

Evaluating Visualizations Based on the Performed Task

Octavio Juarez

Robotics Institute, Carnegie Mellon University, Pittsburgh, PA

Chris Hendrickson, James H. Garrett, Jr.

Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA

Abstract

This paper describes an ongoing project with the goal of designing and implementing a method to evaluate visualizations based on the tasks supported. The method evaluates time to perform the task and the quality of the user solution. This method is described in detail in this paper.

A software tool which uses visualization to support users performing environmental analysis was used to test the first version of this method. The software tool evaluated helps users to perform the computations of environmental impacts of two products. Two versions of the same software tool were used. The first version uses interactive tables and the second version displays the information using different types of interactive visualizations.

The software tool was evaluated with users solving a task, which consists of selecting a product from two alternatives based on environmental parameters. The users were divided into two groups. Each group of users solved the same task but they were using different software tools. A summary of findings and future research is presented at the end of this paper.

1. INTRODUCTION

Information visualization research has produced many techniques and metaphors for visualizing data. However, only some of the techniques developed have been evaluated with users.

The research topics in this area can be categorized as follows: space, interaction, focus+context, document visualization, infosphere, workspace tools, and objects [1]. Visualization evaluation must be included as research topic in information visualization. Some research groups have been concerned with the evaluation of graphics. For example, Larkin and Simon found that graphics are better than words sometimes [4]. Other studies have defined effectiveness of visualizations as measures of how well a

visualization exploits the display media and the user's visual system [3].

The community developing decision support systems has been studying the effects of graphics and tabular presentations on user performance. However, the conclusions are not clear regarding the effectiveness of tables and graphic presentations [2].

While in some situations tables have been shown to be better than graphics, other studies have shown the opposite to be true. Additionally, some studies have shown no difference between tables and graphics. To avoid these controversies some researchers have proposed creating a taxonomy of tasks to interpret the results based on that taxonomy [2].

Vesey tried to explain the differences in these results based on information processing theory:

- Tasks can be classified as either spatial or symbolic;
- Graphic representations emphasize spatial information, and tables emphasize symbolic information;
- The performance is improved when there is a match between the information emphasized by the representation and the information required by the task.

With these findings, Vesey explained the differences in the results of performance using graphics and tables. The theory used to explain the difference is called cognitive fit and it is concerned in the effects on performance of matching the problem representation and the type of task.

In this research, the concern is the evaluation of visualizations in the context of tasks. Although previous research compared tables against graphics, the user was not able to interact with the graphic to transform the representation of the data. Visualization techniques include graphic representations and interaction techniques which makes them different from traditional graphic representations.

The evaluation method under development considers the time to perform the task and the quality of the solution of the task evaluated. In this research, the users generate

their own graphics or tables instead of receiving tables and graphics previously generated.

The first version of this method was used to evaluate the visualizations used to solve environmental problems. These visualizations inform users about the environmental life-cycle assessment (LCA) and they were generated using the software described in the following section.

2. Visual EIO-LCA

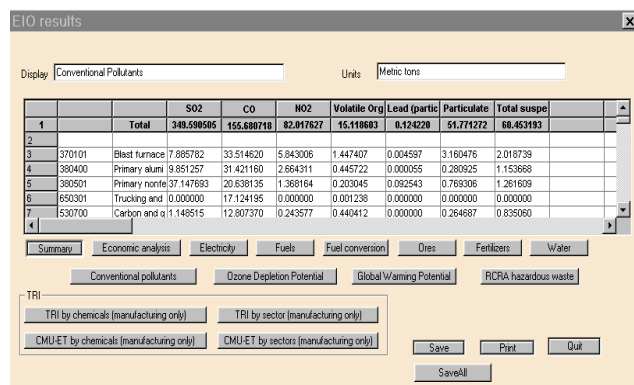
The Economic Input-Output Life-Cycle Assessment (EIO-LCA) helps environmental managers to estimate environmental impacts of products or services. Although a life-cycle assessment (LCA) can take three months, the problems selected to evaluate the system are simple enough to be solved in a maximum of three hours.

2.1 EIO-LCA

Life-Cycle Assessment (LCA) is a tool used by environmental scientists to evaluate the environmental impacts of a product or service during its entire life-cycle [7].

The reason to evaluate the entire life-cycle of a product is that environmental emissions might happen during any life-cycle period, such as manufacturing, material extraction, or end-of-life.

A life-cycle study is divided into the following stages: inventory analysis, impact analysis, and improvement analysis. During the inventory analysis step, the materials, energy, and emissions during the entire life are computed. The impact analysis consists of assessing the effects of emissions and discharges during the inventory step. Finally, the improvement analysis step looks for opportunities to minimize emissions, material usage, and environmental discharges [7].



The screenshot shows the 'EIO results' window. At the top, there's a 'Display' dropdown set to 'Conventional Pollutants' and a 'Units' dropdown set to 'Metric tons'. Below this is a table with 10 columns: '1', 'Total', 'SO2', 'CO', 'NO2', 'Volatile Org', 'Lead (partic)', 'Particulate', 'Total suspe', and a final column with a triangle icon. The table contains 7 rows of data. Below the table are several tabs: 'Summary', 'Economic analysis', 'Electricity', 'Fuels', 'Fuel conversion', 'Ores', 'Fertilizers', and 'Water'. Under the 'Summary' tab, there are sub-tabs for 'Conventional pollutants', 'Ozone Depletion Potential', 'Global Warming Potential', and 'RCRA hazardous waste'. At the bottom, there are buttons for 'Save', 'Print', 'Quit', and 'SaveAll'.

1	Total	SO2	CO	NO2	Volatile Org	Lead (partic)	Particulate	Total suspe	
2	349.595985	155.680716	82.017627	15.118683	0.124228	51.771272	66.453193		
3	370101	Blast furnace 7.685762	33.514620	5.843006	1.447407	0.004597	3.160476	2.016739	
4	380400	Primary steel 9.851257	31.421160	2.864311	0.445722	0.000055	0.200925	1.153688	
5	380501	Primary nonfer 37.147693	20.638135	1.368164	0.203045	0.002543	0.769306	1.261609	
6	550301	Trucking and 0.000000	17.124195	0.000000	0.001238	0.000000	0.000000	0.000000	
7	530700	Carbon and q 1.148515	12.807370	0.243577	0.440412	0.000000	0.264687	0.835060	

Figure 1: Display of EIO-LCA Results

Economic Input-Output Life-Cycle Assessment (EIO-LCA) is a specific method used to perform LCA [8,9,10,

11, 12] based on the economic method named input-output [6]. This method receives as inputs a change in the demand, in millions of dollars, for particular industrial sectors, and it computes the environmental effects due to these changes for the entire supply chain. The data used in this method are publicly available. The economic data used is the commodity-by-commodity matrix that reflects the economic information for a particular country. The toxic release inventory (TRI) for the same year as that of the economic data, and other environmental vectors, such as energy, fuels, and ores consumed for every industrial sector, are components of the EIO-LCA data.

2.2 EIO-LCA Software

An evaluation of the effectiveness of visualizations used to support a task was performed using two different versions of a software tool to perform EIO-LCA. The main difference between them is that Visual EIO-LCA has extra components for results analysis.

EIO-LCA. This version of EIO-LCA software consists of graphic user interfaces based on tables [12]. The graphic user interface receives the problem defined by the user as a vector of million-dollar quantities and the system presents the results which consists of 75 vectors with 485 elements each. The results are displayed using a grid as seen in Figure 1. The data is divided by vector families which can be sorted. The user selects data subsets and then sorts the family based on the environmental vectors. For example, they select conventional pollutants and sort the data set based on the carbon monoxide data.

One of the problems with this interface is that a user can not perform comparisons of several products in the same workspace. Another problem is that a user can not see two data subsets in the same screen. To achieve these tasks, a user saves the data and manipulates them in a spreadsheet application, which is time consuming.

Visual EIO-LCA. The architecture of VisEIO-LCA supports the following three views: chart view, scatter plot view, and data view. The modules were derived from the task analysis performed and the visualization designs based on the tasks.

The chart and the scatter view allows EIO-LCA results analysis to be explored while the data view module allows users to navigate in the data sets used in EIO-LCA.

The graphic interface for the software prototype is shown in Figure 2 and allows users to select any of the three views shown in Figure 2.

The information is presented using the following three windows: the *Project Window*, the *Workspace Window*, and the *Detail Window*.

The *Project Window* is shown in Figure 2 to the left of the window that displays the chart. This window displays the data sets that a user can visualize and at the bottom has three tabs to select the data view. The first tab allows users to display charts, the second tab allows users to display the data sets used for the EIO-LCA software, and the third tab allows users to view the data using scatter plots.

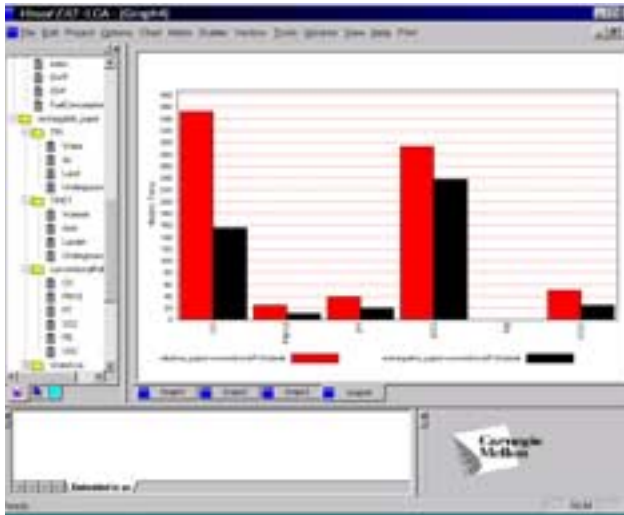


Figure 2: Visual EIO-LCA Graphic User Interface

The *Workspace Window* is the space where the graphics are displayed. In Figure 2, the *Workspace Window* presents a bar chart.

Finally, the *Detail Window*, located to the left of the CMU logo in Figure 2, allows users to see detailed information about the data and scatter views.

The data views that a user sees are described in the following paragraphs.

- **Chart View.** This view allows users to view the data sets in three levels of aggregation: summaries, summaries by family, and detailed data by sector.
- **Scatter View.** This is a view that allows the user to see the results in detail. The complete environmental vector can be displayed.
- **Data View.** This view allows users to explore the data sets used by EIO-LCA instead of data produced by using EIO-LCA. This view is provided to users in order to allow them to obtain some explanations using EIO-LCA data. Figure 3 shows the visualization of four environmental vectors which represent the data for the Toxic Release Inventory (TRI). The data values are divided into ranges of values and are represented using rectangles that are colored based on the value's category.

3. Evaluation Method

The motivation for creating an evaluation method for the visualizations added to a software tool is the production of more efficient software. The method should consider more information in addition to the time to perform a task in a graphic. The evaluation method includes information about the context provided by the task solved and the quality of the solution.

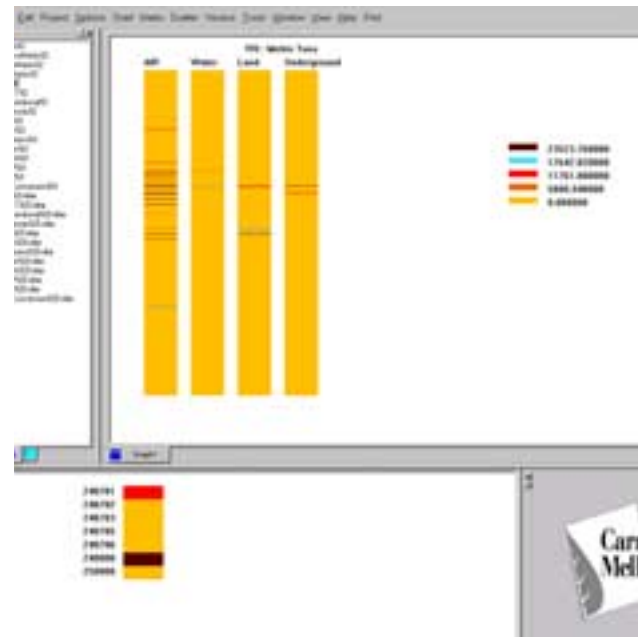


Figure 3: Environmental Data View

The evaluation method designed evaluates a software system in the process of solving a problem and provides qualitative and quantitative information about the solution process followed by users to solve problems.

The variables used in this study were the task performance time and the solution quality.

The features of this method are:

- it requires little time;
- it requires few resources (equipment, time of information procession);
- it supplies information to the software development process;
- it provides more data than one inspection method;
- it provides information about users solving problems.

The evaluation of solution quality combined with time performance is not usually applied when evaluating software to support environmental managers.

3.1 Performance Time

Performance time is the most common variable measured in software system evaluations. No special equipment is required to capture this data component.

In the evaluation of environmental software, time predictions were not performed because the time ranges varied based on the LCA problem complexity.

The total time required to complete a task can be obtained by adding up the set of performance times for each subtask of the general task.

3.2 Solution Quality

Time is a good parameter to use when measuring performance; however, fast answers do not mean high quality decisions.

Solution quality is hard to measure because it is subjective. However, quality might be defined for specific domains by a group of experts as a set of standards in procedures and parameters. Quality criteria are less robust than time measures but they are needed in order to evaluate a system in the context of a problem solution. Time and quality complement information about user task performance.

In this work, quality was measured for every subtask of the task evaluated. Experts assigned a number between 1 and 10 to the different subtasks performed by the users who solved a problem.

Although the methods used to evaluate the solution quality need to be more refined, the approach followed in this study considered how well users solved the problem using visualization. A more standard method to assess visualization quality should be developed in the future.

3.3 The Procedure

The set of steps proposed here to evaluate the visualization software are the following: define the task and set the boundaries to perform the evaluation; design the experiment; perform the experiment; process the results; and provide feedback information to the process of design and implementation of the software tool.

Define the Task and Set the Boundaries to Perform the Evaluation

The software evaluation is performed for each task in the catalog. Only one task is evaluated at a time. The task is evaluated with a problem that allows researchers to test the software system in the problem solution process.

When the number of tasks is too large, only a reduced set of tasks is selected for evaluation. Task subsets are selected based on costs in resources and time.

Once the task being evaluated is selected, it is divided into subtasks. This division is needed to focus attention only on subtasks, which can be performed better by using the software and the visual displays.

Design the Experiment. The experiment planning includes determining the subjects, the tasks to be solved, the time limits, and the data provided to users. Dimensions to measure and control variables should be determined. The number of research participants should be low in order to minimize time and resources required for the experiment.

The description must be detailed in order to document the experiment performed to evaluate the system. The documentation can be used to interpret the results.

The constraints to reducing time and resource costs are:

- reducing the number of experiment participants; and
- selecting simple problems to observe users in reasonable periods of time.

Perform the Experiment. This stage is one of the most time-consuming and requires the scheduling of the participants and the resources used in the research. Some of the activities to be performed in this stage are:

- arrange a physical space to perform the experiments;
- prepare the printing materials to use with each user;
- test the software;
- solve the problems to locate any possible bugs in the coding;
- perform several pilot studies before the final applications;
- apply the test.

Process the Results. The results are analyzed in two groups: quantitative data and qualitative data.

The quantitative data includes the descriptive statistics of performing time and solution quality.

The qualitative data is obtained from the process of solution followed by the users. Qualitative data includes information about user strategies to solve the problem with and without visualization tools.

Provide Information to the Process. Finally, a group of suggestions to improve the computer system performance were identified. The suggestions were grouped based on the structural variables of the model: the user, the computer, the task, and the data. The data was then sent to the design team to be incorporated into a new design.

4. Defining the Task and Setting the Boundaries to Perform the Evaluation

The task used to evaluate the system is the comparison of two alternative designs of a product. To compare two alternative designs a user needs to perform an LCA evaluation for each product being compared.

The comparison of two products might be affected by the performance of structural variables. The structural variables, which are the user, the computer, and the data quality, have a collective impact on how well a user performs the task. Also the size and type of a task affects the system performance.

How well the products are compared depends on user knowledge, user experience, user cognitive strategies, computer user interface, and the data set quality. Also the problem solution is affected by the size and type of task used. In this evaluation the data quality is not considered because the data used to support the decision is the same for the three software versions.

LCA evaluations consist of three tasks: problem formulation, data analysis, and results report. The software tool supports data analysis. The problem formulation is supported partially.

5. Experiment

This experiment was designed to explore the effects of adding information visualization components to the EIO-LCA software in order to influence problem solution. It is likely that users with graphic tools improve the process of results analysis. The selection of the most environmentally-friendly product might be affected by the data set selected to inform the decision. Differences in data sets used to inform the same decision might be produced because users have different criteria regarding the data sets used. The level of detail used to inform their decisions might be affected by how confident a user is about the data sets and how well they understand the model.

5.1 Materials

Subjects were presented with two problems involving a comparison of two products based on their environmental effects, and they received a page with questions about the decision made to select a more environmentally-friendly product. They used a software installed in a computer to perform the data analysis. The software components used were different based on the group.

The software components provided consist of:

- EIO-LCA software to perform LCA studies; and

- The EIO-LCA visual component, which allows users to perform data analysis.

The handouts included the problem description and data needed by users to solve the problem.

5.2 Analysis measures

Subjects were asked to provide the following information:

- Which product was selected based on the environmental effects.
- Reasons why the product model was selected.
- Numerical data considered to support the decisions.

5.3 Subjects

The subjects who participated in this experiment were divided randomly into two groups. Each group used different software versions and each software version had a different number of components to support user tasks. The users were novices because most of them had not performed LCA studies before. The total group of participants in the experiments were enrolled in a course in environmental management.

5.4 Procedure

Participants were given time to familiarize themselves with the software tools before using them to perform the data analysis task. The time used to perform the data analysis was recorded. The grading of the user answers was completed by experts without knowing the software version used.

A number between 1 and 10 was assigned to the participant answers based on domain expert criteria.

6. Results

The results have been analyzed using simple descriptive statistics; however, they are complemented with a qualitative analysis, which provides more information about the solution process.

The population examined consisted of 10 students. The analysis was separated into time analysis and qualitative analysis for each problem component.

6.1 Performing Time

The time data summary is presented in Figure 6, which displays the four variables using boxplots. The vertical axis in the graphic shows the time in minutes. The variable

represented by each box is performing time. The first two boxes belong to the batteries problem and the third and fourth boxes belong to the coffee makers problem.

Every group in Figure 6 represents a group of students solving a problem with one software tool version. The EIO-LCA group is that which used the software without graphic tools, and the VEIO-LCA group used the software with graphic tools. To discriminate the data from different problems the columns are labeled with (B) for the batteries problem and (CM) for coffee makers.

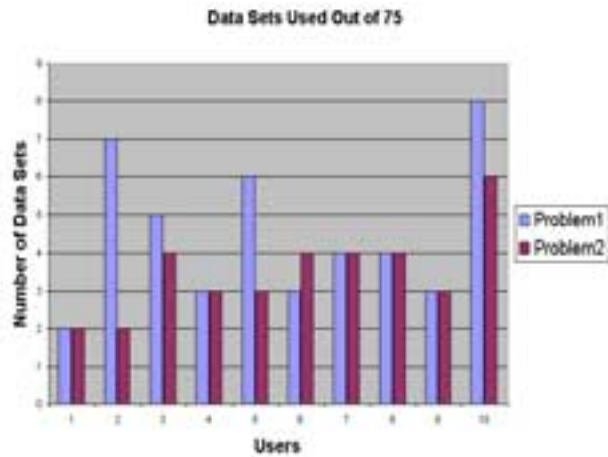


Figure 4: Data Sets Used to Solve the Problems

The groups using the software with visualizations spent less time on the tasks. However, the two groups improved their performance when they used the software tool the second session. The time reduction might be caused by the users learning. However, the group of users with the lower performing time and less variation was the group that used the software with graphics. In Figure 6, every box contains a horizontal line to represent the median value. The shaded area represents an interval around the median with 95% of confidence.

In Figure 6 it can be seen that both groups had a high variation during the solution of the first problem.

6.2 Solution Quality

Figure 7 shows a boxplot that summarizes the quality data for the two groups solving the two problems. The vertical axis represents the grades assigned to the solution by the EIO-LCA experts.

The quality of solution improved from the solution of the batteries problem to the coffee makers problem. The variation of the solution quality variable is too low for the users using the software with graphics. The star in the graphic represents an extreme value.

There is a difference in the quality of solution of users solving the problem using the software with graphics and the software without graphics.

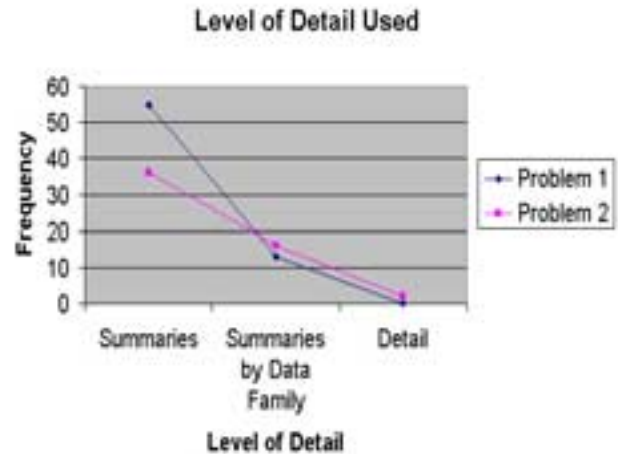


Figure 5: Level of Data Detail Required by Users

6.3 Other Results

Other information that was obtained from this analysis includes the following:

As can be seen in Figure 4, the users made their decisions based on a subset of the data sets. They did not use the complete set of 75 vectors. Most of the users only needed to review 3 or 4 data sets to support their decision.

Figure 5 shows the level of detail used when users analyzed the data sets. As can be seen in Figure 5 users prefer to work in the level of summaries. Only a few of them focus their attention on the data sets details.

7. Providing Information to the Process

The software will be redesigned based on the data produced in this evaluation. The changes consists of:

- displaying the summary views first because users are interested first in summaries;
- allowing users to select the subset of data that they want to use in the analysis.

8. Summary and Future Research

This is an ongoing project. The evaluation method was prototyped with a few students but the results obtained were very useful for obtaining design feedback.

An important piece of information obtained from the evaluation is that users might not need to use complex visualization techniques because they worked in the level of summaries.

However, users that worked with the graphic component performed the task more efficiently.

The method to define the quality of solution needs to be improved in order to get more standard and real results.

Some of the future tasks consist of:

- creating a task taxonomy and a definition of quality solution for each of type of task;
- creating models to perform theoretical analysis of task performance using visualization systems;
- incorporating other performance parameters such as data quality.

9. References

- [1] Card, S., Mackinlay, J., Shneiderman, B., "Readings in Information Visualization: Using Vision to Think," Morgan Kaufmann, 1999.
- [2] Vessey, I., "Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature," *Decision Sciences*, Vol. 22, 219-240, 1991.
- [3] Mackinlay, J. D., "Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics* Vol. 5, 110-141, 1986.
- [4] Larkin, J.H. and Simon, H.A. (1987) "Why a diagram is (sometimes) worth 10,000 words," *Cognitive Science*, vol. 11, pp. 65-100. This is the second reference.
- [5] Juarez, O., Hendrickson, C, Garrett, J., "Domain Analysis: A Technique to Design a User-Centered Visualization Framework," *Proceedings of the 1999 IEEE Symposium on Information Visualization*.
- [6] Leontief, Wassily, "Input-Output Economics." New York: Oxford University Press, 1941.
- [7] Vigon B.W. et al, "Life-Cycle Assessment Inventory Guidelines and Principles." EPA/600/R-92/245, February 1993.
- [8] Cobas E., "Life Cycle Assessment Using Input-Output Analysis," PhD Thesis, Pittsburgh, PA, Carnegie Mellon University, 1996.
- [9] Horvath, A. "Estimation of Environmental Implications of Construction Materials and Designs Using Life-Cycle Assessment Techniques," PhD. Thesis Department of Civil and Environmental Engineering, Carnegie Mellon University, 1997.
- [10] Joshi, S. "Comprehensive Product Life Cycle Analysis Using Life-Cycle Assessment Techniques," PhD. Thesis, Heinz School of Public Policy and Management, Carnegie Mellon University, 1998.
- [11] Hendrickson, Chris.; Horvath, Arpad.; Joshi, Satish., "Economic Input-Output Models for Environmental Life-Cycle Assessment," *Environmental Science & Technology*, v. 32, no. 7, (April 1998) p. 184A-191A.
- [12] Green Design Initiative, "EIO-LCA software," www.eiolca.net, 1999.

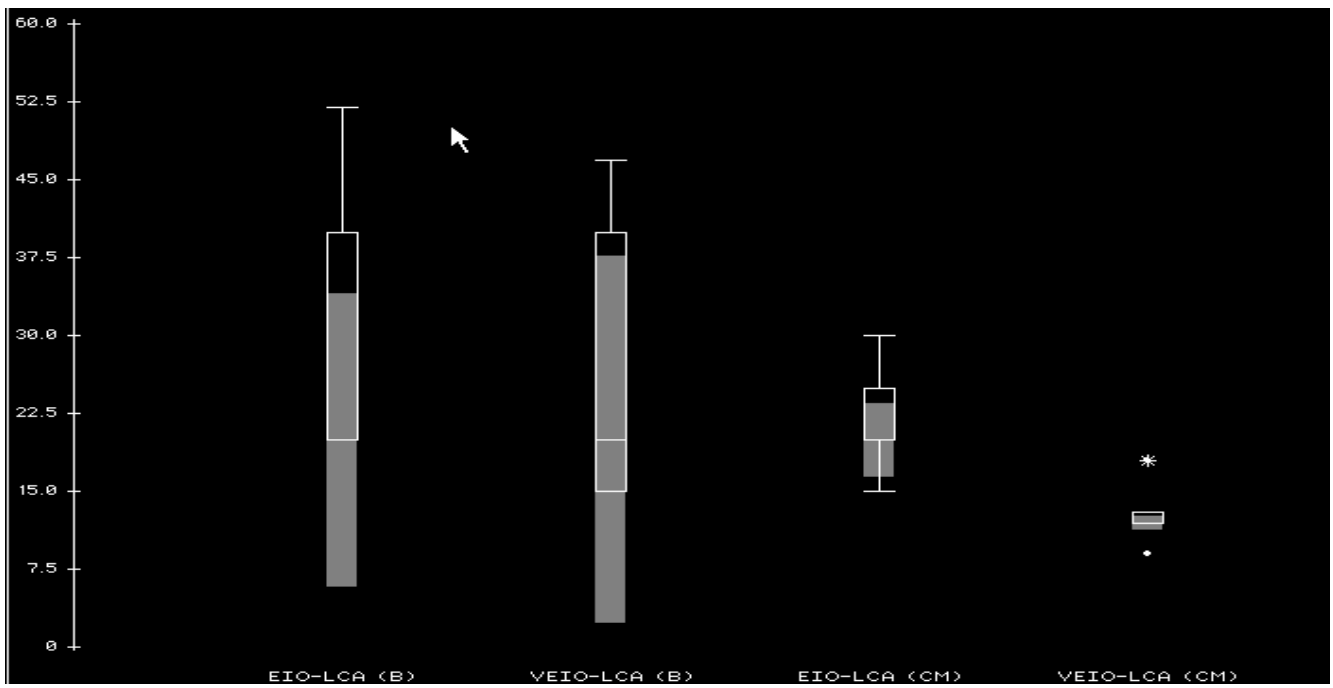


Figure 6: Summary of Performing Times

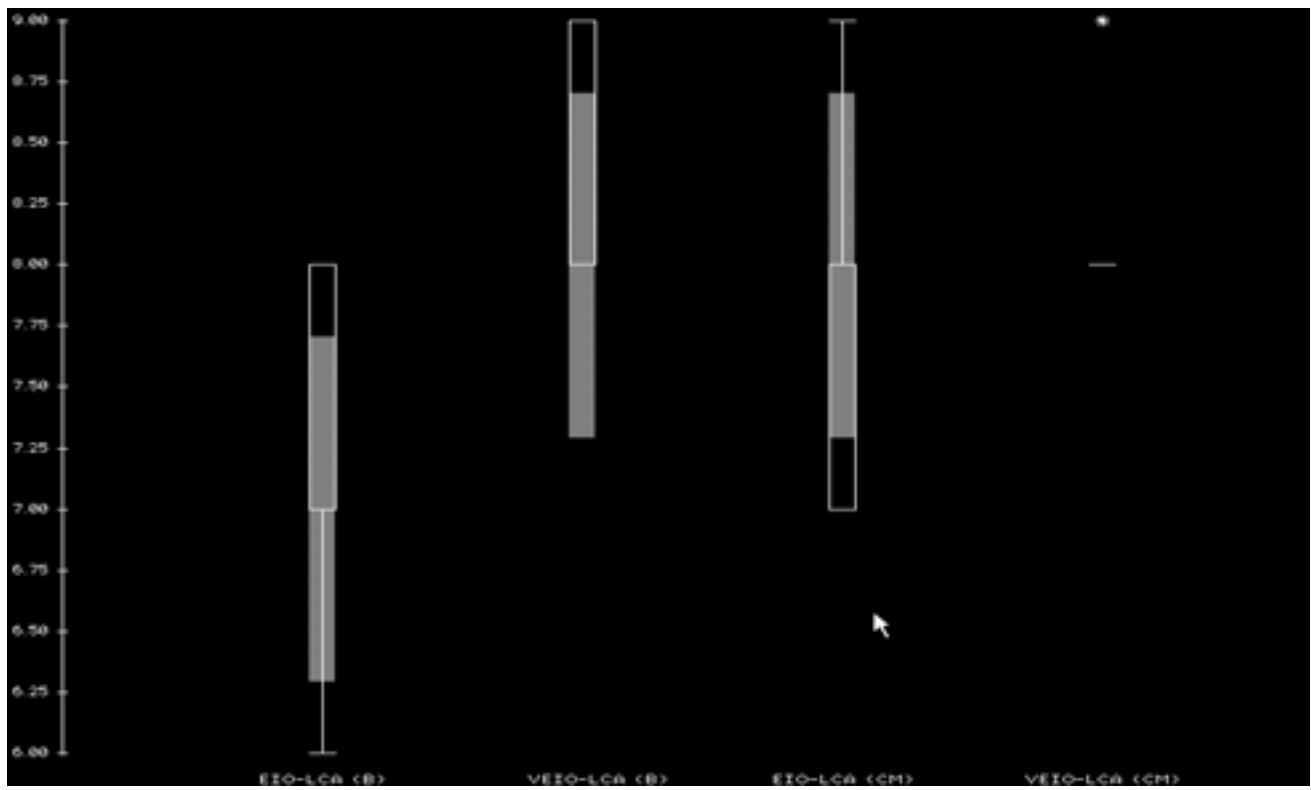


Figure 7: Summary of Solution Quality Data