Portability of materials:* subject materials are appropriate for use in a range of European countries; *Portability of approach:* approach adopted is appropriate for the different curricula and educational traditions/requirements across Europe

Overall judgement for *European portability*: 4 3 2 1

Circle one number: 4 is best, 1 is worst; write your comments overleaf

Criterion 4: Educational materials and approach

Are the materials and the approach educationally sound?

Users and objectives: target users, learning objectives and intended use are clear and adequately defined; *User needs:* project addresses real user (teacher, learner) needs; *Pedagogical approach:* educational approach is appropriate (e.g., at right level, provides appropriate learning activities and feedback, maintains motivation)

Overall judgement for *educational materials & approach*: 4 3 2 1

Circle one number: 4 is best, 1 is worst; write your comments overleaf

Criterion 5: Evaluation of use

Has the software or approach been evaluated, and how good is the evaluation?

Thorough evaluation procedure: product has been thoroughly and appropriately evaluated (e.g. by real users in real-life situations); *Results of evaluation:* evaluation provides evidence of the high quality of the product

Overall judgement for *evaluation of use*: 4 3 2 1

Circle one number: 4 is best, 1 is worst; write your comments overleaf

Appendix 2. Revised evaluation grid (to be used at EASA 2006)

1. Innovation

Recommendation: content innovation should be the only innovation criterion.

New grid: criterion name maintained

1. Novelty
(The product includes activities/approaches not covered by other products)
or
2. Improvement
(The product includes activities/approaches which are already covered by other products, but it does it in a better/more effective way)

2. Installation/access and interface

Recommendation: the installation/access procedure and the interface are checked from the user's point of view.

New grid: change of criterion name from *Design and ease of use*

1. Installation/access (the product is easy to install or access)
2. User interface
 - (a) Design and ease of use: the interface is easy to use and the screen design is attractive, effective. . .)
 - (b) Up-to-date standards
 - (c) Is the interface appropriate for different end-users (i.e., under-grad/post-grad/students, or research end-users, provisions for handicapped)?
3. Documentation (documentation, online help etc., is provided, and is appropriate and of high quality)

3. Europe: language & culture

Recommendation: we need to include multilingual, cultural and curricular aspects which are European specific.

New grid: Change of criterion name from *European portability*

1. Multilingual
 - (a) Software exists in 2 or more languages (n/a only for discipline reasons)
 - (b) Documentation/online help exists in 2 or more languages (glossary)
 - (c) Interface exists in 2 or more languages
2. Culture: differing methods — does the software help bring Europeans together or help understand cultural differences between Europeans?

4. Users, approach and content

Recommendation: the new criterion is more general and should be applicable to the different types of entries (research, teaching/learning, generic tools)

New grid: Change of criterion name from *Educational materials and approach*

1. Users & objectives
 - (a) the intended use of product is clear and adequately defined;
 - (b) the target users of product are clear and adequately defined.
2. User needs (the project addresses real user — teacher, researcher, learner — needs)
3. Approach (the educational/theoretical approach is appropriate)
4. Activities (appropriate activities/content are/is proposed; feedback is provided)

Evaluation of use

Recommendation: We suggest this criterion be eliminated for the following reasons:

- difficult to verify
- other criteria allow for discrimination (and check quality of product) anyway
- users can be checked at stage 1 in document submitted by authors.