

Entropy of Eye Fixations: a Tool for Evaluation of Learning Objects

Carlos Lara-Alvarez¹, Hugo Mitre-Hernandez², Maria Alvarado-Hernandez²

¹ CONACYT Research Fellow – Centro de Investigación en Matemáticas (CIMAT),
Zacatecas, Mexico

² Centro de Investigación en Matemáticas (CIMAT),
Laboratorio de Interacción Humano-Computadora,
Zacatecas, Mexico

carlos.lara@cimat.mx, {hmitre,maria.alvarado}@cimat.mx

Abstract. Learning objects (LOs) are important information resources that support traditional learning methods. To evaluate the impact, effectiveness, and usefulness of learning objects it is necessary a theoretically, reliable, and valid evaluation tool. This paper presents a metric to compare the design of LOs, it uses the information provided by visual fixations measured from a small focus group. We conducted an experiment with children of elementary school (n=23). Results showed that images with higher values of the proposed metric were faster to read (Mean = 0.629 min/image) than those images of LO selected randomly (Mean = 0.782 min/image). The metric is useful to optimize the fluency, this is an important step through obtain a fully automated tool to evaluate LOs.

Keywords: Eye tracking, evaluation of learning objects, entropy.

1 Introduction

There are many definitions of Learning Objects (LOs) in the literature [1,10]. For the purposes of this paper the following definition is adopted:

“Learning objects are information resources or interactive software used in online learning [10]”.

A single image, a page of text, an interactive simulation, or an entire course could all be examples of learning objects. Thousands of LOs are currently available through the web [10]; hence, it is necessary an automatic tool for evaluating the impact, effectiveness, and usefulness of learning them.

In this paper we are interested on evaluating LOs composed of a sequence of images. The order and content of these images are designed to ensure a given learning objective. When designing such a learning object, every person on the design team could have different ideas of what the student needs or wants.

Let us suppose that the content and sequence of a LO is already defined; but, for every position in the sequence there are several options with different

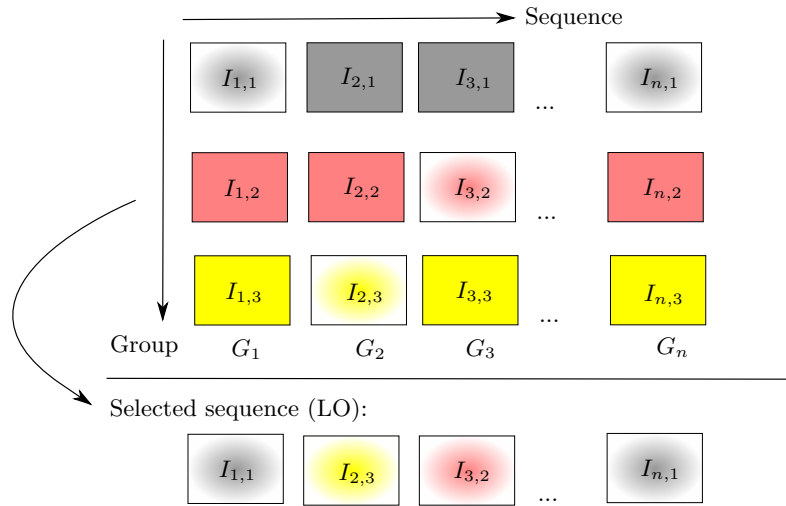


Fig. 1. For each group G_i the designer must select the best image to fill the i -th position of the LO. In this example, image $I_{1,1}$ was selected from $\{I_{1,1}, I_{1,2}, I_{1,3}\}$ to fill position 1 of the LO, image $I_{2,1}$ to fill position 2, etc. This paper proposes an entropy-based metric calculated from eye fixations to select the best image.

graphical design. This problem is illustrated in Fig. 1; for instance, the group G_1 is composed of three images; i.e. $G_1 = \{I_{1,1}, I_{1,2}, I_{1,3}\}$. One of these images must be selected to fill the first position in the LO. Analogously, one image of the group $G_2 = \{I_{2,1}, I_{2,2}, I_{2,3}\}$ must be selected to fill the second position in the LO, and so on. The problem can be stated as follows:

Given a set of candidate images:

$$G_i = \{I_{i,j} \mid j = 1 \dots, N\},$$

where all images $I_{i,j} \in G_i$ has the same information but different design, select the image $\hat{I}_i \in G_i$ that optimize certain predefined evaluation criteria.

The main contribution of this paper is an entropy-based metric able to compare the images of a group. The proposed metric uses the information provided by visual fixations measured from a small focus group; hence, it does not require the experts' intervention, and the information is more accurate as it is originated directly from students. The empirical evaluation shows that images selected by this metric are faster to read.

The rest of this paper is organized as follows: section 2 discusses previous approaches to evaluate LOs, section 3 introduces the entropy-based metric, section 4 presents an experiment with students from primary school, section 5 presents the results and discusses the pros and cons of the proposed technique; finally, section 6 concludes this article.

2 Related Work

In general, LO evaluation approaches can be categorized as:

Indirect methods. These approaches consider that experts, aided with guidelines or other evaluation instruments, can improve the design and content of LOs.

Direct methods These approaches obtain information of the quality of LO from the focus group. A simple strategy is using questionnaires or surveys. Technological advances had open the possibility of automating the evaluation of LOs. A widely used sensor for this purpose is the eye tracker device commonly known as eye-tracker that is capable of obtaining data of when and what the user views on a screen.

There are several studies that suggest guidelines that can be used for indirect evaluation; for instance, Ozcelik et al. [5] study how color coding affects multimedia learning, they suggest that color coding can reduce unnecessary search processes in comparison to black and white material because it guides the student attention by salient information. The Learning Object Review Instrument (LORI) [10] is aimed to evaluate different aspects of a LO; reviewers can rate and comment with respect to nine items: content quality, learning goal alignment, feedback and adaptation, motivation, presentation design, interaction usability, accessibility, reusability, and standards compliance. The metric proposed in this paper is focused on the presentation design; this aspect evaluates the design of visual and auditory information for enhanced learning and efficient mental processing.

An example of a direct instrument is the Learning Object Evaluation Scale for Students (LOES-S) which includes three main categories: learning, quality, and engagement [2]. In this way, after using a LO, students complete a survey to determine their perception of (a) how much they learned, (b) the quality of the learning object, and (c) how much they were engaged with the learning object. We consider that it is possible to generate an automatic method that evaluates these three aspects. The first aspect is easily evaluated through the student's achievements, the second aspect requires to know the emotional state of the player [6] or it can be inferred by the analysis of the data usage [7]. The proposed approach is an effort towards a direct evaluation of the material quality.

Several studies on eye tracking supports the metric proposed in this paper. Tsai et al. [9] state that successful problem-solving students tend to spend more fixation time on inspecting relevant than irrelevant factors. Vatrapu et al. [11] suggest that observation times are longer for harder to understand representations. Considering these facts, we assert that a good design must have an equal distribution of observation times along the image; thus, avoiding hard-to-understand or irrelevant zones. In this sense, an entropy-based metric, such as the proposed in this paper, is a good alternative.

3 Proposed Approach

Entropy can be described qualitatively as a measure of energy dispersal. The concept itself is linked to disorder: entropy is a measure of disorder, and nature tends toward maximum entropy for any isolated system. The information entropy is defined as:

$$H = - \sum_i p_i \log p_i, \quad (1)$$

where p_i is the probability of occurrence of the i -th symbol of an alphabet.

To obtain an entropy-based metric, a grid that covers the region of interest is superimposed on the image. This representation adjusts to a variety of elements and structures that may contain the image (e.g. text or pictures). Because of the size of objects and their spatial relations among them are unknown, two strategies were implemented:

Scale-space. The concept of octave was implemented in order to obtain a robust metric that is invariant to changes of objects' size. Given the initial grid size of $n \times m$ cells, each octave is obtained by multiplying the original grid size by $k = 2^{s-1}$. That is, the grid size G_s for the s -th octave is

$$\begin{aligned} G_s &= m_s \times n_s \\ &= 2^{s-1}m \times 2^{s-1}n. \end{aligned} \quad (2)$$

For instance, given the grid of original size $G_1 = 2 \times 3$ shown in Fig. 2a, the grid for the octave 2 is $G_2 = 4 \times 6$ (Fig. 2b), and the grid for the octave 3 is 8×12 (Fig. 2c).

The cell size for an image of $x \times y$ pixels and a grid of m_s rows and n_s columns is

$$S_x = x/n_s, \quad (3)$$

$$S_y = y/m_s. \quad (4)$$

Sliding window. This concept is used because the position of objects is unknown. The grid is successively moved in both directions using increments of:

$$\Delta_x = S_x/p, \quad (5)$$

$$\Delta_y = S_y/p, \quad (6)$$

where p is the predefined number of steps. As shown in Fig. 3 every pixel in the region of interest is always covered by a single cell. Henceforth, G_s^{hv} denotes a grid of size $G_s = m_s \times n_s$ moved h and v steps in horizontal and vertical directions, respectively.

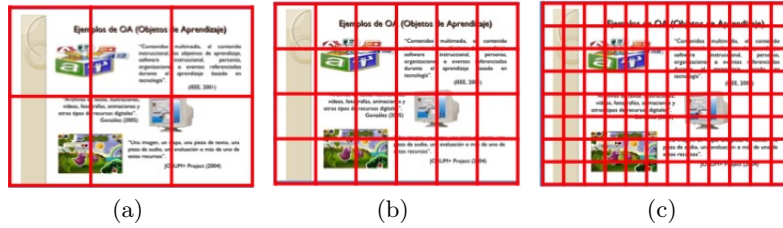


Fig. 2. Illustration of the scale-space strategy: (a) Grid of original size $G_1 = 2 \times 3$ (b) second octave, $G_2 = 4 \times 6$, and (c) third octave $G_3 = 8 \times 12$.

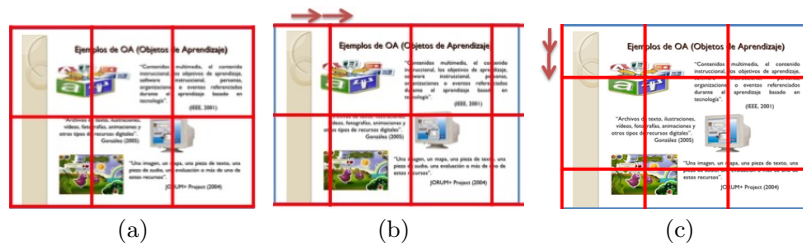


Fig. 3. Illustration of the sliding window strategy: (a) grid at the initial position, (b) grid displaced horizontally, (c) grid displaced vertically.

Entropy-based metric using eye tracking data

Given the observation time over an image (using an eye tracking sensor) and a grid G_s^{hv} , the entropy is calculated as

$$H(G_s^{hv}) = - \sum_{i=1}^{m_s} \sum_{j=1}^{n_s} p(A_{ij}) \log p(A_{ij}), \quad (7)$$

where A_{ij} is the grid cell at row i and column j , and $p(A_{ij})$ is:

$$p(A_{ij}) = \frac{t_{ij}}{\sum_i \sum_j t_{ij}}, \quad (8)$$

here t_{ij} is the total the time cell A_{ij} was observed by the student.

Eye entropy for scale s is

$$H_{\text{eye}}(s) = \frac{1}{p^2} \sum_{h=0}^{p-1} \sum_{v=0}^{p-1} \frac{H(G_s^{hv})}{H'_s}, \quad (9)$$

where $H'_s = -\log \frac{1}{n_s m_s}$ is the maximum entropy for a grid of size $n_s \times m_s$.

The process to calculate the entropy of an Eye tracking data is described in Algorithm 1; finally, Algorithm 2 describes the process of consolidating the LO.

Algorithm 1: EyeEntropy(G, p, τ, E)

Input: $G = m \times n$: initial grid size; p : number of steps, τ : minimum cell size, E : eye tracking data.
Output: H_{eye} : entropy

```

1  $s \leftarrow 1$ 
2  $H_{eye} \leftarrow 0$ 
3 Calculate  $G_s$  (eq. 2)
4 while  $\tau < \min(m_s, n_s)$  do
5    $H_{eye}(s) \leftarrow$  Calculate entropy for scale  $s$  (eq. 9)
6    $H_{eye} \leftarrow H_{eye} + H_{eye}(s)$ 
7    $s \leftarrow s + 1$ 
8   Update  $G_s$  (eq. 2)
9 end
10 return  $H_{eye}$ 

```

Algorithm 2: Consolidating LO by using H_{eye}

Input: $[G_1, G_2, \dots, G_n]$: a proposed sequence, where all images $I_{i,j} \in G_i$ has the same information but different design.
Output: \hat{L} : consolidated LO

```

1 Initialize grid size  $G$ , the minimum cell size  $\tau$ , and the number of steps  $p$ 
2 foreach  $i \in \{1, \dots, n\}$  do
3   foreach  $j \in \{1, \dots, |G_i|\}$  do
4      $H_j \leftarrow 0$ 
5     foreach participant  $p$  do
6        $E \leftarrow$  Eye tracking data of participant  $p$  observing image  $I_{ij}$ 
7        $H_j \leftarrow H_j + \text{EyeEntropy}(G, p, \tau, E)$ 
8     end
9   end
10    $\hat{L}[i] \leftarrow$  Select the image  $I_{ij} \in G_i$  that maximize  $H_j$ 
11 end
12 return  $\hat{L}$ 

```

4 Materials and Methods

The aim of this research is to assess whether H_{eye} is useful as a mechanism to select the best design of a LO.

An eye-tracking device type “Eye Tribe” model ET1000 with 60Hz sampling frequency was used in a screen with 1440×960 pixels resolution. Eye tracker device was located at a distance of 50cm from the student’s face. The device calibration was made with OGAMA [12] (using 12 calibration points). In addition, this tool was used for generating the image sequences for evaluation.

For the calculation of the entropy-based metric, the full image was taken as region of interest and a grid of 2×3 grid was used as initial configuration.

4.1 Participants

Thirty-two fifth grade children from the “Pedro Coronel” Elementary School in Zacatecas, Mexico participated in this study. Nine of them participated in the design phase of LOs, and the other 23 participated in the testing phase.

4.2 Learning objects

Two sets of images were used in this study: “legends” and “theater”. In the design phase, two LO were generated from each image set:

H_{eye} LOs. Two learning objects (one for “legends” and another for “theater”) were generated by following the strategy illustrated in Fig. 1 and using the proposed metric described in (9). The image at each position was selected from a set of three equivalent images. For this aim, each sequence was composed from eye tracking data generated by nine students (each one observing one of the three possible images). This process and some images are shown in Fig. 4.

Random LOs. Two learning objects (one for “legends” and another for “theater”) were selected randomly.

Every LO used in the experiment is a sequence of six images; in every case, the predefined order of the instructional content was preserved. At the end of each LO, a five item questionnaire was included to know the student’s achievements.

4.3 Procedure

For the testing phase, participants were allocated in two groups as follows:

Group 1 Eleven children studied the H_{eye} LO of the “legends” theme, followed by the random LO of the “theater” theme.

Group 2 Twelve children studied the random LO of the “legends” theme, followed by the H_{eye} LO of the “theater” theme.

Each session took approximately 30 minutes per participant, until each participant completed the entire LOs presentation and their respective questions.

4.4 Metrics

The following metrics were used:

Final score. A questionnaire that has five multiple-option questions was used to evaluate student’s achievements.

Median observing speed. The observing speed is the total time that a student observed a given image. The median observing speed is calculated from the data generated by a number of participants who observed an image.

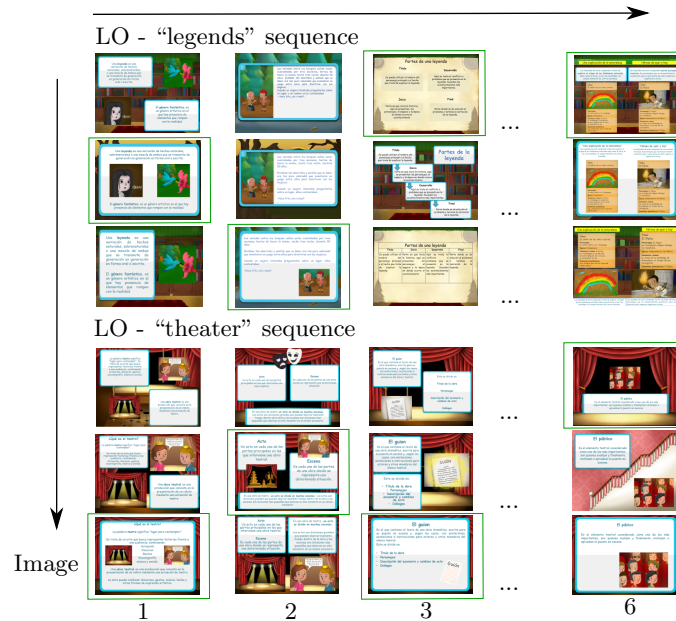


Fig. 4. Illustration of the composition of H_{eye} LOs (design phase). Nine children evaluate groups of three images to obtain the one that maximizes H_{eye} . A green rectangle shows the selected image.

4.5 Statistical analysis

Data are represented as mean \pm SD, and the significance was assessed by Student's t-test for paired data.

5 Results and Discussion

Results of the median reading speed are described in Table 1. There was an extremely significant difference in the median observing speed for H_{eye} LOs (0.629 ± 0.235 min/image) and random LOs (0.782 ± 0.329 min/image); $t(11) = 4.66$, $p = 0.0007$. In other words, the observing speed is faster for those LOs composed of images that maximize H_{eye} . As it is expected, the observing time is correlated to the number of words. One special case is the image 6 of the theater sequence which obtained a lower observing speed compared to the randomly selected image; but, this image also has the lowest number of words. We speculate that the observation time for graphics in this image becomes relevant; i.e. the observation time for graphics is of the same order that the reading time of text.

There was not a significant difference in the final score for H_{eye} LOs (0.513 ± 0.207 points) and random LOs (0.496 ± 0.199 points); $t(22) = 0.7843$. It means that the final score was not affected by the aesthetics. We believe that the students' achievement outcomes are mainly driven by the LO content itself. In

Table 1. Results for the observing speed in minutes per image. Best results are marked in bold.

“legends” sequence						
	1	2	3	4	5	6
# words	45	62	88	78	62	141
H_{eye} LO	0.498	0.723	0.747	0.698	0.607	1.163
random LO	0.772	0.827	0.976	0.988	0.791	1.495

“theatre” sequence						
	1	2	3	4	5	6
# words	65	75	52	43	49	25
H_{eye} LO	0.782	0.719	0.497	0.386	0.475	0.247
random LO	0.988	0.822	0.552	0.438	0.533	0.212

this direction, it is important to choose meaningful content that focuses on the learning objective [8].

Images of H_{eye} LOs have the following qualitative characteristics: (i) their elements (text or graphics) are visually balanced, (ii) element’s placement on the screen establish and strengthen visual relationships between items, and (iii) their color scheme is harmonious. These characteristics agree to guidelines for authors of Learning Objects [8].

Based on the previous facts, we claim that by maximizing the proposed metric, we also maximize the fluency of LO. Fluency as a subjective experience of ease or difficulty associated with a mental process is part of phenomenon obtained from the aesthetics aspects in design [4]. Aesthetics changes as font condition or figure-ground contrast could infer on perceptual fluency.

The advantage of the proposed metric is twofold: (i) it does not require experts intervention, nor surveys applied to students, and (ii) the metric cope different graphical elements. The main drawback is that it is not possible to evaluate dynamic content; e.g. animations.

6 Conclusion and Future Work

This paper presents a metric to compare learning objects with eye tracking. The proposed metric uses the information provided by visual fixations measured from a small focus group. Images selected by maximizing the proposed metric characteristics agree to guidelines for authors of Learning Objects; and they also maximize the fluency. Results show that the aesthetic content can be evaluated by the proposed metric.

This paper opens the opportunity of automating other aspects of the evaluation of LO to obtain an holistic and fully automatic tool. We are also planning a more detailed study to compare the reading speed and the time students spend to analyze graphics. For this aim, it is necessary to detect the reading activity [13,3].

Acknowledgments. We thank all of the children who participated in the study. This research was funded by the National Council of Science and Technology of Mexico (CONACyT) through the project CATEDRAS-3163.

References

1. IEEE, Learning Technology Standards Committee: IEEE standard for learning object metadata. IEEE Standard 1484(1), 2007–04 (2002)
2. Kay, R.H., Knaack, L.: Assessing learning, quality and engagement in learning objects: the learning object evaluation scale for students (loes-s). *Educational Technology Research and Development* 57(2), 147–168 (2009)
3. Kunze, K., Masai, K., Inami, M., Sacakli, Ö., Liwicki, M., Dengel, A., Ishimaru, S., Kise, K.: Quantifying reading habits: counting how many words you read. In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. pp. 87–96. ACM (2015)
4. Oppenheimer, D.M.: The secret life of fluency. *Trends in cognitive sciences* 12(6), 237–241 (2008)
5. Ozcelik, E., Karakus, T., Kursun, E., Cagiltay, K.: An eye-tracking study of how color coding affects multimedia learning. *Computers & Education* 53(2), 445–453 (2009)
6. Peter, C., Urban, B.: Emotion in human-computer interaction. In: *Expanding the Frontiers of Visual Analytics and Visualization*, pp. 239–262. Springer (2012)
7. Sanz-Rodríguez, J., Doderó, J.M.M., Sánchez-Alonso, S.: Ranking learning objects through integration of different quality indicators. *IEEE transactions on learning technologies* 3(4), 358–363 (2010)
8. Smith, R.S.: *Guidelines for authors of learning objects*. New Media Consortium (2004)
9. Tsai, M.J., Hou, H.T., Lai, M.L., Liu, W.Y., Yang, F.Y.: Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers & Education* 58(1), 375–385 (2012)
10. Vargo, J., Nesbit, J.C., Belfer, K., Archambault, A.: Learning object evaluation: computer-mediated collaboration and inter-rater reliability. *International Journal of Computers and Applications* 25(3), 198–205 (2003)
11. Vatrapu, R., Reimann, P., Bull, S., Johnson, M.: An eye-tracking study of notational, informational, and emotional aspects of learning analytics representations. In: *Proceedings of the Third International Conference on Learning Analytics and Knowledge*. pp. 125–134. ACM (2013)
12. Voßkühler, A.: *Ogama description (for version 2.5)*. Berlin, Germany: Freie Universität Berlin, Fachbereich Physik (2009)
13. Yamaya, A., Topić, G., Martínez-Gómez, P., Aizawa, A.: Dynamic-programming-based method for fixation-to-word mapping. In: *Intelligent Decision Technologies*, pp. 649–659. Springer (2015)