XR Composition in the Wild: The Impact of User Environments on Creativity, UX and Flow during Music Production in Augmented Reality

Ruben Schlagowski
HCAI-Lab, University of Augsburg
Augsburg, Germany
ruben.schlagowski@uni-a.de

Silvan Mertes
HCAI-Lab, University of Augsburg
Augsburg, Germany
silvan.mertes@uni-a.de

Dariia Nazarenko
HCAI-Lab, University of Augsburg
Augsburg, Germany
dariia.nazarenko@uni-a.de

Alexander Dauber
HCAI-Lab, University of Augsburg
Augsburg, Germany
alexander-dauber@gmx.de

Elisabeth André
HCAI-Lab, University of Augsburg
Augsburg, Germany
elisabeth.andre@uni-a.de

Figure 1: Musicians being creative using a HoloLens 2 HMD in a creative space of free choice (left) or in a lab environment (right). In what way do their creative experiences differ?

ABSTRACT

With the advent of HMD-based Mixed Reality, or "Spatial Computing" as framed by Apple, creativity- and productivity-related use-cases in XR, such as music production, are rising in popularity. However, even though the importance of environments for creativity is understood, XR applications for creative use cases are often evaluated in laboratories. While the mismatch of familiar creative spaces and lab environments matters little in VR, the effect on evaluation metrics for AR applications needs to be clarified. To this end, we conducted an experiment in which participants composed and produced music on an AR HMD in their preferred creative environments and a typical laboratory environment. Looking at questionnaire data, we observed similar scores for user experience, flow, and creative experience in both conditions, suggesting that supervised evaluation of AR-based creativity support applications does not require mobile demonstrators nor freely selectable environments. Based on qualitative feedback and overall high scores for UX and flow, we discuss our observations and their implications and emphasize the need for field studies in the XR creativity domain.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; **Mixed / augmented reality**; • **Applied computing → Sound and music computing**.

KEYWORDS

Creativity, Creativity Support Tools, Augmented Reality, Musical XR, Head-Mounted Displays, Mixed Reality, Spatial Computing, Generative Adversarial Networks

AM '24, September 18–20, 2024, Milan, Italy

ACM ISBN 979-8-4007-0968-5/24/09
https://doi.org/10.1145/3678299.3678314

ACM Reference Format:
Ruben Schlagowski, Silvan Mertes, Dariia Nazarenko, Alexander Dauber, and Elisabeth André. 2024. XR Composition in the Wild: The Impact of User Environments on Creativity, UX and Flow during Music Production in Augmented Reality. In *Audio Mostly 2024 - Explorations in Sonic Cultures (AM '24), September 18–20, 2024, Milan, Italy*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3678299.3678314

1 INTRODUCTION

Regardless of the field in which one is creative, be it music, fine arts, or poetry, many factors can interact with individual creativity. Be it specific tools, certain lighting conditions, certain natural environments, or inspiration from others: Creative people usually express unique and particular needs when asked about them [45]. One of these factors is the environment in which the creative person finds him- or herself. While some people prefer orderly and minimalist conditions, others prefer more chaotic and pragmatic environments.

These preferences challenge researchers and developers who want to evaluate augmented reality applications for creative people. To ensure the integrity of evaluation metrics, e.g., scores from questionnaires, probands would need to use such applications in their usual creative environments or at least in environments that meet their creative needs. However, especially when novel technologies such as HMDs or complex applications are used, hardware that represents the state of the art is often expensive and not available in large numbers. Consequently, AR applications that use headsets such as the HoloLens 2 are often evaluated in the laboratory, where participants find themselves in unfamiliar creative spaces, rather than in field studies. To assess the impact of such practice, the study presented in this paper addresses the research question of whether it makes an impact to let users freely choose a location in such studies or whether it is sufficient to test AR applications in the laboratory.

For this purpose, we conducted an experiment in which an AR application that enables the composition of drum tracks is used either in a free-choice environment (e.g., in the forest, on campus grounds, or at home) or in a laboratory setting. For both conditions, we used questionnaires to measure typical evaluation metrics such as user experience, creative support, creative experience, and flow. In addition, we evaluated open questions and free-form feedback to learn what is important to users when choosing their creative environment and what criteria study participants used to select their creative environment/space during the study. Our results suggest that the quantitative evaluation metrics derived from the *User Experience Questionnaire* (UEQ), *Creativity Support Index* (CSI), *Flow Short Scale* (FSS), and *Creativity Experience Self Rating Questionnaire* (CESR) neither benefited nor suffered from a freely selectable user environment during our creative music-making sessions in augmented reality. We discuss various potential reasons and factors for this observation, such as interference from experiment setup and the challenges of handling and learning complex applications with new interaction paradigms in the AR domain.

2 RELATED WORK

2.1 Creativity Support Factors

There is a variety of works that focus on how different factors can support users' creativity. Several of these specifically mention environmental cues. For example, contextual cues like furniture or paintings [18, 35], open spaces (e.g., high ceilings) [2, 33], well thought colors [26, 29, 31, 52] or visual details (e.g., "untidiness") [2, 29, 54] were observed to positively affect creativity. Indoor plants [47, 52], windows [5, 18, 29] and random objects (that can serve as inspiring decorative elements) were reported to play a substantial role for creativity as well. While those support factors specifically are relevant for indoor settings, Chulvi et al. [9] stated that being outdoors itself can benefit creativity. Other works suggest the importance of lighting for creativity [21, 23, 31, 49]. Besides purely visual cues, the sound of the environments (i.e., moderate ambient sounds) additionally contributes to feeling creative [30].

Not only environmental factors are relevant for creativity. The communication of positive or embodied metaphors and instructions also increases creativity [28, 37, 55]. Further, a high degree of freedom regarding body posture is desirable. For example, users should be allowed to walk [15, 38, 60], exercise [10, 50], do large and fluid movements [19, 20, 48] and do open and expansive postures [1, 32]. Other possibilities to improve the possibility to feel creative are using the concepts of *Embodied Creativity* (e.g., through avatars) [4, 17] , or *Surprise* [14].

2.2 XR and Creativity

Chandrasekera et al. [6] studied how VR and AR interfaces affect creative design processes in design education, suggesting that the higher tangibility of an AR interface reduces cognitive load compared to VR interfaces. Yilmaz et al. [58] showed that AR technology significantly increases narrative skill and creativity when used as a tool for storytelling. An AR app that translates 3D phone motion and touch pressure into 3D sculptures as well as sound was developed and evaluated by Valer et al. [53]. They reported that participants received good scores on the creativity support index (CSI).

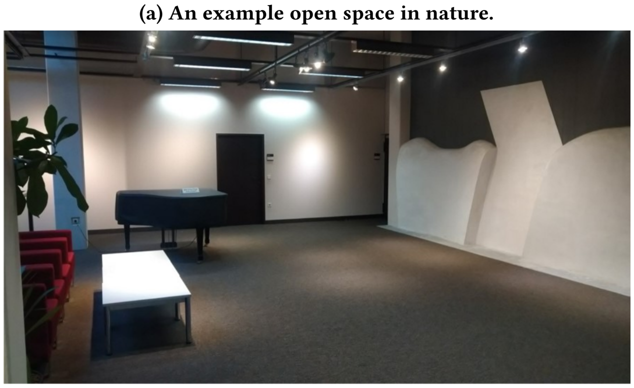
Lin et al. [27] conducted a study to draw findings about the influence of VR on creative self-efficacy. Although they did not find a significant difference there, they observed that their VR-supported project facilitated the participants' efficacy for creative thinking. Fröhlich et al. [16] presented a VR-based landscape sandbox tool incorporating real sand into their application. They reported that all participants of a user study agreed that their system supported creative expression. Obeid et al. [36] investigated the impact of VR technology on design process creativity. They reported that in a user study coping with a design task, participants that made use of a VR system had an increased design process creativity than participants that used a non-immersive setup. For a (product-) design task, Yang et al. [57] developed and assessed a VR-based creative support system. They observed that, compared to a pen-and-paper approach, it performed better in terms of creativity and flow.

Muender et al. [34] included tangible objects into a VR system in order to improve perceptions of correct depth and scale. In an interview study with domain experts, all participants mentioned a possible positive impact on creativity. Further, the participants mentioned that they would integrate such a system into their professional creative workflow. Research by Chang et al. [7] evaluating the potential of VR in a teaching scenario involved 138 seventh-grade students. Participants in a VR group showed more innovative designs and problem-solving skills than users that were working with traditional media (presentations, videos).

In different works, music-related scenarios were used to study the impact of VR on users' creativity. While Ppali et al. [39] investigated how VR environments can be designed to support creativity,

(a) An example open space in nature.

(b) An example covered outdoor space.



(c) An example indoor space with more privacy.

(d) An example public indoor space.

Figure 2: Four example creative spaces that participants could choose in the *FreeSpace* condition.

Schlagowski et al. [45] specifically investigated if customizable VR environments can increase users' creativity when compared to non-customizable VR environments. While they did not find a significant difference here, users predominantly preferred the customizable setting.

3 METHODOLOGY

To investigate the impact of free environment choice, we conducted an experiment during which musically creative and affine participants interacted with an augmented reality application to compose drum beats for existing background tracks. The mobile HMD application and its UI are described in subsection 3.4. As we did not expect large effect sizes and had participants that varied widely regarding musical ability and age, we chose a within-subject study design during which participants saw both conditions in randomized order. The next subsection explains these conditions along with the dependent variables we measured.

3.1 Dependent and Independent Variables

In our experiment we altered the independent variable *environment selection paradigm* in two conditions/steps:

- *FreeSpace:* In this condition, participants conducted creative tasks using our mobile AR application in an environment of their choice (e.g., an open space in nature, see Figure 2).
- *LabSpace:* Within the *LabSpace* condition, participants used our AR prototype in a typical laboratory environment (see the right picture in Figure 1).

The order of the conditions was randomized. In the condition *FreeSpace*, participants were free to either guide the test supervisors to a creative space of their preference or choose a space from a picture catalog presented to them (in this case, test supervisors guided participants to that place). Although creative space choice was not limited to a constraint area (the test leaders also offered to visit subjects at home), all test subjects chose creative places on - or near - the university campus. Figure 2 shows four example places that participants selected. Examples include open areas in a nearby forest, covered outdoor spaces (suited for rainy weather), and private and public indoor spaces at or near the university campus. Participants did not need to stick to a particular creative space for the whole duration of the creative sessions in the *FreeSpace* condition. Instead, they were free to change the location at any time - for instance, if they felt disturbed by pedestrians or saw a more inspiring place nearby. However, in the *LabSpace* condition, participants needed to complete all creative tasks in a specific laboratory environment, a fairly neutral space with mostly seated PC working places. After each creative session in condition *FreeSpace* or *LabSpace*, participants filled out four questionnaires measuring our dependent variables. Table 1 lists these questionnaires with their subscales. Each subscale was considered as a dependent variable for our experiment.

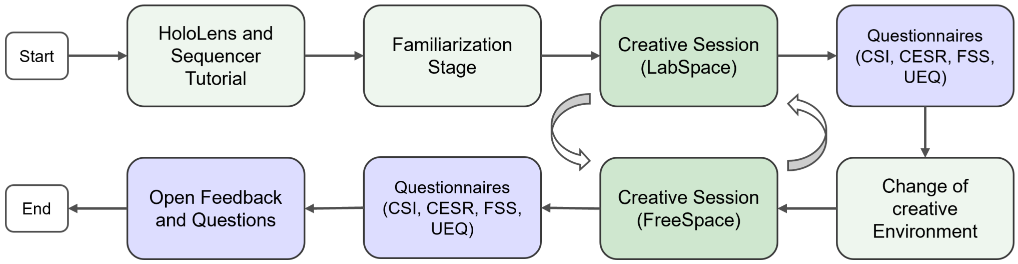
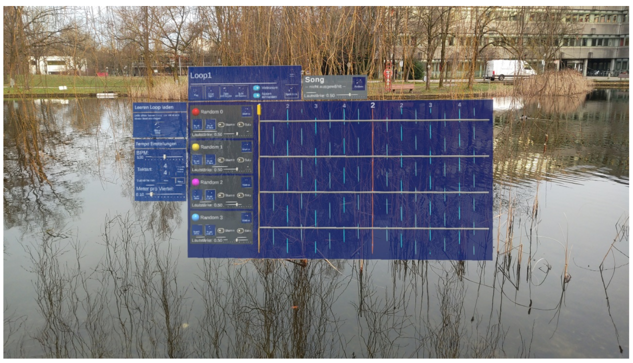
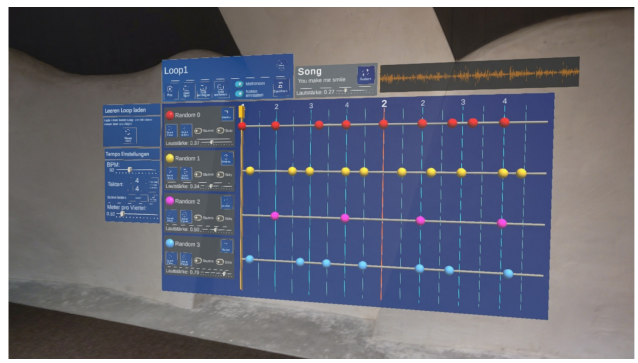


Figure 3: A visualization of the study procedure.

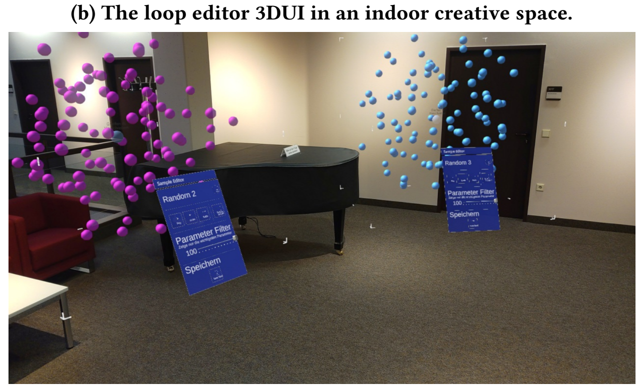


(a) The loop editor 3DUI in an outdoor creative space.



(b) The loop editor 3DUI in an indoor creative space.

(c) A VMM sound design UI in an outdoor creative space.



(d) Two VMM sound design UIs in an indoor creative space.

Figure 4: Screenshots of the HoloLens 2 Application.

3.2 Participants

We acquired 22 participants from Germany through various methods, including social media, mailing lists, and word of mouth. Eighteen of them identified as male, and four as female. Participants ranged in age from 21 to 51 (*M*: 29.0, *SD*: 6.0). Each participant confirmed to be either musically creative or music-affine. One participant stated to have used an AR-HMD before. The others did not have any experience with HMD-based AR. Each participant was paid a compensation of 30 Euros for their efforts.

3.3 Study Procedure

The study procedure is visualized in Figure 3. After arriving at the lab and filling out the data and privacy consent documents, participants first did a HoloLens 2 tutorial using the *HoloLens Tips* app[1]. Subsequently, they watched a tutorial video on how to use the AR drum sequencer application described in subsection 3.4. After watching the video, participants could test the AR app interactively for approximately 10 minutes in the familiarization stage. Then, participants and test supervisors either went to the laboratory environment (if in condition *LabSpace*) or to a freely selectable environment (if in condition *FreeSpace*) to complete their first of two creative sessions. During these creative sessions, participants went through two creative stages:

(1) Exploration: In this stage, participants composed drum beats using the AR drum sequencer UI for three backing tracks which were then stored persistently. They were told that

[1] https://microsoft.com/en-us/p/hololens-tips/9pd4cxkklc47#/

Table 1: Questionnaires employed during the experiment. All subscales are considered as dependent variables.

Name/Source	Measure/Description/Subscales	Range
User Experience Questionnaire (UEQ) [24]	**User Experience**, Experience while using the application, subscales: 'Attractiveness', 'Perspicuity', 'Efficiency', 'Dependability', 'Stimulation', 'Novelty'	1-5
Flow Short Scale (FSS) [40]	**Flow**, Psychological state of being immersed in a task [11], subscales: 'Flow', 'Fluency', 'Absorption', 'Anxiety', 'Challenge'.	1-7
Creativity Support Index (CSI) [8]	**Creativity Support**, Feeling of being creatively supported, subscales: 'Enjoyment', 'Exploration', 'Expressiveness', 'Immersion', 'Results worth Effort'.	0-100
Creativity Experience Self Rating Questionnaire (CESR) [12]	**Creative Experience**, How creative users felt, subscales: 'Competence', 'Autonomy', 'Task Enjoyment'.	0-9

these creations need not be polished, as they could be refined in the next stage.

(2) Refinement: In this stage, participants selected one of their creations and revised them until they were content with the outcome.

Participants were free to invest as much time as they liked in both creative stages. However, if they wished to be in time for a standard study duration (20 minutes for the exploration phase and 10 minutes for the refinement stage), the test supervisors communicated if time went short. After the first creative session, participants filled out the questionnaires and proceeded to session two after changing their creative environment. After the completion of both creative sessions and questionnaires, they were able to provide open feedback regarding both the experimental procedure and the AR drum sequencer application. Furthermore, they were asked to state their decision criteria for their selected creative spaces in condition *FreeSpace*.

3.4 AR Drum Sequencer Application

The AR Drum Sequencer application was built for Microsoft's AR-HMD *HoloLens 2*[2] using the Unity Engine[3] and the *Mixed Reality Toolkit* (MRTK) 3DUI Framework[4]. The sequencer UI was originally designed and developed in a human-centered iterative procedure by Schlagowski and colleagues for VR [45] and adapted by us for the *Universal Windows Platform* (UWP) and HoloLens.

The AR application provides users with two main 3DUI components:

- A loop editor 3DUI that can be used to compose drum loops for a variety of backing tracks by placing spheres on parallel lines that resemble individual audio channels (see Figures 4a and 4b).
- Vector manipulation modules (VMMs, see [43]) that can be instantiated for each drum track and allow customizing the sound of drum samples by manipulating a set of 3D spheres (see Figures 4c and 4d).

As such, besides using the loop editor 3DUI for drum beat composition, users were able to design custom sounds for their drums which opened up the creative space. For a more detailed description of the loop editor interface, please refer to [45] and for a study investigating the VMM modules, to [43].

4 RESULTS

4.1 Quantitative Results

Table 2: Results of the two-tailed Student's t-tests. If data was found to be non-parametric, Wilcoxon test was applied.

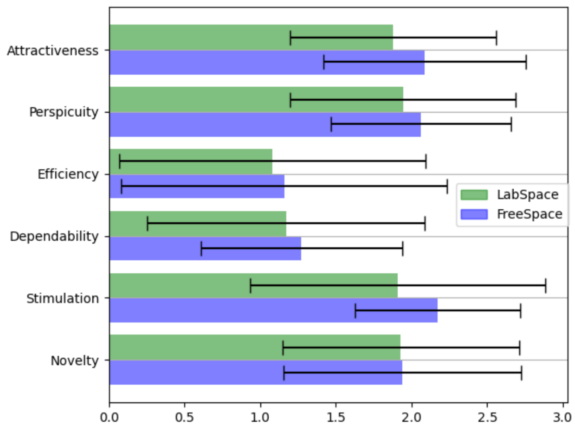
Condition / Variable	LabSpace M(SD)	FreeSpace M(SD)	t(df) / Z*	r	p
UEQ: Attractiveness	1.88(0.69)	2.09(0.68)	$t(21.00) = -1.65$	0.34	.11
UEQ: Perspicuity*	1.94(0.75)	2.06(0.59)	$Z = -0.59$	0.13	.55
UEQ: Efficiency*	1.08(1.01)	1.16(1.08)	$Z = -0.85$	0.18	.39
UEQ: Dependability*	1.17(0.92)	1.27(0.67)	$Z = -0.62$	0.13	.53
UEQ: Stimulation*	1.91(0.98)	2.17(0.55)	$Z = -0.49$	0.10	.62
UEQ: Novelty	1.93(0.78)	1.94(0.78)	$t(21.00) = -0.11$	0.02	.92
FSS: Flow	4.80(0.49)	4.89(0.37)	$t(18.00) = -0.70$	0.16	.49
FSS: Fluency	5.87(0.75)	6.03(0.52)	$t(18.00) = -0.95$	0.22	.35
FSS: Absorption*	5.92(0.88)	6.00(0.62)	$Z = -0.10$	0.02	.94
FSS: Anxiety*	2.07(0.93)	2.14(1.04)	$Z = 0.00$	0.00	.72
FSS: Challenge	4.49(0.91)	4.44(0.94)	$t(18.00) = 0.45$	0.11	.66
CSI: Enjoyment	34.58(23.93)	35.83(23.43)	$t(19.00) = -1.16$	0.26	.26
CSI: Exploration	46.86(17.75)	51.16(23.09)	$t(19.00) = -1.38$	0.30	.18
CSI: Expressiveness	56.56(26.81)	57.78(31.17)	$t(19.00) = -0.33$	0.08	.75
CSI: Immersion	42.68(21.44)	41.42(22.82)	$t(19.00) = 0.50$	0.11	.62
CSI: Worth Effort*	29.34(26.86)	30.93(26.89)	$Z = -1.61$	0.36	.11
CSI: Total*	84.00(19.53)	86.85(22.15)	$Z = -1.46$	0.33	.15
CESR: Competence	5.95(1.90)	6.14(1.52)	$t(18.00) = -0.77$	0.18	.45
CESR: Autonomy	7.16(0.98)	7.34(1.03)	$t(18.00) = -0.88$	0.20	.39
CESR: Enjoyment*	8.06(0.94)	8.29(0.65)	$Z = -1.09$	0.25	.28

*Wilcoxon test applied

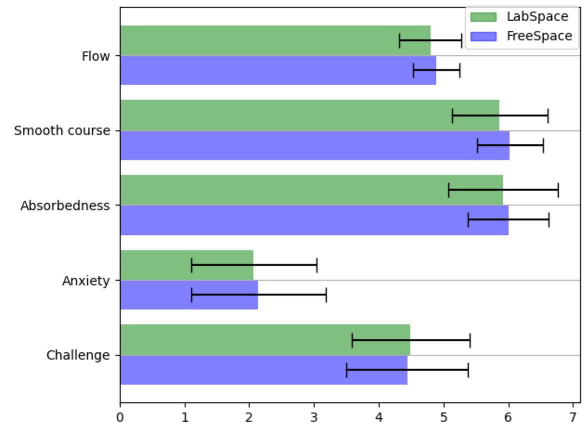
Using both parametric and non-parametric statistical tests, we evaluated all subscale scores of the four questionnaires in Table 1 for differences between conditions *FreeSpace* and *LabSpace*. Mean and Standard Deviations for all of these scores are plotted in Figure 5. Due to a database error, we had to exclude three participants from the FSS and CESR evaluation and two participants from the CSI evaluation and which resulted in $N = 19$ for all flow (FSS) and creative experience (CESR) scores, $N = 20$ for CSI scores and $N = 22$ for all UEQ scores.

Samples were tested for normal distribution using the Shapiro-Wilk test [46] and for equal variances using the Levene test [25]. If the sample data examined turned out to be parametric (this requires that both tests have p values above the significance threshold), a paired-samples t-test [51] was used. If either test failed, the Wilcoxon signed-rank test [56] was used. The significance level was set at $alpha = 0.05$. As we did compare each questionnaire sub-scale individually without any hypotheses, we did not apply any p-value correction as suggested by Rothman [41] and Rubin [42]. Results of t-tests and Wilcoxon tests are reported in Table 2.

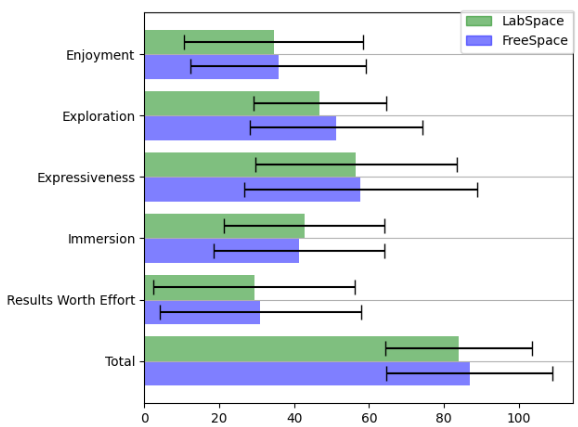
[2] https://microsoft.com/hololens/hardware/
[3] https://unity.com/
[4] https://github.com/microsoft/MixedRealityToolkit-Unity/



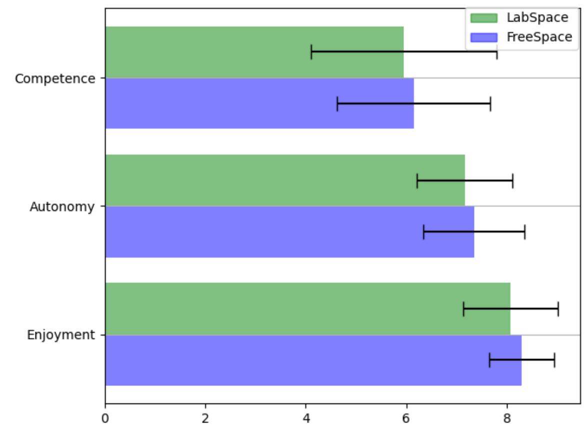
(a) User Experience (UEQ) scores by condition.



(b) Flow (FSS) scores by condition.



(c) Creativity Suppoert (CSI) scores by condition.



(d) Creative Experience (CESR) scores by condition.

Figure 5: Quantitative results of the questionnaires.

As can be seen, we did not observe any statistically significant difference between both conditions for any subscale, as we do not see any p below $alpha = 0.05$. However, we note that for the UEQ subscale *Attractiveness* and the CSI subscales *Exploration*, *Results worth Effort* and the *total* CSI score, even though they are considered as statistically too weak for significance ($0.11 < p < 0.15$), their effect sizes are above the medium threshold ($r > 0.3$) and in favor of the *FreeSpace* condition (compare Figure 5).

4.2 Qualitative Results

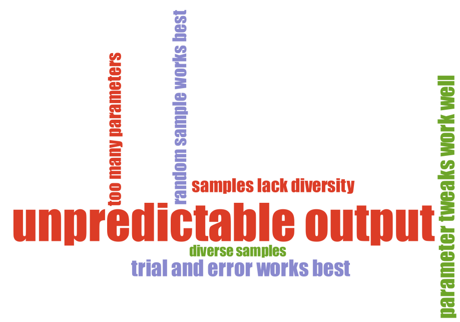
We analyzed the participants' free-form feedback and the open question regarding their creative space preferences through an inductive thematic analysis[3]. Category labels were obtained by summarizing the semantic content of important phrases stated by the participants. The frequencies of category labels were used to create code clouds that can be seen in Figure 6. For the analysis, we used the *MaxQDA*[5] software, which was already used for similar purposes in related work (e.g., [13, 22, 44, 59]). Figure 6a shows the mentioned needs for creative spaces. The most frequently mentioned need was social isolation, which six participants mentioned, followed by concentration and creative stimulation, with four mentions each. Further, one participant stated the need for calmness. Figure 6b depicts the characteristics of creative spaces that participants selected. Here, the most frequently mentioned criterion was *peacefulness*, which was important for twelve participants. Eight participants mentioned *open space*, and six mentioned *nature*. Further, *art-related background* was important to five participants. Three participants mentioned *sunny weather*, while the same number of participants mentioned *avoid bad weather*. *Stimulating background noise*, *fresh air*, *ease of access* and *sense of well-being* each were mentioned two times. Figure 6c and Figure 6d illustrate the open feedback regarding the AR sequencer UI and the VMMs. Here, negative and positive feedback are presented in red and green, respectively. Neutral feedback is colored purple. Seven people expressed *enthusiasm/approval* for the AR sequencer UI. Four participants expressed *excitement* regarding the UI, while three users found the controls *user-friendly*. *Promotes creativity*, *user-friendly UI*, *promotes experimentation*, *unintuitive controls* and *poor UI response* were mentioned two times

[5]https://maxqda.com/

(a) Needs for creative spaces as stated by participants.

(b) Characteristics of selected creative spaces.

(c) Open feedback regarding user experience of the loop editor UI.



(d) Open feedback regarding user experience of the vector manipulation modules (VMMs).

Figure 6: Word clouds/code clouds generated from open open questions (Subfigures 6a and 6b) open feedback (Subfigures 6c and 6d). For the two feedback word clouds, green color means positive, red color negative and purple neutral feedback.

each. For the VMMs, twelve people complained about *unpredictable output*. Three participants stated that the *parameter tweaks work well*. *'trial and error' works best* was reported by three users, while *random sample works best* was mentioned by two users. Two participants said that the *samples lack diversity* and that there were *too many parameters*. One participant acknowledged that there were *diverse samples*.

5 DISCUSSION

Environment choice does not seem to impact quantitative evaluation metrics: We did not observe any significant difference when applying parametric and non-parametric statistics to compare subscale scores between condition *FreeSpace* and *LabSpace*. We did not expect large effect sizes or high statistical effects for User Experience (UEQ) or creative support (CSI) subscales, as the items for both questionnaires are highly focused on the system or device in use, which remained the same between both experimental conditions. However, we were surprised to see no significant effect on Creative Experience (CESR) or Flow (FSS) scales, as these are more focused on the holistic experience of creating music in AR.

Complex musical XR applications for HMD-based AR are feasible: The AR drum sequencer application did perform well in the UEQ scores attractiveness, perspicuity, stimulation, and novelty (all are considered "Excellent" according to Laugwitz et al. [24]). Judging from these scores and the qualitative feedback, people had a positive overall impression of the prototype as they thought it was exciting and novel. Furthermore, they found it easy to get familiar with it. Efficiency and dependability scores, however, are close to average. Similar results were observed by Schlagowski et al. for the original VR application that we ported for HoloLens, where each UEQ scale was rated just slightly higher [45]. Judging from the CESR subscales (See Table 2), users seemed to enjoy the creative experience, regardless of being in the lab or in a freely selectable environment. Further, users felt autonomous and relatively competent while being creative. Hence, we conclude that even though the hardware we used (HoloLens 2) comes with various drawbacks regarding the field of view and occlusion (as it uses optical see-through displays), it can be regarded as a viable testing apparatus that does not pose too many challenges for users and can be a useful creative tool for composing drum beats and designing percussive samples using the VMMs.

Flow may cause a disconnect from the environment, even in AR: In both conditions, users did not feel much anxiety and were slightly above average challenged (see Table 2). Furthermore, they reported a smooth course during the creative sessions and felt heavily absorbed. These subscales contributed to an above-average flow sensation during the creative sessions in both conditions. Ironically, by providing a good user experience and above-average creative support (mean CSI scores of 84.00 and 86.85, see Table 2), our AR prototype might have fostered the likelihood of users entering the state of flow, where users tend to lose connection with time and their surroundings [11]. Thus, the demonstrator might have reduced the effect sizes we saw during the experiment, as users who experienced the state of flow while being creative felt highly absorbed in the activity and thus may have paid little attention to their surroundings.

Familiarity is key when it comes to 3DUI design: While the Loop Editor 3DUI, which could be used to compose drum beats, was primarily met with approval (See Figure 6c), feedback for the Vector Manipulation Modules (VMMs) that could be used for drum sample sound design was mixed (see Figure 6d). A key issue here is that the AI-based drum sample generation is quite chaotic (See [43]) and affordances for such systems remain limited. As such, the audio output of the generative network often remains unpredictable. The loop editor 3DUI however, which was oriented mainly on de-facto standard 2DUI design paradigms for step sequencers [45], did not struggle with such issues. Hence, when designing musical XR systems that are similar to our drum sequencer application, we advise developers and researchers to stick to design paradigms that are well-known to the target user or persona, even if the use of the third dimension that is available for 3DUI design in XR remains limited. By doing so, affordances are better understood and users can concentrate on mastering potentially unused/new interaction techniques while interacting with an unfamiliar device (in our case, HoloLens 2).

Laboratory restrictions are likely to contradict creative needs: Effect sizes may have been further reduced by confounding effects arising from the design and conduct of the study. Even though participants were offered the opportunity to meet them at home and be creative in their familiar creative environment during the *FreeSpace* condition, every participant chose to visit the university campus for both creative sessions and choose creative environments from our catalog of pre-selected spaces. Hence, participants were not only creative while using an entirely novel system but also in unfamiliar environments in both conditions, which may have further reduced the observed effect sizes. Furthermore, laboratory experiments (be they in the lab or not) require personnel to supervise the experiment and be there, e.g., if a technical error occurs or to ensure that the experiment procedure follows the study protocol. Hence, one of the most frequently uttered needs for creative spaces, social isolation (See word cloud in Figure 6a), could not be fulfilled. Furthermore, even though we did not stick to fixed time frames during our experiment, most participants were limited in their time availability. As such, the creative experience, support, flow, and user experience of our AR application could not be verified in an unaltered way during our experiment.

Comfort zones need to be respected: We observed that participants in the *FreeSpace* condition were eager to explore new areas and try out unfamiliar creative spaces. Especially for more introverted participants, the sensation of physically moving out of their "comfort zone" was perceived as inspiring. However, getting out of one's creative habits or spaces does not need to be inspiring for everyone and in every situation. Instead, finding the right balance of familiarity and novelty that results in creative inspiration can be difficult and needs to be adjusted according to situational factors such as mood, time of day, and weather. In order to provide ease of access to familiar creative situations, we highly recommend unsupervised field studies where participants are free to be creative not only *wherever*, but also *whenever* and for *how long* they want.

6 CONCLUSION

We reported on an experiment in which we investigated the impact of free user environment choice on quantitative evaluation metrics for XR-based creative experiences. During this study, participants interacted with an HMD-based drum sequencer to compose drum beats, and designed drum sounds either in a laboratory environment or an open creative space of their choice. Looking at questionnaire data, we did not observe a difference between both conditions for user experience, flow, creativity support, and creative experience scores. Considering these scores and qualitative feedback from participants, we discussed these observations in order to find potential reasons for seeing no effect. Further, we looked at open feedback and selection criteria to derive recommendations for developers in the creative AR domain and researchers who aim to conduct experiments evaluating creative AR experiences. Due to high user experience and flow scores, we encourage further development of mixed reality based applications for music production. Furthermore, we emphasize the importance of unsupervised field studies for the evaluation of such applications.

ACKNOWLEDGMENTS

This paper was partially funded by the DFG through the Leibniz award of Elisabeth André (AN 559/10-1).

REFERENCES

[1] Valentina Rita Andolfi, Chiara Di Nuzzo, and Alessandro Antonietti. 2017. Opening the mind through the body: The effects of posture on creative processes. *Thinking Skills and Creativity* 24 (2017), 20–28. https://doi.org/10.1016/j.tsc.2017.02.012

[2] Akshay Bhagwatwar, Anne Massey, and Alan R. Dennis. 2013. Creative virtual environments: Effect of supraliminal priming on team brainstorming. *Proceedings of the Annual Hawaii International Conference on System Sciences* (2013), 215–224. https://doi.org/10.1109/HICSS.2013.152

[3] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. (2012).

[4] Stéphanie Buisine, Jérôme Guegan, Jessy Barré, Frédéric Segonds, and Améziane Aoussat. 2016. Using avatars to tailor ideation process to innovation strategy. *Cognition, Technology and Work* 18, 3 (2016), 583–594. https://doi.org/10.1007/s10111-016-0378-y

[5] Canan Ceylan, Jan Dul, and Serpil Aytac. 2008. Can the office environment stimulate a manager's creativity? *Human Factors and Ergonomics In Manufacturing* 18, 6 (2008), 589–602. https://doi.org/10.1002/hfm.20128

[6] Tilanka Chandrasekera. 2015. *The effect of augmented reality and virtual reality interfaces on epistemic actions and the creative process.* University of Missouri-Columbia.

[7] Yu Shan Chang, Chia Hui Chou, Meng Jung Chuang, Wen Hung Li, and I. Fan Tsai. 2020. Effects of virtual reality on creative design performance and creative experiential learning. *Interactive Learning Environments* 0, 0 (2020), 1–16. https://doi.org/10.1080/10494820.2020.1821717

[8] Erin Cherry and Celine Latulipe. 2014. Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction* 21, 4 (2014). https://doi.org/10.1145/2617588

[9] Vicente Chulvi, María Jesús Agost, Marta Royo, and Carlos García-García. 2020. The effect of nature on designers' creativity, according to their personality profile. *Alexandria Engineering Journal* 59, 2 (2020), 987–998. https://doi.org/10.1016/j.aej.2020.03.036

[10] Lorenza S. Colzato, Ayca Szapora, Justine N. Pannekoek, and Bernhard Hommel. 2013. The impact of physical exercise on convergent and divergent thinking. *Frontiers in Human Neuroscience* 7, DEC (2013), 1–6. https://doi.org/10.3389/fnhum.2013.00824

[11] Mihaly Csikszentmihalyi. 2000. *Beyond boredom and anxiety.* Jossey-Bass.

[12] Darren W. Dahl and C. Page Moreau. 2007. Thinking inside the box: Why consumers enjoy constrained creative experiences. *Journal of Marketing Research* 44, 3 (2007), 357–369. https://doi.org/10.1509/jmkr.44.3.357

[13] Daniel Diethei, Jasmin Niess, Carolin Stellmacher, Evropi Stefanidi, and Johannes Schöning. 2021. Sharing heartbeats: motivations of citizen scientists in times of crises. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[14] Allan Filipowicz. 2006. From positive affect to creativity: The surprising role of surprise. *Creativity Research Journal* 18, 2 (2006), 141–152. https://doi.org/10.1207/s15326934crj1802_2

[15] Kenton B. Fillingim, Hannah Shapiro, Catherine J. Reichling, and Katherine Fu. 2021. Effect of physical activity through virtual reality on design creativity. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM* 35, 1 (2021), 99–115. https://doi.org/10.1017/S0890060420000529

[16] Thomas Fröhlich, Dmitry Alexandrovsky, Timo Stabbert, Tanja Döring, and Rainer Malaka. 2018. VRBox: A virtual reality augmented sandbox for immersive playfulness, creativity and exploration. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. 153–162.

[17] Jérôme Guegan, Stéphanie Buisine, Fabrice Mantelet, Nicolas Maranzana, and Frédéric Segonds. 2016. Avatar-mediated creativity: When embodying inventors makes engineers more creative. *Computers in Human Behavior* 61 (2016), 165–175. https://doi.org/10.1016/j.chb.2016.03.024

[18] Jérôme Guegan, Julien Nelson, and Todd Lubart. 2017. The Relationship between Contextual Cues in Virtual Environments and Creative Processes. *Cyberpsychology, Behavior, and Social Networking* 20, 3 (2017), 202–206. https://doi.org/10.1089/cyber.2016.0503

[19] Junko Ichino, Aura Pon, Ehud Sharlin, David Eagle, and Sheelagh Carpendale. 2014. Vuzik: The effect of large gesture interaction on children's creative musical expression. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design*. 240–249.

[20] Shu Imaizumi, Ubuka Tagami, and Yi Yang. 2020. Fluid movements enhance creative fluency: A replication of Slepian and Ambady (2012). *PLoS ONE* 15, 7 July (2020), 1–10. https://doi.org/10.1371/journal.pone.0236825

[21] Olga Kombeiz and Anna Steidle. 2018. Facilitation of creative performance by using blue and red accent lighting in work and learning areas. *Ergonomics* 61, 3 (2018), 456–463. https://doi.org/10.1080/00140139.2017.1349940

[22] Veronika Krauß, Alexander Boden, Leif Oppermann, and René Reiners. 2021. Current practices, challenges, and design implications for collaborative AR/VR application development. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[23] Li Lan, Sarra Hadji, Lulu Xia, and Zhiwei Lian. 2021. The effects of light illuminance and correlated color temperature on mood and creativity. *Building Simulation* 14, 3 (2021), 463–475. https://doi.org/10.1007/s12273-020-0652-z

[24] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*. Springer, 63–76.

[25] Howard Levene. 1961. Robust tests for equality of variances. *Contributions to probability and statistics. Essays in honor of Harold Hotelling* (1961), 279–292.

[26] Stephanie Lichtenfeld, Andrew J. Elliot, Markus A. Maier, and Reinhard Pekrun. 2012. Fertile Green: Green Facilitates Creative Performance. *Personality and Social Psychology Bulletin* 38, 6 (2012), 784–797. https://doi.org/10.1177/0146167212436611

[27] Yu-Ju Lin and Hung-chun Wang. 2021. Using virtual reality to facilitate learners' creative self-efficacy and intrinsic motivation in an EFL classroom. *Education and Information Technologies* 26, 4 (2021), 4487–4505.

[28] Alex Marin, Martin Reimann, and Raquel Castaño. 2014. Metaphors and creativity: Direct, moderating, and mediating effects. *Journal of Consumer Psychology* 24, 2 (2014), 290–297. https://doi.org/10.1016/j.jcps.2013.11.001

[29] Janetta Mitchell McCoy and Gary W. Evans. 2002. The potential role of the physical environment in fostering creativity. *Creativity Research Journal* 14, 3-4 (2002), 409–426. https://doi.org/10.1207/S15326934CRJ1434_11

[30] Ravi Mehta, Rui(Juliet) Zhu, and Amar Cheema. 2012. Is noise always bad? exploring the effects of ambient noise on creative cognition. *Journal of Consumer Research* 39, 4 (2012), 784–799. https://doi.org/10.1086/665048

[31] Ravi P. Mehta and Rui Juliet Zhu. 2009. Blue or red? Exploring the effect of color on cognitive task performances. *Advances in Consumer Research* 36, February (2009), 1045–1046.

[32] Nicolas Michinov and Estelle Michinov. 2020. Do Open or Closed Postures Boost Creative Performance? The Effects of Postural Feedback on Divergent and Convergent Thinking. *Psychology of Aesthetics, Creativity, and the Arts* (2020). https://doi.org/10.1037/aca0000306

[33] Randall K. Minas, Alan R. Dennis, and Anne P. Massey. 2016. Opening the mind: designing 3d virtual environments to enhance team creativity. *Proceedings of the Annual Hawaii International Conference on System Sciences* 2016-March (2016), 247–256. https://doi.org/10.1109/HICSS.2016.38

[34] Thomas Muender, Anke V Reinschluessel, Sean Drewes, Dirk Wenig, Tanja Döring, and Rainer Malaka. 2019. Does it feel real? Using tangibles with different fidelities to build and explore scenes in virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[35] Julien Nelson and Jérôme Guegan. 2019. "I'd like to be under the sea": Contextual cues in virtual environments influence the orientation of idea generation. *Computers in Human Behavior* 90, August 2018 (2019), 93–102. https://doi.org/10.1016/j.chb.2018.08.001

[36] Samah Obeid and Halime Demirkan. 2020. The influence of virtual reality on design process creativity in basic design studios. *Interactive Learning Environments* (2020), 1–19.

[37] Linda A. O'Hara and Robert J. Sternberg. 2001. It doesn't hurt to ask: Effects of instructions to be creative, practical, or analytical on essay-writing performance and their interaction with students' thinking styles. *Creativity Research Journal* 13, 2 (2001), 197–210. https://doi.org/10.1207/s15326934crj1302_7

[38] Marily Oppezzo and Daniel L. Schwartz. 2014. Give your ideas some legs: The positive effect of walking on creative thinking. *Journal of Experimental Psychology: Learning Memory and Cognition* 40, 4 (2014), 1142–1152. https://doi.org/10.1037/a0036577

[39] Sophia Ppali, Vali Lalioti, Boyd Branch, Chee Siang Ang, Andrew J Thomas, Bea S Wohl, and Alexandra Covaci. 2022. Keep the VRhythm going: A musician-centred study investigating how Virtual Reality can support creative musical practice. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.

[40] Falko Rheinberg, Regina Vollmeyer, and Stefan Engeser. 2003. Die Erfassung des Flow-Erlebens. In *Diagnostik von Motivation und Selbstkonzept (Tests und Trends – Jahrbuch der pädagogisch-psychologischen Diagnostik - Band 2)*. 261–279.

[41] Kenneth J Rothman. 1990. No adjustments are needed for multiple comparisons. *Epidemiology* (1990), 43–46.

[42] Mark Rubin. 2021. When to adjust alpha during multiple testing: A consideration of disjunction, conjunction, and individual testing. *Synthese* 199, 3 (2021), 10969–11000.

[43] Ruben Schlagowski, Silvan Mertes, and Elisabeth André. 2021. Taming the Chaos: Exploring Graphical Input Vector Manipulation User Interfaces for GANs in a Musical Context. In *Audio Mostly 2021*. ACM, virtual/Trento Italy, 216–223. https://doi.org/10.1145/3478384.3478411

[44] Ruben Schlagowski, Dariia Nazarenko, Yekta Can, Kunal Gupta, Silvan Mertes, Mark Billinghurst, and Elisabeth André. 2023. Wish You Were Here: Mental and Physiological Effects of Remote Music Collaboration in Mixed Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*.

[45] Ruben Schlagowski, Fabian Wildgrube, Silvan Mertes, Ceenu George, and Elisabeth André. 2022. Flow with the beat! Human-centered design of virtual environments for musical creativity support in VR. In *Creativity and cognition*. 428–442.

[46] Samuel Sanford Shapiro and Martin B Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52, 3/4 (1965), 591–611.

[47] Seiji Shibata and Naoto Suzuki. 2004. Effects of an indoor plant on creative task performance and mood. *Scandinavian Journal of Psychology* 45, 5 (2004), 373–381. https://doi.org/10.1111/j.1467-9450.2004.00419.x

[48] Michael L. Slepian and Nalini Ambady. 2012. Fluid movement and creativity. *Journal of Experimental Psychology: General* 141, 4 (2012), 625–629. https://doi.org/10.1037/a0027395

[49] Anna Steidle and Lioba Werth. 2013. Freedom from constraints: Darkness and dim illumination promote creativity. *Journal of Environmental Psychology* 35 (2013), 67–80. https://doi.org/10.1016/j.jenvp.2013.05.003

[50] Hannah Steinberg. 1997. Exercise enhances creativity independently of mood. *British Journal of Sports Medicine* 31, 3 (1997), 240–245. https://doi.org/10.1136/bjsm.31.3.240

[51] Student. 1908. The probable error of a mean. *Biometrika* (1908), 1–25.

[52] Sylvie Studente, Nina Seppala, and Noemi Sadowska. 2016. Facilitating creative thinking in the classroom: Investigating the effects of plants and the colour green on visual and verbal creativity. *Thinking Skills and Creativity* 19 (2016), 1–8. https://doi.org/10.1016/j.tsc.2015.09.001

[53] Rafael Valer, Rodrigo Schramm, and Luciana Porcher Nedel. 2020. Musical Brush: Exploring Creativity in an AR-based Tool Combining Music and Drawing Generation. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* March (2020), 635–636. https://doi.org/10.1109/VRW50115.2020.00167

[54] Kathleen D. Vohs, Joseph P. Redden, and Ryan Rahinel. 2013. Physical Order Produces Healthy Choices, Generosity, and Conventionality, Whereas Disorder Produces Creativity. *Psychological Science* 24, 9 (2013), 1860–1867. https://doi.org/10.1177/0956797613480186

[55] Xinyue Wang, Kelong Lu, Mark A. Runco, and Ning Hao. 2018. Break the "wall" and become creative: Enacting embodied metaphors in virtual reality. *Consciousness and Cognition* 62, October 2017 (2018), 102–109. https://doi.org/10.1016/j.concog.2018.03.004

[56] Frank Wilcoxon. 1992. Individual comparisons by ranking methods. In *Breakthroughs in statistics*. Springer, 196–202.

[57] Xiaozhe Yang, Lin Lin, Pei Yu Cheng, Xue Yang, Youqun Ren, and Yueh Min Huang. 2018. Examining creativity through a virtual reality support system. *Educational Technology Research and Development* 66, 5 (2018), 1231–1254. https://doi.org/10.1007/s11423-018-9604-z

[58] Rabia Meryem Yilmaz and Yuksel Goktas. 2017. Using augmented reality technology in storytelling activities: examining elementary students' narrative skill and creativity. *Virtual Reality* 21 (2017), 75–89.

[59] Tianyi Zhang, Björn Hartmann, Miryung Kim, and Elena L Glassman. 2020. Enabling data-driven api design with community usage data: A need-finding study. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[60] Yanyun Zhou, Yifei Zhang, Bernhard Hommel, and Hao Zhang. 2017. The impact of bodily states on divergent thinking: Evidence for a control-depletion account. *Frontiers in Psychology* 8, SEP (2017), 1–9. https://doi.org/10.3389/fpsyg.2017.01546