EVALUATING INTERACTIVE INTERFACES FOR RELAXATION: AN MRQ-BASED STUDY ON VISUAL BREATHING GUIDES

Uriel Haile Hernandez Belmonte*, Ph.D., professor in the Department of Art, University of Guanajuato, Mexico*
Mitzy Fernanda Vega De La Garza*, Systems Engineering Student, University of Guanajuato, Mexico*

Abstract. *This study rigorously evaluated the cognitive load elicited by four interactive guided breathing interfaces – fixed 4-2-6 and customizable rhythms – using an adapted Multiple Resources Questionnaire (MRQ) on 50 participants. The instrument's reliability was confirmed through Cronbach's alpha and ANOVA. Empirical results demonstrated that Screen A, associated with the customizable pattern, consistently generated the lowest overall cognitive load. The study concludes that the specific visual animation pattern significantly modulates the perceived demand on cognitive resources, offering critical insights for digital wellness tool design.*

Keywords: *cognitive load, breathing interfaces, MRQ, user experience, visual animation.*

Guided breathing applications offer an innovative tool for human well-being and stress management. In User Experience (UX) design, the effectiveness of an interactive interface is not solely determined by its functionality but, critically, by the cognitive load it imposes on the user, especially when aiming for relaxation. An interface that demands excessive mental resources will inherently fail to deliver an optimal user experience. This paper presents a rigorous UX evaluation of visual breathing guides. We use the Multiple Resources Questionnaire (MRQ), a recognized self-report instrument for quantifying cognitive resource demand, and adapt nine of its dimensions for this context. The main objective of this study is to determine whether significant differences exist in the perceived cognitive load across four visually distinct screens of a guided breathing application, and to identify the design that is cognitively most efficient for the end user, thereby contributing to best practices in multimedia interface design.

To evaluate the cognitive load generated by each interface, a protocol was designed that prioritizes sensitivity to specific mental processes over global effort measurement. Unlike scales such as the NASA-TLX or the SWAT, which capture the general magnitude of load without discriminating which cognitive resources are affected, the MRQ allows identification of which particular process generates the greatest demand [1]. This is especially useful in the context of graphical interfaces, where the user processes information of different natures simultaneously. In this way, the results not only indicate how much effort each screen requires, but also which aspect of cognitive processing is most involved in each case.

All participants who evaluated the four screens did so in a different order so that this would not affect the results. This method was used because we needed to ensure that responses changed because the interface was different, not due to fatigue or familiarity with the application. The four interfaces evaluated were grouped into two sets according to the breathing pattern used. Screens A and B correspond to the customizable breathing pattern with configurable times through sliders for inhalation time, retention time, and exhalation time, and differ in their main animated element. Screen A presents a sphere traveling along an arc-shaped curve with a single trajectory; Screen B replaces the arc with a continuous sinusoidal wave along which the sphere moves cyclically. Screens C and D correspond to the fixed 4-2-6 pattern and share a water drop animation as the central element; the main difference between these screens is that in Screen C the character guides the breathing by expanding and contracting, while Screen D incorporates the same sinusoidal wave with a traveling sphere as in Screen B. This design allows comparison, within each pattern, of which visual representation communicates the breathing rhythm with greater clarity and lower cognitive demand (Figure 1).

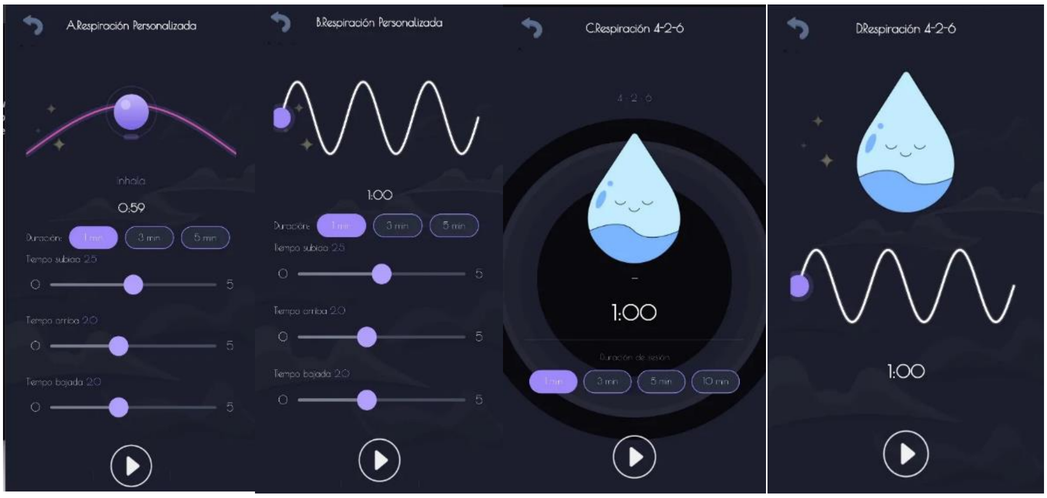

Figure 1 – Visual comparison of the four breathing interfaces evaluated: arc animation (A), sinusoidal wave (B), expanding drop character (C), and drop with sinusoidal wave (D)

The sample consisted of 50 university students from different programs, who had no prior knowledge of the application. The MRQ adapted for this evaluation comprised nine items, each oriented toward a specific cognitive load dimension: eye movement, concentration, animation clarity, comprehension of symbols and colors, readability of instructions, rhythm tracking, ability to anticipate changes, sense of pressure, and frustration. Each item was answered on a scale from 0 (nothing) to 5 (totally). Each participant interacted for one minute with each of the screens and then completed the adapted MRQ questionnaire. The data were analyzed using a one-factor repeated-measures ANOVA for each of the nine instrument dimensions, with the objective of identifying whether statistically significant differences existed among the four screens [4]. A significance level of $\alpha = 0.05$ was established, so items with a p-value below this threshold were considered significant. Additionally, Cronbach's alpha coefficient was calculated to verify the questionnaire's internal consistency [3], both globally and per individual screen, confirming that the instrument reliably measured the proposed cognitive load dimensions. All data processing was carried out in Google Sheets.

The analysis was conducted on 205 responses distributed across the four screens. The overall consistency of the instrument was good ($\alpha = 0.840$) [2, 3], with values per screen ranging from 0.727 for Screen A to 0.871 for Screen B, confirming that the questionnaire reliably measured the evaluated dimensions. The ANOVA revealed statistically significant differences in five of the nine dimensions evaluated [4]: rhythm tracking, animation clarity, symbol comprehension, eye movement, and anticipation of changes. The remaining four dimensions, concentration, readability, pressure, and frustration showed no significant differences among screens, suggesting these aspects are perceived similarly regardless of the animation used. Among the significant dimensions, rhythm tracking showed the greatest difference between screens, followed by eye movement. Screen A consistently obtained the most favorable scores across significant dimensions, particularly in animation clarity and rhythm tracking, while generating the least demand in eye movement and anticipation of changes. Screen D, in contrast, proved the most cognitively demanding, showing the lowest scores in rhythm tracking and the highest in eye movement. Overall, Screen A achieved the highest total score of 736 points, positioning it as the best-evaluated interface, while Screen D achieved the lowest score of 659 points.

The results obtained indicate that the visual design of a guided breathing interface significantly influences how users perceive and process information during use. Of the nine dimensions evaluated, five showed statistically significant differences across the four screens, confirming that not all animations convey the breathing rhythm with the same clarity or elicit the same cognitive demand. Screen A, corresponding to the customizable breathing pattern with arc and sphere animation, was consistently the best evaluated: users reported greater animation clarity, better ability to follow the rhythm, and lower frustration compared to the other interfaces. Screen D, which combines the drop character with the sinusoidal wave, yielded the lowest scores in rhythm tracking and eye movement, suggesting that combining two simultaneous visual elements may increase cognitive demand rather than facilitate comprehension.

The adapted MRQ instrument proved adequate for this type of evaluation, with an overall Cronbach's alpha of 0.840, indicating good internal consistency. This supports its use in future studies evaluating interactive interfaces from a cognitive load perspective. As future work, it is suggested to extend the evaluation to the retention and exhalation phases, and to broaden the sample to include populations with different levels of technological experience to verify whether the results hold.

References

1. Boles, D.B., Bursk, J.H., Phillips, J.B., & Perdelwitz, M.S. (2007). Predicting dual-task performance with the Multiple Resources Questionnaire (MRQ). Human Factors, 49(1), 32-45. https://doi.org/10.1518/001872007779598043.

2. Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), 297-334. https://doi.org/10.1007/BF02310555.

3. Schrepp, M. (2020). On the usage of Cronbach's alpha to measure reliability of UX scales. Journal of Usability Studies, 16(1). https://dl.acm.org/doi/abs/10.5555/3532713.3532717.

4. Strale, F. (2024). Partitioning for enhanced statistical power and noise reduction: Comparing one-way and repeated measures analysis of variance (ANOVA). Cureus. https://doi.org/10.7759/cureus.75322.

5. Wang, X. et al. (2025). Research on the influence of interface visual design features of mobile news on cognitive load. Behavioral Sciences, 16(1), 32. https://doi.org/10.3390/bs16010032.