electronics

MDPI

Article

Response Expiration, Accuracy, and Response Time Across Two Time-Limited Immersive VR Quiz Blocks: An Exploratory Within-Subject Comparison

Fátima Avilés-Castillo [1,2,*], Jorge Buele [3] and Guillermo Palacios-Navarro [2,*]

[1] Facultad de Ingenierías, Universidad Tecnológica Indoamérica, Ambato 180103, Ecuador
[2] Department of Electronic Engineering and Communications, University of Zaragoza, 44003 Teruel, Spain
[3] Centro de Investigación en Mecatrónica y Sistemas Interactivos (MIST), Facultad de Ingenierías, Universidad Tecnológica Indoamérica, Ambato 180103, Ecuador; jorgebuele@uti.edu.ec
* Correspondence: fatimaaviles@uti.edu.ec (F.A.-C.); guillermo.palacios@unizar.es (G.P.-N.)

Abstract

Behavior in time-limited immersive virtual reality (VR) tasks can be characterized by response accuracy, response time among submitted responses, and whether a response is submitted before the available interval expires. This exploratory within-subject study compared outcomes across two non-equivalent VR quiz blocks in 22 adults. The baseline block presented a visual countdown, whereas the auditory-cue block used a different fixed question bank and included prerecorded ticking that accelerated midway through the nominal 15 s response interval. Block order was balanced across participants. Mixed-effects models showed no statistically significant between-block contrast in all-trial accuracy (OR = 0.93, 95% CI [0.22, 3.89], $p = 0.909$) or response time among submitted responses (GMR = 0.93, 95% CI [0.74, 1.16], $p = 0.492$). Response expiration occurred in 33 of 220 baseline trials (15.0%) and eight of 220 auditory-cue trials (3.6%), with an estimated between-block OR of 0.21 (95% CI [0.07, 0.65], $p = 0.009$). Because the question bank and auditory configuration were permanently linked within each block, this contrast cannot be attributed specifically to the auditory cue. The study demonstrates a practical approach for reporting response expiration alongside accuracy and response time among submitted responses in time-limited immersive interaction.

Keywords: immersive virtual reality; auditory deadline cue; response expiration; time-limited interaction; human–computer interaction; mixed-effects models

Check for updates

Academic Editors: Teng-Wen Chang, Marc Aurel Schnabel and Halil I. Erhan

Received: 19 September 2026
Revised: 27 October 2026
Accepted: 29 October 2026
Published: 30 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Time is an integral component of many interactive tasks. Digital assessments, training systems, simulations, games, and decision-support applications often require users to act within a predefined interval [1]. Performance in these settings is commonly described through response accuracy and response time, but a third outcome can also be relevant: whether a response is produced at all before the available interval expires. These dimensions need not vary together, particularly when decisions are made under temporal constraints [2].

Auditory information has long been used to communicate urgency and temporal relevance in human–computer interaction. Classical studies of auditory warnings showed that acoustic characteristics such as repetition rate and rhythm influence perceived urgency [3].

Subsequent work examined how auditory urgency related to behavioral responding, including reaction time and task performance [4]. That literature established that sound can convey information about the need to act, but the behavioral consequences depend on the characteristics of both signal and task.

Time pressure is also closely related to the speed–accuracy relationship. Reducing the time available for a decision can alter information processing, response latency, and correctness [5,6]. However, shortening a deadline and increasing the perceptual salience of an unchanged temporal interval are not equivalent manipulations. Experimental work on temporal preparation has shown that providing countdown information can modify response readiness and reaction time even when the objective interval itself is held constant [7,8]. In the latter case, the amount of available time does not change; rather, the interface provides more explicit information about its progression.

Immersive virtual reality (VR) provides a controlled setting in which such temporal characteristics can be manipulated while behavioral responses are recorded during interaction [9,10]. Immersive VR has been used across education, training, assessment, rehabilitation, and human–computer interaction research [11–14]. The capacity of immersive VR to coordinate visual, auditory, and interactive events also makes it suitable for studying how users respond to temporal information [15].

Temporal interaction has itself become an increasingly relevant topic in VR. Bansal et al. showed that altering the temporal relationship between user movement and virtual events can produce temporal recalibration [16]. More recently, Wei et al. examined expected delays during target selection in VR and demonstrated that temporal characteristics can influence interaction behavior and performance [17]. Although these studies address different forms of temporal manipulation, they illustrate that timing is not merely a background property of an immersive system but can become part of the interaction itself.

One way of representing an approaching deadline is through an auditory deadline cue. In the present study, this consisted of a ticking sound accompanying a visual countdown, with the repetition rate increasing halfway through the nominal response interval. The cue was intended to make the progression toward the deadline perceptually explicit. Importantly, the term auditory deadline cue describes the interface manipulation and does not imply that psychological stress or subjective time pressure was induced. Establishing such psychological responses would require direct self-report, behavioral, or physiological measures [18].

An important methodological issue in timed tasks is that trials ending without a response are often treated differently from completed trials or omitted from response-time analyses altogether. Recent work has emphasized that response omissions under deadlines can themselves contain information about decision behavior and should not necessarily be regarded as ordinary missing data [19]. Likewise, research on speeded-up response tasks has shown that deadlines can influence whether a response is produced as well as how quickly it is produced [20]. Response expiration may therefore provide information not captured by accuracy or response time among submitted responses.

This distinction is particularly relevant in immersive tasks with fixed response windows. A user may respond correctly, respond incorrectly, or fail to commit to any response before the deadline. Analyses limited to completed responses cannot fully represent these alternatives. At the same time, knowledge-based tasks introduce additional variability because individual questions can differ substantially in difficulty and response demands. Accounting for item-level heterogeneity is therefore important when comparing behavior across experimental blocks.

The present exploratory study examined behavioral performance during a time-limited immersive VR quiz implemented on Meta Quest 3. Participants completed two within-

subject blocks presented in balanced order. In the baseline block, a visual countdown was presented without a time-related auditory signal. In the auditory-cue block, the same countdown was accompanied by ticking whose repetition rate increased midway through the nominal response interval. Response accuracy, response time among submitted responses, and response expiration were analyzed at the item level. Because each block used a different fixed question bank, the study was not designed to estimate an isolated causal effect of the auditory cue. The primary analyses therefore focused on descriptive and model-based contrasts between the two implemented block configurations while accounting for participant- and item-level variability. Questions were additionally classified as Ecuador-related or international/general content. Analyses involving that classification were treated as secondary and exploratory because the study was not powered to provide precise estimates of content-category moderation, and the broad classification did not constitute a direct measure of prior knowledge or item familiarity.

The study addressed the following research questions:

1. **RQ1:** What differences in response accuracy, response expiration, and response time among submitted responses are observed between the two implemented VR quiz blocks?
2. **RQ2 (exploratory):** How do participants perceive the usability of the immersive VR application, and is perceived usability associated with overall task performance?

Secondary exploratory analyses examined whether behavioral outcomes varied descriptively across Ecuador-related and international/general content categories and whether the estimated block contrasts differed across those categories.

2. Materials and Methods

2.1. Study Design and Reporting Framework

This exploratory study used a balanced two-period within-subject design to compare behavioral performance across two immersive VR quiz blocks. Each participant completed both implemented block configurations: (A) a baseline block and (B) an auditory deadline-cue block. The block-order schedule contained 11 AB and 11 BA assignments, with the ordering of those assignments pseudorandomized before participant assignment. The design therefore balanced presentation order but did not independently randomize question bank and auditory configuration.

In the baseline block, the task included a visible nominal 15 s countdown without a time-related auditory signal. In the auditory-cue block, the same visual countdown was accompanied by a ticking signal whose repetition rate increased halfway through the nominal response interval. The virtual environment, question-panel position, response interface, interaction technique, and nominal response interval were otherwise maintained across the two blocks.

The two blocks also used different fixed 10-item question banks. Bank and auditory configuration were permanently linked: Bank A was always administered in the baseline block, and Bank B was always administered in the auditory-cue block. Consequently, the within-subject comparison contrasts two complete block configurations rather than independently estimating the effect of the auditory cue. The banks were thematically paired by question position but were not psychometrically calibrated or empirically equated. Accordingly, the ordering of the fixed 11 AB and 11 BA sequence assignments was pseudorandomized, whereas question bank and auditory configuration were not independently crossed.

The study was conducted as a human–computer interaction experiment rather than as a clinical intervention. Reporting was informed by the CONSORT 2025 recommendations

and, where applicable to the two-period within-participant structure, by the CONSORT extension for randomized crossover trials [21,22].

Questions were additionally classified into two content categories: Ecuador-related and international/general. Behavioral observations were recorded at the individual-question level, yielding 20 trials per participant across the two experimental blocks.

2.2. Participants

Participants were recruited by convenience sampling through social media, local dissemination, and institutional communication at Universidad Tecnológica Indoamérica, Ambato, Ecuador. Data collection was conducted between April and May 2026.

Eligible participants were adults able to use an immersive VR headset and with normal or corrected-to-normal vision. Exclusion criteria included photosensitive epilepsy, significant balance disorders, inability to complete the experimental task, previous adverse experiences with VR, and self-reported marked susceptibility to motion sickness. No formal audiometric screening was performed.

Twenty-nine individuals were initially assessed for participation. Seven did not enter the experimental task: one reported a previous negative VR experience, five reported marked susceptibility to motion sickness, and one reported influenza-like symptoms and general malaise and chose not to participate. Twenty-two participants started the experiment, completed both experimental blocks, and were included in the analyses. The analyzed sample comprised 20 males (90.9%) and 2 females (9.1%), with a mean age of 23.91 years (SD = 3.96, range = 18–31 years). Participants received no financial or academic compensation.

No formal a priori sample-size calculation was performed because the study was designed as an exploratory experiment and no single primary behavioral outcome had been prespecified. Accordingly, the precision of the study is characterized by the model-based effect estimates and their 95% confidence intervals. The relatively small sample should therefore be considered when interpreting imprecise or nonsignificant model-based estimates. Figure 1 summarizes recruitment, sequence allocation, completion of both experimental periods, and inclusion in the final analysis.

2.3. Randomization and Experimental Sequence

A fixed balanced sequence-allocation list containing exactly 11 AB and 11 BA assignments was created before the experimental sessions. The ordering of these 22 assignments was pseudorandomized in Microsoft Excel using the `RAND()` function and subsequently linked to numerical participant identifiers. Thus, the procedure was not simple randomization: the number of participants assigned to each sequence was fixed in advance at a 1:1 ratio, whereas the ordering of the AB and BA assignments was randomized. At the time the allocation schedule was generated, the researcher did not know which individual participant would subsequently receive each numerical identifier.

The allocation schedule was generated before participant assignment and was then used without modification during data collection. One researcher generated the sequence schedule, whereas a different researcher conducted the experimental sessions according to the assigned sequence. Because the `RAND()` procedure did not use a prespecified reproducible seed, the pseudorandom ordering cannot be independently regenerated from a recorded seed. Future studies should use a dedicated randomization procedure with a prespecified and archived seed or allocation record to permit exact reproduction of the sequence-generation process.

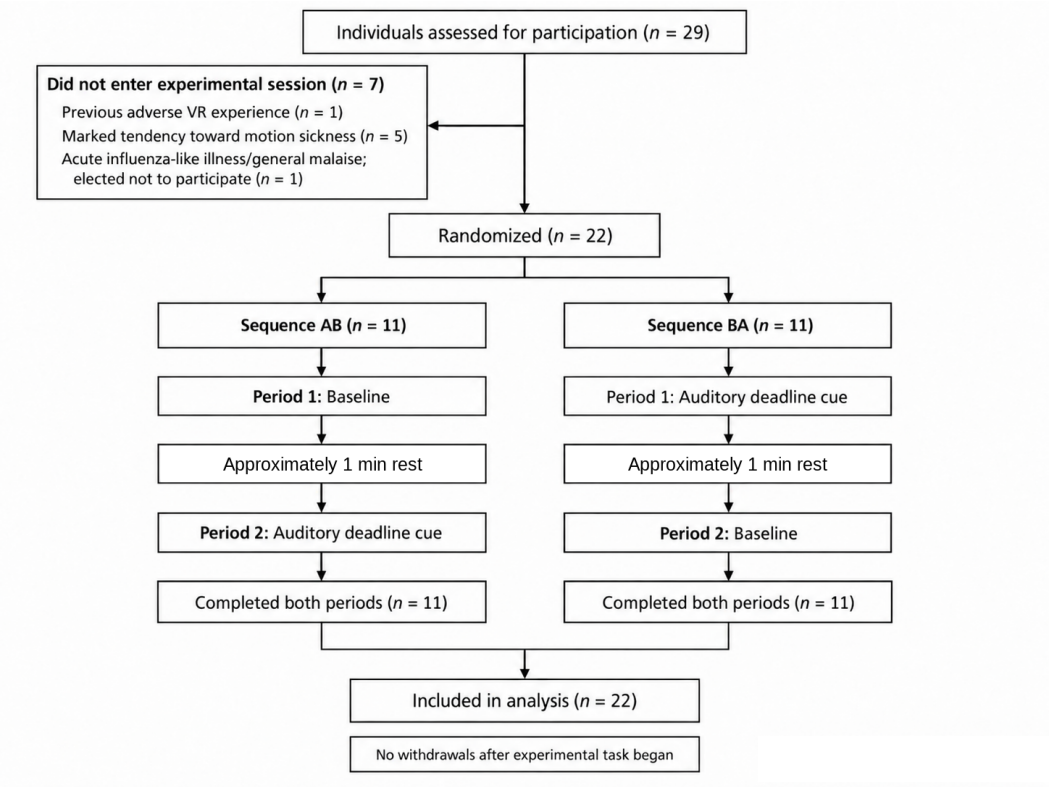


Figure 1. Participant flow through recruitment, balanced sequence allocation, the two experimental periods, and analysis. A denotes the baseline block and B the auditory-cue block.

Participants were informed that they would complete two VR blocks and that they should attend to the visual and auditory elements presented by the system, but they were not informed of the specific hypothesis regarding the auditory manipulation or of an expected direction of effect.

Blinding to the presence of the auditory cue was not possible because the ticking signal was perceptible during the auditory-cue block. Response accuracy, response time, response expiration, and lifeline use were recorded automatically by the application, limiting observer involvement in outcome acquisition.

2.4. Immersive VR System and Application

2.4.1. Hardware and Software Architecture

The immersive application ran directly on a Meta Quest 3 head-mounted display (Meta Platforms, Inc.) as a standalone application, without PC tethering during the experimental session. The headset incorporates LCD displays with a physical resolution of 2064 × 2208 pixels per eye and an approximate field of view of 110° horizontally and 96° vertically. Six-degree-of-freedom head and controller tracking was available through the headset platform.

Participants interacted with the application using two Meta Quest Touch Plus controllers. Either controller could be used during the task. Audio was delivered through the headset's integrated audio system.

The application did not explicitly override the headset display-refresh setting and therefore operated under the Meta Horizon OS default application refresh rate of 72 Hz. Runtime frame rate was not instrumented during data collection; consequently, empirical mean or minimum frame-rate values were not available. Although 72 Hz defined the nominal display-refresh configuration, actual rendering performance during individual sessions could not be verified retrospectively and may have varied with runtime load.

The application was developed in Unity 6, version 6000.0.34f1, using C# and the Universal Render Pipeline (URP). OpenXR was used for XR-device integration, the Unity XR Interaction Toolkit provided controller-based immersive interaction, and the Unity Input System handled controller input. The final software was compiled for Android/ARM64 as an Android Package (APK) and installed directly in the internal storage of the Meta Quest 3.

Firebase Realtime Database was integrated using Firebase Unity SDK version 13.2.0. It served both as the remote source of the experimental question banks and as the destination for session and behavioral data recorded during the experiment. The headset therefore communicated directly with Firebase through a Wi-Fi connection while the VR application itself remained a standalone on-device application.

The software followed a modular architecture. The principal components controlled interface navigation and experimental context, question retrieval, quiz progression, block-specific auditory behavior, response timing, and session metrics. The `QuizManager` coordinated question presentation and response processing; `FirebaseLoader` retrieved the corresponding question bank; `PerturbationManager` and `CountdownPerturbation` controlled block-specific clock audio; and `SessionMetricsManager` recorded behavioral events and transferred the resulting data to Firebase. Figure 2 summarizes the development, runtime interaction, and external data flow.

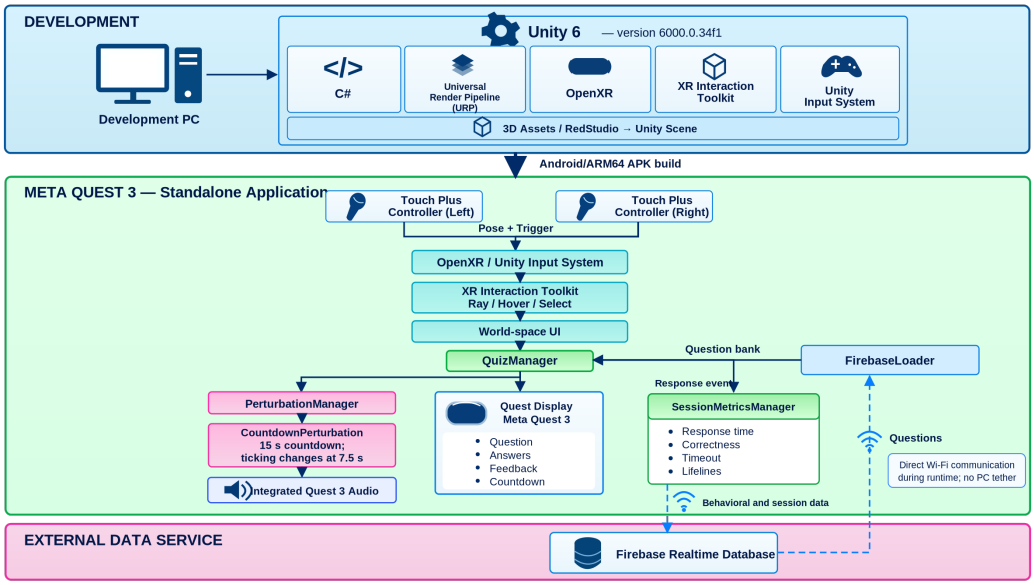


Figure 2. Software architecture and data flow of the standalone immersive VR quiz application.

2.4.2. Virtual Environment and User Interface

The virtual environment was designed as a three-dimensional television game-show-inspired studio. The scene combined reusable three-dimensional objects and prefabs from the RedStudio asset package with custom interface, interaction, experimental-control, timing, quiz, and data-recording components. Environmental elements included studio furniture, virtual screens, lighting elements, cameras, microphones, and other studio objects.

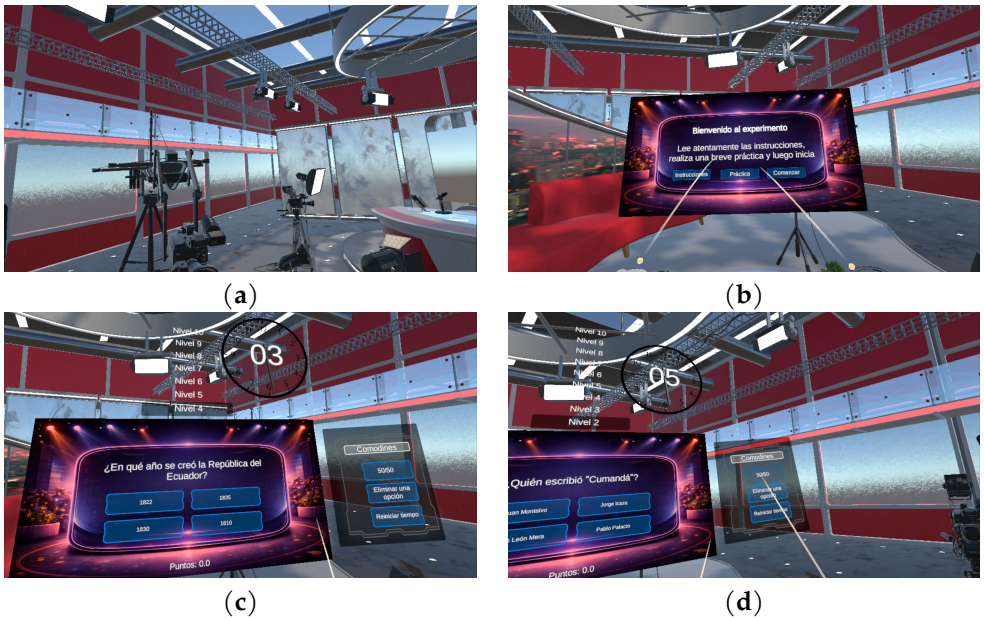
Participants remained seated throughout the experimental task and no virtual locomotion was required. The main question-and-answer interface was implemented as a world-space user interface positioned approximately 1 m in front of the participant. Each trial presented one question, four response alternatives, a visible countdown, progression information, and the available lifelines.

The visual configuration was kept constant between the baseline and auditory deadline-cue blocks. In particular, the virtual environment, question panel, answer-button arrange-

ment, visual countdown, and controller interaction were not intentionally modified as part of the experimental manipulation. Question content nevertheless differed between blocks because each block used its corresponding fixed question bank, as detailed in Section 2.5.

The initial application interface provided access to three components: Instructions, Practice, and Start. Participants reviewed six sequential instruction screens and subsequently completed three practice questions before beginning the experimental blocks.

Figure 3a shows the virtual environment, and Figure 3b presents the initial interface. Figure 3c shows the question-and-answer panel, visual countdown, progression information, and lifeline controls. Figure 3d illustrates controller-based pointing interaction with the lifeline interface.



(**a**) (**b**)

(**c**) (**d**)

Figure 3. Immersive VR environment and user interface. (**a**) General view of the television game-show-inspired virtual environment. (**b**) Initial interface providing access to the instructions, practice session, and experimental task. (**c**) Experimental question interface showing four response alternatives, the visual countdown, progression information, and the three available lifelines. (**d**) Controller-based distal interaction with the lifeline interface using the virtual pointing ray. Interface text is shown in Spanish because the application was administered in Spanish.

2.4.3. Interaction Technique and Task Mechanics

Interaction was based on distal controller pointing. Movement and orientation of either Touch Plus controller defined a virtual ray through the XR Interaction Toolkit. Participants pointed the ray toward an interface element and confirmed the selection by pressing the controller trigger. Direct hand tracking, voice commands, physical touching of interface elements, and application-level haptic feedback were not used.

Each experimental block contained 10 multiple-choice questions. Every question displayed four response alternatives, of which one was correct. Questions were presented in a fixed order within each block, and the ordering of the four response alternatives was also fixed.

Three optional lifelines were available during each experimental block: elimination of two incorrect alternatives (50/50), elimination of one incorrect alternative, and reset of the response timer. Each lifeline could be used once per block.

Use of the timer-reset lifeline restarted the visible countdown at 15 s. In the auditory-cue block, the reset also restarted the auditory sequence from its initial state: the approximately 1 Hz prerecorded ticking pattern resumed and, if no response had been submitted by the

midpoint of the renewed interval, switched again to the approximately 2 Hz pattern when 7.5 s remained. The response-time measure itself was not reset and continued accumulating from the initiation of the original question. Response times greater than 15 s were therefore possible on trials in which the timer-reset lifeline was used.

2.4.4. Auditory Deadline-Cue Manipulation

Participants had a nominal 15 s response interval for each question. The visual countdown was presented together with the question and its four response alternatives.

In the baseline block, participants viewed the 15 s visual countdown without a time-related auditory cue. In the auditory-cue block, the same visual countdown was accompanied by a prerecorded clock-ticking signal reproduced through the Meta Quest 3 integrated audio system. During the first half of each initiated or restarted 15 s response interval (15–7.5 s remaining), a looping ticking pattern with a repetition rate of approximately 1 Hz was presented. When the countdown reached its midpoint (7.5 s remaining), the application switched to a second prerecorded ticking pattern with a repetition rate of approximately 2 Hz. This accelerated pattern continued until a response was submitted, the countdown expired, or the timer-reset lifeline restarted the countdown sequence.

The change from approximately 1 to 2 Hz was selected to create two clearly distinguishable temporal phases within the same nominal response interval. Previous auditory-warning research has shown that repetition rate and the temporal spacing of warning pulses can influence perceived urgency and behavioral responding [3,4,23]. The present 1-to-2 Hz transition should therefore be understood as a temporal interface design choice informed by this broader literature rather than as a previously validated threshold for inducing subjective time pressure. Because perceived urgency or subjective time pressure was not directly measured, the study does not establish that this specific frequency change induced either psychological state.

The two phases were implemented using separate prerecorded audio clips rather than by changing playback pitch or dynamically modifying the repetition rate of a single clip. Analysis of the audio signals indicated inter-tick intervals of approximately 0.998 s for the initial pattern and 0.499 s for the accelerated pattern, corresponding to approximately 1 and 2 ticks/s, respectively. Software-level audio volume was not intentionally changed when the application switched between the two patterns. However, the device-level volume setting was not recorded during individual sessions, acoustic output at the listener's ears was not calibrated in sound-pressure-level units, and the spectral characteristics of the prerecorded signals were not formally quantified as experimental parameters. Consequently, the exact acoustic exposure experienced by each participant cannot be reconstructed retrospectively.

No additional background music or block-specific environmental sound was presented. The implemented comparison therefore combined two auditory features: the presence of repetitive ticking throughout the auditory-cue block and an increase in its repetition rate from approximately 1 to 2 Hz halfway through the nominal response interval. The present design cannot distinguish behavioral responses to the presence of repetitive sound from responses specifically attributable to its temporal acceleration.

If the response interval expired without selection of an answer, the trial was recorded as a response-expiration event, the response was scored as incorrect for the all-trial accuracy outcome, and the application continued to the subsequent feedback and transition sequence. Use of the timer-reset lifeline could restore the visible countdown to 15 s; therefore, 15 s represents the nominal response interval rather than an absolute maximum response time.

The temporal structure of the two experimental blocks and the possible trial outcomes are summarized in Figure 4.

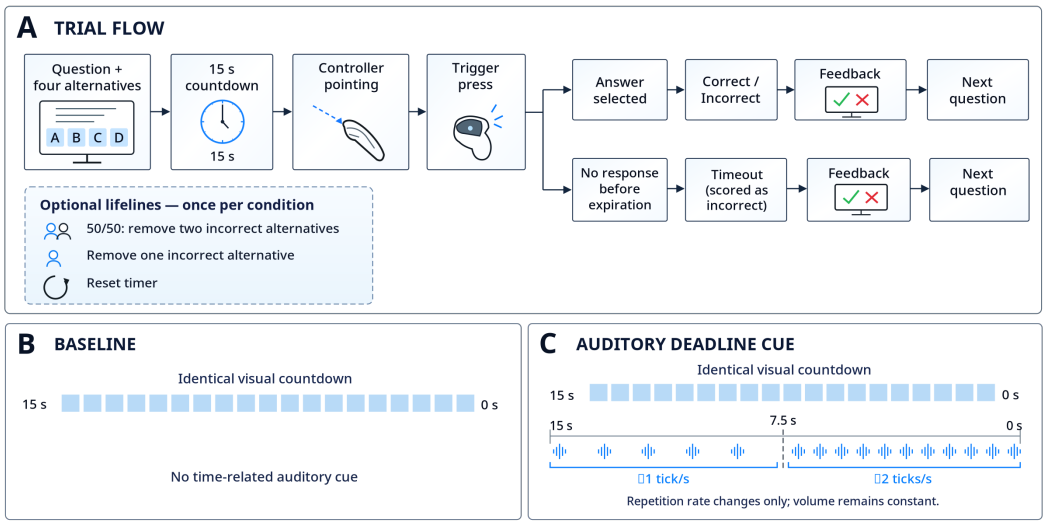


Figure 4. Trial flow and temporal structure of the quiz task. (**A**) Question presentation, controller-based response selection, optional lifelines, and feedback following a submitted answer or response expiration. (**B**) Baseline block with a nominal 15 s visual countdown and no time-related auditory cue. (**C**) Auditory-cue block with the same visual countdown and ticking whose repetition rate increased at 7.5 s remaining. A selected answer terminates the response interval, whereas absence of a response at expiration generates a response-expiration event.

2.4.5. Data Acquisition and Storage

Firebase Realtime Database supported bidirectional communication with the standalone VR application. Before and during each experimental block, the application retrieved the predefined question bank corresponding to the active block. During task execution, behavioral information was recorded automatically and associated with the participant's numerical identifier.

Trial-level records included the question identifier, response time, correct/incorrect status, response expiration, and use of each lifeline. Session-level information included task progression, score, accumulated timing information, and lifeline use. Data were transmitted from the headset to Firebase through the available network connection.

Response time was measured locally within the Unity application from initiation of the question-loading routine to registration of the submitted response. The start timestamp was therefore an application-level software event rather than an instrumented measurement of the physical onset of the rendered question display. The elapsed interval was computed and stored locally in memory before session data were transmitted to Firebase, so network transmission latency was not included in the recorded response-time measure. No external display-onset sensor or frame-synchronized event marker was used to quantify any delay between initiation of the question-loading routine and visual presentation on the headset.

This configuration separated experimental execution from the development computer: the PC was required to develop and compile the Unity project but was not tethered to the Meta Quest 3 during participant testing.

2.5. Question Banks and Content Categories

Two fixed question banks containing 10 items each were used. One bank was permanently associated with the baseline block and the other with the auditory-cue block. All participants therefore received the same question bank in a given block, regardless of whether they completed sequence AB or BA.

In each bank, six items were classified as international/general content and four as Ecuador-related content. Ecuador-related items occupied positions 2, 4, 7, and 9, whereas international/general items occupied positions 1, 3, 5, 6, 8, and 10.

The two banks were organized using corresponding thematic domains by question position. The ten paired domains comprised world geography, Ecuadorian literature, physics, Ecuadorian history, art, computing and technology, Ecuadorian geography, biology, Ecuadorian geography, and environmental knowledge. Thus, question position was thematically matched across blocks, although the specific item presented at each corresponding position differed between banks.

The two question sets were not psychometrically calibrated or empirically equated for difficulty or discrimination before the experiment. Because each bank was observed only under the block configuration with which it was permanently associated, post hoc estimation of bank equivalence would not disentangle item-bank characteristics from the block configuration. In addition, the present sample size was not designed for stable item-response-theory parameter estimation. Consequently, no post hoc psychometric equating analysis was used to claim equivalence between the two banks. Block configuration and question-bank characteristics therefore cannot be separated by the present design. Accordingly, all between-block contrasts in this study should be interpreted as comparisons between the two implemented block configurations rather than as isolated estimates of the causal effect of the auditory deadline cue.

The complete Spanish wording of all questions, response alternatives, correct answers, content-category classifications, thematic pairings, and item-level descriptive results are provided in Supplementary Material S1.

2.6. Experimental Procedure

Experimental sessions were conducted individually. Upon arrival, participants received information about the study, eligibility was confirmed, and written informed consent was obtained. Each participant was associated with the numerical identifier corresponding to the previously randomized experimental sequence.

Participants were seated and fitted with the Meta Quest 3 headset and Touch Plus controllers. They first reviewed the six instruction screens and completed three practice questions to become familiar with question selection, controller pointing, the response deadline, and lifeline use. The practice trials used a 20 s visual countdown and did not include the auditory deadline cue; thus, familiarization with the ticking signal occurred only when participants first entered the auditory experimental block.

Participants then completed the first 10-question experimental block according to their assigned sequence. At the end of the first block, the application displayed an instruction indicating that the block had ended. Participants removed the headset and rested for approximately 1 min before beginning the second experimental block. They then completed the other block using its corresponding question bank.

After completing both experimental blocks, participants removed the headset and completed the System Usability Scale (SUS) digitally using Microsoft Forms on a tablet. The complete session, including instructions, practice, both experimental blocks, the inter-block rest period, and post-task assessment, lasted approximately 15–20 min.

No participant discontinued the experiment after beginning the VR task, and no adverse event requiring termination of the session was observed. One participant who wore prescription glasses reported some physical discomfort associated with wearing the headset over the glasses but completed the experimental procedure.

2.7. Outcomes and Measures

2.7.1. Behavioral Outcomes

The behavioral dataset comprised 440 item-level observations (22 participants × 2 blocks × 10 questions).

Response accuracy was recorded automatically for each question and coded as correct (1) or incorrect (0). For the all-trial accuracy outcome, a trial in which the response interval expired without submission of an answer was scored as incorrect.

Response time was defined as the locally measured elapsed interval from initiation of the question-loading routine to registration of the submitted answer. Response time was observable only for trials in which an answer was submitted; trials ending in response expiration therefore had no response-time value and were not included in this analysis. Because the probability of submitting a response differed between blocks, the set of trials contributing to the response-time analysis also differed between blocks. Accordingly, response time was analyzed only among submitted responses and does not characterize the complete temporal response process. When the timer-reset lifeline was used, the visible countdown and auditory sequence were restarted, but the response-time measure continued accumulating from the initiation of the original question. Response times greater than 15 s were therefore possible.

Response expiration was defined as a binary trial-level outcome indicating whether the available response interval ended before an answer was submitted (0 = response submitted; 1 = response expired). The application stored this event using the technical variable name `timeout`; throughout the manuscript, the behavioral outcome is referred to as response expiration.

Use of the three lifelines was recorded at the trial level. These records were used to characterize interaction behavior and to support sensitivity analyses concerning the timer-reset function and the original 15 s deadline.

2.7.2. System Usability Scale

Perceived usability was evaluated after completion of both experimental blocks using the Spanish-language System Usability Scale (SUS) [24,25]. The SUS comprises 10 statements rated on a five-point Likert scale and yields a global score ranging from 0 to 100, with higher values representing greater perceived usability.

Standard SUS scoring was applied. For odd-numbered items, one point was subtracted from the participant's response; for even-numbered items, the response was subtracted from five. The resulting adjusted values were summed and multiplied by 2.5. Internal consistency of the 10-item scale was assessed using Cronbach's alpha.

2.8. Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics version 26. Descriptive statistics were calculated at both participant and item levels. Continuous variables were summarized using means, standard deviations, ranges, and, where informative, medians; categorical variables were summarized using frequencies and percentages. All inferential tests were two-sided, with statistical significance set at $\alpha = 0.05$. Given the exploratory nature of the study and the absence of a single prespecified primary behavioral outcome, response accuracy, response expiration, and response time among submitted responses were analyzed separately.

To complement the trial-level description of response expiration, participant-level expiration counts were aggregated separately for the baseline and auditory-cue blocks. For each participant, a paired difference was calculated as the baseline count minus the auditory-block count. Participants were then descriptively classified as having fewer, equal,

or more response-expiration events in the auditory-cue block relative to baseline. The paired difference was summarized using its mean, standard deviation, and 95% confidence interval based on the paired-samples *t* distribution. This participant-level summary was treated as a descriptive complement to, rather than a replacement for, the trial-level mixed-effects analysis.

All-trial response accuracy was analyzed using a generalized linear mixed-effects model with a binomial distribution and logit link. Block configuration (baseline vs. auditory-cue block), content category (Ecuador-related vs. international/general), experimental period (first vs. second), and the block-configuration × content-category interaction were specified as fixed effects. Random intercepts for participant and individual question were included to account for repeated observations within participants and heterogeneity among items, consistent with mixed-effects approaches that treat participants and stimuli as crossed sources of variability [26].

Because the block-configuration × content-category interaction model produces an overall marginal block contrast that equally weights the two content-category levels, an additional reduced accuracy model excluding the interaction was fitted as an alternative specification. This reduced model estimated a common block contrast across content categories and was treated as a sensitivity analysis rather than as the main accuracy model.

Response expiration was analyzed using an analogous binomial generalized linear mixed-effects model, with expiration as the binary trial-level outcome and the same fixed- and random-effects structure as the main accuracy model.

Response-time analyses were restricted to trials in which an answer was submitted. Because raw response times showed substantial positive skew, response time was transformed using the natural logarithm before inferential analysis. Log-transformed response time was analyzed using a linear mixed-effects model with a normal distribution and identity link. The fixed-effects structure comprised block configuration, content category, experimental period, and the block-configuration × content-category interaction, with random intercepts for participant and individual question.

Degrees of freedom were estimated using the Satterthwaite approximation. Estimated marginal means were obtained for the two block configurations and for combinations of block configuration and content category. Simple block comparisons within content categories were adjusted using the sequential Bonferroni procedure. For binary outcomes, marginal logit contrasts comparing the auditory-cue block with the baseline block were exponentiated to obtain odds ratios (ORs) and 95% confidence intervals. For log-transformed response time, the corresponding block contrast was exponentiated to obtain a geometric mean ratio (GMR) and 95% confidence interval.

Content-category main effects and block-configuration × content-category interactions were treated as secondary exploratory analyses. The study was not designed or powered to provide precise estimates of content-category moderation, and the broad content classification was not a direct measure of prior knowledge or item familiarity. Accordingly, nonsignificant interaction tests were not interpreted as evidence that the between-block contrast was equivalent or consistent across content categories.

Several supplementary and sensitivity analyses were conducted to evaluate the robustness of the behavioral results. First, because response-expiration trials were coded as incorrect in the all-trial accuracy outcome, accuracy was reanalyzed after restricting the dataset to trials in which an answer had actually been submitted. The same fixed- and random-effects structure as the main accuracy model was retained.

Second, the potential influence of the timer-reset lifeline was examined in two ways. The response-time model was repeated after excluding trials in which the timer-reset lifeline had been used. In addition, a supplementary binary outcome was constructed to indicate

whether a response had been submitted within the original 15 s interval, irrespective of any subsequent timer reset. Trials ending in response expiration and submitted responses occurring after 15 s were both classified as failures to respond within the original deadline. This outcome was analyzed using the same binomial mixed-effects structure used for response expiration.

Third, lifeline use was summarized separately for the 50/50, remove-one-option, and timer-reset functions. Because each lifeline could be used at most once per block, within-participant differences in use of each individual lifeline were evaluated using exact McNemar tests. The total number of lifelines used per block was compared using a paired Wilcoxon signed-rank test.

Fourth, experimental sequence (AB vs. BA) was added as a fixed effect to the accuracy, response-expiration, and response-time models to examine whether the main results depended on presentation order.

Fifth, supplementary models included thematic question pair, experimental period, and experimental sequence as fixed effects to examine whether the overall between-block pattern remained after accounting for corresponding thematic domains across the two question banks. Content category and thematic pair were not entered simultaneously because content category was determined by question position and therefore by thematic pairing. These analyses were treated as robustness checks and not as a means of eliminating the fundamental confounding between block configuration and fixed question bank.

Participant-level random slopes for block configuration were additionally examined as model-robustness checks. These models were used to assess whether allowing the block contrast to vary across participants materially changed the conclusions of the random-intercept models. Because random-slope specifications require sufficient within-participant information to estimate both the average block contrast and its between-participant heterogeneity, models showing convergence problems or a non-positive-definite random-effects covariance matrix or Hessian were considered insufficiently supported by the available data and were not interpreted. The random-intercept specification was retained for the main analyses because it accounted for clustering by participant and item without requiring estimation of an additional participant-specific block-effect variance that the data could not always support reliably.

Because three behavioral outcomes were examined, the three overall between-block p-values from the main behavioral models were additionally subjected to a Holm correction as an exploratory multiplicity sensitivity analysis. This correction was not used to redefine the individual models but to assess whether the overall statistical pattern was sensitive to simultaneous consideration of the three behavioral outcomes.

SUS scores were summarized descriptively, and internal consistency was evaluated using Cronbach's alpha. Two exploratory Spearman rank correlations were performed to examine the association of SUS score with overall response accuracy and overall response time. The corresponding p-values were adjusted using the Holm procedure. No imputation was performed for behavioral or usability outcomes. The SPSS syntax used for the principal and sensitivity analyses is provided in Supplementary Material S3.

2.9. Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of Universidad Tecnológica Indoamérica (approval code UTI-VI-060-2025; 30 July 2025). The approved protocol covered the present immersive VR experiment and the planned participant exposure to both experimental blocks.

All participants provided written informed consent before beginning the experimental session. They were informed of the study procedure, the voluntary nature of participation,

the possibility of discomfort associated with immersive VR, their right to discontinue participation, and the scientific use of the collected data. Experimental records were associated with numerical participant identifiers for analysis.

3. Results

3.1. Participant Characteristics and Experimental Flow

Twenty-nine individuals were initially assessed for participation. Seven did not enter the experimental phase, leaving 22 participants who were assigned according to the predefined balanced block-order schedule. Eleven participants completed the baseline block followed by the auditory-cue block (AB), whereas 11 completed the blocks in the reverse order (BA). All participants who began the experiment completed both periods and were included in the analyses (Figure 1).

The final sample comprised 22 participants with a mean age of 23.91 years (SD = 3.96; range = 18–31). Twenty participants (90.9%) were male and two (9.1%) were female. Participant characteristics are summarized in Table 1.

Table 1. Characteristics of the analyzed sample.

Characteristic	Value
Participants analyzed, *n*	22
Age, years, mean (SD)	23.91 (3.96)
Age range, years	18–31
Male, *n* (%)	20 (90.9)
Female, *n* (%)	2 (9.1)
Sequence AB, *n* (%)	11 (50.0)
Sequence BA, *n* (%)	11 (50.0)
Completed both experimental periods, *n* (%)	22 (100)

3.2. Descriptive Behavioral Performance

The behavioral dataset comprised 440 item-level observations, corresponding to 22 participants completing 10 questions in each of the two experimental blocks. Of these trials, 399 resulted in a submitted response, and 41 ended in response expiration.

At the participant level, mean all-trial accuracy was 47.27% (SD = 17.78) in the baseline block and 49.55% (SD = 18.89) in the auditory-cue block. Mean response time among submitted responses was 7.78 s (SD = 1.56) in the baseline block and 7.29 s (SD = 1.57) in the auditory-cue block. These standard deviations represent variability across participant-level means rather than variability across individual trials.

The most pronounced descriptive contrast concerned response expiration. Thirty-three of 220 baseline trials (15.0%) ended in response expiration, compared with eight of 220 trials (3.6%) in the auditory-cue block. At the participant level, 13 of 22 participants (59.1%) showed fewer response-expiration events in the auditory-cue block than in the baseline block, whereas nine participants (40.9%) showed the same number of expirations in both blocks. No participant showed more expirations in the auditory-cue block than in baseline. The mean paired baseline-minus-auditory difference was 1.14 response-expiration events per participant (SD = 1.17, 95% CI [0.62, 1.65]).

Descriptive behavioral outcomes by block and content category are presented in Table 2.

The fixed effects from the three main mixed-effects models are summarized in Table 3. In the original statistical models, the predictor was labeled “experimental condition”; throughout the results, estimates for this predictor are described as between-block contrasts because question bank and auditory configuration were permanently linked. They therefore compare the two implemented block configurations rather than representing an isolated

auditory-cue effect. For accuracy, the OR reported from the interaction model is an equally weighted marginal contrast across the two content categories; the reduced-model estimate is reported in Section 3.3. Effect estimates are reported with 95% confidence intervals, and exploratory Holm-adjusted *p*-values are shown for the three overall block contrasts.

Table 2. Descriptive behavioral outcomes by experimental block and content category.

Outcome	Content Category	Baseline Block	Auditory-Cue Block
Accuracy (%), mean (SD)	International/general	51.52 (19.18)	61.36 (22.05)
	Ecuador-related	40.91 (25.05)	31.82 (26.93)
	Overall	47.27 (17.78)	49.55 (18.89)
Response time (s), mean (SD)	International/general	7.64 (1.52)	7.38 (1.80)
	Ecuador-related	7.91 (2.43)	7.15 (1.64)
	Overall	7.78 (1.56)	7.29 (1.57)
Submitted responses, *n*/*N* (%)	International/general	111/132 (84.1%)	129/132 (97.7%)
	Ecuador-related	76/88 (86.4%)	83/88 (94.3%)
	Overall	187/220 (85.0%)	212/220 (96.4%)
Response expiration, *n*/*N* (%)	International/general	21/132 (15.9%)	3/132 (2.3%)
	Ecuador-related	12/88 (13.6%)	5/88 (5.7%)
	Overall	33/220 (15.0%)	8/220 (3.6%)

Note: Accuracy and response-time values are participant-level means. For these outcomes, the reported SDs represent variability across participant-level means rather than variability across individual trials. Response time includes only trials with a submitted response. Submitted-response and response-expiration values are trial-level *n*/*N* (%).

Table 3. Fixed effects from the main mixed-effects models comparing the two experimental block configurations.

Outcome	Fixed Effect	*F* (df)	*p*	Effect Estimate (95% CI)	Holm *p*
Accuracy	Block configuration	0.014 (1, 14)	0.909	OR = 0.93 (0.22–3.89)	0.984
	Content category	2.537 (1, 14)	0.134	—	—
	Period	0.375 (1, 435)	0.540	—	—
	Block configuration × content category	0.568 (1, 14)	0.464	—	—
Response expiration	Block configuration	8.148 (1, 23)	0.009	OR = 0.21 (0.07–0.65)	0.027
	Content category	0.705 (1, 23)	0.410	—	—
	Period	0.020 (1, 435)	0.887	—	—
	Block configuration × content category	0.959 (1, 23)	0.338	—	—
Log response time	Block configuration	0.494 (1, 16)	0.492	GMR = 0.93 (0.74–1.16)	0.984
	Content category	0.011 (1, 16)	0.919	—	—
	Period	0.051 (1, 357)	0.821	—	—
	Block configuration × content category	0.009 (1, 16)	0.927	—	—

Note: Block-configuration estimates compare the auditory-cue block with the baseline block. Because block configuration was permanently linked to the question bank, these estimates describe the contrast between the two implemented blocks and should not be interpreted as isolated effects of the auditory cue. OR = odds ratio; GMR = geometric mean ratio. Holm-adjusted *p*-values refer to the exploratory correction across the three overall block-configuration tests.

3.3. Response Accuracy

All-trial response accuracy was analyzed using a binomial generalized linear mixed-effects model including block configuration, content category, experimental period, and the block-configuration × content-category interaction as fixed effects, with random intercepts for participant and individual question.

No statistically significant between-block contrast was detected, $F(1, 14) = 0.014$, $p = 0.909$. Content category was also nonsignificant, $F(1, 14) = 2.537$, $p = 0.134$, as was experimental period, $F(1, 435) = 0.375$, $p = 0.540$. The block-configuration × content-category interaction was not significant, $F(1, 14) = 0.568$, $p = 0.464$.

The model-estimated probabilities of a correct response were 0.540 (95% CI [0.239, 0.815]) for international/general items and 0.402 (95% CI [0.121, 0.766]) for Ecuador-related items in the baseline block. Corresponding probabilities in the auditory deadline-cue block were 0.643 (95% CI [0.327, 0.870]) and 0.273 (95% CI [0.070, 0.651]), respectively.

Sequential Bonferroni-adjusted comparisons showed no significant between-block contrast within international/general items ($p = 0.622$) or Ecuador-related items ($p = 0.580$). Comparisons between content categories were also nonsignificant within the baseline block ($p = 0.555$) and within the auditory-cue block ($p = 0.086$).

The marginal between-block contrast from the main interaction model yielded an auditory-versus-baseline OR of 0.93 (95% CI [0.22, 3.89]). Because this contrast equally weighted the two content-category levels, an additional reduced model without the block-configuration × content-category interaction was fitted as an alternative specification to estimate a common between-block contrast across content categories. In this reduced model, the contrast was nonsignificant, $F(1, 14) = 0.001$, $p = 0.973$, with an auditory-versus-baseline OR of 1.02 (95% CI [0.26, 4.05]). Thus, the alternative specification did not alter the substantive conclusion that no statistically significant between-block contrast in all-trial accuracy was detected.

Substantial heterogeneity was observed across individual questions. The random-intercept variance was 0.470 for participants (SE = 0.241; 95% CI [0.173, 1.282]) and 1.867 for individual items (SE = 0.819; 95% CI [0.790, 4.413]). Observed item accuracy ranged from 4.55% to 100% across the 20 experimental questions. Complete item-level descriptive results are provided in Supplementary Material S1.

3.4. Response Expiration

Response expiration was analyzed using a binomial generalized linear mixed-effects model with the same fixed- and random-effects structure used for all-trial accuracy.

A statistically significant between-block contrast was detected, $F(1, 23) = 8.148$, $p = 0.009$. Averaged over content category, the estimated odds of response expiration in the auditory-cue block were lower than in the baseline block (OR = 0.21, 95% CI [0.07, 0.65]). Because question bank and auditory configuration were permanently linked, this odds ratio represents the contrast between the two implemented block configurations and must not be interpreted as an isolated effect of the auditory cue.

For international/general items, the model-estimated response-expiration probability was 0.133 (95% CI [0.057, 0.278]) in the baseline block and 0.019 (95% CI [0.005, 0.070]) in the auditory-cue block. For Ecuador-related items, the corresponding probabilities were 0.124 (95% CI [0.045, 0.299]) and 0.049 (95% CI [0.014, 0.154]), respectively.

Content category was not significantly associated with response expiration, $F(1, 23) = 0.705$, $p = 0.410$, and experimental period was also nonsignificant, $F(1, 435) = 0.020$, $p = 0.887$. The block-configuration × content-category interaction was not statistically significant, $F(1, 23) = 0.959$, $p = 0.338$.

These content-category analyses were secondary and exploratory. Given the limited sample size and number of items within each category, the nonsignificant interaction should not be interpreted as evidence that the between-block contrast was equivalent or consistent across Ecuador-related and international/general items.

Sequential Bonferroni-adjusted simple comparisons yielded a statistically significant between-block contrast within international/general items ($p = 0.038$), whereas the corresponding comparison within Ecuador-related items was nonsignificant ($p = 0.230$). Because the overall interaction was not statistically significant and the content-category analyses were exploratory, the difference between these two simple-comparison p-values was not interpreted as evidence that the block contrast differed across content categories.

Figure 5 shows the participant-level paired response-expiration counts across the two VR quiz blocks. Thirteen of 22 participants (59.1%) had fewer response-expiration events in the auditory-cue block than in the baseline block, whereas nine participants (40.9%) had the same number in both blocks; no participant had more expirations in the auditory-cue block. The mean number of expirations was 1.50 in the baseline block and 0.36 in the auditory-cue block. Individual observations are connected within participants, and small horizontal offsets are used only to reduce overplotting of identical counts.

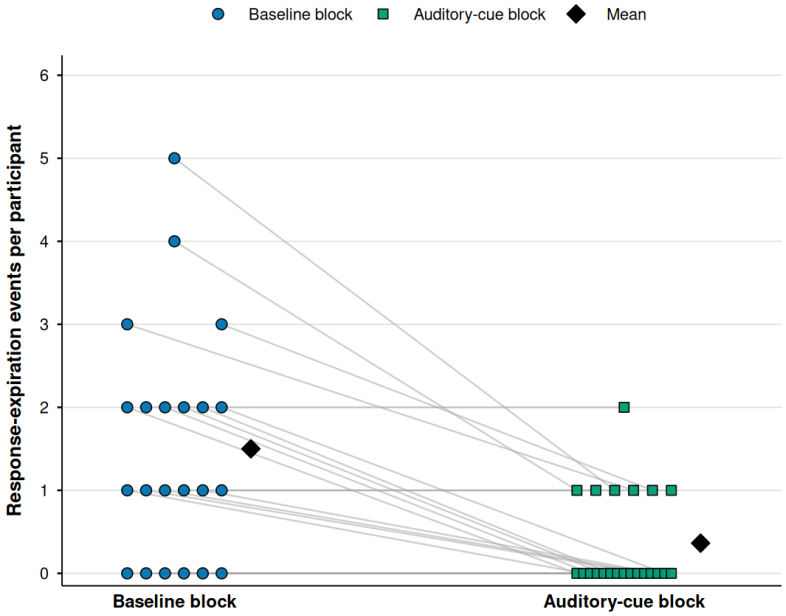


Figure 5. Participant-level response-expiration counts across the baseline and auditory-cue VR quiz blocks.

Random-intercept variances were 1.316 for participants and 1.596 for individual questions, indicating variability in response-expiration probability at both levels.

3.5. Response Time Among Submitted Responses

Response-time analyses were restricted to the 399 trials in which an answer was submitted. Raw response times had a mean of 7.41 s (SD = 3.27; range = 2.24–25.87 s) and showed positive skewness (skewness = 1.535; kurtosis = 4.935). These descriptive statistics were calculated at the individual-trial level and therefore differ from the participant-level response-time SDs reported in Table 2. Following natural-log transformation, the distribution was approximately symmetric ($M = 1.914$, SD = 0.422; skewness = 0.001; kurtosis = −0.200). Inferential analyses therefore used log-transformed response time.

No statistically significant between-block contrast was detected, $F(1, 16) = 0.494$, $p = 0.492$. Content category was also nonsignificant, $F(1, 16) = 0.011$, $p = 0.919$, as was experimental period, $F(1, 357) = 0.051$, $p = 0.821$. The block-configuration × content-category interaction was not significant, $F(1, 16) = 0.009$, $p = 0.927$.

The estimated marginal mean on the log scale was 1.977 (95% CI [1.800, 2.154]) for baseline and 1.902 (95% CI [1.726, 2.079]) for the auditory-cue block. After back-transformation, these estimates corresponded to geometric mean response times of approximately 7.22 s and 6.70 s, respectively. The auditory-versus-baseline geometric mean ratio was 0.93 (95% CI [0.74, 1.16], $p = 0.492$).

For the combinations of block configuration and content category, back-transformed estimated response times were approximately 7.22 s for international/general items and 7.21 s for Ecuador-related items in baseline, compared with 6.77 s and 6.63 s, respectively, in the auditory-cue block. Between-block contrasts were nonsignificant within international/general ($p = 0.637$) and Ecuador-related items ($p = 0.614$).

These estimates describe response time only for trials in which an answer was submitted. Because response completion itself differed between experimental blocks, the response-time model is conditional on a selected subset of trials and should not be interpreted as evidence that the overall temporal response process was equivalent between blocks.

3.6. Sensitivity and Robustness Analyses

Several supplementary analyses examined whether the main behavioral pattern depended on alternative outcome definitions, timer-reset use, lifeline behavior, presentation order, thematic pairing, random-effects specification, or multiplicity.

Accuracy was reanalyzed after excluding trials that ended in response expiration. Among submitted responses, 104 of 187 baseline responses were correct (55.6%), compared with 109 of 212 responses (51.4%) in the auditory deadline-cue block. The corresponding mixed-effects model showed no significant between-block contrast, $F(1, 14) = 0.261$, $p = 0.617$, and no significant block-configuration × content-category interaction, $F(1, 14) = 0.520$, $p = 0.483$. Model-estimated probabilities of a correct response among submitted responses were 0.552 (95% CI [0.307, 0.775]) for baseline and 0.470 (95% CI [0.241, 0.711]) for the auditory-cue block.

The influence of the timer-reset lifeline was examined using the original 15 s deadline. Under this alternative definition, trials ending in response expiration and submitted responses occurring after 15 s were both classified as failures to respond within the original deadline. Thirty-seven of 220 baseline trials (16.8%) and 12 of 220 auditory deadline-cue trials (5.5%) met this criterion. The mixed-effects model yielded a significant Type III test of the between-block contrast, $F(1, 17) = 4.880$, $p = 0.042$. Model-estimated probabilities were 0.131 (95% CI [0.059, 0.266]) for baseline and 0.041 (95% CI [0.016, 0.104]) for the auditory-cue block. The baseline-minus-auditory marginal probability difference was 0.089 (95% CI [−0.011, 0.189], $p = 0.076$), indicating greater uncertainty when the contrast was expressed on the original probability scale.

The response-time model was also repeated after excluding trials in which the timer-reset lifeline had been used. The conclusion remained unchanged: the between-block contrast was not significant, $F(1, 16) = 0.437$, $p = 0.518$. Back-transformed estimated response times were approximately 6.98 s in baseline and 6.55 s in the auditory-cue block.

Lifeline use showed no statistically significant within-participant differences between blocks. The 50/50 lifeline was used by 11 of 22 participants (50.0%) in baseline and 14 of 22 (63.6%) in the auditory-cue block (exact McNemar $p = 0.453$). The remove-one-option lifeline was used by 13 participants (59.1%) and 11 participants (50.0%), respectively ($p = 0.500$), whereas timer reset was used by seven participants (31.8%) and eight par-

ticipants (36.4%), respectively ($p = 1.000$). Participants used a mean of 1.41 lifelines (SD = 1.14) in baseline and 1.50 (SD = 1.06) in the auditory-cue block. The paired Wilcoxon signed-rank test showed no difference in total lifeline use ($Z = -0.368$, $p = 0.713$).

Experimental sequence was added as a fixed effect to each of the three main models. Sequence was not significantly associated with accuracy, $F(1, 19) = 1.376$, $p = 0.256$, response expiration, $F(1, 28) = 0.014$, $p = 0.907$, or response time, $F(1, 20) = 0.431$, $p = 0.519$. Inclusion of sequence did not alter the main between-block contrasts: neither the accuracy contrast ($p = 0.907$) nor the response-time contrast ($p = 0.492$) was significant, whereas the response-expiration contrast remained significant, $F(1, 23) = 8.138$, $p = 0.009$.

Supplementary models accounted for the ten corresponding thematic question pairs together with experimental period and sequence. For accuracy, neither the between-block contrast, $F(1, 8) = 0.005$, $p = 0.948$, nor thematic pair, $F(9, 8) = 0.465$, $p = 0.862$, was statistically significant. For response time, the between-block contrast was nonsignificant, $F(1, 9) = 0.566$, $p = 0.471$, as was thematic pair, $F(9, 9) = 1.023$, $p = 0.487$.

For response expiration, the thematic-pair model retained a significant between-block contrast, $F(1, 8) = 6.046$, $p = 0.040$, whereas thematic pair itself was not significant, $F(9, 6) = 0.376$, $p = 0.909$. The auditory-versus-baseline odds ratio was 0.19 (95% CI [0.04, 0.91]). This analysis provides a robustness check for broad thematic matching but does not remove the confounding between block configuration and fixed question bank.

Participant-level random slopes for the model's original "experimental condition" predictor (coding block configuration) were examined. Random-slope models for all-trial accuracy and response time converged without warnings and produced conclusions consistent with the main random-intercept models. The between-block contrast remained nonsignificant for accuracy, $F(1, 14) = 0.014$, $p = 0.909$, and response time, $F(1, 17) = 0.473$, $p = 0.501$. The corresponding response-expiration model produced a non-positive-definite random-effects covariance matrix and Hessian. Response expiration was relatively sparse, with 41 events across 440 trials and only eight events in the auditory deadline-cue block. Together with the sample of 22 participants, this provided limited information for reliably estimating an additional participant-specific variance for the original "experimental condition" predictor. The non-positive-definite covariance matrix and Hessian therefore indicated that the more complex random-slope specification was not adequately identified by the available data rather than providing a reliable estimate of between-participant heterogeneity in the between-block contrast. Accordingly, the random-slope response-expiration model was considered unstable, its estimates were not interpreted, and inference was retained from the more parsimonious random-intercept model.

Finally, the three overall between-block contrast p-values from the main behavioral models were subjected to an exploratory Holm correction. The response-expiration contrast remained below the conventional significance threshold (Holm-adjusted $p = 0.027$), whereas the adjusted p-values for accuracy and response time were both 0.984.

3.7. System Usability

The mean SUS score was 80.91 (SD = 11.09), with a 95% confidence interval from 75.99 to 85.83. The median score was 81.25, and observed scores ranged from 52.5 to 100. Because the SUS was administered once after completion of both experimental blocks, this score characterizes participants' perception of the VR application as a whole and cannot be attributed separately to either block configuration.

Internal consistency of the 10 SUS items was modest, with Cronbach's $\alpha = 0.611$ (standardized $\alpha = 0.590$). The SUS findings were therefore treated primarily as a descriptive characterization of perceived usability in this sample rather than as a highly precise individual-difference measure.

Two exploratory Spearman correlations examined the relationship between SUS score and overall behavioral performance. Overall participant accuracy had a mean of 48.41% (SD = 15.69), whereas overall response time had a mean of 7.54 s (SD = 1.43). SUS score was positively associated with overall accuracy ($\rho = 0.432$, unadjusted $p = 0.044$), but this association did not remain statistically significant after Holm correction (adjusted $p = 0.088$). No statistically significant association was observed between SUS and overall response time ($\rho = -0.096$, adjusted $p = 0.671$).

4. Discussion

4.1. Principal Findings

This exploratory within-subject study compared behavioral performance across two non-equivalent, time-limited immersive VR quiz blocks. All-trial accuracy was 47.27% in the baseline block and 49.55% in the auditory-cue block, with no statistically significant between-block contrast. Response time among submitted responses likewise showed no statistically significant contrast. Response expiration occurred in 33 of 220 baseline trials (15.0%) and eight of 220 auditory-cue trials (3.6%), corresponding to an estimated between-block OR of 0.21 (95% CI [0.07, 0.65], $p = 0.009$).

At the participant level, 13 of 22 participants showed fewer expirations in the auditory-cue block, nine showed equal counts, and none showed more expirations. The mean paired baseline-minus-auditory difference was 1.14 expiration events per participant (SD = 1.17, 95% CI [0.62, 1.65]). These results characterize the two implemented block configurations rather than an isolated effect of the auditory cue.

4.2. Response Expiration Across the Two VR Quiz Blocks

The methodological relevance of response expiration in the present study does not arise from the claim that deadline-related omissions constitute a newly identified behavioral phenomenon. Previous research on speeded-up and deadline-constrained tasks has already shown that failures to respond within an available interval may contain information about the response process and should not automatically be treated as ordinary missing observations [19,20,27].

The more specific contribution of the present study is to operationalize and report response expiration explicitly within a standalone immersive VR quiz alongside all-trial correctness and response latency among submitted responses. At the trial level, the implementation distinguished among submission of a correct response, submission of an incorrect response, and expiration of the available response interval without submission. This preserved response-expiration events in the behavioral record while allowing completed-response latency to be analyzed using the appropriate subset of submitted trials.

Response expiration and completed-response latency summarize different parts of the observed data: latency is available only when a response is submitted, whereas response expiration identifies trials that reach the response boundary without submission. Reporting both prevents deadline expirations from disappearing from the characterization of temporal task behavior. However, the different statistical patterns observed for these outcomes do not establish that they are psychologically or statistically separable constructs.

The mechanism underlying the observed between-block difference cannot be identified from the present design. The auditory-cue block combined repetitive ticking with an increase in repetition rate, no direct manipulation check of perceived urgency or time pressure was administered, and auditory configuration was linked to question bank. Future studies should independently cross item set and auditory configuration and include acoustically controlled comparison conditions and direct manipulation checks.

4.3. Response Accuracy and Response Time Among Submitted Responses

No statistically significant between-block contrast was detected for all-trial accuracy. In the main model including the block-configuration × content-category interaction, the auditory-versus-baseline marginal OR was 0.93 (95% CI [0.22, 3.89], $p = 0.909$). Because this marginal contrast equally weighted the two content-category levels, a reduced model without the interaction was additionally examined as an alternative specification. That model estimated a common block contrast of OR = 1.02 (95% CI [0.26, 4.05], $p = 0.973$). Although the two specifications produced slightly different point estimates, both were highly uncertain and led to the same substantive conclusion: no statistically significant between-block association with all-trial accuracy was detected.

The conclusion also remained unchanged when accuracy was restricted to the 399 trials in which an answer was submitted. Under this alternative definition, the between-block contrast was nonsignificant ($p = 0.617$), with observed correct-response proportions of 55.6% in the baseline block and 51.4% in the auditory-cue block.

This distinction is important because the all-trial accuracy measure treated response expirations as incorrect responses. Although this is a defensible scoring rule for a time-limited quiz, it combines two qualitatively different events: submitting an incorrect answer and failing to submit an answer before the deadline. The submitted-response analysis separated these components and showed no evidence that the higher completion rate observed in the auditory-cue block was accompanied by greater correctness among submitted responses.

Response time among submitted responses likewise showed no statistically significant between-block contrast. The model-estimated geometric mean was approximately 7.22 s in the baseline block and 6.70 s in the auditory-cue block, with a geometric mean ratio of 0.93 (95% CI [0.74, 1.16], $p = 0.492$).

This result should be interpreted conditionally because response time was observed only when participants submitted an answer, while response completion differed between the two experimental blocks. The response-time analysis therefore characterizes latency among submitted responses rather than the full temporal behavior of all trials. The nonsignificant contrast should not be interpreted as evidence that the complete response process was equivalent between blocks.

This pattern differs from experimental paradigms in which objective time pressure is created by physically reducing the time available for a decision. Such manipulations can alter response latency, accuracy, or confidence [28,29]. In the present experiment, the nominal response interval remained unchanged; instead, the auditory-cue block provided an additional auditory representation of its progression. This distinction cautions against assuming that an auditory representation of an unchanged deadline will produce the same behavioral consequences as shortening the available decision time.

Other studies similarly indicate that temporal constraints do not produce a single universal speed–accuracy response. Hermanns and Teubner reported that time constraints in a human–AI evaluation task affected performance and included trials in which responses were not completed within the available interval [30]. Usai et al. also described situations in simulator and VR settings in which responding faster in high-demand task settings could occur at the expense of accuracy [31].

The present dataset showed different descriptive patterns across the three behavioral outcomes. Response expiration differed more markedly between the two implemented blocks, whereas the estimated between-block contrasts for accuracy and response time among submitted responses were imprecise and their confidence intervals included the null value. A statistically significant contrast for one outcome together with nonsignificant contrasts for the others does not demonstrate differential responses across outcomes, establish equivalence for accuracy or response time, or show that response completion, accuracy,

and response time are empirically separable dimensions. Such conclusions would require a direct statistical comparison or a joint model specifically designed to test differential responses across outcomes.

4.4. Temporal and Auditory Information in Immersive VR

Temporal characteristics increasingly form part of the interaction design of immersive systems. Schatzschneider et al. demonstrated that manipulated temporal cues, cognitive load, and immersion can influence time estimation within immersive virtual environments [32]. More recently, Picard and Botev examined rhythmic stimulation and temporal experience in VR, showing that tempo and sensory characteristics can interact with task context and behavioral performance [33].

Wei et al. showed that expected delays during moving-target selection in VR can alter user behavior and performance, illustrating how temporal characteristics become incorporated into interaction strategies [17]. Bansal et al. likewise demonstrated that manipulating the temporal relationship between movement and virtual events can modify perceived duration [16]. Although these manipulations differ from an auditory countdown, they reinforce the broader view that timing is an active component of immersive interaction rather than merely a background property.

The present experiment examines a different form of temporal manipulation: an auditory representation of an otherwise unchanged response deadline. The visual countdown and nominal 15 s interval were maintained across the two blocks, while the auditory block added ticking whose repetition rate increased at the midpoint of the countdown.

Auditory stimulation should nevertheless not be assumed to facilitate performance simply because it provides additional information. Cucinella et al. found that different visual, auditory, and cognitive demands in an immersive shopping task influenced workload and task behavior differently [34]. Their auditory manipulation consisted of environmental auditory demand rather than a temporally informative signal, so the studies are not directly comparable. The contrast is nevertheless relevant because the behavioral consequence of sound depends on its functional relationship with the task.

A useful direction for future work would therefore be to separate the informational content of the deadline cue from the mere presence of auditory stimulation. A factorial design could compare an accelerated temporal cue, a temporally uninformative sound with matched acoustic characteristics, and a silent baseline while holding the item set constant or independently counterbalanced. Such a design would provide a stronger test of whether any observed behavioral difference arises from temporal information, auditory stimulation, or their combination.

From an immersive HCI design perspective, the present findings do not support recommending accelerated ticking as a general strategy for improving performance in time-limited tasks. Instead, they highlight the need to evaluate auditory deadline information under designs that separate the auditory manipulation from item characteristics.

Future studies should independently manipulate item set and auditory configuration and compare silent, constant-rate, and accelerated auditory conditions under calibrated acoustic and temporal control. Such designs could determine more specifically whether behavioral differences are associated with auditory stimulation, temporal information, signal acceleration, or their combination.

4.5. Exploratory Content-Category Analysis and Item-Level Heterogeneity

The content-category analyses were secondary and exploratory. No statistically significant main effect of content category or block-configuration × content-category interaction was detected for accuracy, response expiration, or response time. Given the small sam-

ple, limited number of items within each category, and uncertainty around several model estimates, these nonsignificant tests should not be interpreted as evidence that the between-block contrast was equivalent or consistent across Ecuador-related and international/general items.

The category labels also require caution. Classification as Ecuador-related does not establish greater familiarity for an individual participant, nor does classification as international/general imply lower familiarity. Because item-specific familiarity, prior exposure, and domain knowledge were not measured, these analyses describe broad content categories rather than direct differences in prior knowledge or cultural familiarity.

Individual-question heterogeneity was nevertheless substantial. In the accuracy model, item-level random-intercept variance exceeded participant-level variance, and observed accuracy across individual items ranged from 4.55% to 100%. This variability reinforces the importance of treating individual questions as a source of heterogeneity in knowledge-based HCI experiments.

Prior knowledge and contextual representations can influence retrieval and learning [35,36], but broad content classifications cannot capture all determinants of item difficulty. Question wording, distractor plausibility, semantic accessibility, prior exposure, and domain-specific knowledge may also contribute to item-level performance.

The thematic-pair analysis therefore serves only as a supplementary robustness check for broad domain matching; it does not eliminate the confounding between question bank and auditory configuration or alter the exploratory status of the content-category findings.

4.6. Lifeline Use and the Original Response Deadline

The availability of lifelines was a distinctive feature of the task because these functions could alter either response difficulty or the effective duration of an individual trial. The 50/50 and remove-one-option lifelines reduced the number of response alternatives, whereas the timer-reset function restored the visible countdown to 15 s.

The observed frequencies did not suggest a pronounced systematic imbalance in lifeline use between blocks. The 50/50 lifeline was used by 50.0% of participants in baseline and 63.6% in the auditory-cue block; the remove-one-option lifeline by 59.1% and 50.0%, respectively; and timer reset by 31.8% and 36.4%. None of the paired comparisons was statistically significant, and the total number of lifelines used was also similar across blocks. Nevertheless, the absence of significant paired differences does not establish that lifeline use had no effect on individual trials.

The sensitivity analysis based on the original 15 s deadline provided a more direct assessment of the timer-reset issue. Reclassifying submitted responses occurring after 15 s as failures to respond within the original deadline increased the number of deadline misses from 33 to 37 in baseline and from eight to 12 in the auditory-cue block. Thus, four additional post-15 s responses were identified in each block.

The Type III test of the between-block contrast from this alternative mixed-effects model remained significant ($p = 0.042$). However, the baseline-minus-auditory marginal probability difference was 0.089 and had greater uncertainty (95% CI [−0.011, 0.189], $p = 0.076$). These two results refer to different scales of inference and should not be interpreted as contradictory.

The timer-reset sensitivity analysis preserved the descriptive between-block pattern, although uncertainty increased when the outcome was expressed on the marginal probability scale. These analyses do not establish that lifeline use was behaviorally irrelevant at the individual-trial level.

4.7. Usability and Its Relationship with Performance

The mean SUS score was 80.91 (SD = 11.09; 95% CI [75.99, 85.83]). Because usability was assessed only once after participants had completed both VR blocks, the score represents a global post-task evaluation of the application rather than a block-specific assessment. It therefore cannot be used to determine whether perceived usability differed between the baseline and auditory-cue configurations.

Interpretation should also account for the modest internal consistency of the SUS in this sample ($\alpha = 0.611$) and the small sample size. Accordingly, the SUS result is reported descriptively rather than used to support a strong categorical claim about usability or precise individual-difference inferences.

The exploratory association between SUS and overall accuracy ($\rho = 0.432$) did not remain statistically significant after Holm correction ($p_{\text{adj}} = 0.088$), and no statistically significant association was observed between SUS and overall response time. The available data therefore do not provide sufficient evidence that participants who reported higher usability ratings also performed better on the quiz task.

4.8. Limitations and Future Research

The principal limitation concerns the experimental design. Each block was permanently linked to a different fixed question bank, and the two banks were thematically paired but were not psychometrically calibrated, empirically equated, or independently crossed with auditory configuration. Consequently, the observed between-block differences may reflect the auditory cue, question-bank characteristics, or both. Item-level random effects and thematic-pair sensitivity analyses account statistically for some heterogeneity but cannot replace an experimental crossing that was absent from the design. The item-level descriptive results illustrate the relevance of this limitation: at question position 1, the baseline item produced eight expirations and three correct responses, compared with zero expirations and 16 correct responses for the corresponding auditory-block item; at position 10, the respective values were seven versus zero expirations and four versus 20 correct responses. These two pairs alone accounted for 15 of the net difference of 25 response-expiration events between blocks. The reported block contrasts, including the odds ratio for response expiration, therefore characterize the two complete block configurations and should not be interpreted as isolated effects of the auditory cue. Future work should independently cross question bank and auditory configuration or use psychometrically calibrated parallel forms.

Interpretation is also limited by the sample. The final sample comprised 22 convenience-recruited participants from a single institution in Ecuador, with a mean age of 23.91 years (SD = 3.96; range = 18–31); the sample was also markedly imbalanced by sex (20 males and two females). These characteristics limit statistical precision and generalizability and preclude meaningful assessment of sex-related moderation. Larger and more demographically diverse samples would permit more precise estimates and more flexible random-effects structures.

Temporal measurement also had technical limitations. Response time was recorded locally in Unity from initiation of the question-loading routine rather than from an instrumented physical display onset. Although network-transmission latency was excluded, runtime frame rate was not recorded; therefore, actual rendering performance under the nominal 72 Hz headset configuration could not be verified. Future studies requiring precise temporal inference should use event- or frame-synchronized timing together with continuous frame-rate monitoring.

The auditory manipulation was likewise incompletely characterized at the participant level. The auditory-cue block used prerecorded ticking at approximately 1 Hz during the first half of the nominal response interval and approximately 2 Hz during the second

half. However, sound-pressure level at the ear was not calibrated, device-volume settings were not recorded for each session, spectral characteristics were not formally quantified as experimental parameters, and formal audiometric screening was not performed. In addition, the manipulation combined the presence of repetitive sound with an increase in repetition rate, so their individual contributions cannot be separated. Future experiments should compare silent, constant-rate, and accelerated auditory block configurations while standardizing acoustic output and documenting participant hearing status.

No validated manipulation check was included for perceived urgency, subjective time pressure, workload, or stress. The study therefore characterizes behavioral responses to the implemented interface configuration rather than a verified psychological state. Future work should include direct manipulation checks and, where theoretically justified, workload or physiological measures to examine whether behavioral differences are associated with perceived deadline salience or other psychological mechanisms. Similarly, the Ecuador-related and international/general classifications represent broad content categories rather than direct measurements of prior knowledge, familiarity, or cultural relevance; these exploratory analyses should therefore not be interpreted as tests of prior knowledge.

Several task-design features may also have influenced individual trials. Questions and response alternatives were presented in fixed order, and the practice phase used a 20 s visual countdown without the auditory cue. Thus, item-position effects and a short-lived novelty or familiarization response to the ticking signal cannot be excluded. Although AB/BA block ordering reduced systematic period imbalance, it did not randomize item order within blocks or eliminate possible carryover in attention to the visual countdown. Future studies should randomize or counterbalance item and response-option order where feasible and include auditory familiarization before experimental measurement.

Finally, lifelines altered either response difficulty or the effective response interval. In particular, timer reset restarted the 15 s countdown and the auditory sequence, while option-elimination functions reduced the number of available alternatives. Sensitivity analyses did not show a pronounced between-block imbalance and preserved the overall descriptive expiration pattern, but lifeline use was self-selected and may itself have formed part of participants' response strategies. Studies designed specifically to isolate temporal behavior should therefore remove, experimentally manipulate, or prespecify lifeline availability.

More broadly, the findings derive from one VR application, one nominal 15 s response interval, one auditory escalation pattern, and one knowledge-based quiz implementation. Generalization to other interfaces, timing parameters, auditory designs, populations, and task domains remains to be established. Taken together, the limitations primarily constrain causal attribution, precision, and generalizability. Stronger inference will require independently crossed item sets and auditory configurations, larger and more diverse samples, direct manipulation checks, and tighter auditory and temporal control.

5. Conclusions

This exploratory study compared response accuracy, response expiration, and response time among submitted responses across two non-equivalent, time-limited immersive VR quiz blocks. Response expiration occurred in 33 of 220 baseline trials (15.0%) and eight of 220 auditory-cue trials (3.6%), whereas no statistically significant between-block contrasts were detected for all-trial accuracy or response time among submitted responses. At the participant level, 13 of 22 participants showed fewer expirations in the auditory-cue block, nine showed equal counts, and none showed more expirations; the mean paired baseline-minus-auditory difference was 1.14 expiration events per participant (SD = 1.17, 95% CI [0.62, 1.65]).

These findings characterize the two implemented block configurations rather than an isolated effect of the auditory deadline cue. Auditory configuration and question bank were permanently linked, and the two banks were not psychometrically equated or independently crossed. The observed item-level heterogeneity further reinforces the need for caution when attributing the between-block expiration difference specifically to the auditory manipulation. Similarly, the nonsignificant contrasts for accuracy and response time should not be interpreted as evidence of equivalence between the blocks.

The methodological contribution of the study is the explicit retention and reporting of response expiration alongside all-trial accuracy and response time among submitted responses in a time-limited immersive task. This approach preserves information from trials in which the available response interval expires without treating those observations as ordinary missing response-time data. Stronger evaluation of auditory deadline cues will require independently crossed item sets or psychometrically calibrated parallel forms, larger and more diverse samples, calibrated auditory stimuli, direct manipulation checks, and event-synchronized temporal measurement.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/electronics15215029/s1, Supplementary Material S1: Question banks, response alternatives, correct answers, content-category classifications, thematic pairings, and item-level descriptive results; Supplementary Material S2: Reporting checklist informed by the CONSORT 2025 statement and, where relevant to the two-period within-participant structure, the CONSORT extension for randomized crossover trials; Supplementary Material S3: SPSS analysis syntax for the principal and sensitivity analyses reported in the manuscript; Audio File S4: Prerecorded ticking signal used during the first half of the auditory-cue response interval (approximately 1 Hz); Audio File S5: Prerecorded ticking signal used during the second half of the auditory-cue response interval (approximately 2 Hz).

Author Contributions: Conceptualization, F.A.-C., J.B. and G.P.-N.; methodology, F.A.-C., J.B. and G.P.-N.; software, F.A.-C. and J.B.; validation, F.A.-C., J.B. and G.P.-N.; formal analysis, F.A.-C. and J.B.; investigation, F.A.-C. and J.B.; resources, F.A.-C. and J.B.; data curation, F.A.-C. and J.B.; writing—original draft preparation, F.A.-C., J.B. and G.P.-N.; writing—review and editing, F.A.-C., J.B. and G.P.-N.; visualization, F.A.-C. and J.B.; supervision, J.B. and G.P.-N.; project administration, F.A.-C.; funding, F.A.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Universidad Tecnológica Indoamérica through the project "Desarrollo y análisis de un entorno virtual con integración de estímulos multisensoriales", Grant Number: IIDI-046-2026.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Universidad Tecnológica Indoamérica (protocol code UTI-VI-060-2025; date of approval: 30 July 2025).

Informed Consent Statement: Written informed consent was obtained from all participants involved in the study. Participation was voluntary, and participants were informed about the study procedures, the possible discomfort associated with immersive virtual reality, their right to discontinue participation, and the scientific use of the collected data.

Data Availability Statement: The de-identified data supporting the findings of this study are available from the corresponding author upon reasonable request. Public release of the participant- and trial-level dataset was not included in the conditions under which informed consent and ethical approval were obtained. Supplementary Material S1 provides question-bank documentation and item-level descriptive results, Supplementary Material S2 provides the reporting checklist, Supplementary Material S3 provides the SPSS analysis syntax used for the reported analyses, and the two prerecorded auditory stimuli are provided as supplementary audio files.

Acknowledgments: The authors thank all participants who voluntarily took part in the immersive virtual reality experiment. The authors also acknowledge Universidad Tecnológica Indoamérica for

providing the laboratory facilities and infrastructure used during data collection. Artificial intelligence tools were used to support language editing, formatting, and organization of the manuscript. The authors reviewed and approved all final content and remain fully responsible for the accuracy, interpretation, and integrity of the work.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bulut, O.; Yildirim-Erbasli, S.N. Assessment Analytics in Digital Assessments. *Encyclopedia* **2026**, *6*, 81. [CrossRef]
2. Alfers, T.; Gittler, G.; Ulitzsch, E.; Pohl, S. Assessing the Speed–Accuracy Trade-Off in Psychological Testing Using Experimental Manipulations. *Educ. Psychol. Meas.* **2025**, *85*, 357–383. [CrossRef] [PubMed]
3. Edworthy, J.; Loxley, S.; Dennis, I. Improving Auditory Warning Design: Relationship between Warning Sound Parameters and Perceived Urgency. *Hum. Factors* **1991**, *33*, 205–231. [CrossRef] [PubMed]
4. Edworthy, J.; Hellier, E.; Walters, K.; Weedon, B.; Adams, A. The Relationship between Task Performance, Reaction Time, and Perceived Urgency in Nonverbal Auditory Warnings. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; SAGE Publications: Los Angeles, CA, USA, 2000; Volume 44, pp. 674–677. [CrossRef]
5. Lozano, D.; Tuquet, J.; Mainer-Pardos, E.; Cartón-Llorente, A.; González, A.; Fábrica, G. Time Pressure in Decision-Making: Effects on Throwing Speed and Accuracy in Handball Jump Shots. *Int. J. Sport. Sci. Coach.* **2026**, *OnlineFirst*. [CrossRef]
6. Gonthier, C. Should Intelligence Tests Be Speeded or Unspeeded? A Brief Review of the Effects of Time Pressure on Response Processes and an Experimental Study with Raven's Matrices. *J. Intell.* **2023**, *11*, 120. [CrossRef] [PubMed]
7. Simon, J.R.; Slaviero, D.P. Differential Effects of a Foreperiod Countdown Procedure on Simple and Choice Reaction Time. *J. Mot. Behav.* **1975**, *7*, 9–14. [CrossRef] [PubMed]
8. Maslovat, D.; Klapp, S.T.; Forgaard, C.J.; Chua, R.; Franks, I.M. The Effect of Response Complexity on Simple Reaction Time Occurs Even with a Highly Predictable Imperative Stimulus. *Neurosci. Lett.* **2019**, *704*, 62–66. [CrossRef] [PubMed]
9. Bhowmik, A.K. Virtual and augmented reality: Human sensory-perceptual requirements and trends for immersive spatial computing experiences. *J. Soc. Inf. Disp.* **2024**, *32*, 605–646. [CrossRef]
10. Pallavicini, F.; Pepe, A.; Minissi, M.E. Gaming in Virtual Reality: What Changes in Terms of Usability, Emotional Response and Sense of Presence Compared to Non-Immersive Video Games? *Simul. Gaming* **2019**, *50*, 136–159. [CrossRef]
11. Makransky, G.; Petersen, G.B. The Cognitive Affective Model of Immersive Learning (CAMIL): A Theoretical Research-Based Model of Learning in Immersive Virtual Reality. *Educ. Psychol. Rev.* **2021**, *33*, 937–958. [CrossRef]
12. Avilés-Castillo, F.; Buele, J.; Salazar, M.; Palacios-Navarro, G. Influence of continuous navigation modes on the immersive experience in a virtual supermarket. *Front. Virtual Real.* **2026**, *7*, 1729505. [CrossRef]
13. Yu, J.; Song, J.; Shen, Q. Effects of Fully Immersive Virtual Reality Training on Cognitive Function in Patients with Mild Cognitive Impairment: A Systematic Review and Meta-Analysis. *Front. Hum. Neurosci.* **2024**, *18*, 1467697. [CrossRef] [PubMed]
14. Ceccato, I.; Ricci, E.; Mazza, C.; Bartolini, E.; Di Crosta, A.; La Malva, P.; Biondi, S.; Colasanti, M.; Mammarella, N.; Palumbo, R.; et al. Influence of Stimuli Emotional Features and Typicality on Memory Performance: Insights from a Virtual Reality Context. *Psychol. Res.* **2024**, *88*, 257–270. [CrossRef] [PubMed]
15. Kim, H.; Lee, I.K. Studying the Effects of Congruence of Auditory and Visual Stimuli on Virtual Reality Experiences. *IEEE Trans. Vis. Comput. Graph.* **2022**, *28*, 2080–2090. [CrossRef] [PubMed]
16. Bansal, A.; Weech, S.; Barnett-Cowan, M. Movement-Contingent Time Flow in Virtual Reality Causes Temporal Recalibration. *Sci. Rep.* **2019**, *9*, 4378. [CrossRef] [PubMed]
17. Wei, Y.; Zhang, S.; Shi, R.; Tu, H.; Jin, H.; Hui, P.; Liang, H.N. D-Model: Understanding the Effect of Expected Delays in Moving Target Selection in VR. *Int. J. Hum.-Comput. Interact.* **2026**, *42*, 17018–17038. [CrossRef]
18. Brunyé, T.T.; Giles, G.E. Methods for Eliciting and Measuring Behavioral and Physiological Consequences of Stress and Uncertainty in Virtual Reality. *Front. Virtual Real.* **2023**, *4*, 951435. [CrossRef]
19. Leng, X.; Fengler, A.; Shenhav, A.; Frank, M.J. The Perils of Omitting Omissions When Modeling Evidence Accumulation. *PLOS Comput. Biol.* **2026**, *22*, e1014667. [CrossRef] [PubMed]
20. Blurton, S.P.; Feifel, J.; Gondan, M. Speeded Response Tasks with Unpredictable Deadlines. *J. Math. Psychol.* **2023**, *115*, 102776. [CrossRef]
21. Hopewell, S.; Chan, A.W.; Collins, G.S.; Hróbjartsson, A.; Moher, D.; Schulz, K.F.; Tunn, R.; Aggarwal, P.R.; Berkwits, M.; Berlin, P.J.A.; et al. CONSORT 2025 Statement: Updated Guideline for Reporting Randomised Trials. *BMJ* **2025**, *389*, e081123. [CrossRef] [PubMed]
22. Dwan, K.; Li, T.; Altman, D.G.; Elbourne, D. CONSORT 2010 statement: Extension to randomised crossover trials. *BMJ* **2019**, *366*, l4378. [CrossRef] [PubMed]

23. Suied, C.; Susini, P.; McAdams, S. Evaluating Warning Sound Urgency with Reaction Times. *J. Exp. Psychol. Appl.* **2008**, *14*, 201–212. [CrossRef] [PubMed]
24. Bangor, A.; Kortum, P.T.; Miller, J.T. An Empirical Evaluation of the System Usability Scale. *Int. J. Hum.-Comput. Interact.* **2008**, *24*, 574–594. [CrossRef]
25. Sevilla-Gonzalez, M.D.R.; Moreno Loaeza, L.; Lazaro-Carrera, L.S.; Bourguet Ramirez, B.; Vázquez Rodríguez, A.; Peralta-Pedrero, M.L.; Almeda-Valdes, P. Spanish Version of the System Usability Scale for the Assessment of Electronic Tools: Development and Validation. *JMIR Hum. Factors* **2020**, *7*, e21161. [CrossRef] [PubMed]
26. Baayen, R.H.; Davidson, D.J.; Bates, D.M. Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items. *J. Mem. Lang.* **2008**, *59*, 390–412. [CrossRef]
27. Damaso, K.A.M.; Castro, S.C.; Todd, J.; Strayer, D.L.; Provost, A.; Matzke, D.; Heathcote, A. A Cognitive Model of Response Omissions in Distraction Paradigms. *Mem. Cogn.* **2022**, *50*, 962–978. [CrossRef] [PubMed]
28. Zhou, Y.B.; Ruan, S.J.; Zhang, K.; Bao, Q.; Liu, H.Z. Time Pressure Effects on Decision-Making in Intertemporal Loss Scenarios: An Eye-Tracking Study. *Front. Psychol.* **2024**, *15*, 1451674. [CrossRef] [PubMed]
29. Wang, Y.; Liu, Y.; Yang, X.J. Experimental Investigation and Queuing Network (QN) Modeling of Speed-Accuracy Tradeoff (SAT) and Speed-Confidence Tradeoff (SCT) in Multi-Task Human-Robot Collaboration (HRC). In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; SAGE Publications: Los Angeles, CA, USA, 2025; Volume 69, pp. 541–547. [CrossRef]
30. Hermanns, L.; Teubner, T. Under Pressure: How Time Constraints, Task Complexity, and AI Reliability Shape Human–AI Interaction. *Behav. Inf. Technol.* **2026**, *45*, 2597–2621. [CrossRef]
31. Usai, M.; Schäfer, P.; Herzberger, N.D.; Flemisch, F. Participants' Speed-Accuracy Trade-Off Behavior in High-Stress Situations in Simulator Studies. In *Proceedings of the Intelligent Human Systems Integration (IHSI 2024): Integrating People and Intelligent Systems*; AHFE International: Palermo, Italy, 2024; Volume 119, pp. 70–78. [CrossRef]
32. Schatzschneider, C.; Bruder, G.; Steinicke, F. Who Turned the Clock? Effects of Manipulated Zeitgebers, Cognitive Load and Immersion on Time Estimation. *IEEE Trans. Vis. Comput. Graph.* **2016**, *22*, 1387–1395. [CrossRef] [PubMed]
33. Picard, S.; Botev, J. Psychophysiology of Rhythmic Stimuli and Time Experience in Virtual Reality. *Comput. Graph.* **2024**, *124*, 104097. [CrossRef]
34. Cucinella, S.L.; de Winter, J.; van den Berg, A.; van den Berg, B.; Hoogsteen, D.; Poutsma, D.; van Wingerden, F.; Marchal-Crespo, L. Shopping in Immersive Virtual Reality: Effects of Visual, Auditory, and Cognitive Demands on Mental Workload. *Appl. Ergon.* **2026**, *137*, 104781. [CrossRef] [PubMed]
35. Heald, J.B.; Lengyel, M.; Wolpert, D.M. Contextual Inference in Learning and Memory. *Trends Cogn. Sci.* **2023**, *27*, 43–64. [CrossRef] [PubMed]
36. Han, J.; Liu, G.; Zheng, Q. Prior Knowledge as a Moderator Between Signaling and Learning Performance in Immersive Virtual Reality Laboratories. *Front. Psychol.* **2023**, *14*, 1118174. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.