Information Research - Vol. 31 No. iConf (2026)

Immersive quality control for 3D data curation

Matt Cook, Hannah Hok Kim, Helen He, and Zack Lischer-Katz

DOI: https://doi.org/10.47989/ir31iConf64136

Abstract

Introduction. We report quantitative findings from a multi-site experiment evaluating the impact of immersive viewing (XR) on routine quality control (QC) tasks in academic labs regularly producing 3D content across the United States.

Method. 29 participants were recruited to complete orientation, error-detection, and comparison tasks under two viewing conditions: Meta Quest 3 (XR) and iPad tablet computer (control). Time-to-completion (TTC) and accuracy performance were measured. Post-condition NASA-TLX and SUS captured perceived cognitive load and system usability.

Analysis. We ran one-tailed within-subjects t-tests for TTC, TLX, and SUS with Holm–Bonferroni correction; accuracy was modeled via logistic regression (participant as random effect); Cohen's *dz* reported.

Results. XR demonstrated faster TTC (M = 11.90 s) than tablet (M = 20.42 s; Holm-corrected p = .0129; dz = .56) and lower cognitive load (TLX) (14.6 vs. 18.92; Holm-corrected p = .026; dz = .46). Accuracy (.76 vs .77) and usability (SUS) (83.9 vs. 80.2) did not differ significantly across conditions.

Conclusion. Immersive review of scholarly 3D assets shows promise as a tool for supporting 3D data curation workflows. Using the XR platform was found to reduce time and perceived effort for 3D creators on QC tasks without impacting accuracy or usability.

Introduction

Scholarly 3D data are regularly modelled (i.e., manually crafted using software such as Blender), procedurally generated (e.g., with AI), and *captured* or *scanned* (using photogrammetry, LiDAR, computed tomography, etc.) through established production pipelines. The end users of 3D data require digital fidelity across physical scale, volume, geometry, color, texture, and any other visuospatial dimensions whose real-world counterparts may be subject to metrological analysis in the field, the laboratory, or the museum (Cook 2024). The resulting scholarly assets are valuable insofar as they facilitate comprehension, expand access, and enable previously impossible analyses that might have otherwise damaged the original or endangered personnel (Blackburn et al., 2024; Boyer, 2016; Hawks & Williams, 1986; Kersten-Oertel et al., 2013; Kluge et al., 2019; Limp et al., 2011; Pfarr-Harfst, 2016; Silvestri et al., 2010).

In 2023, a collaborative team from the University of Arizona and Harvard University began work on a three-year research project funded in part by the Institute for Museum and Library Services (IMLS) to study and improve current 3D data curation efforts at a range of academic institutions across the United States. As of 2025, the investigators have visited 3D production sites across six US states, interviewing a total of 53 3D data creators who represent a range of academic disciplinary practices, and questioning them as to how they constructed *quality* with regards to 3D scanning outputs. In addition to exploring research questions (RQs) related to the construction of 3D data quality (RQ1: '*How do researchers evaluate and document the quality of research data throughout their creation workflows and the 3D data lifecycle?*'), the research team also solicited responses from participants about their downstream uses (e.g., research and instructional) of 3D data (RQ2: '*How do researchers use visual and immersive technologies to analyze their 3D data?*'), and the level of local technical and administrative support they were receiving—and what resources they felt were still lacking—in their overall efforts to ensure data quality (RQ3: '*How can institutions and curation professionals best support 3D data creation, analysis, and curation workflows?*'). The project is also exploring how 3D data creators can use immersive media in their workflows (RQ4: '*How can emerging technologies be used to support 3D data analysis, curation, access, and reuse?*'), which will be the focus of this paper.

This paper reports on the quantitative findings from a quasi-experiment designed to address RQ4 (qualitative data related to RQs 1-3 will be reported on in future publications). While on site, grant investigators selected a subset of participants to evaluate the usefulness of immersive viewing technology (virtual and augmented reality, or XR) to facilitate 3D data curation workflows via the implementation of standardized quality control (QC) protocols. 3D data practitioners were tasked with completing routine QC tasks across two viewing conditions: traditional *flat* media viewing, on a tablet computer, and immersive XR viewing using a Meta Quest 3 head mounted display (HMD or simply '*headset*'). Based on analysis of the combined performance, usability, and cognitive load assessment methods, investigators have documented statistically significant decreases in both time-to-completion (TTC) and cognitive load when participants engage with 3D scans in an immersive viewing environment vs. a 2D viewing device.

Findings suggest the potential benefits of XR for 3D data curation more broadly across fields. Despite specific challenges related to the portable nature of the experiment, the choice of control device, and task logging protocols (see Limitations section), investigators are confident that—with regard to certain spatially relevant task types (e.g., model completeness, resolution, feature detection, comparison, etc.)—3D imaging facilities may benefit from the deployment of XR workstations for ensuring the success (or identifying the failure) of their 3D model outputs. This conclusion and associated protocols advance the project agenda by providing a standard means to ensure 3D model quality in the future while answering affirmatively the question that has guided the investigators from the outset: *Can we help make 3D data citable*?

Related work

XR in the academy

Under discipline-agnostic controlled conditions, researchers in computer science (CS) and related fields (e.g., human-computer interaction [HCI]) have isolated and tested the low-level cognitive affordances associated with stereoscopy (i.e., binocular vision), head tracking, and increased field-of-view (relative to traditional computer monitors), the combination of which define modern XR headsets (Andrews et al, 2010; Batch et al., 2019; Bowman, & McMahan, 2007; Laha et al., 2014; Yang et al., 2021). By preserving everyday perceptual cues related to depth and motion, XR enables participants to use their bodies as part of the analytic process, yielding gains in feature detection, search, and scale judgment (Cutting & Vishton, 1995; Forsberg et al., 2008; Lages & Bowman, 2018; Ragan et al., 2012). In practice, headset-based exploration of digital 3D content mirrors how experts inspect physical specimens in the field (Buckland, 1991; Donalek et al., 2014; Gibson, 2002) and extends these benefits to abstract data, such as scatterplots, text corpora, text embeddings, and guidance maps (Harada & Ohyama, 2020; Lisle et al., 2020; Prouzeau et al., 2019; Whitlock et al., 2020; Yang et al., 2021).

Outside of the laboratory, XR technologies represent an increasingly standard component of research and instruction within the design disciplines (e.g., architecture), medicine, and a handful of physical and applied sciences, including geology, meteorology, and engineering (Giacalone et al., 2019; Helbig et al., 2014; Kuliga et al., 2015; Milovanovic et al., 2017; Mörth et al., 2024; Portman et al., 2015; Safi et al., 2019; Seth et al., 2011; Uhr et al., 2019; Zhao et al., 2019). In these and other studies, researchers have documented improvements associated with communication (between experts and between experts and non-experts), scale perception, feature detection, and have even demonstrated the value of XR as a means to facilitate *new discoveries* within previously overwhelming data sets (Acevedo et al., 2001; Amar et al., 2005; Kingsley et al., 2019; Madura et al., 2015; Van Dam et al., 2002). This literature demonstrates specific performance improvements for spatially relevant task types when using XR.

Beyond laboratory and disciplinary studies, research in instructional design shows that XR can support direct, in-context engagement with complex 3D materials, improving spatial understanding, motivation, and task performance in authentic settings (Cook & Lischer-Katz, 2019; Dede et al., 2017; Jang et al., 2017; Janiszewski et al., 2020; Mills, 2020; Pomerantz, 2018; Trelease & Rosset, 2008; Qin et al., 2021). Meta-reviews similarly report learning benefits when immersive environments let learners inspect at true or near-true scale, alternate fluidly between overview and detail, and compare versions of the same object or space (Makransky & Petersen, 2021; Merchant et al., 2014; Mikropoulos & Natsis, 2011; Pellas et al., 2021; Radianti et al., 2020). These same affordances map directly onto routine curatorial QC judgment tasks, including checking completeness (holes, occluded regions), fidelity (textures, inscriptions), '*scale sanity*' checks (quickly assessing realistic size of 3D models), and version selection (Berger et al., 2017). QC tasks, similar to learning activities, benefit from embodied, walk-around inspection and rapid comparisons in context. Taken together, classroom implementations and syntheses indicate that immersive workstations can standardize and accelerate QC for scholarly 3D assets.

3D data standards

As a platform designed for viewing 3D digital assets, XR technologies are well suited for supporting the 3D data creation process and furthering standards-based practices. The adoption of scientifically accurate 3D data in the form of 3D meshes, CT scans, and other data types, has necessitated the development of new standards and specifications by the wider digital preservation community, including academic libraries, whose technologically savvy users expect access to knowledge in all forms, including 3D material collections (Greene & Groenendyk, 2021; Hall et al., 2019; Lischer-Katz et al., 2019; Lischer-Katz & Cook, 2022). Standards groups, including the Web3D Consortium and the International Image Interoperability Framework (IIIF) 3D subgroup have developed specifications to ensure the interoperability of 3D digital objects, while academic library-based initiatives, such as *Community Standards for 3D Data Preservation* (CS3DP), have synthesized 3D creation knowledge from diverse stakeholder groups and published metadata and preservation standards that ensure 3D data follow FAIR (Findable, Accessible, Interoperable, and Reusable) principles, across disciplines, regardless of their source (Grayburn et al., 2019; Hardesty et al., 2020; Moore et al., 2002; Scheffler et al., 2022; Wilkinson et al., 2016).

Because 3D data creation is a multistep process with each step necessitating decision-making that contributes to the quality of the final 3D output (Publications Office of the European Union [OP], 2022), there have been ongoing concerns about documenting 3D *paradata*, or data that describe the processes of creation and interpretation that shape the resulting quality of 3D scholarly outputs (Huvila, 2022). The *London Charter*, first outlined by the European cultural heritage visualization community in 2006, has laid out principles for documenting 3D data creation while '*ensuring the methodological rigour of computer-based visualization*' (Denard, 2012, p. 1). Documenting paradata is essential to ensuring the future usability of 3D data, since downstream users will want to know how 3D outputs were created before they interpret or reuse them. The European Commission has surveyed 3D capture methods, concluding that however meticulously 3D data are gathered

> complexity does not reside in the geometry of a 3D model or the final number of points and vertices, but derives from the stakeholder requirements, its location and state of condition. Also highly relevant are the set-up of data acquisition, know-how of the operators in place and the integration of multiple datasets from different devices into one archive. (OP, 2022, p. 24)

There is clearly a growing need for '*efficient archival systems able to provide effective search and retrieval functionalities*' (Alliez et al., 2017, p. 10), and metadata and paradata are important components of those systems. Furthermore, museum curators and natural scientists are increasingly leveraging these standards to support the dissemination and analysis of morphological collections at scale as part of federally funded imaging initiatives, and individual scanning outputs are freely displayed for the benefits of remote users on 3D-specific data hosting platforms, such as Morphosource (Blackburn et al., 2024; Ijiri et al., 2018; Lemenager et al., 2023). General data management tools and repositories, such as Open Science Framework and Dataverse could also support 3D curation efforts (Foster & Deardorff, 2017; King, 2007; Sullivan et al., 2019). Research on the benefits of VR found in CS, HCI, and information studies, in conjunction with these three strands of prior development work on 3D data standards (community preservation standards, paradata and documentation standards, and 3D repositories) point directly to the potential for curatorial gains from immersive QC workflows.

First, controlled studies show that stereo depth, head-tracked motion parallax, and wide field-of-view let people use their bodies as part of spatial reasoning, improving feature detection, scale judgment, and search efficiency relative to conventional displays (e.g., Amar et al., 2005; Andrews et al., 2010; Bowman & McMahan, 2007; Cutting & Vishton, 1995; Laha et al., 2014; Lisle et al., 2020). Second, deployments in design, medicine, and the physical sciences report communication and discovery benefits that hinge on rapid, shared inspection of 3D forms at true or near-true scale (e.g., Helbig et al., 2014; Madura et al., 2015; Milovanovic et al., 2017; Portman et al., 2015). Third, research on 3D digital preservation standards and stewardship has demonstrated researcher expectations for fidelity, transparency, and reuse (e.g., London Charter; European Commission guidelines), and a need for fast, repeatable screening for completeness, artifacting, texture integrity, and version selection (Denard, 2012; OP, 2022). Given that XR reproduces the embodied viewpoint that curators use when viewing physical objects, it makes sense that similar QC judgments of 3D digital objects could be improved by leveraging XR's affordances.

Methods

Given the abovementioned benefits of XR and 3D, the investigators sought to test how immersive viewing might impart a relative advantage for practitioners who evaluate the quality of 3D data in their own production facilities on a daily basis. This involved selecting generalizable (i.e., cross-disciplinary) task types associated with documented immersive affordances, like feature detection and comparison, and identifying assessment mechanisms that would allow for cross platform comparison of performance by study participants, all of whom regularly engage with 3D data as a regular part of their professional life. To that end, we proposed four hypotheses, each aligned with an assessment variable:

- H1: XR facilitated 3D QC tasks will have quicker completion times (*efficiency*).
- H2: XR users will make fewer mistakes in 3D QC tasks (*accuracy*).
- H3: XR use will reduce users' cognitive load when completing 3D QC tasks (*cognitive load*).
- H4: XR will be perceived as more usable by users (*usability*).

To test these hypotheses, we designed a within-subject experiment that presented study participants (n = 29) with hypothetical QC tasks that they might encounter if they were actively engaging in curatorial work with 3D scan data within an academic library, museum, or other research context. Each participant completed three tasks using the XR system (Meta Quest 3) and three tasks using the control condition, a touchscreen tablet (iPad). Task sequences were randomly ordered to mitigate order effects. Testing for each condition commenced with a practice task, whose data were excluded from analysis, to familiarize participants with the technology (see Apparatus section, below). Then participants would complete tasks in either XR then control, or control then XR, depending on the randomly assigned sequence. The QC tasks themselves fell broadly into three categories—orientation tasks, error detection tasks, and comparison tasks—all of which correspond to generic task types that have been shown in the literature to benefit from immersive engagement (see Appendix I: Task prompts, below).

Site selection and participant recruitment

For the broader project, the unit of analysis is the 3D creation *lab*, which typically consists of small groups of researchers and technical staff (technicians, librarians, graduate students, etc.) working to create 3D research data. Participants in these labs may support single long-term projects or work on multiple projects. Criteria for selecting sites and study participants include: 1) sites were currently creating 3D research data, 2) sites were practically accessible to the researchers, and 3) site managers were comfortable with having researchers observe their spaces. In order to collect data from a diversity of approaches, labs were recruited from the Northeast, Southeast, Central Plains, and Southwest regions of the US. Labs were selected through the investigators' personal 3D creation contacts in the field, and snowball sampling was used to expand the range of sites recruited (i.e., participants identified other potential participants at their institutions and social networks who were carrying out 3D creation projects).

Because the findings of this study are intended to serve multiple disciplines and 3D creation techniques, site selection was carried out to include a range of disciplines (natural sciences, social sciences, and humanities), scales of production (small and large objects being scanned) and modes of 3D creation (photogrammetry, structured light scanning, computed tomography [CT] scan, and LiDAR 3D creation techniques) in the study (Muenster 2022). This approach also supports cross-case analysis of the data, which supports the identification of the common attributes of all workflows and points of divergence and idiosyncrasies.

Participants were recruited through lab leaders (PIs or managers of labs) by circulating a recruitment email to lab members. 53 participants were recruited for the overall study, while a subset of 29 participants (55%) was selected from this pool for the RQ4 experimental part of the study. In January 2024, prior to participant recruitment and data collection, investigators obtained approval from the lead institution's Institutional Review Board (IRB) for all protocols and instruments.

Apparatus

Campfire3D, a multi-platform, multi-user 3D data visualization software, was deployed as the viewing environment to support RQ4 data gathering activities. With Campfire, test administrators were able to remotely trigger visual scenes representing specific tasks from a laptop computer in a predetermined order, embed task prompts for consistent recitation by administrators, and monitor and record participant progress through the onscreen appearance of ghost-like, transparent avatars: a floating head in the case of the XR participant, a rectangular *screen* avatar in the case of the control condition (see Figure 1, above). The intervention (XR) device was a Meta Quest 3, with *pass through* (mixed reality) rendering capabilities to allow users to view the scholarly 3D models in the context of the local environment, minimizing disorientation. The control condition device, a late generation iPad, was chosen to streamline the multi-site, multi-participant protocol by eliminating the need for peripherals (e.g., mouse, keyboard), at the cost of retaining some immersive quality (i.e., *presence*), which could cause confounding results by mitigating disparity between cross-condition performance (Witmer & Singer, 1998).

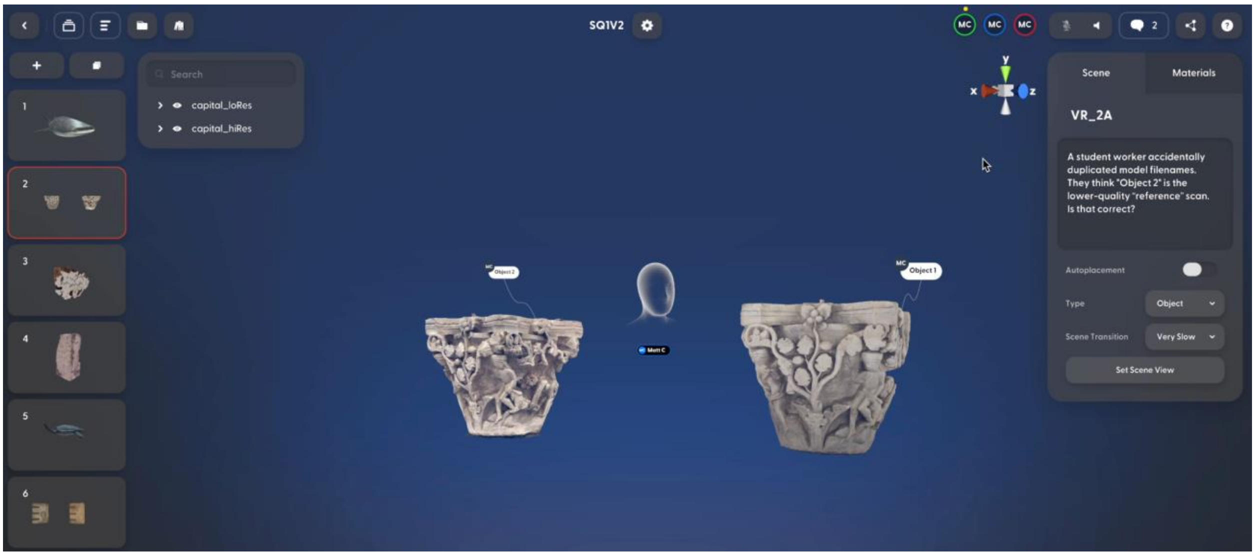

Figure 1. Participant completing 3D scan comparison task in XR using Campfire3D software.

The 3D objects featured in the QC task series represented rare books, mineralogical specimens, architectural motifs, and live animal specimens, all produced by professionals in 3D scanning labs in the Northeast United States. Every object was imported into Campfire3D as a GLB *mesh* file weighing approximately 10MB each, including embedded 4k textures. In the intervention (XR) condition, content was posed for participants at a normalized scale, filling the center of the testing space (a minimum of 6.5ft x 6.5ft), to allow for *room scale* XR. Neither the control nor the intervention conditions required the use of peripherals (e.g., Meta Quest controllers), a decision that reduced the need for pre-test training and minimizes the effect of conflating factors associated with prior (videogame or XR) experience (Richardson et al., 2011).

Participants were instructed to verbally respond "Yes' or 'No' *only*' to each of the six task prompts (including two practice tasks, not included in data collection; one per condition), which were evenly divided across the two conditions. The two dimensions of performance data gathered were time-to-completion (TTC) and response accuracy. Immediately after each experimental condition, participants were asked to answer questions concerning usability and cognitive load, using the System Usability Scale (SUS) for assessing usability and the NASA Task Load Index (NASA-TLX) for measuring cognitive load.

After QC performance testing, a post-test questionnaire was deployed to gather qualitative interview data concerning participants' views on emerging technologies generally and their role in participants' labs. These open-ended questions concerned participants' previous XR experience, challenges associated with local adoption of emerging technologies, and suggestions for emerging technologies that might be useful for supporting 3D creation. Preliminary review of the post-test interview data suggests a trend toward 3D data practitioners considering ways in which emerging Artificial Intelligence (AI) technologies might enhance work in their diverse laboratory settings.

Instruments

Despite allowing for multi-platform synchronization of high-definition 3D content, Campfire3D did not have an *in-app* task response functionality at the time the experiment was conducted, necessitating vocal responses from participants to task prompts. Therefore, time stamped chronology was referenced from a simultaneous screen recording to minimize the influence that an administrator might have on the performance data.

The participants' responses were logged manually by the test administrator during the experiment in an associated Qualtrics survey (see Appendix I, below). Responses were programmatically graded post-experiment against answer keys using a Python script that accounts for the randomized task sequence variation.

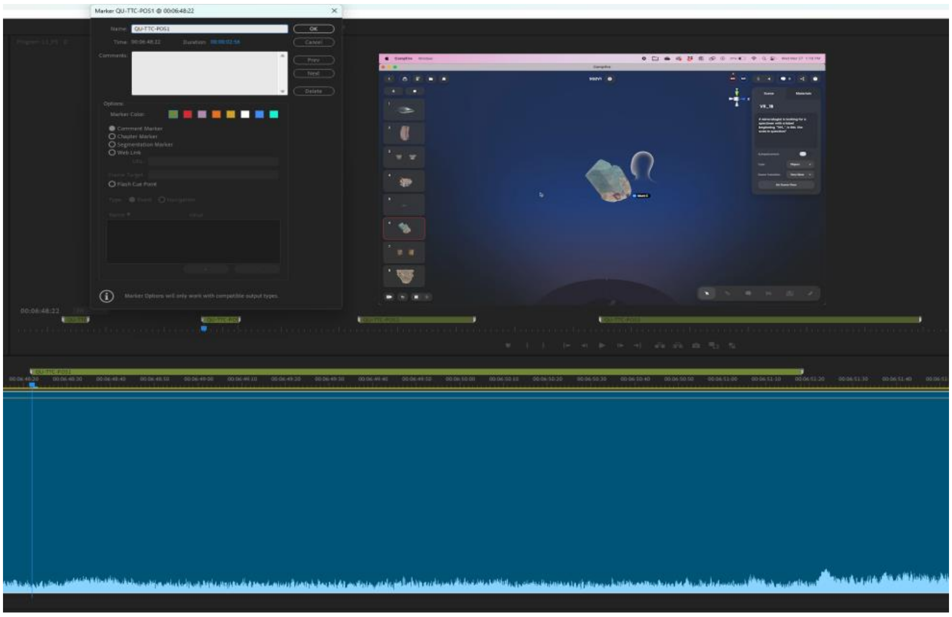

Figure 2. Audio waveform analysis of time-to-completion (TTC) data in Adobe Premiere Pro

TTC was computed post-experiment from simultaneous screen recordings of the experiments. It was measured, with millisecond precision, from the end of the utterance of the last word of the administrator's standardized question to the onset of the utterance of the participant's response word, a process we name '*trough-to-trough recording*' as shown on the sound waveform visualization. In cases where participants requested repetition or clarification of a question, timing was consistently measured from the final iteration of the question to ensure comparability across participants. To extract these intervals, we adopted a semi-automated workflow with video editing software, Adobe Premiere Pro, where frame-accurate timestamps were manually placed by investigators with a visual aid of sound waveforms, durations between timestamps were calculated automatically, and the resulting frame counts were converted to milliseconds using the formula (frames/frame rate)×1000.

Both usability (SUS) and cognitive load measures (NASA-TLX) were programmatically computed from the participants' self-reported survey data. SUS was originally designed as a one-shot tool for collecting usability data on a particular technology and then comparing it to usability scores on other technologies (Bangor et al., 2008; Brooke, 1996). The survey consists of ten statements (five positive and five negative), each rated on a 5-point Likert scale. Following established scoring procedures, raw scores were converted to a 0-100 scale. Higher scores indicate greater perceived usability.

NASA-TLX has been deployed widely, across disciplines and technologies, to measure cognitive load, or the '*mental effort required to process information*' (Wei, pg. 1) (Hart, 2006; Tracy & Albers, 2006; Wei et al., 2025). This includes studies with XR users (Pietschmann et al., 2023). It consists of six subscales: mental, physical, demand, frustration, effort, and performance. Participants were asked to rate each of the variables on a 20-point scale that measures from high to low load. For this study, we calculated the unweighted average of the ratings of the six items for each condition for each participant, converted to a 0-100 scale. Lower scores indicate less cognitive load.

Data analysis and results

We tested four directional hypotheses comparing user performance and experience across devices used. For continuous measures (time, TLX, and SUS), we conducted one-tailed within-subjects t-tests which are reasonably robust to departures from normality given our sample sizes. Conclusions remained unchanged when results were confirmed with Wilcoxon signed-rank tests. Accuracy was binary at the trial level (three items per condition) and analysed with a logistic regression model with accuracy (correct/incorrect) as the outcome variable, condition as the predictor variable, and with participant as a random factor. We additionally applied Holm-Bonferroni correction to account for multiple comparisons across the four outcomes. Effect sizes were calculated through Cohen's d_z.

We hypothesized that tasks completed using the XR platform would be completed more quickly (H1); with fewer mistakes (H2); with reduced cognitive load (H3); and with higher perceived usability (H4) compared to the control (iPad). Figure 3 visualizes the findings supporting each hypothesis. H1 and H3 were supported by the data, while H2 and H4 were not supported by the data.

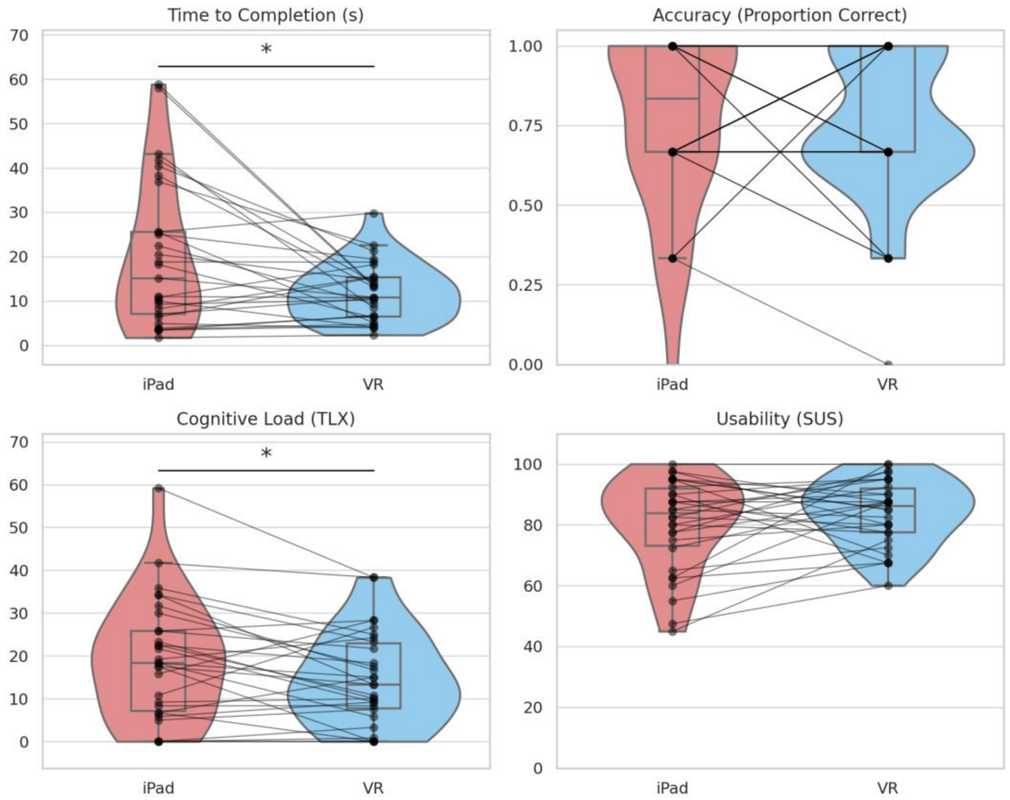

Figure 3. Paired violin boxplots showing task performance across devices. The width indicates density of data, and the length shows the range of the data. Black lines connect each participant's paired score across conditions. Significance markers: $p < .05$ *, $p < .01$ **, $p < .001$ ***.

H1: XR facilitated 3D QC tasks will have quicker completion times (efficiency).

H1 was supported by the data. Participants completed tasks faster using XR (M = 11.90 SD = 6.65) than the iPad (M = 20.42, SD = 16.82), t(27) = -2.95, one-tailed $p = .003$, Holm-corrected $p = .0129$, two-tailed 95% CI = [-14.8, -2.2]. The within-subjects effect size was medium (Cohen's $d_z = -.56$). This result suggests that participants were more efficient at completing QC tasks using XR than when using the iPad control.

H2: XR users will make fewer mistakes in 3D QC tasks (accuracy).

H2 was not supported by the data. There was no significant difference in total accuracy between the XR trials (M = .76, SD = .21) and iPad trials (M = .77, SD = .28). The odds of a correct response did not differ between conditions (OR = 1.06, 95% CI [.52, 2.08], $p = .86$, Holm-corrected $p = 86$).

H3: XR use will reduce users' cognitive load on 3D QC tasks (cognitive load).

H3 was supported by the data. Participants reported significantly lower cognitive load while using XR (M = 14.6, SD = 10.92) than on the iPad (M = 18.92, SD = 14.03), t(29) = -2.53, one-tailed $p = .0085$, Holm-corrected, p = .026, two-tailed 95% CI [-7.80, -.83]. The within-subjects effect size was medium (Cohen's $d_z = -.46$). This result suggests that immersive environments may reduce perceived effort during QC tasks compared to using touchscreens (e.g. iPad).

H4: XR will be perceived as more usable by users (usability).

H4 was not supported by the data. There was no significant difference in perceived usability between XR (M = 83.92, SD = 10.80) and the iPad (M = 80.17, SD = 15.24), t(29) = 1.41, one-tailed $p = .085$, Holm-corrected $p = .170$, two-tailed 95% CI [-1.7, 9.20].

Discussion

The results of this exploratory study provide support for the use of immersive technologies in 3D data curation QC workflows. By revisiting each hypothesis considering our findings, we can situate the contributions of this work within the broader literature on immersive analytics and scholarly 3D data curation. These findings support the hypothesis that immersive viewing reduces task completion times (H1). Participants completed QC tasks significantly faster in the XR condition than on the tablet, suggesting that the embodied affordances of XR facilitate rapid evaluation of spatial data. These results echo prior research in immersive analytics and visualization, which has consistently demonstrated speed gains in tasks involving spatial orientation and comparison. Importantly, these gains occurred despite participants' varied backgrounds and geographical locations, indicating that immersive QC workflows may be broadly accessible, even to practitioners with limited prior XR experience, and across institutions.

Contrary to our expectations, no significant accuracy advantage was observed in the immersive condition (H2). Participants made comparable numbers of errors across both platforms. This finding suggests that while immersive technologies improve efficiency, they may not inherently increase correctness of judgment in QC tasks. However, the relatively high quality of the test models and the simplicity of the binary (yes/no) task design may have minimized opportunities for detectable accuracy differences. Future studies may need to implement more complex or ambiguous QC scenarios—such as detecting subtle texture loss, surface noise, or visual artifacts—to evaluate whether XR can aid in additional varieties of error detection. Incorporating a greater number of tasks with increasing difficulty could provide better data on the influence of XR on task accuracy.

Participants reported lower perceived cognitive load when performing QC tasks in XR than on the tablet (H3). This aligns with evidence from cognitive psychology and human–computer interaction (HCI) that immersive environments reduce the mental transformations required to interpret 3D data on flat displays. In other words, immersive displays reduce the *translation cost* between digital surrogates and real-world embodied perception, freeing cognitive resources for analytic decision-making. These results strengthen the case for XR as a means to reduce fatigue and improve focus in professional 3D data curation workflows, particularly for repetitive or detail-intensive tasks and as 3D services continue to scale up to produce more 3D outputs for a wider variety of scholarly applications.

Perceived usability did not significantly differ between the XR and tablet conditions (H4). This is somewhat surprising given that immersive technologies are often criticized for steep learning curves or discomfort. In this case, the lack of significant usability differences may reflect the decision to not use hand controllers, thereby reducing training overhead and streamlining the experience for all participants (not just the ones familiar with using game controllers). This result also suggests that immersive systems can be at least as usable as familiar 2D platforms for simple QC tasks. This finding is consistent with recent longitudinal studies showing that users quickly adapt to XR for professional tasks (Biener et al., 2022).

Taken together, these results indicate that immersive QC workflows hold promise for addressing a persistent challenge in scholarly 3D data curation: establishing standardized, efficient, and reproducible mechanisms for assessing model quality. By lowering cognitive load and reducing TTC without sacrificing accuracy or usability, immersive QC could provide 3D data producers—particularly those in resource-constrained academic labs—with a practical means to ensure fidelity of their 3D outputs. This finding directly advances the project's broader goal of improving 3D data citability by providing evidence-based guidance for 3D data practitioners.

Through the efficiency and cognitive load improvements described here, immersive QC workflows help 3D data adhere to the FAIR principles (Wilkinson et al., 2016). Standardized immersive workflows offer a reproducible mechanism for verifying model fidelity prior to dissemination. When coupled with established metadata and preservation frameworks (Grayburn et al., 2019; Hardesty et al., 2020; Moore et al., 2022), immersive QC practices support quality assessments that are transparent and function well across academic groups. This convergence of immersive analytics and digital curation represents an important advance toward making 3D data not only technically robust but also epistemically trustworthy, ensuring that curated models can support long-term scholarly reuse and integration with global research agendas.

Limitations

Data gathering activities associated with this immersive QC experiment were conducted onsite, at study participants' home institutions, which introduced variability related to issues of privacy, ambient lighting, physical space, network connectivity (required for Campfire3D usage), and time. Participants were regularly shuttled between multiple investigators, one of whom might be conducting qualitative data gathering (i.e., interviews) in another part of the same room where users were engaged in the XR QC tasks. These variations represent some amount of experimental inconsistency and present an opportunity for future studies to improve study reproducibility. Using the same testing facility for all participants would address this challenge but would limit participant recruitment to those 3D creators who lived nearby.

These challenges with onsite data gathering impacted performance tracking. All experimental data was gathered via Qualtrics except data related to task TTC. As investigators relied on laptop recordings of verbal responses from study participants to measure TTC, audio quality varied significantly by location. Because timestamp logging was performed manually, a small degree of variation may have been introduced.

Investigators also debated the choice of control condition device before finally settling on a tablet (iPad). This device type, while succeeding in presenting media to participants via a *flat* display that clearly contrasted with XR, still retained immersive characteristics which may confound results insofar as the user is able to engage with the 3D targets via hand gestures (e.g., pinch, swipe, drag). The touch-screen capabilities thereby represent some level of embodied interactivity, a defining characteristic of immersive XR, unlike a mouse and keyboard driven interaction experience of a typical desktop computer. Relatedly, investigators chose to design the immersive QC experiment in a way that precluded the use of XR peripherals (e.g., hand controllers). This was both a practical consideration, as additional training would be necessary to prepare the participant for engagement with the target content via controller, and to encourage body-centered interaction with the task objects.

Finally, there was enough variation across target 3D models to warrant further consideration and refinement in future studies. While efforts were made to standardize model resolution and texture size, objects varied geometrically, given their disparate origins (e.g., zoological vs. architectural content). Object orientation relative to the user was similarly inconsistent, with each site of study requiring a fresh spatial mapping for headset configuration. Future studies will benefit from a reduction in model variation—perhaps limiting model selection to a single disciplinary type (e.g., mineralogical models)—and a more considered mapping of task type to target based on geometric, textural, or other attributes.

Future work

Beyond implementing better control over local environment conditions, future research could make use of more rigorous data gathering methodologies, including physiologically grounded cognitive load testing and performance logging schemes that carefully track participant movement in 3D space as they complete QC tasks. Despite the fact that subjective, self-reported measures have been shown to be '*very successful in identifying differences in cognitive load*' (Ayers, 2021, pg.

1), researchers have also concluded that non-invasive, headset mountable physiological measurements like '*pupil dilation may be a valuable measure of mental demands*' (Ayers, 2021, pg. 4) and, more recently, that machine learning analyses of such data may even allow for *realtime measurement* of mental effort (Ayres et al., 2021; Skulmowski & Rey, 2017; Tahmid et al., 2023; Wei et al., 2025).

TTC measurement could also be automated with AI speech recognition models to reduce processing time, save manual labor, and improve precision, especially for studies with large numbers of tasks/participants. Investigators prototyped an automated workflow using Whisper, an open-source speech recognition model published by OpenAI. The workflow transcribes experiment audio recordings with millisecond-level timestamps, extracts the timestamps for target words, and allows investigators to review in a user interface. While not deployed in this study, such an approach shows promise for reducing manual effort and enhancing reproducibility and scalability. Future studies could build on this prototype workflow to fully automate TTC measurement based on verbal responses or integrate more efficient in-app response widgets that do not require verbal reporting.

In the case of advanced XR coding schemes, Wang (et al., 2024) studied 254 students using a '*bottom-up analysis pipeline*' that categorized participant body movements as discrete '*design behaviors (i.e., the amount of horizontal plane movement, frequency of manipulation, choice of tool used) …*', which researchers systematically linked to user output characteristics (pg. 25). In contrast to the current study's reliance on '*manual labeling of verbal protocols and video recordings*', those investigators' approach used XR tracking data that did not require investigator recall or '*depend on what researchers pick[ed] up from video recordings*' (Wang et al., 2024, pg. 3). This granular data gathering and participant output analysis methodology represents a promising direction for future research on immersive QC.

Future work in this area may also reveal emergent constructs that only become apparent to investigators at the data analysis stage. For example, screen recordings of study participants' movements demonstrate fluidity or smoothness of motion in the immersive condition that is lacking in the control condition. Presumably, this relates to the inherent continuity associated with moving one's physical body through space to gain a novel viewpoint. Movement in the touch-screen environment is instantaneous or near instantaneous, requiring only a fast finger motion to move around a model onscreen and resulting in users bypassing potentially useful intermediary target object perspectives. Conversely, traversing a room-scale XR experience to engage with a 3D model may require several footsteps for the users, constituting continuous movement in physical space that mentally constructs the target object.

There are also practical implications of the improvements we see when users engage with 3D outputs using XR. Taking the measurable benefits associated with the abovementioned fluidity of motion as justification, 3D practitioners, including early career or discipline agnostic imaging specialists, might improve existing 3D production throughput via the use of QC protocols that take place in XR. Given the relative low cost and high rendering capabilities of current headset hardware and the availability of open-ended viewing utilities, like Campfire3D, it now makes sense for labs to dedicate space, training, and staff time to formally integrate QC steps in their production processes.

As discussed earlier, this paper addresses the findings from Research Question 4, a part of the broader project that has also collected qualitative data in the form of interviews, documents, and observations from 3D lab sites. Once complete, the analysis of the qualitative data will generate new understanding of 3D data creation and curation practices that will inform how information institutions and digital curators can better support these workflows (RQs 1-3). The findings from this paper already suggest that information institutions can better support 3D data creators by providing XR immersive viewing tools to improve 3D data creation workflows and QC tasks. Academic libraries, in particular, should take notice of the benefits of XR as a visualization platform useful for viewing and analyzing 3D data by researchers in a growing range of disciplines.

Conclusion

This multi-site study provides compelling evidence that immersive QC workflows can make routine curation of scholarly 3D data more efficient and less mentally taxing, without decreasing usability or accuracy. Across our within-subjects design, participants completed QC tasks significantly faster in XR than on a tablet and reported significantly lower cognitive load; accuracy and perceived usability did not differ by condition. Put simply, when the task is inherently spatial—orienting to complex geometry, judging completeness, or detecting features under occlusion, etc.—the embodied affordances of head-tracked, stereoscopic viewing help 3D data creators and users.

In sum, our study demonstrates that the embodied interactions XR affords can be strategically leveraged to standardize and accelerate QC steps that support scholarly validity of 3D data. Immersive QC yields a practical, discipline-agnostic pathway toward FAIR 3D objects. By moving quality assessment closer to how experts naturally interrogate physical artifacts, we can help close the gap between the capture and downstream citation of scholarly 3D assets.

Acknowledgements

This project was funded in part by a grant from the Institute for Museum and Library Services (IMLS), through the National Leadership Grants for Libraries Program, Grant# LG-254830-OLS-23.

About the authors

Matt Cook is Digital Scholarship Program Manager at Harvard Library. He was trained in philosophy of mind and has published R&D in multiple fields, including architecture, chemistry, information science, and a range of humanities.

Hannah Hok Kim is a Postdoctoral Researcher at the Brain and Cognitive Sciences department at MIT and the Department of Psychology at Harvard University studying social cognitive development. She also works with the Center for Research on Equitable and Open Scholarship (CREOS) at MIT Libraries.

Helen He is an undergraduate researcher at Harvard University and a Rhodes Scholar, jointly concentrating in Computer Science and East Asian Studies with a secondary in Classics. Her research centers on using 3D computer vision to reimagine and reconstruct cultural heritage.

Zack Lischer-Katz is currently an assistant professor at the College of Information Science, University of Arizona, where he co-founded the co/lab: Critical Archives & Curation Collaborative. His research explores visual information practices, media preservation, and digital curation of emerging media formats, including 3D, virtual reality, and volumetric video.

References

Acevedo, D., Vote, E., Laidlaw, D. H., & Joukowsky, M. S. (2001). Archaeological data visualization in VR: Analysis of lamp finds at the Great Temple of Petra, a case study. *Proceedings Visualization*. VIS'01., 493-597. https://doi.org/10.1109/VISUAL.2001.964560

Alliez, P., Bergerot, L., Bernard, J. F., Boust, C., Bruseker, G., Carboni, N., Chayani, M., Dellepiane, M., Dell'unto, N., Dutailly, B. and Gautier, H., Guidi, G., Guillem, A., Joffres, A., Laroche, F., Manuel, A., Manzetti, M. C., Michel, A. P., Pamart, A., … Tournon-Valiente, S. (2017). Digital 3D objects in art and humanities: Challenges of creation, interoperability and preservation. *White Paper*, 71 [Report, European Commission ; Horizon H2020 Projects]. https://inria.hal.science/hal-01526713

Amar, R., Eagan, J., & Stasko, J. (2005). Low-level components of analytic activity in information visualization. *IEEE Symposium on Information Visualization, INFOVIS 2005*, 111-117. https://doi.org/10.1109/INFOVIS.2005

Andersen, S. A. W., Mikkelsen, P. T., Konge, L., Cayé-Thomasen, P., & Sørensen, M. S. (2016). Cognitive load in mastoidectomy skills training: Virtual reality simulation and traditional dissection compared. *Journal of Surgical Education*, 73(1), 45-50. https://doi.org/10.1016/j.jsurg.2015.09.010

Andrews, C., Endert, A., & North, C. (2010). Space to think: Large high-resolution displays for sensemaking. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 55–64. https://doi.org/10.1145/1753326.1753336

Ayres, P., Lee, J.Y., Paas, F., & Van Merrienboer, J.J.G. (2021). The validity of physiological measures to identify differences in intrinsic cognitive load. *Frontiers in Psychology*, 12. https://doi.org/10.3389/fpsyg.2021.702538

Bangor, A., Kortum, P.T., & Miller, J.T. (2008). An empirical evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(6), 574–594. https://doi.org/10.1080/10447310802205776

Batch, A., Cunningham, A., Cordeil, M., Elmqvist, N., Dwyer, T., Thomas, B. H., & Marriott, K. (2019). There is no spoon: Evaluating performance, space use, and presence with expert domain users in immersive analytics. *IEEE Transactions on Visualization and Computer Graphics*, 26(1), 536-546. https://doi.org/10.1109/TVCG.2019.2934803

Berger, M., Tagliasacchi, A., Seversky, L. M., Alliez, P., Guennebaud, G., Levine, J. A., Sharf, A., & Silva, C. T. (2017). A survey of surface reconstruction from point clouds. *Computer Graphics Forum*, 36(1), 301-329. https://doi.org/10.1111/cgf.12802

Biener, V., Kalamkar, S., Nouri, N., Ofek, E., Pahud, M., Dudley, J. J., Hu, J., Kristensson, P. O., Weerasinghe, M., Pucihar, K. Č., & Kljun, M. (2022). Quantifying the effects of working in VR for one week. *IEEE Transactions on Visualization and Computer Graphics*, 28(11), 3810-3820. https://doi.org/10.1109/TVCG.2022.3203103

Blackburn, D. C., Boyer, D. M., Gray, J. A., Winchester, J., Bates, J. M., Baumgart, S. L., Braker, E., Coldren, D., Conway, K. W., Rabosky, A. D., de la Sancha, N., Dillman, C. B., Dunnum, J. L., Early, C. M., Frable, B. W., Gage, M. W., Hanken, J., Maisano, J. A., Marks, B. D., … Stanley, E. L. (2024). Increasing the impact of vertebrate scientific collections through 3D imaging: The openVertebrate (oVert) Thematic Collections Network. *BioScience*, 74(3), 169-186. https://doi.org/10.1093/biosci/biad120

Bowman, D. A., & McMahan, R. P. (2007). Virtual reality: How much immersion is enough? *Computer*, 40(7), 36–43. https://doi.org/10.1109/MC.2007.257

Boyer, D. M., Gunnell, G. F., Kaufman, S., & McGeary, T. M. (2016). Morphosource: Archiving and sharing 3-D digital specimen data. *The Paleontological Society Papers*, 22, 157-181. doi:10.1017/scs.2017.13

Boyer, D. M. (2016, February 25). *Virtual fossils revolutionize the study of human evolution*. Aeon. https://aeon.co/ideas/virtual-fossils-revolutionise-the-study-of-human-evolution (Archived by the Internet Archive at https://web/20250915001832/https://aeon.co/ideas/virtual-fossils-revolutionise-the-study-of-human-evolution

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. Jordan, B. Thomas, & B. Weerdmeester (Eds.), *Usability evaluation in industry* (pp. 189–194). London, UK: Taylor & Francis.

Buckland, M. K. (1991). Information as thing. *Journal of the American Society for Information Science*, 42(5), 351-360. https://doi.org/10.1002/(SICI)1097-4571(199106)42:5<351::AID-ASI5>3.0.CO;2-3

Cook, M. (2024). The lifecycle of 3D data in academic libraries: A survey of methods and implications for information professionals. *The Journal of Academic Librarianship*, 50(5), 102933. https://doi.org/10.1016/j.acalib.2024.102933

Cook, M., & Lischer-Katz, Z. (2019). Integrating 3D and virtual reality into research and pedagogy in higher education. In K.J. Varnum (Ed.), *Beyond reality: Augmented, virtual, and mixed reality in the library* (pp. 69-85). ALA Editions.

Cook, M., Lischer-Katz, Z., Hall, N., Hardesty, J., Johnson, J., McDonald, R., & Carlisle, T. (2019). Challenges and strategies for educational virtual reality. *Information Technology and Libraries*, 38(4), 25-48. https://doi.org/10.6017/ital.v38i4.11075

Creswell J.W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches.* Los Angeles: Sage Publications.

Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein & S. Rogers (Eds.), *Perception of space and motion* (pp. 69–117). Academic Press. https://doi.org/10.1016/B978-012240530-3/50005-5

Dede, C. J., Jacobson, J., & Richards, J. (2017). Introduction: Virtual, augmented, and mixed realities in education. In *Virtual, augmented, and mixed realities in education* (pp. 1-16). Springer, Singapore.

Denard, H. (2012). A new introduction to the London Charter. In A. Bentkowska-Kafel, D. Baker, & H. Denard (Eds.), *Paradata and transparency in virtual heritage* (pp. 57–71). Routledge. https://www.london-charter.org/media/files/ch6_denard.pdf (Archived by the Internet Archive at https://web.archive.org/web/20250915011142/https://www.london-charter.org/media/files/ch6_denard.pdf)

Donalek, C., Djorgovski, S. G., Cioc, A., Wang, A., Zhang, J., Lawler, E., Yeh, S., Mahabal, A., Graham, M., Drake, A., & Davidoff, S. (2014). Immersive and collaborative data visualization using virtual reality platforms. In *Proceedings of 2014 IEEE International conference on big data* (pp. 609-614). IEEE. https://doi.org/10.1109/BigData.2014.7004282

European Commission: Directorate-General for Communications Networks, Content and Technology. (2022). *Study on quality in 3D digitisation of tangible cultural heritage: mapping parameters, formats, standards, benchmarks, methodologies, and guidelines: executive summary*. Publications Office of the European Union. https://data.europa.eu/doi/10.2759/581678

Forsberg, A., Katzourin, M., Wharton, K., & Slater, M. (2008). A comparative study of desktop, fishtank, and cave systems for the exploration of volume rendered confocal data sets. *IEEE Transactions on Visualization and Computer Graphics*, 14(3), 551-563. https://doi.org/10.1109/tvcg.2007.70433

Foster, E. D. & Deardorff, A. (2017). Open science framework (OSF). *Journal of the Medical Library Association*, 105(2), 203. https://doi.org/10.5195/jmla.2017.88

Giacalone, G., Yamamoto, T., Belva, F., Hayashi, A., Dori, Y., Zviman, M. M., Gysen, M., Nam, H. H., Jolley, M. A., & Kato, M. (2019). The application of virtual reality for preoperative planning of lymphovenous anastomosis in a patient with a complex lymphatic malformation. *Journal of Clinical Medicine*, 8(3), 371. https://doi.org/10.3390/jcm8030371

Gibson, J.J. (2002). A theory of direct visual perception. In A. Noë & E. Thompson (Eds.), *Vision and Mind: Selected Readings in the Philosophy of Perception* (pp. 77–90). MIT Press. https://doi.org/10.7551/mitpress/7111.003.0007

Grayburn, J., Lischer-Katz, Z., Golubiewski-Davis, K., & Ikeshoji-Orlati, V. (2019). *3D/VR in the academic library: Emerging practices and trends*. Council on Library and Information Resources. https://www.clir.org/wp-content/uploads/sites/6/2019/02/Pub-176.pdf (Archived by the Internet Archive at https://web.archive.org/save/https://www.clir.org/wp-content/uploads/sites/6/2019/02/Pub-176.pdf

Greene, D., & Groenendyk, M. (2021). An environmental scan of virtual and augmented reality services in academic libraries. *Library Hi Tech*, 39(1), 37-47. https://doi.org/10.1108/LHT-08-2019-0166

Hall, N., Hardesty, J., Lischer-Katz, Z., Cook, M., Xie, Z., Hardesty, J., Johnson, J., Ogier, A., & McDonald, R. (2019) Challenges and directions in 3D and VR data curation: Findings from a nominal group study. *International Journal of Digital Curation*, 14(1), 150-167. https://doi.org/10.2218/ijdc.v14i1.588

Harada, Y., & Ohyama, J. (2020). Evaluations of visual guidance effects for 360-degree directions around users: VR experimental psychology. *Proceedings of the Annual Convention of the Japanese Psychological Association, Bunkyō, Japan*, 84, PQ-019. https://doi.org/10.4992/pacjpa.84.0_PQ-019

Hardesty, J. L., Johnson, J., Wittenberg, J., Hall, N., Cook, M., Lischer-Katz, Z., Xie, Z., & McDonald, R. (2020). 3D Data repository features, best practices, and implications for preservation models: Findings from a national forum. *College & Research Libraries*, 81(5), 789. https://doi.org/10.5860/crl.81.5.789

Hart, S.G. (2006). NASA-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9), 904–908. https://doi.org/10.1177/154193120605000909

Hawks, C. A., & Williams, S. L. (1986). Arsenic in natural history collections. *Leather Conservation News*, 2(2), 1-4.

Helbig, C., Bauer, H.-S., Rink, K., Wulfmeyer, V., Frank, M., & Kolditz, O. (2014). Concept and workflow for 3D visualization of atmospheric data in a virtual reality environment for analytical approaches. Environmental *Earth Sciences* 72, 3767–3780. https://doi.org/10.1007/s12665-014-3136-6

Huvila, I. (2022). Improving the usefulness of research data with better paradata. *Open Information Science*, 6(1), 28-48. https://doi.org/10.1515/opis-2022-0129

Ijiri, T., Todo, H., Hirabayashi, A., Kohiyama, K., & Dobashi, Y. (2018). Digitization of natural objects with micro-CT and photographs. *PLOS ONE*, *13*(4), e0195852. https://doi.org/10.1371/journal.pone.0195852

Jang, S., Vitale, J. M., Jyung, R. W., & Black, J. B. (2017). Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment. *Computers & Education*, 106, 150–165. https://doi.org/10.1016/j.compedu.2016.12.009

Janiszewski, M., Uotinen, L., Merkel, J., Leveinen, J., & Rinne, M. (2020). Virtual Reality Learning Environments for Rock Engineering, Geology and Mining Education. 54*th U.S. Rock Mechanics/Geomechanics Symposium*, *28 June - 1 July: Physical Event Cancelled*, ARMA. https://www.onepetro.org/conference-paper/ARMA-2020-1101

Kersten-Oertel, M., Chen, S. J.-S., & Collins, D. L. (2014). An evaluation of depth enhancing perceptual cues for vascular volume visualization in neurosurgery. *IEEE Transactions on Visualization and Computer Graphics*, 20(3), 391–403. https://doi.org/10.1109/TVCG.2013.240

King, G. (2007). An Introduction to the Dataverse Network as an Infrastructure for Data Sharing. *Sociological Methods & Research*, 36(2), 173–199. https://doi.org/10.1177/0049124107306660

Kingsley, L. J., Brunet, V., Lelais, G., McCloskey, S., Milliken, K., Leija, E., Fuhs, S. R., Wang, K., Zhou, E., & Spraggon, G. (2019). Development of a virtual reality platform for effective communication of structural data in drug discovery. *Journal of Molecular Graphics & Modelling*, 89, 234–241. https://doi.org/10.1016/j.jmgm.2019.03.010

Kluge, S., Gladisch, S., von Lukas, U. F., Staadt, O., & Tominski, C. (2019). Virtual Lenses as Embodied Tools for Immersive Analytics. *arXiv preprint arXiv*:1911. 10044. https://doi.org/10.48550/arXiv.1911.10044

Kuliga, S. F., Thrash, T., Dalton, R. C., & Hölscher, C. (2015). Virtual reality as an empirical research tool—Exploring user experience in a real building and a corresponding virtual model. *Computers, Environment and Urban Systems*, 54, 363–375. https://doi.org/10.1016/j.compenvurbsys.2015.09.006

Lages, W. S., & Bowman, D. A. (2018). Move the Object or Move Myself? Walking vs. Manipulation for the Examination of 3D Scientific Data. *Frontiers in ICT*, 5. https://doi.org/10.3389/fict.2018.00015

Laha, B., Bowman, D. A., & Socha, J. J. (2014). Effects of VR system fidelity on analyzing isosurface visualization of volume datasets. *IEEE Transactions on Visualization and Computer Graphics*, 20(4), 513–522. https://doi.org/10.1109/TVCG.2014.20

Leménager, M., Burkiewicz, J., Schoen, D. J., & Joly, S. (2023). Studying flowers in 3D using photogrammetry. *New Phytologist*, 237(5), 1922–1933. https://doi.org/10.1111/nph.18553

Limp, F., Payne, A., Simon, K., Winters, S., & Cothren, J. (2011). Developing a 3-D digital heritage ecosystem: From object to representation and the role of a virtual museum in the 21st century. *Internet Archaeology*, 30(30), 1-38. https://doi.org/10.11141/IA.30.1

Lischer-Katz, Z., Cook, M. N., Hall, N., Hardesty, J., Wittenberg, J., Johnson, J., & Carlisle, T. (2019). Supporting virtual reality and 3D in academic libraries: Defining preservation and curation challenges. In *Proceedings of the 16th International Conference on Digital Preservation iPRES*

2019 (pp. 114-125), Amsterdam, The Netherlands (Sept. 16-20). https://ipres2019.org/static/proceedings/iPRES2019.pdf (Archived by the Internet Archive at https://web.archive.org/web/20250630112557/https://ipres2019.org/static/proceedings/iPRES2019.pdf

Lischer-Katz, Z., & Cook, M. (2022). Virtual Reality and the Academic Library of the Future. *Transactions of the American Philosophical Society*, 110(3), 185-210. https://www.jstor.org/stable/45420507

Lischer-Katz, Z., Cook, M., & Boulden, K. (2018). Evaluating the impact of a virtual reality workstation in an academic library: Methodology and preliminary findings. *Proceedings of the Association for Information Science and Technology*, 55(1), 300–308. https://doi.org/10.1002/pra2.2018.14505501033

Lisle, L., Chen, X., Edward Gitre, J. K., North, C., & Bowman, D. A. (2020). Evaluating the Benefits of the Immersive Space to Think. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops* (VRW), 331–337. https://wevr.adalsimeone.me/2020/WEVR2020_Lisle.pdf (Archived by the Internet Archive at https://web.archive.org/web/20250915032442/https://wevr.adalsimeone.me/2020/WEVR2020_Lisle.pdf

Madura, T. I., Clementel, N., Gull, T. R., Kruip, C. J. H., & Paardekooper, J.-P. (2015). 3D printing meets computational astrophysics: Deciphering the structure of η Carinae's inner colliding winds. Monthly Notices of the Royal Astronomical Society, 449(4), 3780–3794. https://doi.org/10.1093/mnras/stv422

Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): A Theoretical Research-Based Model of Learning in Immersive Virtual Reality. *Educational Psychology Review*, 33(3), 937–958. https://doi.org/10.1007/s10648-020-09586-2

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education*, 70, 29–40. https://doi.org/10.1016/j.compedu.2013.07.033

Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education*, 56(3), 769–780. https://doi.org/10.1016/j.compedu.2010.10.020

Mills, N. (2020). Engagement and Immersion in Virtual Reality Narratives. *Student engagement in the language classroom*, 11.

Milovanovic, J., Moreau, G., Siret, D., & Miguet, F. (2017). Virtual and augmented reality in architectural design and education. *17th International Conference, CAAD Futures 2017*, Gülen Çağdaş, Mine Özkar, Leman F. Gül and Ethem Gürer, Jul 2017, Istanbul, Turkey.

Moore, J., Rountrey, A., & Kettler, H. S. (Eds.). (2022). *3D data creation to curation: Community standards for 3D data preservation*. Association of College and Research Libraries.

Mörth, E., Nielsen, C., Turner, M., Liu, X., Keller, M. S., Beyer, J., Pfister, H., Chen, Z.-T., & Gehlenborg, N. (2024). A mixed reality and 2D display hybrid approach for visual analysis of 3D tissue maps. OSF Preprints. https://doi.org/10.31219/osf.io/zka2j

Muenster, S. (2022). Digital 3D technologies for humanities research and education: An overview. *Applied Sciences*, 12(5), 2426. https://doi.org/10.3390/app12052426

Pellas, N., Mystakidis, S., & Kazanidis, I. (2021). Immersive Virtual Reality in K-12 and Higher Education: A systematic review of the last decade scientific literature. *Virtual Reality*, 25(3), 835–861. https://doi.org/10.1007/s10055-020-00489-9

Pfarr-Harfst, M. (2016). Typical Workflows, Documentation Approaches and Principles of 3D Digital Reconstruction of Cultural Heritage. In S. Münster, M. Pfarr-Harfst, P. Kuroczyński, & M. Ioannides (Eds.), *3D Research Challenges in Cultural Heritage II* (pp. 32–46). Springer International Publishing. https://doi.org/10.1007/978-3-319-47647-6_2

Pietschmann, L., Zürcher, P.-D., Bubík, E., Chen, Z., Pfister, H., & Bohné, T. (2023). Quantifying the impact of XR visual guidance on user performance using a large-scale virtual assembly experiment. In *2023 IEEE Visualization and Visual Analytics (VIS) Short Papers* (pp. 211–215). IEEE. https://doi.org/10.48550/arXiv.2308.03390

Pomerantz, J. (2018). Learning in three dimensions: Report on the EDUCAUSE/HP Campus of the Future project. EDUCAUSE. https://library.educause.edu/resources/2018/8/learning-in-three-dimensions-report-on-the-educause-hp-campus-of-the-future-project (Archived by the Internet Archive at https://web.archive.org/web/20250915035931/https://www.educause.edu/ecar/research-publications/learning-in-three-dimensions-report-on-the-educause-hp-campus-of-the-future-project/executive-summary-key-findings-acknowledgments

Portman, M. E., Natapov, A., & Fisher-Gewirtzman, D. (2015). To go where no man has gone before: Virtual reality in architecture, landscape architecture and environmental planning. *Computers, Environment and Urban Systems*, 54, 376-384. https://doi.org/10.1016/j.compenvurbsys.2015.05.001

Prouzeau, A., Cordeil, M., Robin, C., Ens, B., Thomas, B. H., & Dwyer, T., 2019, May. Scaptics and highlight-planes: Immersive interaction techniques for finding occluded features in 3D scatterplots. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (1-12). https://doi.org/10.1145/3290605.33005

Qin, T., Cook, M., & Courtney, M. (2020) Exploring chemistry with wireless, PC-less portable virtual reality laboratories. *Journal of Chemical Education* 98(2), 521-529. https://doi.org/10.1021/acs.jchemed.0c00954

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147, 103778. https://doi.org/10.1016/j.compedu.2019.103778

Ragan, E. D., Kopper, R., Schuchardt, P., & Bowman, D. A. (2012). Studying the effects of stereo, head tracking, and field of regard on a small-scale spatial judgment task. *IEEE transactions on visualization and computer graphics*, 19(5), 886-896. https://doi.org/10.1109/TVCG.2012.163

Richardson, A. E., Powers, M. E., & Bousquet, L. G. (2011). Video game experience predicts virtual, but not real navigation performance. *Computers in Human Behavior*, 27(1), 552-560. https://doi.org/10.1016/j.chb.2010.10.003

Safi, M., Chung, J., & Pradhan, P. (2019). Review of augmented reality in aerospace industry. *Aircraft Engineering and Aerospace Technology*. 91(9), 1187-1194. https://doi.org/10.1108/AEAT-09-2018-0241

Scheffler, M., Aeschlimann, M., Albrecht, M., Bereau, T., Bungartz, H. J., Felser, C., Greiner, M., Groß, A., Kock, C. T., Kremer, K., Nagel, W. E., Scheidgen, M., Wöll, C., & Draxl, C. (2022). FAIR data enabling new horizons for materials research. *Nature*, 604, 635-642. https://doi.org/10.1038/s41586-022-04501-x

Seth, A., Vance, J. M., & Oliver, J. H. (2011). Virtual reality for assembly methods prototyping: a review. *Virtual reality*, 15(1), 5-20.

Silvestri, C., Motro, R., Maurin, B., & Dresp-Langley, B. (2010). Visual spatial learning of complex object morphologies through the interaction with virtual and real-world data. *Design Studies*, *31*(4), 363-381. https://doi.org/10.1016/j.destud.2010.03.001

Skulmowski, A., & Rey, G. D., 2017. Measuring cognitive load in embodied learning settings. *Frontiers in psychology*, 8, 1191. https://doi.org/10.3389/fpsyg.2017.01191

Sullivan, I., DeHaven, A., & Mellor, D. (2019). Open and reproducible research on open science framework. *Current Protocols Essential Laboratory Techniques*, 18(1), e32. https://doi.org/10.1002/cpet.32

Tahmid, I. A., Lisle, L., Davidson, K., Whitley, K., North, C., & Bowman, D. A. (2023). Evaluating the Feasibility of Predicting Information Relevance During Sensemaking with Eye Gaze Data. In *2023 IEEE International Symposium on Mixed and Augmented Reality* (ISMAR), *Sydney, Australia* (pp. 713-722). IEEE. https://doi.org/10.1109/ISMAR59233.2023.00086

Tracy, J. P., & Albers, M. J. (2006). Measuring cognitive load to test the usability of web sites. In *Annual Conference-society for technical communication* (Vol. 53, p. 256).

Trelease, R. B., & Rosset, A. (2008). Transforming clinical imaging data for virtual reality learning objects. *Anatomical sciences education*, 1(2), 50-55. https://doi.org/10.1002/ase.13

Uhr, M., Haselmann, S., Steep, L., Eickhoff, J., & Steinicke, F. (2019). Designing an Interactive Visualization for Road Works Coordination using Virtual Reality. In *Proceedings of the Mensch und Computer 2019 Workshop on VR and AR in Everyday Context* (VARECo).

Van Dam, A., Laidlaw, D. H., & Simpson, R. M. (2002). Experiments in immersive virtual reality for scientific visualization. *Computers & Graphics*, 26(4), 535-555.

Wang, P., Miller, M. R., Han, E., DeVeaux, C., & Bailenson, J. N. (2024). Understanding virtual design behaviors: A large-scale analysis of the design process in Virtual Reality. *Design Studies*, 90, 101237. https://doi.org/10.1016/j.destud.2023.101237

Wei, J., Siegel, E., Sundaramoorthy, P., Gomes, A., Zhang, S., Vankipuram, M., Smathers, K., Ghosh, S., Horii, H., & Bailenson, J. (2025). Cognitive Load Inference Using Physiological Markers in Virtual Reality. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces* (VR) (pp. 759-769). IEEE. https://doi.org/10.1109/VR59515.2025.00098

Whitlock, M., Smart, S., & Szafir, D. A. (2020). Graphical perception for immersive analytics. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces* (VR) (pp. 616-625). IEEE. https://doi.org/10.1109/VR46266.2020.00084

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., & Bouwman, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9. https://doi.org/10.1038/sdata.2016.18

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3), 225–240. https://doi.org/10.1162/105474698565686

Yang, Y., Cordeil, M., Beyer, J., Dwyer, T., Marriott, K., & Pfister, H. (2021). Embodied Navigation in Immersive Abstract Data Visualization: Is Overview+ Detail or Zooming Better for 3D Scatterplots? *IEEE Transactions on Visualization and Computer Graphics*, 27(2) 1214-1244 https://doi.org/10.1109/TVCG.2020.3030427

Zhao, J., Wallgrün, J. O., LaFemina, P. C., Normandeau, J., & Klippel, A. (2019). Harnessing the power of immersive virtual reality-visualization and analysis of 3D earth science data sets. *Geo-spatial Information Science*, 22(4), 237-250. https://doi.org/10.1080/10095020.2019.1621544

Appendix I: Task prompts

Orientation task (OR_A):

A professor wants to confirm the accuracy of a scan before showing it to his marine biology class. Does this whale shark have the requisite 5 gill slits?

Correct answer = Yes

Navigation/Identification task (VR_1A):

An unclaimed mineral specimen was located in the scanning lab. Inspect the model to identify the owner. Does Yale University own this scan?

Correct answer = No

Comparison task (VR_2A):

A student worker accidentally duplicated model filenames. They think "Object 2" is the lower quality “reference” scan. Is that correct?

Correct answer = Yes

Error detection task (VR_3A):

The scanning team are concerned about missing data within meteorite cavities, which appears as blurriness. Is scanning complete?

Correct answer = No

Orientation task (OR_B):

A professor is comparing Leatherback turtle "skylights”, prominent spots atop their heads. Is this scan an example of a white-coloured skylight?

Correct answer = Yes

Confirmation task (CO_1B):

A mineralogist is looking for a specimen with a label beginning "144…". Is this the scan in question?

Correct answer = Yes

Comparison task (CO_2B):

A librarian believes a mis-catalogued rare book might be among recent scans. Is volume "F" present in this scene?

Correct answer = No

Error detection task (CO_3B):

An historian is worried that detail may be missing from a recent capital scan. Missing data appears as blur. Is this scan complete?

Correct answer = No