Promoting Scientific Creativity by Utilising Web-based Research Objects

Project acronym: Dr Inventor

Deliverable No. 8.1
Initial Version of Evaluation Methodology
Grant agreement no: 611383

Dissemination Level

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PU</td>
<td>Public</td>
</tr>
<tr>
<td>PP</td>
<td>Restricted to other programme participants (including the Commission Services)</td>
</tr>
<tr>
<td>RE</td>
<td>Restricted to a group specified by the consortium (including the Commission Services)</td>
</tr>
<tr>
<td>CO</td>
<td>Confidential, only for members of the consortium (including the Commission Services)</td>
</tr>
</tbody>
</table>

COVER AND CONTROL PAGE OF DOCUMENT

Project Acronym: Dr Inventor

Project Full Name: Promoting Scientific Creativity by Utilising Web-based Research Objects

Deliverable No.: D8.1

Document name: Initial Version of Evaluation Methodology

Nature (R, P, D, O)¹ R

Dissemination Level (PU, PP, RE, CO)² RE

Version: 1

Actual Submission Date: 19/12/2014

Editor: Diarmuid P. O'Donoghue
Institution: (NUI) Maynooth University
E-Mail: diarmuid.odonoghue@nuim.ie

ABSTRACT:

This deliverable addresses the challenge of evaluation in computational creativity. The approach we adopt is to focus on the role of Dr Inventor as a creativity support tool (CST) and to assess it within this context. We adopt the newest version of (an independently developed and validated) psychometric assessment test as the main means of assessing the creative impact of Dr Inventor.

¹ R=Report, P=Prototype, D=Demonstrator, O=Other
² PU=Public, PP=Restricted to other programme participants (including the Commission Services), RE=Restricted to a group specified by the consortium (including the Commission Services), CO=Confidential, only for members of the consortium (including the Commission Services)
KEYWORD LIST:
Evaluation methodology, creativity support tool, creativity support index, creative achievement questionnaire.

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7-ICT-2013.8.1) under grant agreement n° 611383.

The author is solely responsible for its content, it does not represent the opinion of the European Community and the Community is not responsible for any use that might be made of data appearing therein.

MODIFICATION CONTROL

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Status</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>28/11/2014</td>
<td>Draft</td>
<td>Dr Yalemisew Abgaz</td>
</tr>
<tr>
<td>2.0</td>
<td>2/12/2014</td>
<td>Draft</td>
<td>Dr Yalemisew Abgaz</td>
</tr>
<tr>
<td>3.0</td>
<td>12/12/2014</td>
<td>Draft</td>
<td>Dr Yalemisew Abgaz</td>
</tr>
<tr>
<td>4.0</td>
<td>15/12/2014</td>
<td>Draft</td>
<td>Dr Diarmuid O'Donoghue</td>
</tr>
</tbody>
</table>

List of contributors
- Dr Yalemisew Abgaz, (NUI) Maynooth University, Ireland.
- Dr Donny Hurley, (NUI) Maynooth University, Ireland.
- Dr Diarmuid P. O'Donoghue, (NUI) Maynooth University, Ireland.
Contents

1 EXECUTIVE SUMMARY ........................................................................................................5
2 INTRODUCTION ..................................................................................................................6
  2.1 PURPOSE OF THIS DOCUMENT ..................................................................................6
3 CREATIVITY ASSESSMENT TECHNIQUES ......................................................................7
  3.1 CREATIVE ACHIEVEMENT QUESTIONNAIRE (CAQ) ..............................................7
  3.2 CREATIVE SUPPORT INDEX (CSI) .............................................................................8
  3.3 ADMINISTRATION OF CSI FOR DR INVENTOR .........................................................11
4 CONCLUSION ......................................................................................................................12
5 REFERENCES .........................................................................................................................12
APPENDIX 1 – ABBREVIATIONS AND ACRONYMS ...............................................................13
APPENDIX 2. A SAMPLE OF CREATIVE ACHIEVEMENT QUESTIONNAIRE ......................14
1 Executive Summary

This deliverable addresses the topic of evaluation in the Dr Inventor project, focused within the target area of computer graphics. It addresses the evaluation process and particularly focuses on metrics which may be useful in both the final evaluation, as well as guiding development of the creative element of the project. It should be acknowledged that evaluation is a noted problem within Computational Creativity – but one that really needs to be addressed.

Part of the overall objective in this deliverable is forward looking, being cognisant of the fact that the final evaluation will involve "T8.4 Evaluation of the system usability". This task is scheduled for months 25-36; a duration of 11 months. This duration gives scope for evaluation to feed-back into system development and refinement, with the objective of improving the evaluation. One of the implications of this leads to one desire implicit in this document, for an evaluation process and an evaluation metric that offers sufficiently fine-grained feedback in order to support incremental improvements.

These considerations have led us to choose the Creativity Assessment Index (CSI) (Cherry and Latulipe, 2014) as the central metric with which we shall evaluate Dr Inventor. The CSI looks primarily at the factors that are relevant to supporting creative work processes. This is a new psychometric test but initial versions were first published in 2009 (Carroll and Latulipe, 2009), showing it to be reasonably well developed and “fit for purpose”. Importantly, a number of significant improvements have been made to the metric since its initial presentation. Furthermore, this metric has been used and has also been validated by a number of trials. It is quick and easy to administer and takes the form of a user survey answering a number of rating questions followed by a number of choice questions. An open source tool exists that can be easily adapted to the specific needs of Dr Inventor.
2 Introduction

2.1 Purpose of this document

The purpose of this document is to identify the initial work that has been conducted into “T8.1 Investigation of evaluation methodology”. This task has a number of specific objectives, foremost of which is the identification of an appropriate process for evaluating the Dr Inventor project. Evaluation of computational creativity is a widely discussed problem, but one that is only beginning to attract significant amounts of attention from the academic community. A recent review of creativity assessment publications (Silva et al, 2012) identified that much of evaluation has focused on the artistic and other styles of creativity – not necessarily even focusing on computational creativity. This paper focuses on evaluating scientific computational creativity and on evaluating creativity support tools.

This initial report identifies a newly published metric to support the evaluation process. This is a new psychometric test that has been independently developed and validated for evaluating creativity support tools. It is a development of an earlier version of this work and contains a number of improvements.
3 Creativity Assessment Techniques

Evaluation in the Dr Inventor project is aimed at evaluating the ability of the Dr Inventor tool to act in its role of creativity support. This will focus on Dr Inventor and its effectiveness as a Creativity Support Tool (CST) - rather than its operation as a model of human creative reasoning. CST tools often function simultaneously as productivity tools and as creativity support tools. Evaluation in Dr Inventor aims to focus primarily on its role in creativity support tools. Thus we wish to identify metrics that measure how well Dr Inventor supports open-ended creative tasks. Shneiderman (in 2007) and others note that for CST's there are no obvious metrics to quantify and this problem lies at the core of creativity assessment and evaluation. Others have noted that there is no one-size-fits-all approach and so a variety of metrics are required for a proper evaluation.

3.1 Creative Achievement Questionnaire (CAQ)

The Creative Achievement Questionnaire (CAQ) (Carson, 2005) is a commonly used method of assessing creativity. CAQ provides a good scale to evaluate creativity using questions from low creative task to high creative task. It further allows the users to provide numeric values to quantify their achievements. The CAQ is very general and measures creative achievement of individuals in 10 different domains. The domains include: Visual arts, Music, Dance, Architectural design, Creative writing, Humour, Inventions, Scientific discovery, Theatre and film, and Culinary Arts.

The CAQ takes the form of a questionnaire and is completed as a self-reporting exercise. Each domain has 8 items rated from 0 - 7 where 0 indicates no training, experience or accomplishment and 7 indicate a higher level of creativity. Users are asked to tick the most applicable item to their creative activity. Users can provide further numeric values to quantify their number of achievements. Each question indicates a progressive creative achievement, so endorsing a high-ranked item implies endorsing all of the prior items. For each domain, the scale corresponding to the user’s response will be used. In the case of supplied numeric values the scale will be multiplied by the number. The overall result is calculated by taking the sum of the results of the individual domains. The larger scale the respondent answers the better the result. Sample questions from CAQ are listed in Appendix 2.

While much of the CAQ is not relevant to the needs of Dr Inventor, the CAQ also covers a few items that are interesting and relevant to assessing Dr Inventor’s creativity. Some of these items are listed under the heading of Inventions and Scientific Discovery domains that include:

Note: the Write number instruction on the following items indicates that the respondent is required to list the number of times that this particular item applies to them. Thus, for the first question below, the respondent is required to indicate the number of patents that have been awarded to them.
Inventions

I have received a patent for one of my inventions. *Write number.*
I have sold one of my inventions to a manufacturing firm. *Write number.*

Scientific Discovery

I have been author or co-author of a study published in a scientific journal.
I have received a grant to pursue my work in science or medicine. *Write number.*
My work has been cited by other scientists in national publications.

However, the use of the CAQ as a general evaluation tool for Dr Inventor is not particularly useful for the following reasons. Firstly, the CAQ categories are very general and evaluate the overall creativity rather than creativity in a specific domain. Secondly, CAQ focuses on the individual’s creativity and does not evaluate the creative assistance provided by a tool. Thirdly, the response categories of the CAQ are very broad and do not provide good coverage of lower levels of creative achievement (before publishing in a scientific journal) that would be useful in guiding the development of Dr Inventor. Finally, the majority of the domains in the CAQ are not relevant to the graphics domain of Dr Inventor.

If the final stages of development of Dr Inventor are successful and it becomes adopted by many users, then the CAQ may be useful in quantifying how many publications refer to Dr Inventor as, perhaps a contributor to a publication. We do not foresee Dr Inventor being listed as an author of a publication, as writing journal papers is not a direct objective of the project.

### 3.2 Creative Support Index (CSI)

The CAQ might be useful for the final evaluation of Dr Inventor – possibly even outside the lifespan of the project funding. However, the more immediate goal is an evaluation methodology and metric to assist during development and early evaluation of the system. Ideally, this should be focused on creativity supported by the tools, rather than evaluating Dr Inventor as a model of human creativity.

The Creative Support Index (CSI) is an assessment model specifically designed to evaluate the creativity support provided by a tool or a system. The CSI is composed of two sections; a rating scale section and a paired-factor comparison section. CSI identified 6 major factors of creativity, namely: *Enjoyment, Exploration, Expressiveness, Immersion, Results Worth Effort* and *Collaboration*. Under each of the factors, CSI asks two questions that are rated between 0 and 10. 0 indicates the lowest value and 10 indicate the highest achievement.
The 12 questions are:

<table>
<thead>
<tr>
<th>Collaboration</th>
<th>Enjoyment</th>
<th>Exploration</th>
<th>Expressiveness</th>
<th>Immersion</th>
<th>Results Worth Effort</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The system or tool allowed other people to work with me easily.</td>
<td>1. I would be happy to use this system or tool on a regular basis.</td>
<td>1. It was easy for me to explore many different ideas, options, designs, or outcomes, using this system or tool.</td>
<td>1. I was able to be very creative while doing the activity inside this system or tool.</td>
<td>1. My attention was fully tuned to the activity, and I forgot about the system or tool that I was using.</td>
<td>1. I was satisfied with what I got out of the system or tool.</td>
</tr>
<tr>
<td>2. It was really easy to share ideas and designs with other people inside this system or tool.</td>
<td>2. I enjoyed using the system or tool.</td>
<td>2. The system or tool was helpful in allowing me to track different ideas, outcomes, or possibilities.</td>
<td>2. The system or tool allowed me to be very expressive.</td>
<td>2. I became so absorbed in the activity that I forgot about the system or tool that I was using.</td>
<td>2. What I was able to produce was worth the effort I had to exert to produce it.</td>
</tr>
</tbody>
</table>

The paired-factor comparison section consists of each factor paired against every other factor for a total of 15 comparisons. The result will be used as a weighting for the individual factors. The authors of the CSI note that it can be completed in 5 minutes and “does not place any significant burden on the participants to complete the study”.

CSI is a suitable tool for evaluating Dr Inventor as a creativity assistance tool due to the following reasons. Firstly, CSI is tailored towards evaluating Creativity Assessment Tool. Secondly, the six factors are general to evaluate creativity and are applicable to our domain. Further, the authors of CSI have used CSI to evaluate computer applications that support creativity. Finally, its ease of use and the customizability of the factors (paired-factor comparison that allows weighting of the factors) make it a preferable tool. The authors of CSI further provide a java application to administer the question. Thus, we will use CSI for evaluating Dr Inventor with some allowed customization of the questions to fit our target tool and the extension of a web based application to allow wider participation. Our customized version of the questions is presented in Figure 1 and Figure 2. Samples of the paired-factor comparisons are presented in Figure 3 and Figure 4.
We have customised the questions by changing the “System or tool” phrase by “Dr Inventor” where it is clearer or more expressive. For example, the first question on Figure 1 was “I enjoyed using the system or tool” and this is changed to “I enjoyed using Dr Inventor”.

Figure 1. Screenshot of the second six of the 12 customized agreement statements

Figure 2. Screenshot of six of the 12 customized agreement statements

Figure 3. A screenshot of the paired-factor comparison 2/15

Figure 4. A screenshot of the paired-factor comparison 1/15
3.3 Administration of CSI for Dr Inventor

Each participant of the Dr Inventor evaluation will be given a chance to use Dr Inventor. Once the participants have used the system, they will be asked to fill out the online version of these 12 agreement statements and the 15 paired-factor comparisons. Participants will be given “sufficient time” to learn and use Dr Inventor before they are asked to evaluate the system. “Sufficient time” will include the time it takes to learn the Dr Inventor tool and to exploit it for the proposed purpose. This will depend on different factors such as the ease of use of the system, the learning curve etc. Users will be asked to use Dr Inventor to perform use cases such as “create new idea” discussed under D2.1 “user requirement and use case report”. At the end of the task, they will fill the CSI agreement statements and the paired-factor comparison. The data from each respondent will be collected and calculated accordingly to the CSI specifications (Carroll and Latulipe, 2014) and will be stored. Before conducting the evaluation, respondents will be informed about the purpose of the evaluation and will be given assurance to the confidentiality of their response and the restricted use of the data only to the intended purpose.

The CSI score is on a scale of 0-100 with a score of 100 indicating the highest levels of creativity support. An individual CSI score appears to be a combination of influences arising from the tool, the task and the expertise level of the user. As Dr Inventor will support users with different levels of expertise, this factor in particular will have to be controlled for during the evaluation process. Analysis on the individual factors can also be performed and Dr Inventor can be evaluated against each individual factor. We also point out that Cherry and Latulipe (2014) note that the CSI appears to rate complex CST tools as higher than simple ones – but that this is somewhat moderated by the difficulty of the creative task itself.
4 Conclusion

Evaluation is a crucial part of the Dr Inventor project, and is a particular challenge faced by many computationally creative systems. Dr Inventor adopts the CSI as its main psychometric test for evaluation purposes. While this is a newly published test, earlier versions were published and have led to the improvements embodied in the 2014 version of this metric. It is quick and easy to administer and allows for relatively fine-grained judgements of the creative impact of Dr Inventor on its users.

5 References


Appendix 1 – Abbreviations and acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAQ</td>
<td>Creative Achievement Questionnaire</td>
</tr>
<tr>
<td>CSI</td>
<td>Creative Support Index</td>
</tr>
<tr>
<td>CST</td>
<td>Creativity Support Tool</td>
</tr>
</tbody>
</table>
Appendix 2. A Sample of Creative Achievement Questionnaire

Place a check mark beside sentences that apply to you. Next to sentences with an asterisk (*), write the number of times this sentence applies to you.

A. Visual Arts (painting, sculpture)
   ____ 0. I have no training or recognized talent in this area. (Skip to Music).
   ____ 1. I have taken lessons in this area.
   ____ 2. People have commented on my talent in this area.
   ____ 3. I have won a prize or prizes at a juried art show.
   ____ 4. I have had a showing of my work in a gallery.
   ____ 5. I have sold a piece of my work.
   ____ 6. My work has been critiqued in local publications.
   *___ 7. My work has been critiqued in national publications. (Write number.)

B. Music
   ____ 0. I have no training or recognized talent in this area (Skip to Dance).
   ____ 1. I play one or more musical instruments proficiently.
   ____ 2. I have played with a recognized orchestra or band.
   ____ 3. I have composed an original piece of music.
   ____ 4. My musical talent has been critiqued in a local publication.
   ____ 5. My composition has been recorded.
   ____ 6. Recordings of my composition have been sold publicly.
   *___ 7. My compositions have been critiqued in a national publication. (Write number.)

C. Dance
   ____ 0. I have no training or recognized talent in this area (Skip to Architecture)
   ____ 1. I have danced with a recognized dance company.
   ____ 2. I have choreographed an original dance number.
   ____ 3. My choreography has been performed publicly.
   ____ 4. My dance abilities have been critiqued in a local publication.
   ____ 5. I have choreographed dance professionally.
   ____ 6. My choreography has been recognized by a local publication.
   *___ 7. My choreography has been recognized by a national publication. (Write number.)

D. Architectural Design
   ____ 0. I do not have training or recognized talent in this area (Skip to Writing).
   ____ 1. I have designed an original structure.
   ____ 2. A structure designed by me has been constructed.
   ____ 3. I have sold an original architectural design.
   ____ 4. A structure that I have designed and sold has been built professionally.
   ____ 5. My architectural design has won an award or awards.
   ____ 6. My architectural design has been recognized in a local publication.
   *___ 7. My architectural design has been recognized in a national publication. (Write number.)