Investigating the Effects of Individual Spatial Abilities on Virtual Reality Object Manipulation

Tobias Drey
Institute of Media Informatics, Ulm University
Ulm, Germany
tobias.drey@uni-ulm.de

Michael Montag
Department of General Psychology, Magdeburg-Stendal University of Applied Science
Magdeburg, Germany
michael.montag@h2.de

Andrea Vogt
Institute of Psychology and Education, Ulm University
Ulm, Germany
andrea.vogt@uni-ulm.de

Nico Rixen
Institute of Media Informatics, Ulm University
Ulm, Germany
nico.rixen@uni-ulm.de

Tina Seufert
Institute of Psychology and Education, Ulm University
Ulm, Germany
tina.seufert@uni-ulm.de

Steffi Zander
Department of General Psychology, Magdeburg-Stendal University of Applied Science
Magdeburg, Germany
steffi.zander@h2.de

Michael Rietzler
Institute of Media Informatics, Ulm University
Ulm, Germany
michael.rietzler@uni-ulm.de

Enrico Rukzio
Institute of Media Informatics, Ulm University
Ulm, Germany
enrico.rukzio@uni-ulm.de

ABSTRACT

Object manipulation in 3D space, meaning translating, rotating, and scaling, is ubiquitous in virtual reality (VR), and several interaction techniques have been developed in the past to optimize the task performance and usability. However, preliminary research indicates that individual spatial abilities also have an impact. Yet, it was never investigated if users' spatial abilities influence VR object manipulation. We assessed this in a user study (N=66) using 21 manipulation tasks defined in a Fitts' law-related approach. As interaction techniques, we chose gizmos for simultaneously manipulating 1 and 3 degrees of freedom (DOF) and a handle bar metaphor for 7 DOF. Higher spatial abilities resulted in significantly shorter task completion time and more targeted manipulations, while task accuracy was unaffected. However, an optimized interaction technique could compensate individual disadvantages. We propose seven guidelines on spatial abilities in interaction technique design and research to personalize and improve VR applications.

CCS CONCEPTS

• **Human-centered computing** → *HCI design and evaluation methods*; **Virtual reality**; **Mixed / augmented reality**; *Interaction devices*; **Interaction techniques**; **Interaction design theory, concepts and paradigms**.

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3581004

KEYWORDS

spatial abilities; virtual reality; mixed reality; object manipulation; interaction technique; individual characteristics; docking task; Fitts' law

ACM Reference Format:
Tobias Drey, Michael Montag, Andrea Vogt, Nico Rixen, Tina Seufert, Steffi Zander, Michael Rietzler, and Enrico Rukzio. 2023. Investigating the Effects of Individual Spatial Abilities on Virtual Reality Object Manipulation. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 24 pages. https://doi.org/10.1145/3544548.3581004

1 INTRODUCTION

Manipulating objects, meaning translating, rotating, and scaling, are one of the primary and most frequent interactions in virtual reality (VR) [51]. They are used to move simple objects, solve riddles in (serious) games, and create complex three-dimensional construction drawings using computer-aided design (CAD). Therefore, the design of VR object manipulation techniques is essential for the interface's task performance and perceived usability. This is why multiple works have proposed optimized solutions in the past (e.g., comparing a new interaction technique against a baseline [6, 35, 51, 58]), as the overviews of Mendes et al. [58] and Bergström et al. [6] show.

Another important factor influencing the interface's task performance and perceived usability could be the individual spatial abilities, meaning a person's ability to understand their surroundings, as first studies show superior results for individuals with higher spatial abilities for such measures [3–5, 12, 49]. Based on these results, we expect that spatial abilities also significantly affect the VR object manipulation tasks and their typical measures, such as task completion time and accuracy [6]. Despite these assumed

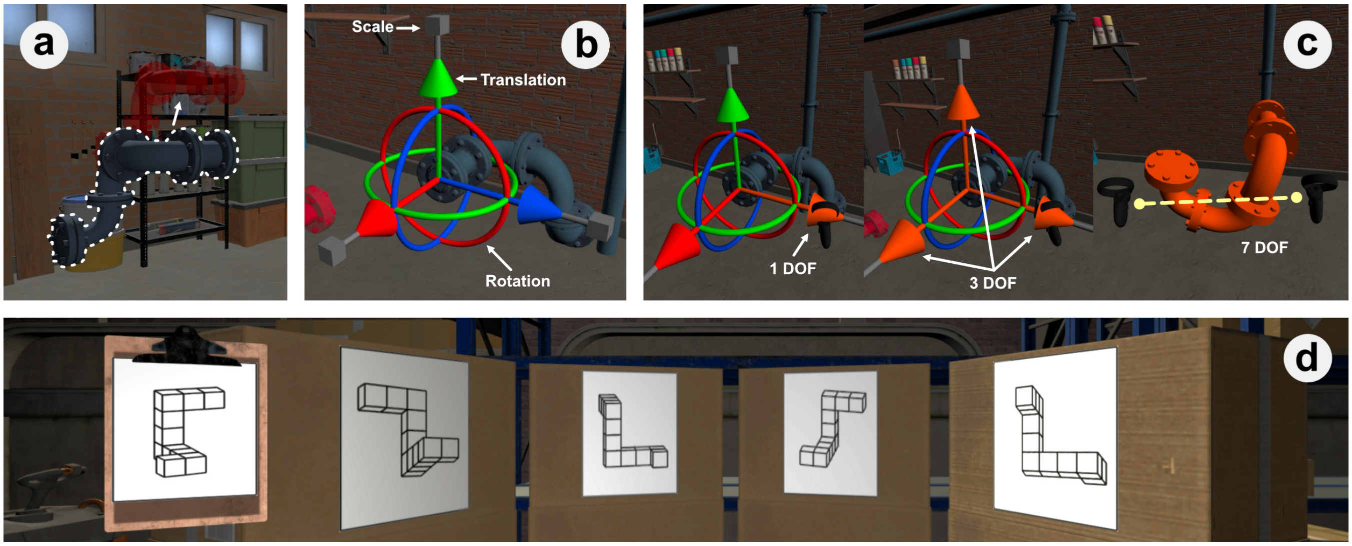

Figure 1: We developed a virtual reality application to investigate in a user study (N=66) how the individual spatial abilities, measured with the test by Vandenberg & Kuse [80] and Peters et al. [64] (d), influence the task performance in object manipulation tasks (a). The objects could be manipulated (translated, rotated, scaled, (b)) with either gizmos for 1 degrees of freedom (DOF) (left picture in (c)) and 3 DOF manipulations (middle picture in (c)) as well as a handle bar metaphor for 7 DOF manipulations (right picture and yellow line in (c)). The currently manipulable axes (1 and 3 DOF) or the whole object (handle bar, 7 DOF) are highlighted in orange. Our findings show a significant effect of increasing spatial abilities on reducing task completion time and using a more targeted problem-solving approach in the object manipulation tasks (a).

performance effects on nearly every VR application using object manipulation, this was never systematically investigated in past research. Therefore, we defined the following exploratory research question (RQ): *"Do individuals with higher spatial abilities outperform individuals with lower spatial abilities in VR mid-air object manipulation tasks?"*

To answer our RQ, we conducted an experimental comparison of mid-air object manipulation techniques in VR with 66 participants and correlated the results with their individual spatial abilities (see Fig. 1 / 3) by exploratory analyzing several dependent variables. We determined their spatial abilities by measuring their mental rotation capabilities with a test based on the well-established Shepard and Metzler [69] as well as Vandenberg & Kuse [80] tests (see Fig. 1d). Participants were required to complete 21 manipulations/docking tasks of varying difficulty, similar as proposed by Bergström et al. [6]. We further present an own Fitts' law adaption for that as nothing suitable for controlling the difficulty of our tasks was defined unit now [78] (see Fig. 1a). Our two main metrics were task completion time and task accuracy (final error/deviation to the target) to measure our participants' performance. To provide generalizable and broadly relevant results, we focused on interaction techniques applicable to a standard head-mounted display (HMD) setup based on mid-air controller input [18]. In contrast to previous works that focused on subsets [6, 51, 58], we wanted to provide holistic findings regarding all possible degrees of freedom (DOF) provided by object manipulation. The DOF are nine in total: three for translation, three for rotation, and three for scaling. We chose as the first interaction technique the established gizmos/widgets as used in Unity [79] or Blender [8] for simultaneous 1 DOF and 3 DOF object manipulation (see Fig. 1b/c). Difficult manipulations must be decomposed and performed sequentially when the task needs more manipulations than simultaneously possible with the respective gizmos. We further chose the handle bar metaphor [71] for simultaneous 7 DOF object manipulation (see Fig. 1c). It can handle all manipulations simultaneously, and participants needed no decomposition, even for the most difficult task in our study.

Due to its high level of standardization as well as its generalizability, our study design is a methodological contribution according to Wobbrock and Kientz [85]. It can serve as a template and ease researchers' work in defining and conducting highly standardized object manipulation studies based on docking tasks in a Fitts' law-like manner (possible target group: [7, 9, 14, 30, 35, 41, 44, 47, 59–61, 75, 77, 83]). Furthermore, we demonstrated how individual characteristics such as spatial abilities should be considered by design. Both provisions will improve the comparability of docking tasks results across studies in the future (see Kulik et al. [48]), which is why we have added them to our guidelines.

The findings of our exploratory study show that, as assumed, higher spatial abilities significantly improve task completion time and result in more targeted manipulations. We found no effect on task accuracy. Very important is also our interpretation of the results, that an optimized interaction technique can compensate for lower spatial abilities. We saw this by the results of the handle bar, which, in our study, resulted in the best performance - independent of spatial ability. It also shows that providing multiple simultaneous DOF (for the handle bar 7 DOF) can be advantageous. To summarize, when designing a VR application, it is critical to consider both individual spatial abilities and an optimized interaction technique.

We advise researchers and designers to do so in the future and, therefore, created seven guidelines for interaction technique design, implementation, and user research.

Our main contributions are as follows:

(1) Findings from an exploratory user study (N=66) regarding the correlations of individual spatial abilities and task completion time, task accuracy, targeted object manipulations, cognitive load, usability, interaction technique, and task difficulty measured by eleven dependent and five control variables.
(2) The definition of seven guidelines on how to consider the individual spatial abilities of participants in VR interaction technique design as well as associated user studies.
(3) The definition of a study design template that considers users' individual characteristics for object manipulation interaction technique studies based on docking tasks which are defined by formulas derived from Fitts' law. It will ease defining and conducting studies in a standardized way to improve across-study result comparability.

2 RELATED WORK

Our work is based on previous work regarding spatial abilities and object manipulation, focusing on how to measure and compare interaction techniques with docking tasks.

2.1 Spatial Abilities

We will first explain spatial abilities before we show selected examples where spatial abilities were considered in human-computer interaction (HCI) research.

2.1.1 Introduction to Spatial Abilities. Spatial abilities are an inherent characteristic of humans and describe their ability to understand their surrounding environment as well as objects inside it. Most definitions assume three fundamental abilities: spatial perception, spatial visualization, and mental rotation [52, 82]. Other definitions include other abilities, such as spatial orientation, spatial relation, and spatial transformation as a separate dimension of spatial abilities or as a subcategory of one of the first three abilities [13, 38, 40, 52, 86]. Spatial abilities are further not isolated capabilities and are linked to the working memory and cognitive load [43, 76]. Previous works have shown that spatial abilities correlate with participants' performance in tasks where spatial skills are required, such as assembly tasks [12] or immersive 3D drawing [3]. Several tests were developed to measure spatial abilities or sub-categories. One of the first were Shepard and Metzler [69] who developed a test based on objects made of cubes placed side by side in different orientations to measure mental rotation capabilities. Participants had to decide whether these cube objects were identical or not. This work will use this principle by using the test of Vandenberg & Kuse [80], which was frequently used as a mental rotation test in the past (see Zander et al. [86], and Peters et al. [64]). The test shows one reference cube object and four other cube objects, where two are identical to the reference object but rotated, and two are different (see Fig. 1d). Both objects identical to the reference object have to be selected. Depending on the version, the test consists of 20 to 24 trials. We used a 12 task subset of the redrawn 24 trial version of Peters et al. [64]. Using a subset reduced its duration and eased its feasibility, as Peters et al. [65] did for digital versions.

2.1.2 Spatial Abilities in HCI Research. Bowman et al. [10] defined five categories influencing performance in virtual environment (VE): the interaction technique, the task, the environment, the system, and the users with, e.g., different spatial abilities. While the firsts can be defined and controlled in a study, spatial abilities are always user-dependent. In this work, we will focus on the influence of individual spatial abilities on object manipulation task performance by controlling the other four in our study.

Despite spatial abilities being an essential factor for interaction in HCI, previous research has seldom investigated their influence, as the survey papers and overviews of LaViola et al. [51], Mendes et al. [58], and Bergström et al. [6] show. Two works doing so were conducted by Barrett and Hegarty [4, 5], who considered spatial abilities in their studies as moderators for performance in virtual molecule manipulation tasks and varied the display dimensionality (stereo vs. mono) and the hand-held device location (co-located vs. displaced). Only participants with lower spatial abilities benefited from co-located interfaces and stereoscopic view. Carlson et al. [12] conducted a study about learning assembly tasks and showed that higher spatial abilities lead to higher assembly performance. Barrera-Machuca et al. [3] showed a positive effect of spatial abilities on shape likeliness in a 3D mid-air sketching task. However, line precision was not affected in their work. Lages and Bowman [49] investigated if there were differences in VR 3D data examination when walking around or moving the data itself. They found that high spatial abilities could improve task completion time when walking. These works show why we expect a similar positive effect of increasing spatial abilities on object manipulation in VR, which we will systematically investigate in this work.

2.2 Object Manipulation in VR

In this section, we will first define what we will consider as object manipulation in this work before we explain how it is related to the DOF, show the state of the art of VR object manipulation techniques, and how they can be evaluated with docking tasks.

2.2.1 Canonical Object Manipulation Tasks. According to LaViola et al. [51], object interaction can be defined with the canonical object manipulation tasks: selection, positioning/translation, rotation, and scaling. These four basic tasks are building blocks for more complex interactions [51, 62]. Previous research has mainly focused on object selection, followed by object manipulation with rotation and sometimes translation, as the survey papers and overviews of LaViola et al. [51], Mendes et al. [58], and Bergström et al. [6] show. Their works further show that scaling is scarcely investigated, and object manipulation with combined rotation, translation, and scale manipulation also only in very seldom cases.

This work provides a holistic evaluation of object manipulation and focuses, therefore, on all three canonical manipulation tasks, namely rotation, translation, and scale, and investigates the influence of spatial abilities on them. As selection does not manipulate objects, we excluded it from our evaluation.

2.2.2 DOF in Object Manipulation. The degrees of freedom (DOF) in object manipulation can be defined (1) as the absolute number of control dimensions in a task and (2) as the number of simultaneously controllable dimensions by an interaction technique [51]. Translation, rotation, and scale have each three manipulation axes resulting in nine possible DOF. The absolute number of control dimensions of a task (1) describes if translation, rotation, and scale are possible, which results in one (one axis possible) or up to nine (all axis possible) DOF. In the following, we will refer to this as *absolute task DOF*. The number of simultaneously controllable dimensions by an interaction technique (2) describes how many of the absolute DOF of the task can be manipulated by the interaction technique at the same time [51]. In the following, we will refer to this as *simultaneous interaction technique DOF*. In our study, we use both (1) the *absolute task DOF* and (2) the *simultaneous interaction technique DOF* as independent variables.

As 3D interaction is difficult and offers potential for improvement [24, 39, 41, 61], previous works have investigated how many simultaneous DOF for an interaction technique are appropriate for object manipulation [51]. For example, LaViola et al. [51] proposed to reduce the interaction technique's DOF to reduce effort and increase precision. This was investigated by Veit et al. [81] with a rotation task where they compared a simultaneous 3 DOF interaction technique with a technique with only 1 DOF. Participants using only 1 DOF were faster, while the achieved precision was the same. Decomposing the task, meaning rotating the axis sequentially, was very important for the participants. Contrary results exist by Kulik et al. [48]. They showed that an interaction technique combining translation and rotation (6 DOF simultaneously) leads to a shorter task completion time than decomposing them, meaning performing all necessary translations and rotations sequentially. Similar results from Mendes et al. [61] exist that show that participants preferred simultaneous 3 DOF over 1 DOF interaction techniques for object manipulation. This is in line with interaction techniques such as the handle bar [71], which show that object manipulation techniques for manipulating up to 7 simultaneous DOF can be created and successfully used. Techniques for 8 or 9 simultaneous DOF object manipulation for VR do not exist at the moment [58].

These works have in common that they do not consider the spatial abilities of the participants in their study, which may have an influence on the results and may explain why contrary results for how many DOF are best for interaction exist. This assumption is the base for our research question.

2.2.3 Object Manipulation Techniques. Object manipulation techniques for VR were frequently investigated in the past [6, 58]. However, their study results comparability is limited, as the used hardware and interaction techniques vary strongly and are often optimized for specific use cases. This could be another explanation for the contradicting study results presented in Section 2.2.2. To prevent such effects in this work, we used our own criteria to select appropriate VR interaction techniques for this work (see Section 3.1.1). Using criteria was necessary, as Mendes et al. [58] showed, that there were plenty of techniques investigated in the past, even for our use-case mid-air object manipulation for VR. Different tracking technologies (e.g., controllers/handheld devices, wearable devices, or hand-tracking), multiple input mappings and metaphors (e.g., grabbing or handles), direct and indirect interactions (e.g., in the vicinity of the user or out-of-reach), and scaffold or manipulation aids to increase precision (e.g., snapping, scale down movements, or widgets) exist [58]. Besides this brief summary, we will not provide a holistic overview of manipulation techniques, which is a topic on its own, and refer for a broader overview to corresponding survey papers (see Mendes et al. [58], and Bergström et al. [6]).

2.2.4 Docking Tasks and Fitts' Law in 3D. Whereas Fitts' law [34] exists for comparing object selection interaction techniques, no equally standardized model exists for comparison of 3D object manipulation. Nevertheless, many previous works have used so-called docking task as a setup for comparisons [14, 44, 58, 60]. The goal of a docking task is to manipulate a given 3D source object by applying translation, rotation, and scale manipulations so that it matches a target object. The target object is often visualized as semitransparent, and the 3D object has to fit inside the target perfectly. Docking tasks can be evaluated regarding completion time and/or accuracy. Task completion time is often measured until the 3D object is fitted within a predefined threshold, which should be done as fast as possible [44]. In contrast, accuracy means that not the time counts it takes to manipulate the object, but the highest achievable accuracy should be reached [60].

Similar to Fitts' law, which uses the index of difficulty [54, 55, 72], docking tasks need to be definable with different levels of difficulty, which makes them further comparable if multiple trials are used within a study. The difficulty of docking tasks can be varied by changing the amount objects have to be translated, rotated, or scaled [6, 14, 41]. As this is a somehow similar approach to changing the target distance and width for the index of difficulty in Fitts' law studies [55, 72], previous works focused on transforming Fitts' law into 3D to apply it to translation and rotation object manipulation [74, 78]. This work will build on the formulas proposed by Stoelen and Akin [74] (see Section 3.1.2). They adapted Fitts' law and its index of difficulty to translation and rotation manipulation tasks to define docking tasks with comparable difficulty. Kulik et al. [48] confirmed these formulas as appropriate based on a user study.

3 USER STUDY

According to LaViola et al. [51], object manipulation can be defined by (1) the absolute number of control dimensions as well as (2) the number of simultaneously controllable dimensions (see Section 2.2.2). In our case, these are defined by (1) the object manipulation task and (2) the interaction technique, which we used as independent variables in the following. Further, we will show how our study design is linked to Fitts' law, our dependent and control variables, our system's design and technical details, the study procedure, and details about our N=66 participants. We oriented the design of our study on the recommendations and the checklist for *VR object selection and manipulation studies* provided by Bergström et al. [6].

3.1 Independent Variables

In the following, we will show our two independent variables *interaction techniques* and *manipulation tasks* and explain how we have chosen the conditions. This is followed by a study design overview, including a power analysis.

3.1.1 Interaction Techniques. The three canonical object manipulations (translation, rotation, scale; see LaViola et al. [51]) can have up to 3 DOF each, which results in a spectrum from 1 to 9 DOF (see Section 2.2.2). To inspect our RQ (see Section 1), we investigate the whole range of the spectrum. As we did not want to invent a new interaction technique but investigate already existing and well-established ones (see LaViola et al. [51]), we selected them based on the overview of Mendes et al. [58].

We started our selection on the lower end with 1 DOF, allowing to manipulate only one axis of the object simultaneously (either translation/rotation/scale). Therefore, conducting complex object manipulation tasks requires decomposing the task and performing several manipulations sequentially (e.g., first conduct the x-axis translation, then the y-axis translation, and then the x-axis rotation). 1 DOF manipulations are possible in applications such as Unity [79] or Blender [8] by using widgets or gizmos (see Fig. 1b/c), as we will call them in this work. They are well-established and known to designers and developers [17, 42, 58, 63, 70] and work in VR as well [58] (see Fig. 1b/c). As they work with the default HMD controllers, our results will be generalizable for the vast majority of VR HMDs. As the second interaction technique, we selected one that allows the manipulation of 3 DOF simultaneously. 3 DOF are the maximum for one of the canonical tasks translation, rotation, and scale when considered separately. We again selected the gizmos, as they also allow simultaneous 3 DOF manipulation (see Fig. 1b/c). Further, gizmos do not mix the canonical tasks, and the 3 DOF interaction for translation and rotation are in VR quite similar to an interaction and object manipulation with an object in the real world. We further selected an interaction technique on the higher end of the DOF spectrum with the handle bar [71] metaphor, which allows manipulating 7 DOF simultaneously (3 DOF translation, 3 DOF rotation, 1 DOF uniform scaling) without any context switch or additional input (see Fig. 1c). This technique is related to real-world object manipulation as well (two-hand interaction with an imaginary intersection that spikes objects). It is also realized using the default HMD controllers. We could not select an 8 or 9 DOF interaction technique, as all currently proposed interaction techniques for object manipulation in mid-air only support uniform scaling [58].

We deliberately decided to select only three interaction techniques for our comparison according to the guidelines of Bergström et al. [6], which state that the number of independent variables should be kept low. With the selected 1 DOF, 3 DOF, and 7 DOF interaction techniques, we covered the whole DOF spectrum.

3.1.2 Manipulation Tasks. Similar to the interaction techniques, object manipulation tasks can also be divided into the same 1 to 9 DOF spectrum (see Section 2.2.2). As for many object manipulation studies (see Section 2.2.4) and as suggested by Bergström et al. [6], we used docking tasks to inspect our RQ (see Section 1) and will describe their design and the choice of our manipulation objects in the following (see Fig. 1a).

Docking Tasks based on Fitts' Law: To define the difficulty of our docking tasks (see Section 2.2.4), we extended the formulas of Stoelen and Akin [74], which are based on the original Fitts' law formula:

$$ID_{translation} = log_2(\frac{A}{W} + 1)$$

$$ID_{rotation} = log_2(\frac{\alpha}{\omega} + 1)$$

As we wanted to investigate all three canonical manipulations (see LaViola et al. [51]), we extended this principle to scaling and used a similar formula:

$$ID_{scale} = log_2(\frac{s}{\Delta s} + 1)$$

A, α, and s define the distance of the translation, rotation, and scale docking task, whereas W, ω, and Δs represent the tolerance of how accurate the position, orientation, and scale of the target object have to be met. This way, an ID can be calculated for docking tasks similar to Fitts' law tasks. As stated by Stoelen and Akin [74] as well as Kulik et al. [48], the IDs of the three formulas can be combined. However, results exist by Triantafyllidis and Li [78] that question the generalizability of combining these IDs. Therefore, and as we did not evaluate the generalizability for ID_{scale}, we used the previously provided formulas not to compare the difficulty of, e.g., a translation with a rotation task, but to formally define our tasks to have similar difficulties inside the canonical task categories translation, rotation, and scaling.

To define our tasks, we decided to choose tasks narrowed down to separated *translation*, *rotation*, and *scale* as our simplest tasks on the DOF spectrum (3 DOF tasks). As a more difficult task level, we combined two of the canonical tasks to 6 DOF tasks resulting in the three tasks *rotation+translation*, *rotation+scale*, and *scale+translation*. The last level combined all three, resulting in the task *rotation+translation+scale*. We limited this task to uniform scaling to match it with our handle bar interaction technique (7 DOF task). This resulted in seven tasks, similar to our interaction techniques, covering the whole DOF spectrum. Our seven tasks are part of three groups (3 DOF/6 DOF/7 DOF), which keeps the independent variables low, similar to the interaction techniques (see Bergström et al. [6]).

For the translation, we defined $A = 0.5m$ and randomly defined distinct start positions around the target object, one for each translation task (see Bergström et al. [6]). We chose a distance of $0.5m$ so that the displaced object is within arm's reach of the target object. As the previously introduced formulas reduce the manipulation to a theoretical 1D movement, W could also be defined for this 1D movement. However, as we are always remapping this theoretical 1D movement to our 3D environment, this could lead to minimal and difficult tolerances on one or several of the x, y, and z axes. Our goal was to measure the time until this tolerance threshold was reached, as well as the final time-independent accuracy participants match the target object. Therefore, the threshold had to be difficult enough to challenge the participants. On the other side, we further wanted that there is still some distance left to improve the positioning to measure the final time-independent accuracy as well. By clearly separating the time and accuracy measure this way by using a threshold, we followed the guidelines of Bergström et al. [6] and prevented a time and accuracy trade-off present if measured simultaneously [41]. Therefore, we defined a tolerance of $W = \pm 0.05m$ for each axis, based on try-runs, and deviated with this 3D definition slightly from the previously shown 1D formula.

We defined the rotation in a similar way ($\alpha = 170°; \omega = \pm 10°$). The uniform scaling would always be similar if defined with a static s. We have chosen a different approach than for translation and rotation and defined several s but Δs dependent on s: $\Delta s = \pm 0.1s$. This approach also ensured a similar difficulty for the different scaling tasks. Using these task definitions, the 6 DOF tasks should be about two times as difficult as the 3 DOF tasks, and the 7 DOF task should be three times as difficult. As we already varied the *ID* with the 3 DOF, 6 DOF, and 7 DOF tasks, and further *ID* variations would not help to investigate our RQ, we decided to keep our study design small and defined not more than one trial for each of the seven tasks.

Manipulation Object Design: Besides defining the task difficulty, the manipulation objects have to be defined as well. Previous works have often used primitives, such as triangles or cubes [30, 47, 48, 59]. Still, as we include rotation tasks, it is important that the objects are unambiguous in how they have to be rotated (e.g., non-symmetrical objects) to not confuse participants (see Bergström et al. [6]). For this, we oriented our object design on the objects used by the test of Vandenberg & Kuse [80] and created seven individual objects so that every task has a unique one (see Fig. 1a). This way, there was no learning effect based on already knowing the object on the later tasks, and imitating the Vandenberg & Kuse objects allowed us to create similar difficult to understand objects as well.

3.1.3 Study Design Overview. Based on the previously explained design decisions, we had the two independent variables *interaction technique* with the conditions 1 DOF gizmos, 3 DOF gizmos, and 7 DOF handle bar (which resulted in increasing complexity) as well as *manipulation task* with the conditions 3 DOF task, 6 DOF task, and 7 DOF task (which resulted in increasing difficulty). This resulted in a 3*x*3 study design shown in Fig. 2. As each *interaction technique* had to use all defined tasks, this resulted in 21 tasks in total. Following the guidelines of Bergström et al. [6], we did a power analysis with G*Power 3.1.9.7 and decided to conduct a within-subject study as this would need about 54 participants (MANOVA, repeated measures, within factors; $f = 0.25; \alpha = 0.05; 1 - \beta = 0.95; groups = 9; measures = 21$) [29]. We selected as effect size $f = 0.25$, as we wanted to find at least medium-sized effects within our sample (see Cohen [15, 16]).

3.2 Dependent Variables

To investigate our RQ *"Do individuals with higher spatial abilities outperform individuals with lower spatial abilities in VR mid-air object manipulation tasks?"*, we thought of performance characteristics for object manipulation tasks that could be linked with the individual spatial abilities extending the typical measures of task completion time [7, 14, 30, 35, 41, 44, 47, 48, 56, 59, 61, 66, 81, 87] and accuracy [7, 35, 41, 44, 48, 61, 66, 81] (see Section 2.2.4). As first step, we split the time measures up to the several phases of the docking tasks (see Poupyrev et al. [66]) to be able to make more precise statements. The same applied to the accuracy measures, which were split into the different canonical operations (translation, rotation, scale). As shown by Barrera-Machuca et al. [3] for sketching, we could also expect more targeted manipulations to solve a manipulation task by higher spatial ability participants. We, therefore, counted the needed interactions and the absolute length

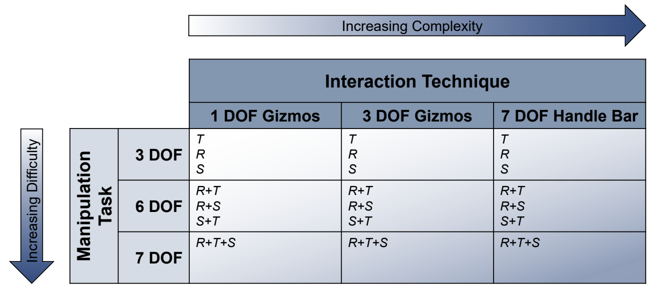

Figure 2: The study design consisted of the independent variables *interaction technique* with the conditions 1 DOF gizmos, 3 DOF gizmos, and 7 DOF handle bar that provide increasing complexity and *manipulation task* with the conditions 3 DOF task, 6 DOF task, and 7 DOF task that provide increasing difficulty based on a derived Fitts' law approach. We defined seven individual tasks and used a within-subject design which resulted in 21 trials for each participant. *T=translation; R=rotation; S=scale*

of the required manipulation way. Cognitive load is further a way to measure problem-solving skills [76]. We expected that higher spatial abilities should ease problem-solving of object manipulation tasks (see Just and Carpenter [43]), which is why we expected reduced cognitive load. Measuring the usability could further help to investigate how different interaction techniques are perceived and rated by participants with different spatial abilities. Table 1 shows all dependent variables.

We analyzed our results on three different levels, the overall level (OV) not distinguishing the interaction techniques and manipulation tasks but only investigating the influence of the covariates, e.g., the spatial abilities (see Section 3.3), the interaction technique level (IT) considering the interaction technique and the covariates, and the manipulation task level (MT) considering the manipulation task and the covariates. In the following, we will add the prefixes *OV*, *IT*, or *MT* to our dependent variables to indicate the level of investigation.

3.3 Spatial Ability Measure and Control Variables

We used the test of Vandenberg & Kuse [80] to measure the mental rotation capabilities of our participants as our *spatial abilities* measure and selected this instead of other measures (e.g., spatial orientation; see Section 2.1), as mental rotation is directly linked to our object manipulation docking tasks.

To further characterize our participants, we collected demographics (*age* and *gender*) as well as their *3D modeling experience*, which showed us if they were familiar with our interaction techniques, and their *VR experience*, which told us how advanced they were in using VR HMDs. As we measured both time and accuracy/error as dependent variables, we also asked participants to self-assess how accurately/perfectionist they, in general, solve tasks with the measurement *task accuracy self-assessment* stating this. This could tell us if participants would rush through a task or diligently solve it, which influences the achieved time and accuracy.

Table 1: We measured several dependent variables and analyzed them on three levels: overall level (OV), interaction technique level (IT), and manipulation task level (MT). We created sums or averaged values if multiple trials were combined (see Mendes et al. [59]), and streamlined the dependent variables on the IT and MT level to focus on our main measures. m=mean; s=sum; o=original

Dependent Variable	*OV*	*IT*	*MT*	Unit	Explanation
threshold time	m	m	m	seconds (s)	the time until an accuracy threshold was reached
first interaction time	m			seconds (s)	the time from the task start until the first interaction (initial thoughts and planning time)
task time	s	s	s	seconds (s)	the time participants actually conducted the task (= threshold time - first interaction time)
interaction time	s			seconds (s)	the time, only measured when an interaction was performed
think time	s			seconds (s)	the time, only measured when no interaction was performed
trans./rot./scal. error	s	s	s	meters (m) / degree (°) / -	the final error/deviation/accuracy to the target split up to the canonical tasks translation, rotation, and scale
interactions	s			-	the number of conducted/needed interactions
trans./rot./scal. sum	s			meters (m) / degree (°) / -	the absolute length of the needed manipulation way split up to the canonical tasks translation, rotation, and scale
SUS		o		-	the System Usability Scale (SUS) questionnaire [11]
active cognitive load	m	o		-	single-item question according to Klepsch and Seufert [46]
passive cognitive load	m	o		-	single-item question according to Klepsch and Seufert [46]

3.4 System

We used a plumber’s environment as our application’s setting as it worked perfectly for our setup and gamified our study. The chosen manipulation objects (see Section 3.1.2) based on the Vandenberg & Kuse [80] mental rotation test could be perfectly modeled with pipe sections (see Fig. 1a/d). Laying pipes further fitted into our story as we mapped this to our object manipulation docking tasks (see Section 3.1.2; see Fig. 1a). The solid source pipe had to be manipulated to match the red transparent target pipe. If the defined threshold (see Section 3.1.2) was met, a bell sound as well as a spark animation indicated that now the speed-dependent part was over, and the task switched to the accuracy-dependent one. Participants then had as much time as they wanted to manipulate and fit the object perfectly and then proceed to the next task. As described in Section 3.1.1, we implemented 1 DOF and 3 DOF gizmos and a 7 DOF handlebar (see Fig. 1b/c) as interaction techniques. To disregard the handedness of our participants, interactions could be performed equally and interchangeably with the right as well as the left controller.

The Vandenberg & Kuse [80] mental rotation test consisted of four stacks of boxes, aka the test objects (see Fig. 1d). The reference object was placed on a clipboard with limited mobility to fixate the perspective on the object as in the original test. This way, the two out of four identical to the clipboard but rotated objects on the boxes had to be selected, which is identical to the original test (see Fig. 1d). We used the redrawn version of Peters et al. [64] and used the first half of the test to limit its duration to 3 minutes[1] to ease its feasibility for our digital version (see Peters et al. [65]). All other questionnaires were also implemented in-game.

We implemented our application based on Unity [79] and developed it for the Meta Quest 1 and 2 [28], which have only minor hardware differences and similar controller designs and worked, therefore, interchangeably for our study [84]. We chose to implement for the Quests, as they are the most popular HMDs on Steam [18], and we also planned to collect data via online distribution of our application. This was also why we gamified our application, as we wanted to motivate a wide range of persons to participate. We restricted gamification to the plumber story as well as the setting of the environment but did not change any specifics necessary by our study design.

3.5 Procedure

We designed our app in a way that it could run as an unsupervised online study considering the best practices of Radiah et al. [68] for conducting remote VR studies. This way, we could recruit remote participants online and also local participants with different local study supervisors without affecting our results. Therefore, we included an in-game examiner that guided the participants and provided all necessary information with texts and voice-overs.

In the first scene, participants entered a hallway, where they got an initial introduction of our study’s purpose and the consent form. When this was accepted, the first game scene, the warehouse (see Fig. 1d), started where the participants conducted the Vandenberg & Kuse mental rotation test [80] and answered our demographic questions (see Section 3.3). After this, the first interaction technique was randomly selected, and the first basement scene (see Fig. 1a/b/c) started. We implemented a tutorial where participants got familiar with the manipulation tasks, as well as the interaction techniques. The tutorial included four example manipulation tasks, one for translation, rotation, and scale each, and one including all three. After this, the first seven study tasks allocated to the interaction technique started in the order as shown in Fig. 2. Most of our dependent variables (see Table 1) were unobtrusively measured by our app while the participants conducted the tasks. When the seven tasks were finished, the participants answered the *SUS* as well as the *cognitive load* questionnaires. Then the next basement scene started with the next randomly selected interaction technique, beginning once again with a tutorial. After all three interaction

[1]We used the MRT-A set.

technique basement scenes were finished, the data was uploaded to our university's Nextcloud [36] server, and the app closed itself. We ensured that the online version of our app could only be started once as a study run. Beginning with the second run, our app entered the free-to-play mode, where only the basement scenes were loaded, but no data was collected anymore.

3.6 Participants

We recruited 66 participants (see power analysis in Section 3.1.3) in four different ways. The largest group (n=47) was collected using the app during the university's curriculum. Another 14 participants were recruited through convenience sampling. We also used Prolific [67] to recruit four participants online. Prolific participants were compensated with 7.50 €. Our 66 participants were between 18 and 61 years (M = 23.70, SD = 8.56) old. 52 identified themselves as female, 13 as male, and one participant preferred not to answer. The overall *VR experience* of the participants was low, with 58 participants stating to use VR HMDs less than 1 hour per month (1-2h: n=2; 3-5h: n=1; 5-10h: n=3; >10h: n=2). The same applied to the *3D modeling experience*, with 58 participants stating to use 3D modeling applications less than 1 hour per month (1-2h: n=3; 3-5h: n=2; 5-10h: n=0; >10h: n=3). Their rating for the *task accuracy self-assessment* was slightly above the midpoint of the scale ranging from "strongly disagree (1)" to "strongly agree (5)" (M = 3.55, SD = .85). We did not normalize these participants' characteristics when used for statistical analysis. The vision of the participants was normal or corrected to normal.

We analyzed the results of the Vandenberg & Kuse [80] mental rotation test to measure the *spatial abilities* of our participants. For scoring, we used the approach where one point is given for each correct answer, and one point is subtracted for each missing or incorrect answer[2]. We further normalized the scores to range from 0 (no correct answer in the test) to 1 (everything answered correctly in the test). This resulted in the histogram shown in Fig. 3 (M = .46, SD = .20, Mdn = .42). It shows that the *spatial abilities* of our participants have a broad range with a peak near the mean. As we have N > 30, we treat this data as approximately normally distributed [32]. During our statistical analysis, we tested this data and its effects on our dependent variables with multiple linear regressions but also conducted a median split to create two nearly equally sized groups (see Barrera-Machuca et al. [3] and Lages and Bowman [49]), one with lower *spatial abilities* (n=37) and one with higher *spatial abilities* (n=29) for group comparisons.

4 RESULTS

In the following, we explain our exploratory statistical method and present our study results. They are presented following our study design (see Fig. 2/Table 1).

4.1 Statistical Methods

We conducted as our main analysis multiple linear regressions to analyze whether the *spatial abilities*, *age*, *VR experience*, *3D modeling experience*, and the *task accuracy self-assessment* of the participants contribute significantly to explaining the variance of the dependent variables (see Table 1). A linear statistical method allowed us to consider the linearity of our data, e.g., the *spatial abilities* (see Fig. 3), and predict trends showing linear relationships not possible with group comparisons (e.g., MANOVA) which would reduce our linear data to predefined artificial groups and just test for significant differences between the groups. Using multiple linear regressions instead of correlations further allowed us to investigate the effect of our proposed main predictor *spatial abilities* together with the covariates in one model [32]. This method allowed us to show if our RQ *"Do individuals with higher spatial abilities outperform individuals with lower spatial abilities in VR mid-air object manipulation tasks?"*, predicting the dominant effect of *spatial abilities* is correct and if there are further dominant covariates we considered in our study design. For our analysis, it was not important to create exact models that perfectly predict the dependent variables but to find significant influencing factors and show their effects. We, therefore, also made no model selection and always used all control variables for all dependent variables.

Based on our study design (see Fig. 2/Table 1), we started our analysis on the overall level (OV), followed by the interaction technique level (IT) and the manipulation task level (MT) and use these abbreviations for the dependent variables in the following. As the interaction technique (IT) and the manipulation task (MT) levels provide groups (the conditions of the independent variables) by study design, we further conducted as a second-level analysis group comparisons in addition to the multiple linear regressions. To consider the *spatial abilities* in these group comparisons as well, we performed a median split for the *spatial abilities* (Mdn = .42; see Fig. 3; see Barrera-Machuca et al. [3] and Lages and Bowman [49]). We created two groups, one with lower (n = 37) and one with higher (n = 29) spatial abilities. Conducting a median split to create these two groups is appropriate due to the approximately normal distribution of the *spatial abilities* displayed in the histogram [32] (see Section 3.6/Fig. 3). This way, we created two nearly equally sized groups, one with below-average and one with above-average *spatial abilities*. We conducted for the group comparisons MANOVA analysis with univariate post hoc tests (ANOVA) using Bonferroni correction for the pairwise comparisons [33]. We used SPSS 27.0.1.0. We deliberately did not compare the two groups below-average and above-average spatial abilities with MANOVA analysis against each other as this was investigated already by multiple linear regressions, as previously explained, which had the benefit of not reducing an

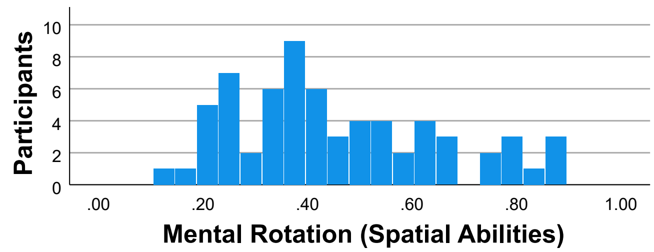

Figure 3: This histogram shows the *mental rotation capabilities (spatial abilities)* of our participants based on the test of Vandenberg & Kuse [80]. We normalized the test scores to range from 0 (no correct answer in the test) to 1 (everything answered correctly in the test) (M = .46, SD = .20, Mdn = .42).

[2]We explain the test in Section 2.1. Its implementation is described in Section 3.4.

interval scaled variable as spatial abilities is (see Fig. 3) into two artificial groups. We did not want a second and, as explained, weaker analysis for the same question. The MANOVA analysis focused on comparing the conditions of the independent variables.

For all our results, we report the effect sizes (f^2, $partial\ \eta^2$, or β; see Acock [1], Cohen [15, 16], Dragicevic [21], and Fey et al. [31]) ($f^2 :> .02$ = small, $> .15$ = medium, $> .35$ = large effect; $partial\ \eta^2 :> .01$ = small, $> .06$ = medium, $> .14$ = large effect; $\beta :< .2$ = small, $> .2$ = medium, $> .5$ = large effect).

4.2 Overall Influence of Spatial Abilities

When analyzing the overall level with multiple regressions, we found that the *spatial abilities* significantly decrease *OV threshold time*, *OV task time*, *OV think time*, *OV interactions*, *OV translation sum*, *OV rotation sum*, and *OV scale sum*. The *task accuracy self-assessment* significantly decreased *OV threshold time*, *OV task time*, *OV translation error*, *OV rotation error*, *OV scale error*, and *OV rotation sum*. The *age* significantly increased *OV think time*. Table A1 provides detailed reporting. We discuss these results in Section 5.1.

4.3 Influence of Spatial Abilities on Interaction Technique Level

We also analyzed the interaction technique level with multiple regressions for the 1 DOF and 3 DOF gizmos and the 7 DOF handle bar. We further performed group comparisons.

4.3.1 Interaction Technique Multiple Regressions. We found for the 1 DOF gizmos that *spatial abilities* significantly decrease *IT 1DOF threshold time*, *IT 1DOF task time*, and *IT 1DOF rotation error*. The *task accuracy self-assessment* significantly decreased *IT 1DOF threshold time*, *IT 1DOF task time*, *IT 1DOF translation error*, *IT 1DOF rotation error*, and *IT 1DOF scale error*. The *VR experience* significantly increased *IT 1DOF translation error*. For the 3 DOF gizmos *task accuracy self-assessment* significantly decreased *IT 3DOF rotation error* and increased *IT 3DOF SUS*. The *VR experience* significantly decreased, and the *3D modeling experience* significantly increased *IT 3DOF SUS*. For the 7 DOF handle bar *task accuracy self-assessment* significantly decreased *IT 7DOF threshold time*, *IT 7DOF task time*, *IT 7DOF scale error*, and increased *IT 7DOF SUS*. The *3D modeling experience* significantly increased *IT 7DOF threshold time* and decreased *IT 7DOF SUS*. Table A2 provides detailed reporting. We discuss these results in Section 5.2.

4.3.2 Interaction Technique Group Comparisons. We found significant differences for *IT threshold time* and *IT task time* for participants with lower and higher spatial abilities for *1DOF* vs. *3DOF* vs. *7DOF* (see Fig. 4a-d). For *IT rotation error*, there was a significant difference for higher spatial ability participants for *1DOF* vs. *7DOF* (see Fig. 4e). Higher spatial ability participants further had a significant difference for *IT scale error* for *3DOF* vs. *7DOF* (see Fig. 4f). For *IT SUS*, we found significant differences for both lower and higher spatial ability groups for *1DOF* vs. *7DOF* (see Fig. 4g-h). The *IT passive cognitive load* was significantly different for lower and higher spatial ability participants for *1DOF* vs. *7DOF* (see Fig. 4i-j). Table A3 provides detailed reporting. We discuss these results in Section 5.2.

4.4 Influence of Spatial Abilities on Manipulation Task Level

Similar as before, we also analyzed the manipulation task level with multiple regressions for the 3 DOF, 6 DOF, and the 7 DOF tasks and also performed group comparisons.

4.4.1 Manipulation Task Multiple Regressions. We found for the 3 DOF tasks that *spatial abilities* significantly decrease *MT 3DOF threshold time* and *MT 3DOF task time*. The *task accuracy self-assessment* significantly decreased *MT 3DOF rotation error*. For the 6 DOF tasks *task accuracy self-assessment* significantly decreased *MT 6DOF threshold time*, *MT 6DOF task time*, *MT 6DOF translation error*, *MT 6DOF rotation error*, and *MT 6DOF scale error*. For the 7 DOF tasks *task accuracy self-assessment* significantly decreased *MT 7DOF rotation error* and *MT 7DOF scale error*. Table A4 provides detailed reporting. We discuss these results in Section 5.3.

4.4.2 Manipulation Task Group Comparisons. We found significant differences for *MT threshold time* for participants with lower and higher spatial abilities for *3DOF* vs. *6DOF* and *3DOF* vs. *7DOF* (see Fig. 5a-b). For *MT task time*, there was a significant difference for the lower spatial abilities participants for *3DOF* vs. *6DOF* vs. *7DOF* and for the higher spatial abilities participants for *3DOF* vs. *6DOF* and *6DOF* vs. *7DOF* (see Fig. 5c-d). The *MT translation error* was significant for participants with lower and higher spatial abilities for *3DOF* vs. *7DOF* and *6DOF* vs. *7DOF* (see Fig. 5e-f). For *MT rotation error*, there was a significant difference for participants with lower and higher spatial abilities for *3DOF* vs. *6DOF* vs. *7DOF* (see Fig. 5g-h). The *MT scale error* was significant for participants with lower spatial abilities for *3DOF* vs. *7DOF* and *6DOF* vs. *7DOF* and for the higher spatial abilities participants for *3DOF* vs. *6DOF* vs. *7DOF* (see Fig. 5i-j). Table A5 provides detailed reporting. We discuss these results in Section 5.3.

5 DISCUSSION

To answer our RQ *"Do individuals with higher spatial abilities outperform individuals with lower spatial abilities in VR mid-air object manipulation tasks?"*, we will now discuss our results (see Section 4). They are split into the three levels, overall, interaction technique, and manipulation task (see Fig. 2/Table 1), which guided this section's structure. This discussion is the base for our guidelines presented in Section 6.

5.1 Discussing the Overall Influence of Spatial Abilities

For the overall level, we grouped our discussion by the dependent variables (see Section 3.2).

5.1.1 OV Task Completion Time: Our results show that *spatial abilities* significantly reduced the *OV threshold time* (the time to achieve the task's threshold), *OV task time* (the time needed to actually solve the task), and *OV think time* (the time participants think about a solution) (see Table A1). The effect sizes are all medium, according to Cohen [15, 16]. This confirms our RQ and shows that participants with higher spatial abilities completed the tasks in a shorter time. Interestingly, we found a further significant reduction of *OV threshold time* and *OV task time* by the *task accuracy self-assessment*

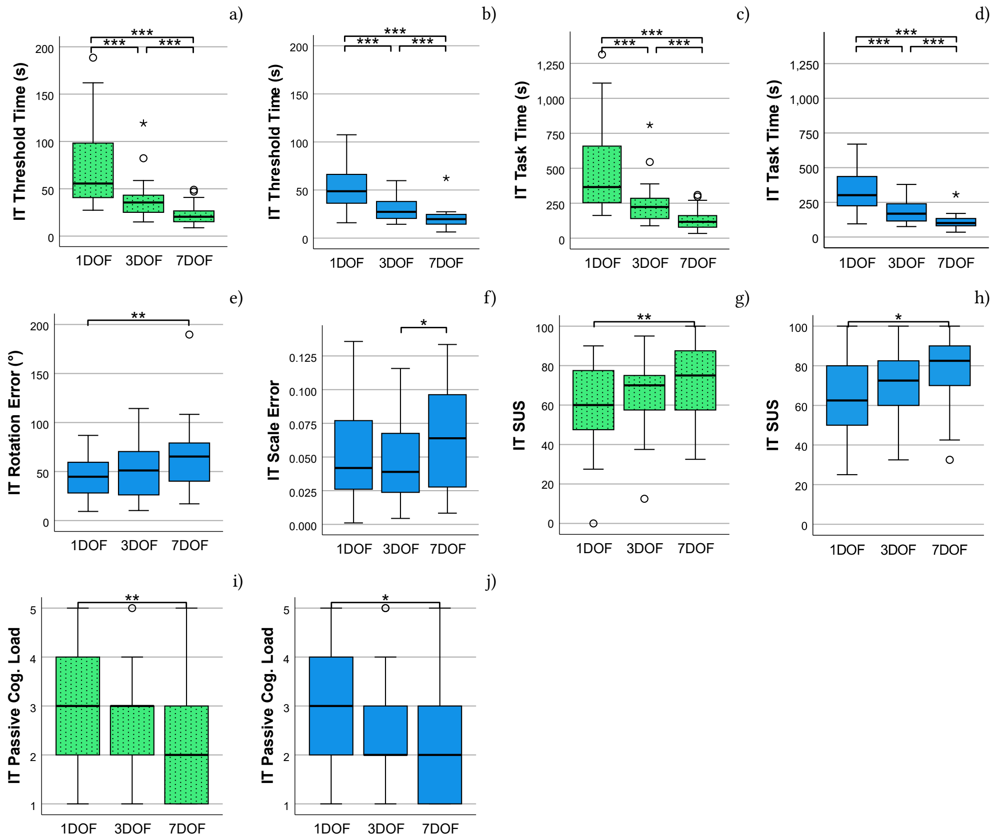

Figure 4: We found significant differences for the three interaction techniques for our dependent variables based on a median split for the *spatial abilities* of the participants. The green box plots with patterns represent the results for the participants with lower *spatial abilities*, and the blue ones without patterns the results for the participants with higher *spatial abilities*. Asterisk (*) indicates a statistically significant difference between conditions: p < .05 (*); p < .01 (**); p < .001 (***).

covariate. This means that participants who judged themselves to work more accurately could finish the tasks faster. One possible explanation could be that they work more concentrated, which would relate to the theory that perfectionists have a greater work engagement [73]. According to the results of the multiple regression (see β values in Table A1), the *task accuracy self-assessment* has even a slightly higher influence on these two measures than the spatial abilities. This means that besides spatial abilities, the accuracy individuals usually solve tasks is also a very important factor that should be considered when it comes to task completion times. These results have a strong influence on object manipulation technique design as well as the correlated user studies.

5.1.2 OV Task Accuracy: We found no significant contribution of *spatial abilities* to *OV translation error*, *OV rotation error*, and *OV scale error*, which describe the final accuracy of the docking tasks. This means that spatial abilities did not influence the achievable accuracy. Users with lower spatial abilities can, therefore, perform the same high-precision tasks as users with higher spatial abilities. However, as the discussion of *task completion time* (see Section 5.1.1) shows, they may need more time for the same result. Though, as we separated our time and accuracy measure as proposed by Bergström et al. [6] (see Section 3.1.2), we can make this statement only for the threshold times, but not for the final accuracy times. These results are linked to Barrera-Machuca et al. [3] who showed that spatial abilities have a positive effect on shape-likeliness in 3D freehand sketching but not on line precision. As expected, we found an influence of the *task accuracy self-assessment* on these three variables with small and medium effect sizes.

5.1.3 OV Problem-solving Approach: We measured the performed *OV interactions* and the in total performed manipulations of the object with *OV translation sum*, *OV rotation sum*, and *OV scale sum*, to predict how targeted and problem-oriented participants solved the docking tasks. We found a significant reduction for all of them, which means that users with higher spatial abilities conducted 3D manipulations more targeted and less based on a trial-and-error approach (see Table A1). We think our high spatial ability participants needed fewer interactions, as they better knew which ones were necessary. As a result, they could choose a more direct and, therefore, shorter manipulation path. The effect sizes were primarily medium and only small for the *OV rotation sum*.

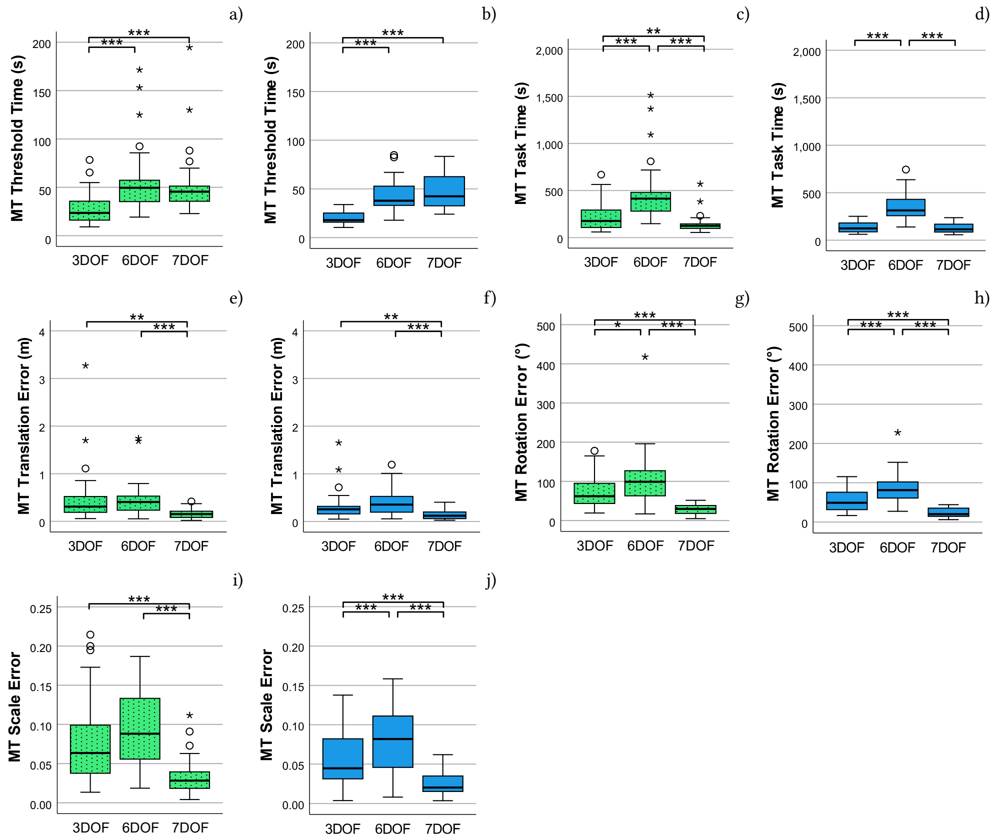

Figure 5: We found significant differences for the three manipulation task levels for our dependent variables based on a median split for the *spatial abilities* of the participants. The green box plots with patterns represent the results for the participants with lower *spatial abilities*, and the blue ones without patterns the results for the participants with higher *spatial abilities*. Asterisk (*) indicates a statistically significant difference between conditions: p < .05 (*); p < .01 (**); p < .001 (***).

Our findings extend the work of Zander et al. [86], who found a similar improvement of target rotation when spatial abilities were trained first at a tablet-based object rotation study. They are also linked to Barrera-Machuca et al. [3], who stated that participants with higher spatial abilities moved more systematically during sketching. Besides the *OV rotation sum*, where we also found a significant contribution by the *task accuracy self-assessment*, the *spatial abilities* were the only significant contributor. This shows the significant importance of spatial abilities when it comes to efficient object manipulations, especially compared to the other covariates, and shows why these user characteristics should be considered for interface design in general and, in particular, for interaction technique design.

5.1.4 OV Cognitive Load: Interestingly we found no significant result for the cognitive load on our overall level, as we would have expected a correlation with the spatial abilities based on the usage of the working memory [43, 76]. However, our participants rotated the object in VR and not mentally as in these previous works, which could explain the different results. It is also possible that a more differentiated cognitive load questionnaire measuring intrinsic cognitive load (ICL), germane cognitive load (GCL), and extraneous cognitive load (ECL) such as the one by Klepsch et al. [45] would provide further information, especially regarding learning aspects (see e.g., Drey et al. [22], or Albus et al. [2]). There was also no linear relationship with one of the other covariates. These results show why it is essential to specify the interaction technique for cognitive load analysis, as we found significant differences on the interaction technique level (see Fig. 4). We will discuss this in the following section.

5.2 Discussing the Influence of Spatial Abilities on the Interaction Techniques

The discussion of the interaction technique level is once again grouped by the dependent variables.

5.2.1 IT Task Completion Time: Our results show that *IT 1DOF threshold time* and *IT 1DOF task time* are, with a medium effect, significantly decreasing with higher spatial abilities of the participants (see Table A2). However, we did not find these effects for the 3 DOF gizmos and the 7 DOF handle bar. Interestingly, we found that the 7 DOF handle bar significantly outperformed the 3 DOF gizmos, which significantly outperformed the 1 DOF gizmos for

IT 1DOF threshold time and *IT 1DOF task time* for both users with higher and lower spatial abilities (see Fig. 4). The effect sizes were large in all cases. This shows a dominant effect of the interaction technique with better results for more simultaneously controllable DOF. This is similar to the findings of Kulik et al. [48], Mendes et al. [61], and Hinckley et al. [41] and suggests that this is a universal finding not influenced by the individual spatial abilities. Therefore, the interaction technique can compensate for lower spatial abilities, which we considered in our guidelines.

5.2.2 IT Task Accuracy: Similar to the overall level (see Section 5.1) and as expected, the effect of *task accuracy self-assessment* on the task accuracy was more prominent than that of the *spatial abilities*, as we found only one significant contribution for *spatial abilities* on *IT 1DOF rotation error* (see Table A2). However, for *task accuracy self-assessment*, we found significant contributions on the *IT 1DOF translation error*, *IT 1DOF rotation error*, *IT 1DOF scale error*, *IT 3DOF rotation error*, and *IT 7DOF scale error*. Similar to the IT task completion time (see Section 5.2.1), we see that the effect of *task accuracy self-assessment* drops with multiple DOF interaction techniques as either the effect sizes become smaller or it is not present at all. Nevertheless, we see no significant improvement for the task accuracy dependent variables for multiple DOF interaction techniques (see Fig. 4), which is in line with Veit et al. [81]. Therefore, and similar as stated for the *spatial abilities* (see Section 5.1.2), the interaction technique has no influence on the general task accuracy.

5.2.3 IT Usability: We measured the usability of every interaction technique with the SUS questionnaire and found no significant contribution of the *spatial abilities*. We saw that, independent of the spatial abilities, the usability was rated higher with more DOF simultaneously possible to manipulate. However, only the difference between the 1 DOF and 7 DOF interaction techniques was significant, but these effects were large (see Fig. 4). This shows once again as for the IT task completion time (see Section 5.2.1) the dominant effect of the interaction technique with a favor for multiple DOF, which is a highly valuable result for our guidelines.

5.2.4 IT Cognitive Load: Consistent with our findings for OV cognitive load (see Section 5.1.4), we found no effects for *spatial abilities* considering the used interaction techniques for *active cognitive load* and *passive cognitive load*. Nevertheless, we found a decrease of the *passive cognitive load* independent of the *spatial abilities* for interaction techniques with more simultaneously manipulable DOF (see Fig. 4). This was significant for the 1 DOF gizmos vs. the 7 DOF handle bar, but not for a comparison with the 3 DOF gizmos. It means that especially the highly sequential 1 DOF gizmos were more exhausting for the participants than the parallel interaction technique 7 DOF handle bar. This could be due to the technique itself but also linked to the shorter task completion time discussed in IT task completion time (see Section 5.2.1) as both could have effects on the working memory [43, 76]. This finding is very important and consistent with the findings of the other dependent variables showing the strong influence of the used interaction technique. It is further essential that we did not find significant differences for the *active cognitive load*, as it shows that our participants were similarly engaged in conducting all the trials for the different interaction techniques, which shows that our data is sound.

5.3 Discussing the Influence of Spatial Abilities on the Manipulation Tasks

And again, the discussion of the manipulation task level is grouped by the dependent variables.

5.3.1 MT Task Completion Time: We see a significant contribution of the *spatial abilities* for the task completion time only for *MT 3DOF threshold time* and *MT 3DOF task time* (see Table A4). This means, as we stated for OV task completion time (see Section 5.1.1), that *spatial abilities* positively influence the task completion time, but participants with higher spatial abilities have no over-proportional advantage the more difficult the tasks get. Nevertheless, we saw that the *MT threshold time* significantly increased when comparing the 3 DOF tasks with the 6 DOF and 7 DOF tasks (see Fig. 5). We expected this, as it reflects that they have increased difficulty as intended and predicted by our used formulas based on Stoelen and Akin [74] (see Section 3.1.2). As the effect sizes are large, this is a strong verification of our study design's soundness.

5.3.2 MT Task Accuracy: Regarding the task accuracy, we saw no influence of the individual *spatial abilities* on the task accuracy measures *MT translation error*, *MT rotation error*, and *MT scale error*. But as discussed for OV task accuracy (see Section 5.1.2) and IT task accuracy (see Section 5.2.2), we found once again an influence of *task accuracy self-assessment* on these dependent variables. Furthermore, the 7 DOF task had the smallest error, which was significant compared to the 3 DOF and 6 DOF tasks. This applied similarly for translation, rotation, and scaling (see Fig. 5). As the 7 DOF task is the most difficult one, this is surprising but could be explained by the task order. The 7 DOF task was always conducted last for each interaction technique, which means that the participants have mastered it the best.

5.4 Summary

Answering our RQ, we can state that higher spatial abilities are significantly correlated with lower task completion time as well as more targeted manipulations. Our further results show that independent of the spatial abilities, all users can achieve similar task accuracy, and the task accuracy is affected by the individual self-assessed perfectionism but neither task nor interaction technique dependent. The cognitive load depends not on the individual spatial abilities but on the interaction technique and is lowest if multiple DOF can be manipulated simultaneously and no sequential task execution is necessary. In general, we found a dominant effect of the interaction technique that could compensate for the differences in spatial abilities with better performance results for more simultaneously manipulable DOF. We conclude that higher spatial abilities help to solve object manipulation tasks faster and more targeted, but the interaction technique has a more dominant effect.

As one might say that ~20% for our R^2_{adj} values of the multiple linear regressions is a low model fit, we want to emphasize that the opposite is the case. With our exploratory RQ, we wanted to investigate if spatial abilities influence object manipulation in VR. We never intended to create exact models (see Section 4.1), which probably have a lot of further influencing factors as indicated by previous works (e.g., satisfaction, simulator sickness, presence, fatigue, immersion; see survey of Bergström et al. [6]). Our current models only consider spatial abilities and other demographic characteristics and can explain ~20% of the manipulation task performances. We think that ~20% contribution of a factor, which was completely neglected in such studies in the past (see Sections 2.1.2 and 2.2.2), is a very high value and combined with the effect sizes which were mostly medium or large, a powerful result of our work, showing how important it is to consider spatial abilities in future work (see guidelines Section 6). Combining spatial abilities with the previously named other influencing factors could create a model fit >20%, which could be investigated in the future.

These findings are the basis for our guidelines presented in the following.

6 GUIDELINES FOR OBJECT MANIPULATION

Extending the guidelines of LaViola et al. [51] and Bergström et al. [6] based on our results, we define the following guidelines for VR and controller-based interaction techniques used for object manipulation. They are split into ones addressing general interaction technique design and implementation and ones for user research. The guidelines should be self-explaining, which is why we deliberately repeat some of the previous findings in the explanations. They are based and formulated on our VR study but could be an inspiration for other systems as well and we encourage researchers to extend them further. We recommend considering them when appropriate but advise researchers and designers, especially when transferring them to other domains, to always tailor them to their specific needs and consider the exploratory nature of our work.

6.1 Guidelines for Interaction Technique Design

I1: *The individual spatial abilities should be considered for interaction technique design as users' performance is linked to them.*
Our results show that the task completion time significantly drops with increasing spatial abilities (see Table A1). We further saw that higher spatial abilities contribute to a more targeted solution, leading to fewer necessary interactions and manipulations. Therefore, we suggest that interaction techniques should at least (1) provide means to help users with lower spatial abilities to solve tasks faster and more targeted/problem-oriented. This could be done with hints that provide a solution approach or visualizations that show the current accuracy/error. Further support, such as grids or snapping, is possible, too. But we also state that this is maybe not always necessary for users with higher spatial abilities and suggest (2) allowing these users to use the interaction technique unrestricted and without support.

I2: *The individual spatial abilities should be measured in applications to adapt interaction techniques appropriately.*
Our guideline I1 suggests adapting the interaction technique to the individual spatial abilities. However, to do so, they have to be measured. This is possible with a standardized test such as the one of Vandenberg & Kuse [80], which we used. However, others also exist, as discussed in Section 2.1. It is possible to integrate a gamified version inside the application as we proposed in this work (see Section 3.3). Further possibilities to assess the individual spatial abilities could be to use time thresholds as our results show that task completion time is linked to the spatial abilities (see Table A1/A2) or to use a self-assessment. Doing so will allow adaptive interaction techniques which keep the flow high [19, 20] and prevent users from getting stuck in their work [23, 25, 26].

I3: *Input device and task optimized interaction techniques should be preferred over universal but maybe well-known ones.*
We found a dominant effect of the interaction technique, compensating for lower spatial abilities. Our interpretation is that with an interaction technique ideally suited for the task and the current input device/hardware, even users with lower spatial abilities can outperform users with higher spatial abilities that use an inappropriate and not optimized interaction technique (see Fig. 4). Therefore, choosing the proper interaction technique should always come first before it is optimized regarding the spatial abilities with I1 and I2. Following LaViola et al. [51], we suggest that this does not have to mean necessarily always inventing a new interaction technique. As we for example saw an improvement from the 1 DOF gizmos to the 3 DOF gizmos, it is also a valid option to use the most appropriate of the existing ones or optimize existing ones for a specific task.

I4: *Independent of their individual characteristics such as spatial abilities or perfectionism, users should not be restricted in the usable DOF, neither for the used interaction technique nor for the task.*
Our results for the interaction technique show that the task completion time and the measured usability are positively affected by more simultaneously controllable DOF. We also found positive implications for the passive cognitive load (see Fig. 4). Even more complex tasks with more DOF can have similar or even better results for accuracy as easier ones (see Fig. 5). This guideline is in line with the results of Kulik et al. [48], Mendes et al. [61], and Hinckley et al. [41]. Therefore, we state that it is generalizable and not dependent on individual characteristics such as spatial abilities or perfectionism.

6.2 Guidelines for Research and User Studies

R1: *Interaction technique user studies should always consider and control participants' individual characteristics such as spatial abilities or perfectionism, as this could significantly influence the dependent variables.*
Our results show that spatial abilities significantly influenced the task completion time, while the self-assessed task accuracy/perfectionism significantly influenced task accuracy (see Section 5.1). This means that these individual characteristics significantly change the results in object manipulation studies and possibly in similar affiliated ones such as object selection studies. For within-subject designs, this could mean that the own sample does not represent the population, which would cause non-generalizable results [50]. Examples are studies at universities with students of STEM disciplines, which tend to have above-average spatial abilities [53]. For between-subject studies, the implications are even higher,

as this could mean imbalanced groups, which also could cause non-generalizable results [50]. Such effects could be an explanation for contradicting results in previous works (see Section 2.2.2). Using these individual characteristics as covariates and ensuring they are normally distributed could prevent such errors. We showed that this was not done very often in the past (see Section 2.1.2), but based on our results, we encourage researchers to do so in the future.

R2: *Study results should not always strive to create one result for the whole population but investigate it in a differentiated way, as the entire population is highly diverse due to multiple individual characteristics.*
Research always tries to find generalizable results, which is a good thing. However, our results show the strong influence of individual characteristics on standard measures such as task completion time and accuracy (see Table A1). As these individual characteristics scatter strongly (see Fig. 3), we encourage authors to also state findings only valid for a specific part of a population. This guideline is related to the advice of Lance and Hattori [50] about how to sample a population representatively. It is also linked to the findings of Lages and Bowman [49], who also advise considering individual characteristics such as spatial abilities or game experience, as they significantly influenced their 3D data examination task. Our guideline should create awareness for this, especially for the individual spatial abilities and self-assessed perfectionism, where we found a significant influence on object manipulation, and encourage authors to consider this in future research.

R3: *Object manipulation studies in VR should use Fitts' law-derived approaches to adjust task difficulty.*
We presented in Section 3.1.2 our study design template to adjust the task difficulty of our docking tasks based on formulas derived from Stoelen and Akin [74] which are based on Fitts' law. They divide the task into the canonical object manipulations translation, rotation, and scale [51]. Similar to Kulik et al. [48], our results showed that these formulas can be used to increase and decrease the task difficulty appropriately, which resulted in higher and lower task completion times in our case (see Fig. 5). This shows that they are accurate for similar object manipulation studies, are an own methodological contribution according to Wobbrock and Kientz [85], and can be used as a distance/difficulty metric. Using our study design template will ease researchers' work in defining and conducting such studies in a standardized way to improve across-study result comparability in a Fitts' law-like manner (see Kulik et al. [48]; possible target group: [7, 9, 14, 30, 35, 41, 44, 47, 59–61, 75, 77, 83]).

7 LIMITATIONS

We explained in Section 2.1 that multiple different types of spatial abilities exist. Nevertheless, we decided to limit our analysis to mental rotation capabilities and only measured them as they are contrary to the others directly linked to our object manipulation tasks as described in Section 3.3. We also limited the test to a subset to ease its digital feasibility (see Peters et al. [65]). However, this way, our results do not show possible effects of other spatial abilities, which could be investigated in future work as we discuss in Section 8.

The results for our covariates *3D modeling experience* and *VR experience* are quite limited, as they are not normally distributed, and they have their peaks at the lowest level of experience (see Section 3.6). This means that our participants were primarily novices and that repeating the study with expert users (e.g., gamers, designers, or developers) may lead to different results. We further have a gender imbalance that could also limit our results' generalizability. Conducting a study in the wild further has a trade-off between external and internal validity as well as to gamify a study application to motivate participants [27, 37]. There is further a trade-off using two very similar device types during data collection, Quest 1 and 2 [84] (see Section 3.4), to increase available participants. However, our main research focus was our participants' mental rotation capabilities, which we could accurately measure, and where we had an approximately normal distribution [32] (see Section 3.6). This ensured a proper sample for our RQ (see Lance and Hattori [50]).

This work is exploratory, analyzing several dependent variables. We advise considering this when interpreting our results (see McDonald [57]) and, therefore, provide all conducted tests in the appendix for transparency.

8 FUTURE WORK

Our results show that spatial abilities can significantly influence, e.g., task completion time. But as discussed in the limitations (see Section 7), our results only focused on mental rotation capabilities. However, other spatial abilities, such as spatial orientation or spatial relation (see Section 2.1), exist, which could have potential further effects. It is also possible that effects exist for other interaction techniques than the ones we used (see Section 2.2.3) as well as when using gestures or locomotion or other non-immersive devices such as 2D screens with touch input. Different tasks could also provide promising future research direction, such as selection tasks, which a classic Fitts' law study could investigate. As discussed in Section 5.4, further influencing factors may exist that could be used together with the spatial abilities to build a model with a higher performance prediction. These ideas show that there are plenty of future research directions considering individual spatial abilities.

9 CONCLUSION

Spatial abilities are highly individual and can influence how users interact. However, such individual characteristics were seldom considered for interaction technique design nor corresponding user research. We exploratory investigated if the spatial ability sub-type mental rotation has a significant effect on the dependent variables task completion time as well as task accuracy in a VR-based object manipulation study. Each of our 66 participants conducted 21 docking tasks with varying difficulty, similar to a Fitts' law design, and used three interaction techniques in a randomized order (either 1 DOF gizmos, 3 DOF gizmos, or a 7 DOF handle bar metaphor). We found that task completion time significantly depends on the individual spatial abilities and decreases if they get higher. Furthermore, participants with higher spatial abilities perform more targeted object manipulations and need fewer interactions. Our findings further show that lower spatial abilities can be compensated by an appropriate interaction technique. This was evident in our case for the handle bar, which also shows that multiple simultaneously usable DOF can be beneficial. Our second main dependent variable, task accuracy, was not influenced by spatial abilities or the interaction technique. All participants could achieve similar accuracy. Our findings highlight the significance of individual spatial abilities and the selection of an optimized interaction technique, which we recommend that researchers consider in future work. We, therefore, defined seven guidelines for interaction techniques on how to consider spatial abilities for VR object manipulation design, implementation, and user research. Our work will assist researchers and designers in personalizing and improving object manipulation in VR.

ACKNOWLEDGMENTS

This work was conducted within the project "AuCity 2 - AR in University Teaching Using the Example of Civil Engineering" and within the project "AuCity 3 - Collaborative and Adaptive MR in Higher Education Teaching Using Civil Engineering as an Example" both funded by the Federal Ministry of Education and Research (BMBF).

REFERENCES

[1] A.C. Acock. 2018. *A Gentle Introduction to Stata.* Stata Press.

[2] Patrick Albus, Andrea Vogt, and Tina Seufert. 2021. Signaling in virtual reality influences learning outcome and cognitive load. *Computers & Education* 166 (2021), 104154. https://doi.org/10.1016/j.compedu.2021.104154

[3] Mayra Donaji Barrera Machuca, Wolfgang Stuerzlinger, and Paul Asente. 2019. The Effect of Spatial Ability on Immersive 3D Drawing. In *Proceedings of the 2019 on Creativity and Cognition* (San Diego, CA, USA) *(C&C '19)*. Association for Computing Machinery, New York, NY, USA, 173–186. https://doi.org/10.1145/3325480.3325489

[4] Trevor Barrett and Mary Hegarty. 2014. Interaction design and the role of spatial ability in moderating virtual molecule manipulation performance. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 36.

[5] Trevor J. Barrett and Mary Hegarty. 2016. Effects of interface and spatial ability on manipulation of virtual models in a STEM domain. *Computers in Human Behavior* 65 (2016), 220–231. https://doi.org/10.1016/j.chb.2016.06.026

[6] Joanna Bergström, Tor-Salve Dalsgaard, Jason Alexander, and Kasper Hornbæk. 2021. How to Evaluate Object Selection and Manipulation in VR? Guidelines from 20 Years of Studies. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 533, 20 pages. https://doi.org/10.1145/3411764.3445193

[7] Lonni Besançon, Paul Issartel, Mehdi Ammi, and Tobias Isenberg. 2017. Mouse, Tactile, and Tangible Input for 3D Manipulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 4727–4740. https://doi.org/10.1145/3025453.3025863

[8] Blender Foundation. [n.d.]. *Home of the Blender project.* https://www.blender.org/ (Accessed on 12/02/2022).

[9] Benoît Bossavit, Asier Marzo, Oscar Ardaiz, Luis Diaz De Cerio, and Alfredo Pina. 2014. Design Choices and Their Implications for 3D Mid-Air Manipulation Techniques. *Presence: Teleoperators and Virtual Environments* 23, 4 (11 2014), 377–392. https://doi.org/10.1162/PRES_a_00207 arXiv:https://direct.mit.edu/pvar/article-pdf/23/4/377/1625398/pres_a_00207.pdf

[10] Doug A. Bowman, Donald B. Johnson, and Larry F. Hodges. 2001. Testbed Evaluation of Virtual Environment Interaction Techniques. *Presence: Teleoperators and Virtual Environments* 10, 1 (2001), 75–95. https://doi.org/10.1162/105474601750182333 arXiv:https://doi.org/10.1162/105474601750182333

[11] John Brooke. 1996. SUS-A Quick and Dirty Usability Scale. *Usability evaluation in industry* 189, 194 (1996).

[12] P. Carlson, A. Peters, S. B. Gilbert, J. M. Vance, and A. Luse. 2015. Virtual Training: Learning Transfer of Assembly Tasks. *IEEE Transactions on Visualization and Computer Graphics* 21, 6 (2015), 770–782.

[13] John B Carroll et al. 1993. *Human cognitive abilities: A survey of factor-analytic studies.* Number 1. Cambridge University Press.

[14] E. Chapoulie, T. Tsandilas, L. Oehlberg, W. Mackay, and G. Drettakis. 2015. Finger-based manipulation in immersive spaces and the real world. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*. 109–116.

[15] J Cohen. 1988. Statistical power analysis for the behavioral sciences. *Hoboken: Taylor and Francis* (1988).

[16] Jacob Cohen. 1992. Quantitative methods in psychology: A power primer. In *Psychological bulletin.* Citeseer.

[17] Brookshire D. Conner, Scott S. Snibbe, Kenneth P. Herndon, Daniel C. Robbins, Robert C. Zeleznik, and Andries van Dam. 1992. Three-Dimensional Widgets. In *Proceedings of the 1992 Symposium on Interactive 3D Graphics* (Cambridge, Massachusetts, USA) *(I3D '92)*. Association for Computing Machinery, New York, NY, USA, 183–188. https://doi.org/10.1145/147156.147199

[18] Valve Corporation. 2022. *Steam Hardware and Software Survey.* https://store.steampowered.com/hwsurvey/Steam-Hardware-Software-Survey-Welcome-to-Steam (Accessed on 12/02/2022).

[19] Mihaly Csikszentmihalyi. 1975. Beyond boredom and anxiety: Experiencing flow in work and play. *San Francisco/Washington/London* (1975).

[20] Mihaly Csikszentmihalyi. 1990. *Flow: The psychology of optimal experience.* Vol. 1990. Harper & Row New York.

[21] Pierre Dragicevic. 2016. *Fair Statistical Communication in HCI.* Springer International Publishing, Cham, 291–330. https://doi.org/10.1007/978-3-319-26633-6_13

[22] Tobias Drey, Patrick Albus, Simon der Kinderen, Maximilian Milo, Thilo Segschneider, Linda Chanzab, Michael Rietzler, Tina Seufert, and Enrico Rukzio. 2022. Towards Collaborative Learning in Virtual Reality: A Comparison of Co-Located Symmetric and Asymmetric Pair-Learning. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 610, 19 pages. https://doi.org/10.1145/3491102.3517641

[23] Tobias Drey, Fabian Fischbach, Pascal Jansen, Julian Frommel, Michael Rietzler, and Enrico Rukzio. 2021. To Be or Not to Be Stuck, or Is It a Continuum? A Systematic Literature Review on the Concept of Being Stuck in Games. *Proc. ACM Hum.-Comput. Interact.* 5, CHI PLAY, Article 229 (oct 2021), 35 pages. https://doi.org/10.1145/3474656

[24] Tobias Drey, Jan Gugenheimer, Julian Karlbauer, Maximilian Milo, and Enrico Rukzio. 2020. VRSketchIn: Exploring the Design Space of Pen and Tablet Interaction for 3D Sketching in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, Hawai'i, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, Article 501, 14 pages. https://doi.org/10.1145/3313831.3376628

[25] Tobias Drey, Jessica Janek, Josef Lang, Dietmar Puschmann, Michael Rietzler, and Enrico Rukzio. 2022. SpARklingPaper: Enhancing Common Pen- And Paper-Based Handwriting Training for Children by Digitally Augmenting Papers Using a Tablet Screen. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 3, Article 113 (sep 2022), 29 pages. https://doi.org/10.1145/3550337

[26] Tobias Drey, Pascal Jansen, Fabian Fischbach, Julian Frommel, and Enrico Rukzio. 2020. Towards Progress Assessment for Adaptive Hints in Educational Virtual Reality Games. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3334480.3382789

[27] J.N. Druckman, D.P. Greene, J.H. Kuklinski, and A. Lupia. 2011. *Cambridge Handbook of Experimental Political Science.* Cambridge University Press. https://books.google.de/books?id=ES5PUvvShF8C

[28] LLC. Facebook Technologies. 2022. *Oculus Quest 2.* https://www.oculus.com/quest-2/ (Accessed on 12/02/2022).

[29] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. Statistical Power Analyses Using G*Power 3.1: Tests for Correlation and Regression Analyses. *Behavior research methods* 41 (11 2009), 1149–60. https://doi.org/10.3758/BRM.41.4.1149

[30] Jinbo Feng, Isaac Cho, and Zachary Wartell. 2015. Comparison of Device-Based, One and Two-Handed 7DOF Manipulation Techniques. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction* (Los Angeles, California, USA) *(SUI '15)*. Association for Computing Machinery, New York, NY, USA, 2–9. https://doi.org/10.1145/2788940.2788942

[31] Carl F. Fey, Tianyou Hu, and Andrew Delios. 2022. The Measurement and Communication of Effect Sizes in Management Research. *Management and Organization Review* (2022), 1–22. https://doi.org/10.1017/mor.2022.2

[32] Andy Field. 2013. *Discovering statistics using IBM SPSS statistics.* sage.

[33] Holmes Finch. 2005. Comparison of the performance of nonparametric and parametric MANOVA test statistics when assumptions are violated. *Methodology* 1, 1 (2005), 27–38.

[34] Paul M Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47, 6 (1954), 381.

[35] P. Christopher Gloumeau, Wolfgang Stuerzlinger, and JungHyun Han. 2021. PinNPivot: Object Manipulation Using Pins in Immersive Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics* 27, 4 (2021), 2488–2494. https://doi.org/10.1109/TVCG.2020.2987834

[36] Nextcloud GmbH. 2022. *Nextcloud*. https://nextcloud.com/ (Accessed on 12/02/2022).
[37] Juho Hamari, Jonna Koivisto, and Harri Sarsa. 2014. Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. In *2014 47th Hawaii International Conference on System Sciences*. 3025–3034. https://doi.org/10.1109/HICSS.2014.377
[38] Justin Harris, Kathy Hirsh-Pasek, and Nora S Newcombe. 2013. Understanding spatial transformations: similarities and differences between mental rotation and mental folding. *Cognitive processing* 14, 2 (2013), 105–115.
[39] Devamardeep Hayatpur, Seongkook Heo, Haijun Xia, Wolfgang Stuerzlinger, and Daniel Wigdor. 2019. Plane, Ray, and Point: Enabling Precise Spatial Manipulations with Shape Constraints. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1185–1195. https://doi.org/10.1145/3332165.3347916
[40] Mary Hegarty and David A Waller. 2005. *Individual differences in spatial abilities*. Cambridge University Press.
[41] Ken Hinckley, Joe Tullio, Randy Pausch, Dennis Proffitt, and Neal Kassell. 1997. Usability analysis of 3D rotation techniques. In *Proceedings of the 10th annual ACM symposium on User interface software and technology*. 1–10.
[42] Stephanie Houde. 1992. Iterative Design of an Interface for Easy 3-D Direct Manipulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, USA) *(CHI '92)*. Association for Computing Machinery, New York, NY, USA, 135–142. https://doi.org/10.1145/142750.142772
[43] Marcel Adam Just and Patricia A Carpenter. 1976. Eye fixations and cognitive processes. *Cognitive Psychology* 8, 4 (1976), 441–480. https://doi.org/10.1016/0010-0285(76)90015-3
[44] MyoungGon Kim and JungHyun Han. 2019. Effects of Switchable DOF for Mid-Air Manipulation in Immersive Virtual Environments. *International Journal of Human–Computer Interaction* 35, 13 (2019), 1147–1159. https://doi.org/10.1080/10447318.2018.1514163 arXiv:https://doi.org/10.1080/10447318.2018.1514163
[45] Melina Klepsch, Florian Schmitz, and Tina Seufert. 2017. Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. *Frontiers in Psychology* 8 (2017), 1997. https://doi.org/10.3389/fpsyg.2017.01997
[46] Melina Klepsch and Tina Seufert. 2021. Making an Effort Versus Experiencing Load. *Frontiers in Education* 6 (2021). https://doi.org/10.3389/feduc.2021.645284
[47] Max Krichenbauer, Goshiro Yamamoto, Takafumi Taketom, Christian Sandor, and Hirokazu Kato. 2018. Augmented Reality versus Virtual Reality for 3D Object Manipulation. *IEEE Transactions on Visualization and Computer Graphics* 24, 2 (2018), 1038–1048. https://doi.org/10.1109/TVCG.2017.2658570
[48] A. Kulik, A. Kunert, and B. Froehlich. 2020. On Motor Performance in Virtual 3D Object Manipulation. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (2020), 2041–2050.
[49] Wallace S. Lages and Doug A. Bowman. 2018. Move the Object or Move Myself? Walking vs. Manipulation for the Examination of 3D Scientific Data. *Frontiers in ICT* 5 (2018), 15. https://doi.org/10.3389/fict.2018.00015
[50] P Lance and Aiko Hattori. 2016. Sampling and evaluation. *Chapel Hill, North Carolina: MEASURE Evaluation, University of North Carolina* (2016).
[51] Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. 2017. *3D user interfaces: theory and practice*. Addison-Wesley Professional.
[52] Marcia C. Linn and Anne C. Petersen. 1985. Emergence and Characterization of Sex Differences in Spatial Ability: A Meta-Analysis. *Child Development* 56, 6 (1985), 1479–1498. http://www.jstor.org/stable/1130467
[53] David Lubinski. 2010. Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences* 49 (09 2010), 344–351. https://doi.org/10.1016/j.paid.2010.03.022
[54] I. Scott MacKenzie. 1989. A Note on the Information-Theoretic Basis for Fitts' Law. *Journal of Motor Behavior* 21, 3 (1989), 323–330. https://doi.org/10.1080/00222895.1989.10735486 arXiv:https://doi.org/10.1080/00222895.1989.10735486 PMID: 15136269.
[55] I. Scott MacKenzie. 1992. Fitts' Law as a Research and Design Tool in Human-Computer Interaction. *Human–Computer Interaction* 7, 1 (1992), 91–139. https://doi.org/10.1207/s15327051hci07013 arXiv:https://doi.org/10.1207/s15327051hci07013
[56] Anthony Martinet, Géry Casiez, and Laurent Grisoni. 2010. The Effect of DOF Separation in 3D Manipulation Tasks with Multi-Touch Displays. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology* (Hong Kong) *(VRST '10)*. Association for Computing Machinery, New York, NY, USA, 111–118. https://doi.org/10.1145/1889863.1889888
[57] John H McDonald. 2009. *Handbook of biological statistics*. Vol. 2. sparky house publishing Baltimore, MD.
[58] D. Mendes, F. M. Caputo, A. Giachetti, A. Ferreira, and J. Jorge. 2019. A Survey on 3D Virtual Object Manipulation: From the Desktop to Immersive Virtual Environments. *Computer Graphics Forum* 38, 1 (2019), 21–45. https://doi.org/10.1111/cgf.13390 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.13390
[59] Daniel Mendes, Fernando Fonseca, Bruno Araùjo, Alfredo Ferreira, and Joaquim Jorge. 2014. Mid-air interactions above stereoscopic interactive tables. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. 3–10. https://doi.org/10.1109/3DUI.2014.6798833
[60] Daniel Mendes, Filipe Relvas, Alfredo Ferreira, and Joaquim Jorge. 2016. The Benefits of DOF Separation in Mid-Air 3D Object Manipulation. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology* (Munich, Germany) *(VRST '16)*. Association for Computing Machinery, New York, NY, USA, 261–268. https://doi.org/10.1145/2993369.2993396
[61] Daniel Mendes, Maurício Sousa, Rodrigo Lorena, Alfredo Ferreira, and Joaquim Jorge. 2017. Using Custom Transformation Axes for Mid-Air Manipulation of 3D Virtual Objects. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (Gothenburg, Sweden) *(VRST '17)*. Association for Computing Machinery, New York, NY, USA, Article 27, 8 pages. https://doi.org/10.1145/3139131.3139157
[62] Marvin E Mundell. 1978. *Motion and time study: Improving productivity*. Prentice Hall,.
[63] Gregory M. Nielson and Dan R. Olsen. 1987. Direct Manipulation Techniques for 3D Objects Using 2D Locator Devices. In *Proceedings of the 1986 Workshop on Interactive 3D Graphics* (Chapel Hill, North Carolina, USA) *(I3D '86)*. Association for Computing Machinery, New York, NY, USA, 175–182. https://doi.org/10.1145/319120.319134
[64] M. Peters, B. Laeng, K. Latham, M. Jackson, R. Zaiyouna, and C. Richardson. 1995. A Redrawn Vandenberg and Kuse Mental Rotations Test - Different Versions and Factors That Affect Performance. *Brain and Cognition* 28, 1 (1995), 39–58. https://doi.org/10.1006/brcg.1995.1032
[65] Michael Peters, John T Manning, and Stian Reimers. 2007. The effects of sex, sexual orientation, and digit ratio (2D: 4D) on mental rotation performance. *Archives of sexual behavior* 36, 2 (2007), 251–260.
[66] Ivan Poupyrev, Suzanne Weghorst, Mark Billinghurst, and Tadao Ichikawa. 1997. A Framework and Testbed for Studying Manipulation Techniques for Immersive VR. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (Lausanne, Switzerland) *(VRST '97)*. Association for Computing Machinery, New York, NY, USA, 21–28. https://doi.org/10.1145/261135.261141
[67] Prolific. [n.d.]. *Prolific | Online participant recruitment for surveys and market research*. https://prolific.co/ (Accessed on 12/02/2022).
[68] Rivu Radiah, Ville Mäkelä, Sarah Prange, Sarah Delgado Rodriguez, Robin Piening, Yumeng Zhou, Kay Köhle, Ken Pfeuffer, Yomna Abdelrahman, Matthias Hoppe, Albrecht Schmidt, and Florian Alt. 2021. Remote VR Studies: A Framework for Running Virtual Reality Studies Remotely Via Participant-Owned HMDs. *ACM Trans. Comput.-Hum. Interact.* 28, 6, Article 46 (nov 2021), 36 pages. https://doi.org/10.1145/3472617
[69] Roger N. Shepard and Jacqueline Metzler. 1971. Mental Rotation of Three-Dimensional Objects. *Science* 171, 3972 (1971), 701–703. https://doi.org/10.1126/science.171.3972.701 arXiv:http://science.sciencemag.org/content/171/3972/701.full.pdf
[70] Ken Shoemake. 1992. ARCBALL: A user interface for specifying three-dimensional orientation using a mouse. In *Proceedings of Graphics Interface '92* (Vancouver, British Columbia, Canada) *(GI '92)*. Canadian Human-Computer Communications Society, Toronto, Ontario, Canada, 151–156. http://graphicsinterface.org/wp-content/uploads/gi1992-18.pdf
[71] Peng Song, Wooi Boon Goh, William Hutama, Chi-Wing Fu, and Xiaopei Liu. 2012. A Handle Bar Metaphor for Virtual Object Manipulation with Mid-Air Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1297–1306. https://doi.org/10.1145/2207676.2208585
[72] R. William Soukoreff and I. Scott MacKenzie. 2004. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies* 61, 6 (2004), 751–789. https://doi.org/10.1016/j.ijhcs.2004.09.001 Fitts' law 50 years later: applications and contributions from human-computer interaction.
[73] Joachim Stoeber and Lavinia E. Damian. 2016. *Perfectionism in Employees: Work Engagement, Workaholism, and Burnout*. Springer International Publishing, Cham, 265–283. https://doi.org/10.1007/978-3-319-18582-8_12
[74] Martin F. Stoelen and David L. Akin. 2010. Assessment of Fitts' Law for Quantifying Combined Rotational and Translational Movements. *Human Factors* 52, 1 (2010), 63–77. https://doi.org/10.1177/0018720810366560 arXiv:https://doi.org/10.1177/0018720810366560 PMID: 20653226.
[75] Junwei Sun and Wolfgang Stuerzlinger. 2019. Extended Sliding in Virtual Reality. In *25th ACM Symposium on Virtual Reality Software and Technology* (Parramatta, NSW, Australia) *(VRST '19)*. Association for Computing Machinery, New York, NY, USA, Article 38, 5 pages. https://doi.org/10.1145/3359996.3364251
[76] John Sweller. 1988. *Cognitive Science* 12, 2 (1988), 257–285. https://doi.org/10.1207/s15516709cog1202_4 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog1202
[77] Robert J. Teather and Wolfgang Stuerzlinger. 2007. Guidelines for 3D Positioning Techniques. In *Proceedings of the 2007 Conference on Future Play* (Toronto, Canada) *(Future Play '07)*. Association for Computing Machinery, New York, NY, USA,

61–68. https://doi.org/10.1145/1328202.1328214

[78] Eleftherios Triantafyllidis and Zhibin Li. 2021. *The Challenges in Modeling Human Performance in 3D Space with Fitts’ Law*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411763.3443442

[79] Unity Technologies. [n.d.]. *Unity Real-Time Development Platform | 3D, 2D VR & AR Visualizations*. https://unity.com/ (Accessed on 12/02/2022).

[80] Steven G. Vandenberg and Allan R. Kuse. 1978. Mental Rotations, a Group Test of Three-Dimensional Spatial Visualization. *Perceptual and Motor Skills* 47, 2 (Dec. 1978), 599–604. https://doi.org/10.2466/pms.1978.47.2.599

[81] Manuel Veit, Antonio Capobianco, and Dominique Bechmann. 2009. Influence of Degrees of Freedom’s Manipulation on Performances during Orientation Tasks in Virtual Reality Environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology* (Kyoto, Japan) *(VRST ’09)*. Association for Computing Machinery, New York, NY, USA, 51–58. https://doi.org/10.1145/1643928.1643942

[82] Daniel Voyer, Susan Voyer, and M Philip Bryden. 1995. Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychological bulletin* 117, 2 (1995), 250.

[83] Vanessa Vuibert, Wolfgang Stuerzlinger, and Jeremy R. Cooperstock. 2015. Evaluation of Docking Task Performance Using Mid-Air Interaction Techniques. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction* (Los Angeles, California, USA) *(SUI ’15)*. Association for Computing Machinery, New York, NY, USA, 44–52. https://doi.org/10.1145/2788940.2788950

[84] Adrian Willings. 2022. *Meta Quest 2 vs Oculus Quest: What’s the difference between these VR headsets?* https://www.pocket-lint.com/ar-vr/buyers-guides/oculus-rift/153716-oculus-quest-2-vs-oculus-quest-whats-the-difference (Accessed on 11/28/2022).

[85] Jacob O. Wobbrock and Julie A. Kientz. 2016. Research Contributions in Human-computer Interaction. *Interactions* 23, 3 (April 2016), 38–44. https://doi.org/10.1145/2907069

[86] Steffi Zander, Stefanie Wetzel, and Sven Bertel. 2016. Rotate it! – Effects of touch-based gestures on elementary school students’ solving of mental rotation tasks. *Computers & Education* 103 (Dec. 2016), 158–169. https://doi.org/10.1016/j.compedu.2016.10.007

[87] Shumin Zhai and Paul Milgram. 1998. Quantifying Coordination in Multiple DOF Movement and Its Application to Evaluating 6 DOF Input Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Los Angeles, California, USA) *(CHI ’98)*. ACM Press/Addison-Wesley Publishing Co., USA, 320–327. https://doi.org/10.1145/274644.274689

APPENDIX

Table A1: Multiple regressions overall level: This table shows the results of the multiple regressions and how the control variables contributed to the dependent variables on the overall level (OV). Significant regressions, as well as significant contributions, are highlighted in bold.

Dependent Variables	Control Variable	Statistics
OV threshold time		$\mathbf{F(5,60) = 4.075, p = .003, R^2_{adj} = .191, f^2 = .236}$
	spatial abilities	$\boldsymbol{\beta} = \mathbf{-.290, t = -2.531, p = .014}$
	age	$\beta = .077, t = .613, p = .542$
	VR experience	$\beta = -.025, t = -.203, p = .840$
	3D modeling experience	$\beta = .014, t = .105, p = .917$
	task accuracy self-assessment	$\boldsymbol{\beta} = \mathbf{-.368, t = -3.134, p = .003}$
OV first interaction time		$F(5,60) = 2.165, p = .070, R^2_{adj} = .082, f^2 = .089$
	spatial abilities	$\beta = -.035, t = -.285, p = .777$
	age	$\beta = .122, t = .911, p = .366$
	VR experience	$\beta = .043, t = .330, p = .743$
	3D modeling experience	$\beta = .301, t = 2.187, p = .033$
	task accuracy self-assessment	$\beta = .025, t = .202, p = .840$
OV task time		$\mathbf{F(5,60) = 4.219, p = .002, R^2_{adj} = .198, f^2 = .247}$
	spatial abilities	$\boldsymbol{\beta} = \mathbf{-.287, t = -2.513, p = .015}$
	age	$\beta = .064, t = .512, p = .610$
	VR experience	$\beta = -.030, t = -.242, p = .810$
	3D modeling experience	$\beta = -.019, t = -.144, p = .886$
	task accuracy self-assessment	$\boldsymbol{\beta} = \mathbf{-.371, t = -3.175, p = .002}$
OV interaction time		$F(5,60) = 1.728, p = .142, R^2_{adj} = .053, f^2 = .056$
	spatial abilities	$\beta = -.255, t = -2.054, p = .044$
	age	$\beta = .277, t = 2.032, p = .047$
	VR experience	$\beta = -.145, t = -1.084, p = .283$
	3D modeling experience	$\beta = -.092, t = -.660, p = .512$
	task accuracy self-assessment	$\beta = -.006, t = -.050, p = .960$
OV think time		$\mathbf{F(5,60) = 3.452, p = .008, R^2_{adj} = .159, f^2 = .189}$
	spatial abilities	$\boldsymbol{\beta} = \mathbf{-.259, t = -2.214, p = .031}$
	age	$\boldsymbol{\beta} = \mathbf{.377, t = 2.933, p = .005}$
	VR experience	$\beta = -.199, t = -1.581, p = .119$
	3D modeling experience	$\beta = -.013, t = -.096, p = .924$
	task accuracy self-assessment	$\beta = .202, t = 1.685, p = .097$
OV translation error		$\mathbf{F(5,60) = 2.387, p = .048, R^2_{adj} = .096, f^2 = .106}$
	spatial abilities	$\beta = -.094, t = -.773, p = .443$
	age	$\beta = -.234, t = -1.754, p = .085$
	VR experience	$\beta = .175, t = 1.342, p = .185$
	3D modeling experience	$\beta = -.018, t = -.132, p = .895$
	task accuracy self-assessment	$\boldsymbol{\beta} = \mathbf{-.253, t = -2.036, p = .046}$
OV rotation error		$\mathbf{F(5,60) = 5.153, p < .001, R^2_{adj} = .242, f^2 = .319}$
	spatial abilities	$\beta = -.091, t = -.820, p = .416$
	age	$\beta = -.159, t = -1.301, p = .198$
	VR experience	$\beta = .078, t = .654, p = .516$
	3D modeling experience	$\beta = -.102, t = -.814, p = .419$
	task accuracy self-assessment	$\boldsymbol{\beta} = \mathbf{-.437, t = -3.841, p < .001}$
OV scale error		$\mathbf{F(5,60) = 2.849, p = .023, R^2_{adj} = .125, f^2 = .143}$
	spatial abilities	$\beta = -.101, t = -.846, p = .401$
	age	$\beta = -.121, t = -.921, p = .361$
	VR experience	$\beta = .053, t = .415, p = .680$
	3D modeling experience	$\beta = -.118, t = -.881, p = .382$
	task accuracy self-assessment	$\boldsymbol{\beta} = \mathbf{-.320, t = -2.615, p = .011}$
OV interactions		$\mathbf{F(5,60) = 3.031, p = .017, R^2_{adj} = .135, f^2 = .156}$
	spatial abilities	$\boldsymbol{\beta} = \mathbf{-.338, t = -2.847, p = .006}$
	age	$\beta = .261, t = 2.002, p = .050$
	VR experience	$\beta = -.185, t = -1.447, p = .153$
	3D modeling experience	$\beta = -.145, t = -1.083, p = .283$
	task accuracy self-assessment	$\beta = -.066, t = -.542, p = .590$
OV translation sum		$\mathbf{F(5,60) = 3.301, p = .011, R^2_{adj} = .150, f^2 = .176}$
	spatial abilities	$\boldsymbol{\beta} = \mathbf{-.272, t = -2.312, p = .024}$
	age	$\beta = -.138, t = -1.071, p = .289$
	VR experience	$\beta = .164, t = 1.291, p = .202$
	3D modeling experience	$\beta = -.121, t = -.915, p = .364$
	task accuracy self-assessment	$\beta = -.212, t = -1.757, p = .084$

continued on next page...

Table A1 – continued from previous page

Dependent Variables	Control Variable	Statistics
OV rotation sum		$\mathbf{F(5,60) = 2.926, p = .020, R^2_{adj} = .129, f^2 = .148}$
	spatial abilities	$\boldsymbol{\beta = -.268, t = -2.247, p = .028}$
	age	$\beta = -.049, t = -.375, p = .709$
	VR experience	$\beta = .052, t = .402, p = .689$
	3D modeling experience	$\beta = -.058, t = -.434, p = .666$
	task accuracy self-assessment	$\boldsymbol{\beta = -.271, t = -2.222, p = .030}$
OV scale sum		$\mathbf{F(5,60) = 3.019, p = .017, R^2_{adj} = .134, f^2 = .155}$
	spatial abilities	$\boldsymbol{\beta = -.351, t = -2.958, p = .004}$
	age	$\beta = .240, t = 1.839, p = .071$
	VR experience	$\beta = .119, t = .930, p = .356$
	3D modeling experience	$\beta = -.043, t = -.320, p = .750$
	task accuracy self-assessment	$\beta = -.111, t = -.913, p = .365$
OV active cognitive load		$F(5,60) = 1.479, p = .210, R^2_{adj} = .036, f^2 = .037$
	spatial abilities	$\beta = -.068, t = -.542, p = .590$
	age	$\beta = .282, t = 2.046, p = .045$
	VR experience	$\beta = -.143, t = -1.057, p = .295$
	3D modeling experience	$\beta = -.257, t = -1.823, p = .073$
	task accuracy self-assessment	$\beta = .029, t = .226, p = .822$
OV passive cognitive load		$F(5,60) = 1.302, p = .275, R^2_{adj} = .023, f^2 = .024$
	spatial abilities	$\beta = -.235, t = -1.863, p = .067$
	age	$\beta = .169, t = 1.223, p = .226$
	VR experience	$\beta = -.072, t = -.531, p = .597$
	3D modeling experience	$\beta = -.193, t = -1.358, p = .180$
	task accuracy self-assessment	$\beta = .079, t = .613, p = .542$

Table A2: Multiple regressions interaction technique level: This table shows the significant results of the multiple regressions and how the control variables contributed to the dependent variables on the interaction technique level (IT). Significant regressions, as well as significant contributions, are highlighted in bold.

Dependent Variables	Control Variable	Statistics
IT 1DOF threshold time		$\mathbf{F(5,60) = 4.071, p = .003, R^2_{adj} = .191, f^2 = .236}$
	spatial abilities	$\boldsymbol{\beta = -.263, t = -2.289, p = .026}$
	age	$\beta = -.009, t = -.073, p = .942$
	VR experience	$\beta = .005, t = .039, p = .969$
	3D modeling experience	$\beta = -.093, t = -.717, p = .476$
	task accuracy self-assessment	$\boldsymbol{\beta = -.346, t = -2.943, p = .005}$
IT 1DOF task time		$\mathbf{F(5,60) = 4.321, p = .002, R^2_{adj} = .203, f^2 = .255}$
	spatial abilities	$\boldsymbol{\beta = -.263, t = -2.308, p = .024}$
	age	$\beta = -.004, t = -.030, p = .976$
	VR experience	$\beta = -.007, t = -.053, p = .958$
	3D modeling experience	$\beta = -.101, t = -.788, p = .434$
	task accuracy self-assessment	$\boldsymbol{\beta = -.356, t = -3.049, p = .003}$
IT 1DOF translation error		$\mathbf{F(5,60) = 4.751, p = .001, R^2_{adj} = .224, f^2 = .289}$
	spatial abilities	$\beta = -.162, t = -1.443, p = .154$
	age	$\beta = -.202, t = -1.634, p = .107$
	VR experience	$\boldsymbol{\beta = .309, t = 2.553, p = .013}$
	3D modeling experience	$\beta = -.121, t = -.960, p = .341$
	task accuracy self-assessment	$\boldsymbol{\beta = -.305, t = -2.649, p = .010}$
IT 1DOF rotation error		$\mathbf{F(5,60) = 6.020, p < .001, R^2_{adj} = .279, f^2 = .387}$
	spatial abilities	$\boldsymbol{\beta = -.227, t = -2.092, p = .041}$
	age	$\beta = -.070, t = -.590, p = .557$
	VR experience	$\beta = .212, t = 1.818, p = .074$
	3D modeling experience	$\beta = -.207, t = -1.696, p = .095$
	task accuracy self-assessment	$\boldsymbol{\beta = -.370, t = -3.338, p = .001}$
IT 1DOF scale error		$\mathbf{F(5,60) = 2.640, p = .032, R^2_{adj} = .112, f^2 = .126}$
	spatial abilities	$\beta = -.169, t = -1.405, p = .165$
	age	$\beta = -.056, t = -.425, p = .672$
	VR experience	$\beta = .054, t = .414, p = .680$
	3D modeling experience	$\beta = -.153, t = -1.129, p = .263$
	task accuracy self-assessment	$\boldsymbol{\beta = -.276, t = -2.241, p = .029}$
IT 1DOF SUS		$F(5,60) = 1.574, p = .181, R^2_{adj} = .042, f^2 = .044$
	spatial abilities	$\beta = .029, t = .230, p = .819$
	age	$\beta = .138, t = 1.009, p = .317$
	VR experience	$\beta = -.121, t = -.901, p = .371$
	3D modeling experience	$\beta = .201, t = 1.429, p = .158$
	task accuracy self-assessment	$\beta = .153, t = 1.196, p = .236$

continued on next page...

Table A2 – continued from previous page

Dependent Variables	Control Variable	Statistics
IT 1DOF active cognitive load		$F(5,60) = 1.429, p = .227, R^2_{adj} = .032, f^2 = .033$
	spatial abilities	$\beta = -.037, t = -.293, p = .771$
	age	$\beta = .186, t = 1.346, p = .184$
	VR experience	$\beta = -.101, t = -.746, p = .459$
	3D modeling experience	$\beta = -.286, t = -2.025, p = .047$
	task accuracy self-assessment	$\beta = -.082, t = -.635, p = .528$
IT 1DOF passive cognitive load		$F(5,60) = 1.580, p = .180, R^2_{adj} = .043, f^2 = .045$
	spatial abilities	$\beta = -.237, t = -1.898, p = .063$
	age	$\beta = -.007, t = -.049, p = .961$
	VR experience	$\beta = -.052, t = -.389, p = .699$
	3D modeling experience	$\beta = -.214, t = -1.520, p = .134$
	task accuracy self-assessment	$\beta = .104, t = .812, p = .420$
IT 3DOF threshold time		$F(5,60) = 2.089, p = .079, R^2_{adj} = .077, f^2 = .083$
	spatial abilities	$\beta = -.248, t = -2.024, p = .047$
	age	$\beta = .123, t = .910, p = .367$
	VR experience	$\beta = -.046, t = -.351, p = .727$
	3D modeling experience	$\beta = .035, t = .255, p = .800$
	task accuracy self-assessment	$\beta = -.254, t = -2.028, p = .047$
IT 3DOF task time		$F(5,60) = 2.128, p = .074, R^2_{adj} = .080, f^2 = .087$
	spatial abilities	$\beta = -.250, t = -2.042, p = .046$
	age	$\beta = .121, t = .903, p = .370$
	VR experience	$\beta = -.050, t = -.379, p = .706$
	3D modeling experience	$\beta = .019, t = .140, p = .889$
	task accuracy self-assessment	$\beta = -.256, t = -2.042, p = .046$
IT 3DOF translation error		$F(5,60) = 1.211, p = .315, R^2_{adj} = .016, f^2 = .016$
	spatial abilities	$\beta = -.068, t = -.538, p = .592$
	age	$\beta = -.218, t = -1.570, p = .122$
	VR experience	$\beta = .127, t = .932, p = .355$
	3D modeling experience	$\beta = -.014, t = -.098, p = .923$
	task accuracy self-assessment	$\beta = -.150, t = -1.158, p = .252$
IT 3DOF rotation error		$\mathbf{F(5,60) = 3.026, p = .017, R^2_{adj} = .135, f^2 = .156}$
	spatial abilities	$\beta = -.014, t = -.119, p = .905$
	age	$\beta = -.098, t = -.753, p = .454$
	VR experience	$\beta = -.062, t = -.483, p = .631$
	3D modeling experience	$\beta = -.072, t = -.540, p = .591$
	task accuracy self-assessment	$\boldsymbol{\beta} \mathbf{= -.382, t = -3.142, p = .003}$
IT 3DOF scale error		$F(5,60) = 1.587, p = .178, R^2_{adj} = .043, f^2 = .045$
	spatial abilities	$\beta = -.136, t = -1.092, p = .279$
	age	$\beta = -.082, t = -.598, p = .552$
	VR experience	$\beta = .041, t = .305, p = .762$
	3D modeling experience	$\beta = -.175, t = -1.247, p = .217$
	task accuracy self-assessment	$\beta = -.157, t = -1.226, p = .225$
IT 3DOF SUS		$\mathbf{F(5,60) = 3.665, p = .006, R^2_{adj} = .170, f^2 = .205}$
	spatial abilities	$\beta = .089, t = .762, p = .449$
	age	$\beta = -.096, t = -.755, p = .453$
	VR experience	$\boldsymbol{\beta} \mathbf{= -.271, t = -2.167, p = .034}$
	3D modeling experience	$\boldsymbol{\beta} \mathbf{= .284, t = 2.170, p = .034}$
	task accuracy self-assessment	$\boldsymbol{\beta} \mathbf{= .288, t = 2.418, p = .019}$
IT 3DOF active cognitive load		$F(5,60) = 1.242, p = .301, R^2_{adj} = .018, f^2 = .018$
	spatial abilities	$\beta = -.019, t = -.151, p = .880$
	age	$\beta = .279, t = 2.009, p = .049$
	VR experience	$\beta = -.025, t = -.182, p = .857$
	3D modeling experience	$\beta = -.283, t = -1.991, p = .051$
	task accuracy self-assessment	$\beta = .019, t = .150, p = .881$
IT 3DOF passive cognitive load		$F(5,60) = 1.276, p = .286, R^2_{adj} = .021, f^2 = .021$
	spatial abilities	$\beta = -.163, t = -1.293, p = .201$
	age	$\beta = .212, t = 1.532, p = .131$
	VR experience	$\beta = .039, t = .288, p = .775$
	3D modeling experience	$\beta = -.250, t = -1.757, p = .084$
	task accuracy self-assessment	$\beta = -.051, t = -.392, p = .696$
IT 7DOF threshold time		$\mathbf{F(5,60) = 4.698, p = .001, R^2_{adj} = .221, f^2 = .284}$
	spatial abilities	$\beta = -.196, t = -1.741, p = .087$
	age	$\beta = .232, t = 1.876, p = .066$
	VR experience	$\beta = -.071, t = -.584, p = .562$
	3D modeling experience	$\boldsymbol{\beta} \mathbf{= .338, t = 2.664, p = .010}$
	task accuracy self-assessment	$\boldsymbol{\beta} \mathbf{= -.309, t = -2.677, p = .010}$

continued on next page...

Table A2 – continued from previous page

Dependent Variables	Control Variable	Statistics
IT 7DOF task time		$\mathbf{F(5, 60) = 3.327, p = .010, R^2_{adj} = .152, f^2 = .179}$
	spatial abilities	$\beta = -.198, t = -1.684, p = .097$
	age	$\beta = .165, t = 1.280, p = .205$
	VR experience	$\beta = -.056, t = -.443, p = .659$
	3D modeling experience	$\beta = .251, t = 1.898, p = .063$
	task accuracy self-assessment	$\boldsymbol{\beta} \mathbf{= -.326, t = -2.709, p = .009}$
IT 7DOF translation error		$F(5, 60) = 1.513, p = .199, R^2_{adj} = .038, f^2 = .040$
	spatial abilities	$\beta = -.018, t = -.146, p = .884$
	age	$\beta = -.214, t = -1.553, p = .126$
	VR experience	$\beta = .028, t = .209, p = .835$
	3D modeling experience	$\beta = .106, t = .749, p = .457$
	task accuracy self-assessment	$\beta = -.260, t = -2.033, p = .047$
IT 7DOF rotation error		$F(5, 60) = 1.650, p = .161, R^2_{adj} = .048, f^2 = .050$
	spatial abilities	$\beta = -.024, t = -.196, p = .846$
	age	$\beta = -.196, t = -1.432, p = .157$
	VR experience	$\beta = .103, t = .768, p = .446$
	3D modeling experience	$\beta = .010, t = .070, p = .945$
	task accuracy self-assessment	$\beta = -.260, t = -2.039, p = .046$
IT 7DOF scale error		$\mathbf{F(5, 60) = 2.379, p = .049, R^2_{adj} = .096, f^2 = .106}$
	spatial abilities	$\beta = .055, t = .452, p = .653$
	age	$\beta = -.167, t = -1.253, p = .215$
	VR experience	$\beta = .039, t = .299, p = .766$
	3D modeling experience	$\beta = .033, t = .240, p = .811$
	task accuracy self-assessment	$\boldsymbol{\beta} \mathbf{= -.370, t = -2.979, p = .004}$
IT 7DOF SUS		$\mathbf{F(5, 60) = 2.754, p = .026, R^2_{adj} = .119, f^2 = .135}$
	spatial abilities	$\beta = .200, t = 1.673, p = .100$
	age	$\beta = .071, t = .536, p = .594$
	VR experience	$\beta = .164, t = 1.271, p = .209$
	3D modeling experience	$\boldsymbol{\beta} \mathbf{= -.323, t = -2.397, p = .020}$
	task accuracy self-assessment	$\boldsymbol{\beta} \mathbf{= .271, t = 2.210, p = .031}$
IT 7DOF active cognitive load		$F(5, 60) = 1.718, p = .144, R^2_{adj} = .052, f^2 = .055$
	spatial abilities	$\beta = -.121, t = -.978, p = .332$
	age	$\beta = .264, t = 1.932, p = .058$
	VR experience	$\beta = -.246, t = -1.837, p = .071$
	3D modeling experience	$\beta = -.087, t = -.623, p = .536$
	task accuracy self-assessment	$\beta = .144, t = 1.131, p = .262$
IT 7DOF passive cognitive load		$F(5, 60) = .983, p = .436, R^2_{adj} = -.001, f^2 = -.001$
	spatial abilities	$\beta = -.152, t = -1.189, p = .239$
	age	$\beta = .198, t = 1.408, p = .164$
	VR experience	$\beta = -.147, t = -1.068, p = .290$
	3D modeling experience	$\beta = -.001, t = -.009, p = .993$
	task accuracy self-assessment	$\beta = .123, t = .940, p = .351$

Table A3: MANOVAs interaction technique level: This table shows the MANOVAs for the interaction technique level (IT) as well as the univariate post hoc tests (ANOVA) using Bonferroni correction for the pairwise comparisons. Significant tests are highlighted in bold. It complements Fig. 4.

Dependent Variables	Pairwise Comparisons	Statistics
Lower Spatial Abilities Group (MANOVA)		$\mathbf{F(16, 21) = 6.787, p < .001, Wilks' \Lambda = .162, partial\ \eta^2 = .838}$
IT threshold time		$\mathbf{F(1.310, 47.145) = 50.223, p < .001, partial\ \eta^2 = .582}$
	1DOF* vs. *3DOF	$\mathbf{p < .001}$
	1DOF* vs. *7DOF	$\mathbf{p < .001}$
	3DOF* vs. *7DOF	$\mathbf{p < .001}$
IT task time		$\mathbf{F(1.285, 46.271) = 47.773, p < .001, partial\ \eta^2 = .570}$
	1DOF* vs. *3DOF	$\mathbf{p < .001}$
	1DOF* vs. *7DOF	$\mathbf{p < .001}$
	3DOF* vs. *7DOF	$\mathbf{p < .001}$
IT translation error		$F(2, 72) = .642, p = .529, partial\ \eta^2 = .018$
	1DOF vs. *3DOF*	$p = 1.000$
	1DOF vs. *7DOF*	$p = .904$
	3DOF vs. *7DOF*	$p = 1.000$
IT rotation error		$F(1.478, 53.215) = .805, p = .419, partial\ \eta^2 = .022$
	1DOF vs. *3DOF*	$p = 1.000$
	1DOF vs. *7DOF*	$p = .251$
	3DOF vs. *7DOF*	$p = 1.000$
IT scale error		$F(2, 72) = 1.271, p = .287, partial\ \eta^2 = .034$
	1DOF vs. *3DOF*	$p = .662$
	1DOF vs. *7DOF*	$p = 1.000$
	3DOF vs. *7DOF*	$p = .498$

continued on next page...

Table A3 – continued from previous page

Dependent Variables	Pairwise Comparisons	Statistics
IT SUS		**$F(2,72) = 6.250$, $p = .003$, partial $\eta^2 = .148$**
	1DOF vs. *3DOF*	$p = .251$
	1DOF* vs. *7DOF	**$p = .005$**
	3DOF vs. *7DOF*	$p = .225$
IT active cognitive load		$F(2,72) = 1.800, p = .173, partial\ \eta^2 = .048$
	1DOF vs. *3DOF*	$p = 1.000$
	1DOF vs. *7DOF*	$p = .233$
	3DOF vs. *7DOF*	$p = .373$
IT passive cognitive load		**$F(2,72) = 5.855$, $p = .004$, partial $\eta^2 = .140$**
	1DOF vs. *3DOF*	$p = .108$
	1DOF* vs. *7DOF	**$p = .009$**
	3DOF vs. *7DOF*	$p = .602$
Higher Spatial Abilities Group (MANOVA)		**$F(16,13) = 4.691$, $p = .004$, Wilks' $\Lambda = .148$, partial $\eta^2 = .852$**
IT threshold time		**$F(1.324, 37.082) = 38.430$, $p < .001$, partial $\eta^2 = .579$**
	1DOF* vs. *3DOF	**$p < .001$**
	1DOF* vs. *7DOF	**$p < .001$**
	3DOF* vs. *7DOF	**$p < .001$**
IT task time		**$F(1.332, 37.307) = 42.514$, $p < .001$, partial $\eta^2 = .603$**
	1DOF* vs. *3DOF	**$p < .001$**
	1DOF* vs. *7DOF	**$p < .001$**
	3DOF* vs. *7DOF	**$p < .001$**
IT translation error		$F(2,56) = .766, p = .470, partial\ \eta^2 = .027$
	1DOF vs. *3DOF*	$p = .614$
	1DOF vs. *7DOF*	$p = 1.000$
	3DOF vs. *7DOF*	$p = 1.000$
IT rotation error		**$F(2,56) = 6.623$, $p = .003$, partial $\eta^2 = .191$**
	1DOF vs. *3DOF*	$p = .262$
	1DOF* vs. *7DOF	**$p = .007$**
	3DOF vs. *7DOF*	$p = .160$
IT scale error		**$F(2,56) = 3.919$, $p = .026$, partial $\eta^2 = .123$**
	1DOF vs. *3DOF*	$p = 1.000$
	1DOF vs. *7DOF*	$p = .329$
	3DOF* vs. *7DOF	**$p = .018$**
IT SUS		**$F(2,56) = 5.336$, $p = .008$, partial $\eta^2 = .160$**
	1DOF vs. *3DOF*	$p = .211$
	1DOF* vs. *7DOF	**$p = .028$**
	3DOF vs. *7DOF*	$p = .279$
IT active cognitive load		$F(2,56) = 1.481, p = .236, partial\ \eta^2 = .050$
	1DOF vs. *3DOF*	$p = 1.000$
	1DOF vs. *7DOF*	$p = .391$
	3DOF vs. *7DOF*	$p = .809$
IT passive cognitive load		**$F(2,56) = 4.480$, $p = .016$, partial $\eta^2 = .138$**
	1DOF vs. *3DOF*	$p = .635$
	1DOF* vs. *7DOF	**$p = .029$**
	3DOF vs. *7DOF*	$p = .249$

Table A4: Multiple regressions manipulation task level: This table shows the significant results of the multiple regressions and how the control variables contributed to the dependent variables on the manipulation task level (MT). Significant regressions, as well as significant contributions, are highlighted in bold.

Dependent Variables	Control Variable	Statistics
MT 3DOF threshold time		**$F(5,60) = 3.513$, $p = .007$, $R^2_{adj} = .162$, $f^2 = .193$**
	spatial abilities	**$\beta = -.419$, $t = -3.584$, $p < .001$**
	age	$\beta = .151, t = 1.179, p = .243$
	VR experience	$\beta = -.041, t = -.326, p = .746$
	3D modeling experience	$\beta = -.074, t = -.563, p = .575$
	task accuracy self-assessment	$\beta = -.138, t = -1.152, p = .254$
MT 3DOF task time		**$F(5,60) = 3.708$, $p = .005$, $R^2_{adj} = .172$, $f^2 = .208$**
	spatial abilities	**$\beta = -.412$, $t = -3.550$, $p < .001$**
	age	$\beta = .119, t = .930, p = .356$
	VR experience	$\beta = -.019, t = -.150, p = .881$
	3D modeling experience	$\beta = -.142, t = -1.089, p = .280$
	task accuracy self-assessment	$\beta = -.140, t = -1.174, p = .245$
MT 3DOF translation error		$F(5,60) = 1.595, p = .175, R^2_{adj} = .044, f^2 = .046$
	spatial abilities	$\beta = -.081, t = -.653, p = .516$
	age	$\beta = -.190, t = -1.387, p = .171$
	VR experience	$\beta = .207, t = 1.543, p = .128$
	3D modeling experience	$\beta = -.005, t = -.036, p = .972$
	task accuracy self-assessment	$\beta = -.193, t = -1.510, p = .136$

continued on next page...

Table A4 – continued from previous page

Dependent Variables	Control Variable	Statistics
MT 3DOF rotation error		$\mathbf{F(5,60) = 3.443, p = .008, R^2_{adj} = .158, f^2 = .188}$
	spatial abilities	$\beta = -.159, t = -1.358, p = .180$
	age	$\beta = -.128, t = -.997, p = .323$
	VR experience	$\beta = .196, t = 1.556, p = .125$
	3D modeling experience	$\beta = -.166, t = -1.261, p = .212$
	task accuracy self-assessment	$\boldsymbol{\beta} = \mathbf{-.285, t = -2.381, p = .020}$
MT 3DOF scale error		$F(5,60) = 1.625, p = .167, R^2_{adj} = .046, f^2 = .048$
	spatial abilities	$\beta = -.198, t = -1.592, p = .117$
	age	$\beta = -.018, t = -.132, p = .895$
	VR experience	$\beta = .008, t = .057, p = .955$
	3D modeling experience	$\beta = -.084, t = -.600, p = .551$
	task accuracy self-assessment	$\beta = -.208, t = -1.627, p = .109$
MT 6DOF threshold time		$\mathbf{F(5,60) = 3.861, p = .004, R^2_{adj} = .180, f^2 = .220}$
	spatial abilities	$\beta = -.167, t = -1.449, p = .153$
	age	$\beta = -.014, t = -.114, p = .910$
	VR experience	$\beta = -.047, t = -.375, p = .709$
	3D modeling experience	$\beta = .066, t = .507, p = .614$
	task accuracy self-assessment	$\boldsymbol{\beta} = \mathbf{-.440, t = -3.716, p < .001}$
MT 6DOF task time		$\mathbf{F(5,60) = 3.938, p = .004, R^2_{adj} = .184, f^2 = .225}$
	spatial abilities	$\beta = -.164, t = -1.426, p = .159$
	age	$\beta = -.021, t = -.164, p = .871$
	VR experience	$\beta = -.056, t = -.452, p = .653$
	3D modeling experience	$\beta = .051, t = .394, p = .695$
	task accuracy self-assessment	$\boldsymbol{\beta} = \mathbf{-.441, t = -3.735, p < .001}$
MT 6DOF translation error		$\mathbf{F(5,60) = 2.666, p = .031, R^2_{adj} = .114, f^2 = .129}$
	spatial abilities	$\beta = -.082, t = -.680, p = .499$
	age	$\beta = -.246, t = -1.861, p = .068$
	VR experience	$\beta = .113, t = .874, p = .386$
	3D modeling experience	$\beta = -.005, t = -.036, p = .971$
	task accuracy self-assessment	$\boldsymbol{\beta} = \mathbf{-.290, t = -2.361, p = .021}$
MT 6DOF rotation error		$\mathbf{F(5,60) = 3.733, p = .005, R^2_{adj} = .174, f^2 = .211}$
	spatial abilities	$\beta = -.012, t = -.108, p = .915$
	age	$\beta = -.133, t = -1.042, p = .302$
	VR experience	$\beta = -.014, t = -.112, p = .911$
	3D modeling experience	$\beta = -.034, t = -.261, p = .795$
	task accuracy self-assessment	$\boldsymbol{\beta} = \mathbf{-.434, t = -3.653, p < .001}$
MT 6DOF scale error		$\mathbf{F(5,60) = 2.941, p = .019, R^2_{adj} = .130, f^2 = .149}$
	spatial abilities	$\beta = .013, t = .108, p = .914$
	age	$\beta = -.206, t = -1.579, p = .120$
	VR experience	$\beta = .109, t = .852, p = .397$
	3D modeling experience	$\beta = -.138, t = -1.031, p = .307$
	task accuracy self-assessment	$\boldsymbol{\beta} = \mathbf{-.301, t = -2.474, p = .016}$
MT 7DOF threshold time		$F(5,60) = 1.412, p = .233, R^2_{adj} = .031, f^2 = .032$
	spatial abilities	$\beta = -.204, t = -1.621, p = .110$
	age	$\beta = .180, t = 1.301, p = .198$
	VR experience	$\beta = .092, t = .682, p = .498$
	3D modeling experience	$\beta = -.030, t = -.209, p = .835$
	task accuracy self-assessment	$\beta = -.147, t = -1.143, p = .258$
MT 7DOF task time		$F(5,60) = 1.527, p = .195, R^2_{adj} = .039, f^2 = .041$
	spatial abilities	$\beta = -.217, t = -1.738, p = .087$
	age	$\beta = .195, t = 1.421, p = .160$
	VR experience	$\beta = .068, t = .502, p = .618$
	3D modeling experience	$\beta = -.034, t = -.241, p = .811$
	task accuracy self-assessment	$\beta = -.155, t = -1.212, p = .230$
MT 7DOF translation error		$F(5,60) = 2.113, p = .076, R^2_{adj} = .079, f^2 = .086$
	spatial abilities	$\beta = -.112, t = -.915, p = .364$
	age	$\beta = -.202, t = -1.503, p = .138$
	VR experience	$\beta = .074, t = .564, p = .575$
	3D modeling experience	$\beta = -.110, t = -.795, p = .430$
	task accuracy self-assessment	$\beta = -.199, t = -1.589, p = .117$
MT 7DOF rotation error		$\mathbf{F(5,60) = 3.351, p = .010, R^2_{adj} = .153, f^2 = .181}$
	spatial abilities	$\beta = -.132, t = -1.128, p = .264$
	age	$\beta = -.162, t = -1.256, p = .214$
	VR experience	$\beta = .049, t = .387, p = .700$
	3D modeling experience	$\beta = -.089, t = -.676, p = .502$
	task accuracy self-assessment	$\boldsymbol{\beta} = \mathbf{-.329, t = -2.739, p = .008}$

continued on next page...

Table A4 – continued from previous page

Dependent Variables	Control Variable	Statistics
MT 7DOF scale error		$\mathbf{F(5, 60) = 2.851, p = .022, R^2_{adj} = .125, f^2 = .143}$
	spatial abilities	$\beta = -.039, t = -.329, p = .744$
	age	$\beta = -.087, t = -.666, p = .508$
	VR experience	$\beta = .000, t = .001, p = .999$
	3D modeling experience	$\beta = -.067, t = -.497, p = .621$
	task accuracy self-assessment	$\boldsymbol{\beta} = \mathbf{-.382, t = -3.127, p = .003}$

Table A5: MANOVAs manipulation task level: This table shows the MANOVAs for the manipulation task level (MT) as well as the univariate post hoc tests (ANOVA) using Bonferroni correction for the pairwise comparisons. Significant tests are highlighted in bold. It complements Fig. 5.

Dependent Variables	Pairwise Comparisons	Statistics
Lower Spatial Abilities Group (MANOVA)		$\mathbf{F(10, 27) = 17.119, p < .001, Wilks' \Lambda = .136, partial\ \boldsymbol{\eta}^2 = .864}$
MT threshold time		$\mathbf{F(2, 72) = 16.470, p < .001, partial\ \boldsymbol{\eta}^2 = .314}$
	3DOF* vs. *6DOF	$\mathbf{p < .001}$
	3DOF* vs. *7DOF	$\mathbf{p < .001}$
	6DOF vs. *7DOF*	$p = 1.000$
MT task time		$\mathbf{F(1.352, 48.659) = 40.441, p < .001, partial\ \boldsymbol{\eta}^2 = .529}$
	3DOF* vs. *6DOF	$\mathbf{p < .001}$
	3DOF* vs. *7DOF	$\mathbf{p = .002}$
	6DOF* vs. *7DOF	$\mathbf{p < .001}$
MT translation error		$\mathbf{F(1.376, 49.541) = 12.946, p < .001, partial\ \boldsymbol{\eta}^2 = .264}$
	3DOF vs. *6DOF*	$p = 1.000$
	3DOF* vs. *7DOF	$\mathbf{p = .005}$
	6DOF* vs. *7DOF	$\mathbf{p < .001}$
MT rotation error		$\mathbf{F(1.452, 52.282) = 32.476, p < .001, partial\ \boldsymbol{\eta}^2 = .474}$
	3DOF* vs. *6DOF	$\mathbf{p = .019}$
	3DOF* vs. *7DOF	$\mathbf{p < .001}$
	6DOF* vs. *7DOF	$\mathbf{p < .001}$
MT scale error		$\mathbf{F(2, 72) = 37.086, p < .001, partial\ \boldsymbol{\eta}^2 = .507}$
	3DOF vs. *6DOF*	$p = .188$
	3DOF* vs. *7DOF	$\mathbf{p < .001}$
	6DOF* vs. *7DOF	$\mathbf{p < .001}$
Higher Spatial Abilities Group (MANOVA)		$\mathbf{F(10, 19) = 20.172, p < .001, Wilks' \Lambda = .086, partial\ \boldsymbol{\eta}^2 = .914}$
MT threshold time		$\mathbf{F(2, 56) = 40.959, p < .001, partial\ \boldsymbol{\eta}^2 = .594}$
	3DOF* vs. *6DOF	$\mathbf{p < .001}$
	3DOF* vs. *7DOF	$\mathbf{p < .001}$
	6DOF vs. *7DOF*	$p = 1.000$
MT task time		$\mathbf{F(1.276, 35.725) = 73.730, p < .001, partial\ \boldsymbol{\eta}^2 = .725}$
	3DOF* vs. *6DOF	$\mathbf{p < .001}$
	3DOF vs. *7DOF*	$p = .327$
	6DOF* vs. *7DOF	$\mathbf{p < .001}$
MT translation error		$\mathbf{F(2, 56) = 17.340, p < .001, partial\ \boldsymbol{\eta}^2 = .382}$
	3DOF vs. *6DOF*	$p = .466$
	3DOF* vs. *7DOF	$\mathbf{p = .001}$
	6DOF* vs. *7DOF	$\mathbf{p < .001}$
MT rotation error		$\mathbf{F(1.616, 45.247) = 61.837, p < .001, partial\ \boldsymbol{\eta}^2 = .688}$
	3DOF* vs. *6DOF	$\mathbf{p < .001}$
	3DOF* vs. *7DOF	$\mathbf{p < .001}$
	6DOF* vs. *7DOF	$\mathbf{p < .001}$
MT scale error		$\mathbf{F(2, 56) = 43.179, p < .001, partial\ \boldsymbol{\eta}^2 = .607}$
	3DOF* vs. *6DOF	$\mathbf{p < .001}$
	3DOF* vs. *7DOF	$\mathbf{p < .001}$
	6DOF* vs. *7DOF	$\mathbf{p < .001}$