

Specific Immersive Competence in VR-Based Assessments: Development, Psychometric Evaluation and Associations with Medical Performance

Verena Schreiner, Joy Backhaus, Marco Lindner, Melina Heinisch, Sarah König,
Sebastian Oberdörfer, Tobias Mühling

Submitted to: JMIR Medical Education
on: August 10, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 29

 Figures 30

 Figure 1..... 31

 Figure 2..... 32

 Figure 3..... 33

 Multimedia Appendixes 34

 Multimedia Appendix 1..... 35

 Multimedia Appendix 2..... 35

Specific Immersive Competence in VR-Based Assessments: Development, Psychometric Evaluation and Associations with Medical Performance

Verena Schreiner¹; Joy Backhaus¹ MSc; Marco Lindner¹ MD; Melina Heinisch² MSc; Sarah König¹ MD, MME; Sebastian Oberdörfer² PhD; Tobias Mühling¹ MD

¹ University Hospital Würzburg Institute of Medical Teaching and Medical Educational Research Würzburg DE

² Julius Maximilian University of Würzburg Chair for Human-Computer-Interaction Würzburg DE

Corresponding Author:

Tobias Mühling MD

University Hospital Würzburg
Institute of Medical Teaching and Medical Educational Research
Joseph-Schneider-Strasse 2
Würzburg
DE

Abstract

Background: Virtual reality (VR) is increasingly used in medical education for training and examination purposes, yet learners' performance in VR-based assessments may be influenced by more than clinical competence alone. Immersive competence (IC) has been proposed as a relevant factor in VR-based performance assessment. While general IC captures application-independent VR interaction skills, domain-specific applications may require an additional construct - specific IC - reflecting situational interaction proficiency.

Objective: This pilot study aimed (1) to examine the psychometric properties of a newly developed in situ checklist for assessing specific IC, (2) to investigate the relationship between specific and general IC as well as related human abilities and characteristics, and (3) to explore associations between both IC measures and medical performance in a VR-based OSCE station.

Methods: This observational follow-up study was conducted with 21 final-year medical students who had previously completed a curricular OSCE including a VR-based emergency medicine station. General IC was assessed using the VR competence app, and specific IC using a checklist embedded in the original VR simulation. Additional measures included self-reported technological affinity, spatial ability, and OSCE performance scores. Item difficulty, item-total correlations, and internal consistency were calculated for the specific IC checklist. Pearson's r was used to examine associations among variables.

Results: The final 13-item specific IC checklist demonstrated acceptable internal consistency (Cronbach's $\alpha = .79$) and balanced item difficulty (mean $P = 0.56$). Specific IC showed a strong correlation with general IC ($r = 0.56$, $p = .008$) and with prior 3D application experience ($r = 0.57$, $p = .007$), but no relevant association with spatial ability. Both general and specific IC correlated moderately with VR-OSCE performance ($r = 0.41$, $p = .06$ and $r = 0.37$, $p = .09$, respectively), but neither was related to overall analog OSCE performance.

Conclusions: Specific IC appears to be a psychometrically accessible and context-sensitive dimension of VR user readiness. The proposed in situ approach offers a feasible and scalable method to assess this construct within domain-specific VR applications. While correlations with medical performance were limited in this small sample, the findings underscore the importance of considering IC in VR-based high-stakes assessments to support fairness and validity.

(JMIR Preprints 10/08/2025:82136)

DOI: <https://doi.org/10.2196/preprints.82136>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <https://www.jmir.org/>

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://www.jmir.org/>



Original Manuscript

Specific Immersive Competence in VR-Based Assessments: Development, Psychometric Evaluation and Associations with Medical Performance

Verena Schreiner¹, Joy Backhaus¹, Marco Lindner¹, Melina Heinisch², Sarah König¹, Sebastian Oberdörfer², Tobias Mühling^{1*}

¹ Institute of Medical Teaching and Medical Educational Research, University Hospital Würzburg, Würzburg, Germany

² Chair for Human-Computer-Interaction, Julius Maximilian University of Würzburg, Würzburg, Germany

* Corresponding author

Abstract

Background: Virtual reality (VR) is increasingly used in medical education for training and examination purposes, yet learners' performance in VR-based assessments may be influenced by more than clinical competence alone. Immersive competence (IC) has been proposed as a relevant factor in VR-based performance assessment. While general IC captures application-independent VR interaction skills, domain-specific applications may require an additional construct - specific IC - reflecting situational interaction proficiency.

Objective: This pilot study aimed (1) to examine the psychometric properties of a newly developed in situ checklist for assessing specific IC, (2) to investigate the relationship between specific and general IC as well as related human abilities and characteristics, and (3) to explore associations between both IC measures and medical performance in a VR-based OSCE station.

Methods: This study was conducted with 21 final-year medical students who had previously completed a curricular OSCE including a VR-based emergency medicine station. General IC was assessed using the VR competence app, and specific IC using a checklist embedded in the original VR simulation. Additional measures included self-reported technological affinity, spatial ability, and OSCE performance scores. Item difficulty, item-total correlations, and internal consistency were calculated for the specific IC checklist. Pearson's r was used to examine associations among variables.

Results: The final 13-item specific IC checklist demonstrated acceptable internal consistency (Cronbach's $\alpha = .79$) and balanced item difficulty (mean $P = 0.56$). Specific IC showed a strong correlation with general IC ($r = 0.56$, $p = .008$) and with prior 3D application experience ($r = 0.57$, $p = .007$), but no relevant association with spatial ability. Both general and specific IC correlated moderately with VR-OSCE performance ($r = 0.41$, $p = .06$ and $r = 0.37$, $p = .09$, respectively), but neither was related to overall analog OSCE performance.

Conclusions: Specific IC appears to be a psychometrically accessible and context-sensitive dimension of VR user readiness. The proposed in situ approach offers a feasible and scalable method to assess this construct within domain-specific VR applications. While correlations with medical performance were limited in this small sample, the findings underscore the importance of considering IC in VR-based high-stakes assessments to support fairness and validity.

Keywords: virtual reality, immersive competence, OSCE, medical education, performance assessment, psychometrics

Introduction

Virtual reality (VR) is increasingly being used in medical education, providing students with

immersive and interactive environments that support skill acquisition, deliberate practice, and feedback-driven learning [1]. VR-based training formats can support the development of both technical skills, such as those required in laparoscopic surgery [2], as well as procedural competencies, including clinical reasoning [3], teamwork [4], and the management of rarely encountered scenarios such as medical emergencies [5] or brain death diagnostics [6]. Beyond its role in simulation-based training, VR is progressively being integrated into performance assessment, offering authentic and interactive scenarios for the structured evaluation of clinical competencies [7–9]. A growing number of studies have explored the use of VR for high-stakes evaluations, such as virtual Objective Structured Clinical Examinations (OSCEs), in which learners' clinical decisions and actions are assessed within pre-defined clinical scenarios in an immersive environment [10–12]. Such formats allow for precise logging of user behavior, consistent scenario delivery across candidates, and automated scoring options. Moreover, VR-based assessments have been recognized for their scalability, objectivity, and ability to replicate complex clinical environments that are challenging or resource intensive to recreate in conventional OSCE settings [8, 10–12]

As the boundaries between learning and testing environments become increasingly fluid, the question emerges as to whether learners' performance in VR accurately reflects their actual clinical competence, or whether it is confounded by interface-related factors. This concern is reinforced by the fact, that VR users must adapt to abstract and often non-standardized metaphorical representations [13], such as administering medication via 2D menus or receiving feedback through controller vibrations. Steed et al. [14] introduced the concept of *Immersive Competence* (IC), defined as a user's familiarity and skill with immersive controls. IC increases with prior VR experience [15, 16] but also appears to be trainable [17], thereby reducing cognitive load and enhancing multitasking performance in VR environments [17]. From a validity perspective, IC may constitute a source of construct-irrelevant variance [18], as it can influence performance in VR-based assessments without being part of the intended construct of clinical competence. Indeed, marked differences in item discrimination parameters between VR-based and content-equivalent analog OSCE stations already suggested a systematic bias in VR-based examinations, potentially driven by differences in IC [12]. To capture the construct of IC, based on previous approaches [19, 20], a novel assessment tool was developed: the VR competence app [16]. This application presents users with a range of prototypical interactions commonly required in immersive

environments, such as picking up and placing objects, spatial navigation, and the activation of user interface (UI) elements. Importantly, it is content-agnostic and does not reference any specific educational or clinical scenario. As such, it enables a standardized and domain-independent estimation of a user's general interactional proficiency in VR (general IC).

However, VR interfaces vary considerably across applications and hardware platforms [21]. This heterogeneity can be confusing for users and hinder the effective implementation of this technology [22, 23]. Differences include customized controller mappings, unique interface layouts, and spatial or logistical features specific to a simulation [21, 24]. For instance, general familiarity with UI activation in VR environments does not necessarily translate into proficiency with context-specific tasks such as adjusting infusion rates or navigating electronic menus in a VR-based emergency room simulation. The importance of distinguishing between general (medium-related) and specific (content- or task-related) operational competences has been highlighted by studies from other fields: For example, general internet skills e.g. like navigating websites or conducting basic research have been shown to be a poor predictor of more specific internet skills such as selecting appropriate websites for information retrieval or critically evaluating information sources [25]. Similarly, principles derived from the acquisition of simple motor tasks often fail to generalize to complex, context-rich skill environments [26]. While usability measures focus on the system's design quality [27] and general IC captures abstract interaction skills [16], neither adequately reflects a user's ability to operate a specific VR interface under real-world constraints. Technological and digital literacy are typically self-reported and too general to capture the specific, context-bound interaction skills required in VR-based assessments [28, 29]. This underscores the need for a complementary measure that reflects application-specific motor, cognitive, and interface-related skills. Here, we introduce the concept of *Specific Immersive Competence* (specific IC) - a context-sensitive construct that reflects a user's ability to perform interactional tasks within a particular, domain-relevant VR application. Specific IC should ideally be measured directly *in situ*, that is, directly within the target application e.g. by using guided, non-clinical tasks designed to isolate interactional skills from domain-specific knowledge. This approach maximizes contextual congruence between training, task, and testing environments, in line with best practices in simulation-based assessment [30, 31]. In the presented implementation here, users are guided through tasks composed of exemplary interaction steps (e.g., connecting an IV line or

confirming a medication dose via UI). Task completion within a given time frame is logged using an objective checklist format. This enables a targeted assessment of users' interactional readiness for high-stakes VR-based training and assessments.

To examine the relationship between specific and general immersive IC and their association with students' performance, as well as to gather initial validity evidence for the specific IC construct, we conducted a follow-up study within an existing curricular assessment framework. Specifically, we recruited participants from a cohort of advanced clinical-phase medical students who had previously completed a VR-based station as part of their OSCE [12]. This post-hoc recruitment strategy allowed us to link newly collected data on both general and specific IC with previously obtained medical performance scores - both for the VR station and for the OSCE as a whole. In doing so, this pilot study pursued three objectives:

- (1) We first examined the psychometric properties of the newly developed in-situ checklist for assessing specific IC, including item difficulty, discrimination (via item-total correlations), and internal consistency.
- (2) Second, we explored the relationship between specific and general IC as well as established human abilities and characteristics to explore whether both constructs reflect distinct aspects of VR interaction competence.
- (3) Finally, we analyzed the associations between both IC measures and students' medical performance in a VR-based OSCE station as well as the overall OSCE score, to provide initial evidence of IC's potential impact on VR-based performance assessment.

Methods

Apparatus

The *VR competence app* (developed by the Chair of Human-Computer-Interaction, University of Wuerzburg, Germany) was used for the assessment of general IC. *STEP-VR* (Simulation-based Training of Medical Emergencies for Physicians Using Virtual Reality; version 0.11b; ThreeDee GmbH, Munich, Germany) was used as a VR simulation of complex emergencies for in-situ assessment of specific IC [32]. The application places users in a virtual emergency department, where they are required to diagnose and treat patients using a variety of medical procedures. *STEP-VR* is a joint development between a 3D visualization company and the University Hospital of Würzburg and had previously been used for the VR-OSCE station [12].

The hardware setup consisted of a Schenker XMG Core 15 Laptop (Schenker Technologies GmbH; chipset: Intel Core i7-9750H, 6×2.6 GHz; graphics adapter: Nvidia GeForce GTX 1650, 4 GB GDDR6 VRAM) and a Meta Quest III VR head-mounted display.

Study Design

This observational follow-up study was conducted at the skills lab of a German medical school with a 6-year medicine degree curriculum between July 2024 and May 2025. Participants were retrospectively recruited from a cohort of advanced clinical-phase medical students who had completed a curricular OSCE in July 2023, including a VR-based emergency medicine station (VR-OSCE station). Recruitment was through semester mailing lists. The study appointments took place outside regular teaching hours in the form of individual sessions. To examine the relationship between general and specific IC and students' medical performance, we performed individual post-hoc sessions with students that had previously completed the VR-OSCE station. However, at the time of the study, participants were in their final year of medical training and thus dispersed across various clinical rotations and institutions, making re-contact and scheduling particularly challenging. From the original cohort of 57 students, a total of 21 (37%) could be recruited for the follow-up.

Each session lasted approximately 30 minutes. After the study procedure had been explained and informed consent was given, participants received a short basic introduction to the operation of the VR headset and controllers. Following, general IC was assessed using the *VR competence app* (10 minutes). Specific IC was tested using the domain-specific application (6 minutes) which was already used during the original VR-OSCE

station. At last, participants completed a questionnaire assessing demographic data, technological affinity, and spatial ability (10 minutes). Figure 1 illustrates the study procedure and data collection.

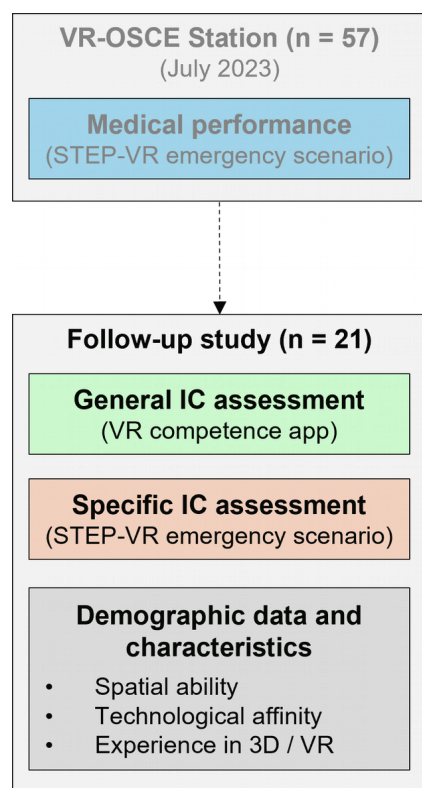


Figure 1: Study Procedure and Data Collection. The present study was conducted with participants who completed the VR-OSCE station in July 2023.

Specific IC Assessment

Specific IC was assessed using a checklist-based performance measure embedded within the VR emergency room scenario previously employed for summative OSCE assessment. This approach ensured high ecological validity and close alignment between IC measurement and the application-specific interaction demands encountered in high-stakes VR-based settings. To isolate immersive competence from medical knowledge, all tasks were deliberately designed to be independent of clinical reasoning. Participants received standardized verbal instructions (e.g., “Measure oxygen saturation”), which were decomposed into discrete, technically feasible interaction steps (e.g., “Pick up and attach SpO₂ clip” and “Read out SpO₂ value”). Each step was classified as object manipulation, spatial navigation, or UI activation, in accordance with established VR interaction typologies [16] (see Figure 2). Steps involving more than one category were assigned to all relevant categories for analytical purposes.

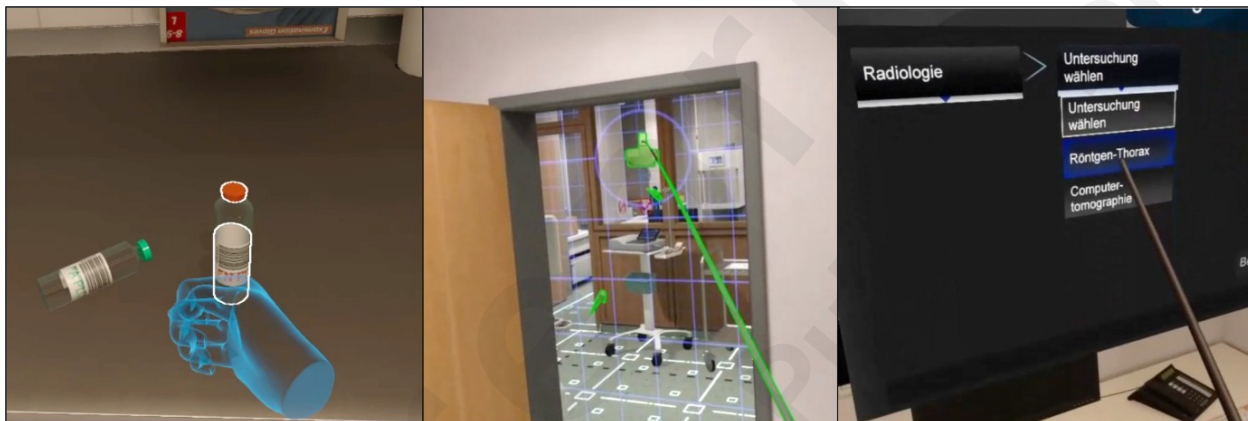


Figure 2: Exemplary interaction steps in the native control logic of STEP-VR. Right: Picking up blood culture bottles (object manipulation). Middle: Teleporting to the nurse station (navigation). Left: Selecting diagnostics and consults in the virtual computer menu (UI activation).

Initially, a pool of 20 interaction steps was defined based on expert walkthroughs of the VR scenario and a systematic mapping to the application's core control logic. Steps were selected to reflect a range of difficulty levels and to ensure coverage of all three interaction categories. This initial item pool (Multimedia Appendix 1) was reviewed by one VR developer (SO), one clinician with specialized qualification in medical education (SK) and one clinician experienced in simulation-based training (TM) to confirm technical feasibility and independence from medical knowledge. During pilot testing with five medical students, six steps were consistently completed by all participants and thus showed no variance. These were excluded from further analysis, as they lacked discrimination. One additional item (selecting 'Surgical Consult') showed negative item-total correlation ($r = -0.06$) in our data, likely due to a software usability issue in menu navigation that was identified retrospectively and was therefore removed. The final IC checklist included 13

dichotomously scored steps (0 = incorrect or incomplete; 1 = correctly completed within time limit), with a maximum raw score of 13 points (Table 2). Time limits for each step were calibrated based on internal pilot runs to reflect escalating difficulty, ensuring that early steps were accessible for users with low IC, while later steps challenged higher proficiency levels.

General IC Assessment

General IC was measured essentially as previously described [16] using the VR competence app. This app comprises eight standardized challenges reflecting common VR interactions from the categories of object manipulation, spatial navigation, and UI activation. For one minute, the user is asked to complete as many repetitions of the respective interaction as possible while task difficulty increased. Notably, this application is content-agnostic and independent of any specific educational or clinical scenario. Each task was scored automatically and summed into a total IC score.

Technological Affinity

Participants' attitudes toward and confidence with technology were assessed using two subscales of the validated German TA-EG (Affinity for Technology- Electronic Gadgets) questionnaire [29], namely "enthusiasm" (e.g., "I enjoy trying out new technical devices") and "competence" (e.g., "I am confident using complex devices without help"). Each scale comprised five items rated on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). Scores were averaged separately for each scale.

Spatial Ability

Spatial ability was assessed via a German version of a paper-based mental rotation task used in psychometric aptitude testing [33] following a similar principle as the original version [34]. Participants were shown 10 three-dimensional tube-shaped figures and asked to identify matching rotated versions from four options. Each correct response was awarded one point; thus, scores ranged from 0 to 10.

Medical Performance

Students' medical performance data were extracted from the original OSCE scores [12]. Two performance indicators were analyzed: The VR-OSCE station score (range: 0-100%) and the overall OSCE score (calculated as the mean of all 9 other OSCE stations, range 0-100%).

Statistical Analysis

Descriptive statistics including means (M) and standard deviations (SD) as well as

difficulties (P), were used to characterize the sample and items of specific IC assessment. Given the small sample size, which precluded common factor analysis, we followed suggestions enhance the validation of education instruments [35]. Pearson's r was calculated to examine associations between general/specific IC and medical performance outcomes. Item discrimination (r') representing item-total correlation for the dichotomous items of the specific IC assessment was calculated using point-biserial correlations. Point-biserial correlations were calculated in R (Version 4.3.1, RRID:SCR_001905; R Core Team), all other statistical analyses were conducted in GraphPad Prism (Version 10.5.0, RRID:SCR_002798; GraphPad Software Inc). A significance level of $\alpha = .05$ was used throughout.

Ethics approval

The local institutional review and ethics board (Ethikkommission der Universität Würzburg) judged the project as not to constitute medical or epidemiological research on human subjects and as such adopted a simplified assessment protocol. The project was approved without any reservation under the proposal number 20240208-01.

Informed Consent and Compensation Details

Students were informed about the study, and their participation was voluntary. Informed consent was obtained from all participants, who were also provided with information on data processing for the analysis and the publication of results. Contact details were supplied for participants wishing to withdraw their consent to data processing. The decision to participate or had no consequences on the students' academic progress. A 30-euro (~33.96\$) book voucher given to the participants at the end of the study appointment.

Privacy and Confidentiality

Data were stored in pseudonymized form on university servers with restricted access, and no personal identifiers were retained beyond the matching of OSCE ID codes. Survey data from the questionnaires were collected anonymously using the EvaSys® platform (Evasys GmbH, Lüneburg, Germany). Data were processed and stored in accordance with local data protection laws.

Results

Descriptive Statistics

The final sample included 21 medical students (age: $M = 26.0$ years, $SD = 2.7$), with 10 identifying as female and 11 as male. Both exposure to 3D applications and prior experience with VR applications were very limited (Table 1). Participants reported moderate levels of technology enthusiasm ($M = 2.38$, $SD = 0.87$) and above average self-rated technology competence ($M = 3.46$, $SD = 0.68$, both on a 5-point Likert scale) on the TA-EG. Spatial ability, measured via a standardized mental rotation test, showed a mean score of 5.48 out of 10 ($SD = 2.29$). General immersive competence, assessed via the VR-Competence App, yielded a mean performance score of 57.27 ($SD = 7.38$), while specific immersive competence, measured in the VR emergency room simulation, showed a mean of 7.33 out of 13 ($SD = 2.62$). Medical task performance among participants in the original VR OSCE station averaged 67.73% ($SD = 14.62$), and the overall OSCE score across all stations was 80.30% ($SD = 4.83$).

		n	%
Age, M ± SD		26.5 ± 2.6	
Gender	Female	10	48
	Male	11	52
	Diverse	0	0
Frequency of using desktop 3D-applications	Never	14	67
	Rarely (<1 / month)	4	19
	Sometimes (<1 / week)	3	14
	Frequently (>1 / week)	0	0
	Daily	0	0
Cumulative experience with VR applications	None	4	19
	0-1 hours	11	52
	1-10 hours	6	29
	10-50 hours	0	0
	>50 hours	0	0

Table 1: Participant characteristics including age, gender and experience with 3D and VR applications (total n = 21). VR: virtual reality

Performance Checklist and Psychometric Properties

Overall, item completion rates ranged from 0.05 to 0.90 ($M: 0.56$), indicating a broad range of item difficulty (P). While several basic object manipulation tasks (e.g., picking up the oxygen saturation clip or retrieving an infusion bottle) were completed by nearly all

participants, more complex UI-dependent steps - such as setting the exact infusion rate or accessing the digital consultation menu - showed substantially lower success rates. Discrimination ranged from 0.14 to 0.60 (mean: 0.37), with a subset of UI-related and navigation-heavy tasks displaying particularly good discrimination. In contrast, items with extreme difficulty (either too easy or too hard) showed limited ability to differentiate between participants. All performance data are displayed in Table 2.

The internal consistency across all 13 items, calculated using Cronbach's alpha, was $\alpha = 0.79$, suggesting good reliability for an early-stage checklist instrument. Removing low-discrimination items only slightly improved internal consistency (maximum $\alpha = 0.80$), supporting the checklist's robustness in its current form. For interaction categories Cronbach's alpha was $\alpha = 0.60$ ($n = 8$) for navigation, $\alpha = 0.64$ for object manipulation ($n = 8$) and $\alpha = 0.65$ for UI activation ($n = 3$). Solely for navigation removal of the item #1 "Attach SpO₂ clip to patient" would improve reliability markedly improved reliability to 0.64.

Table 2: Tasks, instructions and individual interaction steps with their corresponding time limits, assigned interaction categories (O = object manipulation, N = navigation, and U = UI activation), and the resulting difficulty (P) and discrimination values (r').

Task	Instruction	Time (s)	#	Interaction step	Interaction category	P	r'
Measuring SpO ₂	Take the SpO ₂ clip from the desk and attach it to the patient's right hand. Then read the SpO ₂ value from the vital signs monitor aloud.	10	1	Attach SpO ₂ clip to patient	O, N	0.71	0.14
			2	Read out SpO ₂ value	U	0.52	0.31
Administering saline	Take an IV catheter out of the box in the labelled drawer and place it on the patient's arm. Then pick up a 1 liter IV fluid bag from the drawer and hang it on the IV pole. Increase the rate of the infusion to 1000 ml per hour.	30	3	Pick up IV catheter	O, N	0.90	0.29
			4	Pick up IV fluid bag	O, N	0.90	0.22
			5	Set desired infusion rate	U	0.48	0.48
Taking blood cultures	Take a blood culture bottle from the "Microbiology" drawer and connect it to a cannula. Use this blood culture bottle to draw blood from one of the patient's arms. Then place it on the desk.	15	6	Pick up blood culture bottle	O, N	0.90	0.14
			7	Connect bottles to cannula	O	0.57	0.48
			8	Draw blood	O	0.43	0.54
			9	Place filled bottles on desk	N	0.05	0.23
Administering oral medication	Take a box of clopidogrel from the "oral medication" drawer. Drop the medication onto the patient's mouth.	10	10	Pick up oral medication	O, N	0.86	0.22
			11	Drop medication on patient's mouth	N	0.48	0.55

Administering IV medication	Take a box of aspirin from the "IV medication" drawer. Fill it into a 50ml bottle drawer and confirm the dose. Then hang the infusion on the IV pole.	15	12	Prepare IV push and confirm dose	U	0.33	0.59
			13	Hang infusion on IV pole	O, N	0.19	0.60
			Mean			0.56	0.37

Correlation Analyses

Specific IC showed a strong correlation with general IC ($r = 0.56$, $p = .008$), supporting their conceptual relatedness. In contrast, correlations between general and specific IC within corresponding interaction categories were low and non-significant for object manipulation ($r = 0.17$, $p = .50$) and navigation ($r = 0.17$, $p = .14$). The highest correspondence was observed for UI activation ($r = 0.41$, $p = .07$).

Moreover, specific IC correlated moderately to strongly with prior 3D experience ($r = 0.57$, $p = .007$) and self-rated technical competence ($r = 0.46$, $p = .04$). While both general and specific IC were moderately associated with performance in the VR-OSCE station at a borderline level of significance ($r = 0.41$, $p = .06$ and $r = 0.37$, $p = .09$, respectively), neither correlated meaningfully with overall OSCE performance ($r = 0.16$, $p = .48$ and $r = 0.02$, $p = .92$, respectively). Similarly, no relevant correlations were found between specific IC and technology enthusiasm or spatial ability.

When decomposing specific IC into interaction categories, object manipulation correlated most strongly with the total specific IC score ($r = 0.94$, $p < .001$). Object manipulation and - most notably - UI activation also showed strong associations with 3D experience ($r = 0.51$, $p = .02$ and $r = 0.67$, $p = .001$) and technical competence ($r = 0.54$, $p = .01$ and $r = 0.35$, $p = .12$). All correlations are listed in Figure 3 and significance values are listed in Multimedia Appendix 2.

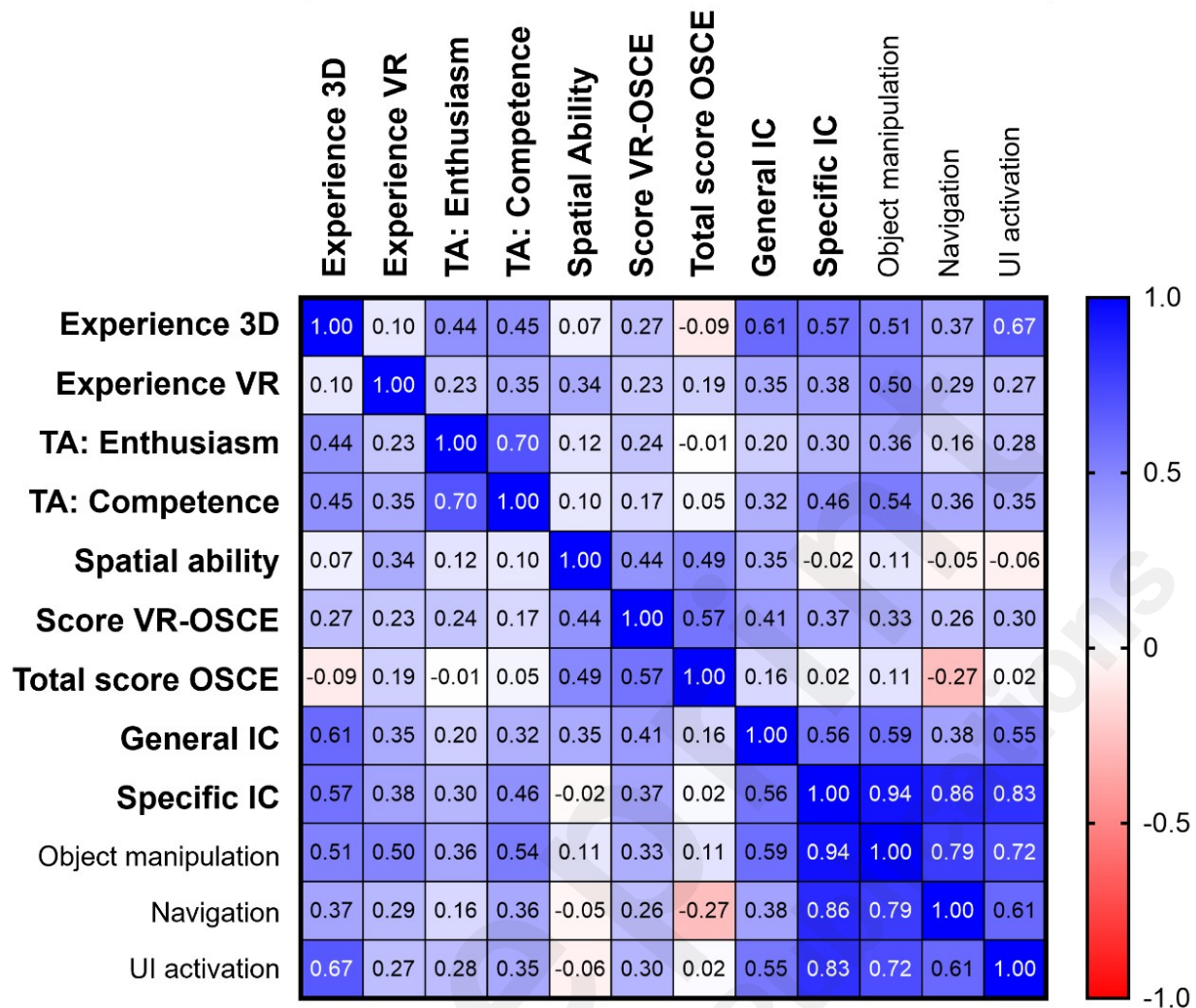


Figure 3: Correlation matrix between prior experience in 3D and VR applications, technical affinity, spatial ability, and students' performance scores of the VR-OSCE station and the total OSCE as well as general and specific IC (the latter subdivided into object manipulation, navigation and UI activation). The heatmap scale (right panel) indicates the color coding for positive (blue) and negative (red) correlations. The more intense the color, the higher the correlation coefficient. Corresponding significance values are listed in Multimedia Appendix 2. IC: immersive competence, OSCE: objective structured clinical examination, TA: technical affinity questionnaire, UI: user interface, VR: virtual reality

Discussion

This pilot study introduces and empirically examines the construct of specific IC as a context-sensitive dimension of user readiness, thereby laying the groundwork for further exploration of its role in predicting performance in VR-based medical assessments. Building on prior theoretical definitions of IC [14] and on established approaches for assessing broad, application-independent VR interaction skills such as object manipulation, spatial navigation, and UI activation [16, 19, 20] we argue that domain-specific VR systems may demand an additional, situationally anchored level of interaction proficiency. In response, we evaluated a simple, objective and reproducible approach for measuring this construct referred here as “specific IC”, directly within a domain-specific VR application.

Our first objective addressed the psychometric evaluation of our new method for measuring specific IC. The checklist used in this study demonstrated satisfactory psychometric properties for an experimental prototype: overall internal consistency was acceptable (Cronbach's $\alpha = .79$), meeting common thresholds for exploratory research, where α values above 0.70 are typically considered adequate for group-level comparisons [36]. However, internal consistency for subscales measured by Cronbach's α requires further exploration. Since the α coefficient solely is based solely on inter-item correlation, it is possible that lower values were caused by subgroups, which could not be investigated due to the small sample. Overall difficulty levels were balanced and largely within the desired range, which is typically between 0.4 and 0.8 [37]. Most items demonstrated good item-total correlations (> 0.3), indicating that they were able to differentiate well between participants with varying levels of specific IC. An item discrimination above 0.3 is usually considered indicative of a useful item, especially in early test development [37]. One exception was item #6, which displayed poor discrimination and an item difficulty of 0.90. Since our small sample size did not allow conclusions about the specific IC range of the general population, we retained this item to avoid potential floor effects in future studies with different populations. Future refinements may benefit from larger, more representative samples and from user interviews, for example, to identify subtle misunderstandings of task instructions.

Our second research objective was to examine its relationship with general IC and related human abilities to determine whether the two constructs reflect distinct aspects of VR readiness. We observed a strong and significant correlation between general and specific IC scores while their subscales did not correlate. This suggests that while both constructs may partially share underlying cognitive resources, they also reflect distinct aspects of user

competence, similar to findings from other domains [25, 26]. Both general and specific IC were significantly correlated with participants' prior exposure to 3D applications, which is consistent with previous findings on (general) IC that describe these abilities as (formally or informally) trainable [15–17]. The comparatively stronger effect of 3D experience over prior VR exposure in the present study is likely due to the limited variance in the latter, as most participants had minimal VR experience, thereby reducing its explanatory power.

While there is some conceptual overlap with constructs such as self-reported technological competence, correlations with these measures remained weak and non-significant - both in this and in previous studies [16]. The absence of a significant correlation with spatial ability in the present study is noteworthy, especially given that similar - although not identical - mental rotation tasks were used in the earlier study [16]. One possible explanation could be the different disciplinary background of the present sample and participants' potential familiarity with the mental rotation test, which for some students forms part of the medical school aptitude test, potentially leading to biased results.

At the level of interaction categories, object manipulation and UI activation showed strong associations with general and specific IC as well as prior 3D experience, while spatial navigation appeared less indicative. One plausible explanation is the simplicity and uniformity of the navigation tasks in this study, which relied on basic teleportation in a confined virtual space. In more complex simulations requiring advanced spatial orientation or locomotion under time pressure, navigation may pose high cognitive demands [38, 39] and could be more substantially associated with IC or related human abilities.

With regard to the third research objective, both general and specific IC showed moderate, borderline-significant correlations with students' medical performance as measured by the VR-OSCE station. Beyond the relatively small sample size, several methodological and contextual factors may have limited the observable correlation between IC and medical performance. First, the VR-OSCE station allowed for technical assistance from faculty members. Such instructor-mediated compensation strategies can mitigate usability barriers, particularly for users with lower digital competence, as noted in digital learning contexts [40]. However, this likely reduced the influence of IC on medical performance in our setting. Furthermore, the original medical checklist in the VR-OSCE station also included two verbal questions (about the correct diagnosis and the correct sequence of actions), which could be answered without interaction (and thus without engaging IC) and may therefore have weakened the observed correlations. Although the correlations did not reach conventional

significance thresholds, the consistent direction and convergence of effects across both types of IC - despite differing operationalizations - support the construct relevance of IC in performance-based VR assessment. This is in line with arguments emphasizing that near-significant effects, particularly when consistent across measures, may still reflect meaningful relationships [41]. The lack of correlation between both types of IC and overall OSCE medical performance suggests that IC is unlikely to be a relevant factor in traditional analog assessment settings. Apparently, IC as a technical, practical skill forms a psychometrically distinct construct from analytical or academic intelligence - a distinction supported by seminal intelligence theories [42]. In line with this, empirical studies have shown that practical intelligence correlates poorly with general cognitive ability, yet predicts job-relevant outcomes, even after controlling for general intelligence [43]. Taken together, our findings underscore the need for systematic assessment in larger cohorts to better understand the role of IC in VR-based assessment of medical performance: From a psychometric perspective, addressing potential construct-irrelevant variance is essential to uphold the validity and interpretability of performance assessments [18], particularly as immersive technologies are scaled for high-stakes examinations. From an equity standpoint, ignoring individual differences in IC may unintentionally privilege those with prior exposure to similar technologies - such as through gaming - thus reinforcing a *second-order digital divide* [44]. Therefore, integrating specific IC measures in VR-based examinations may help distinguish true knowledge deficits from challenges in VR interaction skills. This could inform both targeted training interventions and adjustments to assessment design to ensure fairness.

Several limitations of this study should be noted. First, the sample size was small, which limits the generalizability and statistical power of the findings. Second, although the checklist for specific IC showed promising internal consistency and item-level characteristics, external validation by independent raters is still pending. Since the measure was assessed using a single domain-specific application (STEP-VR), the generalizability of a checklist-based approach to other VR systems or medical domains also remains to be established. Third, retrospective recruitment may have introduced selection bias, as participants who volunteered for follow-up testing may have had above-average motivation or technical interest. Fourth, participants had previously been exposed to the VR-OSCE station and preparatory training. While we assumed that exposure effects had subsided after six months, residual familiarity with the application might still have facilitated specific

IC performance, potentially inflating scores. Similarly, frequent interim use of VR could have led to changes in IC since the original performance assessment. However, this appears unlikely considering the low levels of VR exposure reported by participants. Finally, the correlational design precludes causal conclusions, which future studies could address by manipulating IC (e.g., through targeted training) and evaluating its impact on medical performance.

Conclusion

Despite its limitations, this pilot study provides initial evidence that general and specific IC are distinct, meaningful, and psychometrically accessible constructs in the context of VR-based medical assessments. The proposed in-situ approach - via guided, domain-relevant interaction tasks - demonstrated adequate item characteristics for early-stage evaluation and offers a scalable, pragmatic method to gauge user readiness. While correlations with medical performance in the current sample were limited, this may reflect ceiling effects, instructor compensation, or the independence of certain verbally addressed OSCE items from IC. Nevertheless, the increasing adoption of immersive technologies in high-stakes training and assessment highlights the practical relevance of IC as a determinant of user performance.

Future research should refine the item design, validate IC measures in larger and more diverse populations and conduct interventional studies to clarify causal pathways. Such work could inform strategies to mitigate performance inequities related to operational readiness and ultimately enhance the fairness and validity of VR-based assessments in medical education.

Acknowledgments

None

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request. The VR competence app used in this study was designed by the Chair of Human-Computer-Interaction at the University of Würzburg. Requests regarding its use for research purposes should be directed to the corresponding author. The software STEP-VR (Simulation-based Training of Medical Emergencies for Physicians Using Virtual Reality) used in this study is licensed by ThreeDee GmbH. Requests regarding its use for research purposes should be directed to the company.

Authors' Contributions

TM conceived the study, acquired funding, supervised the project, performed data analysis, and wrote the original draft. VS contributed to the study design and conducted the investigation. ML supported participant recruitment and assisted with data collection. JB performed the statistical analyses. SO and MH contributed to the study conceptualization and contributed to the review and editing of the manuscript. SK also contributed to the study design and participated in writing, reviewing, and editing the manuscript.

Conflicts of Interest

TM was involved in the software development of STEP-VR (Simulation-based Training of Medical Emergencies for Physicians Using Virtual Reality). All other authors declare no conflicts of interest.

Multimedia Appendix 1: Original item set for specific IC.

Multimedia Appendix 2: Significance values for correlation data

References

1. Mergen M, Graf N, Meyerheim M. Reviewing the current state of virtual reality integration in medical education - a scoping review. *BMC Med Educ.* 2024;24:788. doi:10.1186/s12909-024-05777-5.
2. Alaker M, Wynn GR, Arulampalam T. Virtual reality training in laparoscopic surgery: A systematic review & meta-analysis. *Int J Surg.* 2016;29:85–94. doi:10.1016/j.ijsu.2016.03.034.
3. Jans C, Bogossian F, Andersen P, Levett-Jones T. Examining the impact of virtual reality on clinical decision making - An integrative review. *Nurse Educ Today.* 2023;125:105767. doi:10.1016/j.nedt.2023.105767.
4. Wespi R, Schwendimann L, Neher A, Birrenbach T, Schaubert SK, Manser T, et al. TEAMS go VR-validating the TEAM in a virtual reality (VR) medical team training. *Adv Simul (Lond).* 2024;9:38. doi:10.1186/s41077-024-00309-z.
5. Lindner M, Leutritz T, Backhaus J, König S, Mühling T. Knowledge Gain and the Impact of Stress in a Fully Immersive Virtual Reality-Based Medical Emergencies Training With Automated Feedback: Randomized Controlled Trial. *J Med Internet Res.* 2025;27:e67412. doi:10.2196/67412.
6. Junga A, Kockwelp P, Valkov D, Schulze H, Bozdere P, Hätscher O, et al. Teach the Unteachable with a Virtual Reality (VR) Brain Death Scenario - 800 Students and 3 Years of Experience. *Perspect Med Educ.* 2025;14:44–54. doi:10.5334/pme.1427.
7. Neher AN, Bühlmann F, Müller M, Berendonk C, Sauter TC, Birrenbach T. Virtual reality for assessment in undergraduate nursing and medical education - a systematic review. *BMC Med Educ.* 2025;25:292. doi:10.1186/s12909-025-06867-8.
8. Keicher F, Backhaus J, König S, Mühling T. Virtual reality for assessing emergency medical competencies in junior doctors - a pilot study. *Int J Emerg Med.* 2024;17:125. doi:10.1186/s12245-024-00721-2.
9. Kman NE, Panchal AR, Patterson J, Danforth D, Mani A, Babbitt D, et al. Virtual Reality Simulation for Assessment of Hemorrhage Control and SALT Triage Accuracy in First Responders. *Disaster med. public health prep.* 2024. doi:10.1017/dmp.2024.233.
10. Knudsen MH, Breindahl N, Dalsgaard T-S, Isbye D, Mølbak AG, Tiwald G, et al. Using Virtual Reality Head-Mounted Displays to Assess Skills in Emergency Medicine: Validity Study. *J Med Internet Res.* 2023;25:e45210. doi:10.2196/45210.
11. Rodríguez-Matesanz M, Guzmán-García C, Oropesa I, Rubio-Bolivar J, Quintana-Díaz M, Sánchez-González P. A New Immersive Virtual Reality Station for Cardiopulmonary Resuscitation Objective Structured Clinical Exam

- Evaluation. *Sensors* (Basel) 2022. doi:10.3390/s22134913.
12. Mühling T, Schreiner V, Appel M, Leutritz T, König S. Comparing Virtual Reality-Based and Traditional Physical Objective Structured Clinical Examination (OSCE) Stations for Clinical Competency Assessments: Randomized Controlled Trial. *J Med Internet Res*. 2025;27:e55066. doi:10.2196/55066.
 13. LaViola JJ, Kruijff E, McMahan RP, Bowman D, Poupyrev IP. *3D User Interfaces: Theory and Practice*. 2nd ed. Boston: Addison-Wesley; 2017.
 14. Steed A, Archer D, Izzouzi L, Numan N, Shapiro K, Swapp D, et al. Immersive competence and immersive literacy: Exploring how users learn about immersive experiences. *Front. Virtual Real*. 2023. doi:10.3389/frvir.2023.1129242.
 15. Junga A, Schulze H, Scherzer S, Hätscher O, Bozdere P, Schmidle P, et al. Immersive learning in medical education: analyzing behavioral insights to shape the future of VR-based courses. *BMC Med Educ*. 2024;24:1413. doi:10.1186/s12909-024-06337-7.
 16. Oberdörfer S, Heinisch M, Mühling T, Schreiner V, König S, Latoschik ME. Ready for VR? Assessing VR Competence and Exploring the Role of Human Abilities and Characteristics 2025. doi:10.3389/frvir.2025.1608593.
 17. Marsh WE, Hantel T, Zetsche C, Schill K. Is the user trained? Assessing performance and cognitive resource demands in the Virtusphere. 2013:15–22. doi:10.1109/3DUI.2013.6550191.
 18. MESSICK S. THE PSYCHOLOGY OF EDUCATIONAL MEASUREMENT. *J Educational Measurement*. 1984;21:215–37. doi:10.1111/j.1745-3984.1984.tb01030.x.
 19. Griffiths G, Sharples S, Wilson JR. Performance of new participants in virtual environments: The Nottingham tool for assessment of interaction in virtual environments (NAÏVE). *International Journal of Human-Computer Studies*. 2006;64:240–50. doi:10.1016/j.ijhcs.2005.08.008.
 20. Lampton DR, Knerr BW, Goldberg SL, Bliss JP, Moshell JM, Blau BS. The Virtual Environment Performance Assessment Battery (VEPAB): Development and Evaluation. *Presence: Teleoperators & Virtual Environments*. 1994;3:145–57. doi:10.1162/pres.1994.3.2.145.
 21. Tusher HM, Mallam S, Nazir S. A Systematic Review of Virtual Reality Features for Skill Training. *Tech Know Learn*. 2024;29:843–78. doi:10.1007/s10758-023-09713-2.
 22. Lie SS, Helle N, Sletteland NV, Vikman MD, Bonsaksen T. Implementation of Virtual Reality in Health Professions Education: Scoping Review. *JMIR Med Educ*. 2023;9:e41589. doi:10.2196/41589.
 23. Mühling T, Backhaus J, Demmler L, König S. How Personality and Affective Responses Are Associated with Skepticism Towards Virtual Reality in Medical Training-A Pre-Post Intervention Study. *Cyberpsychol Behav Soc Netw*. 2025;28:335–41. doi:10.1089/cyber.2024.0567.

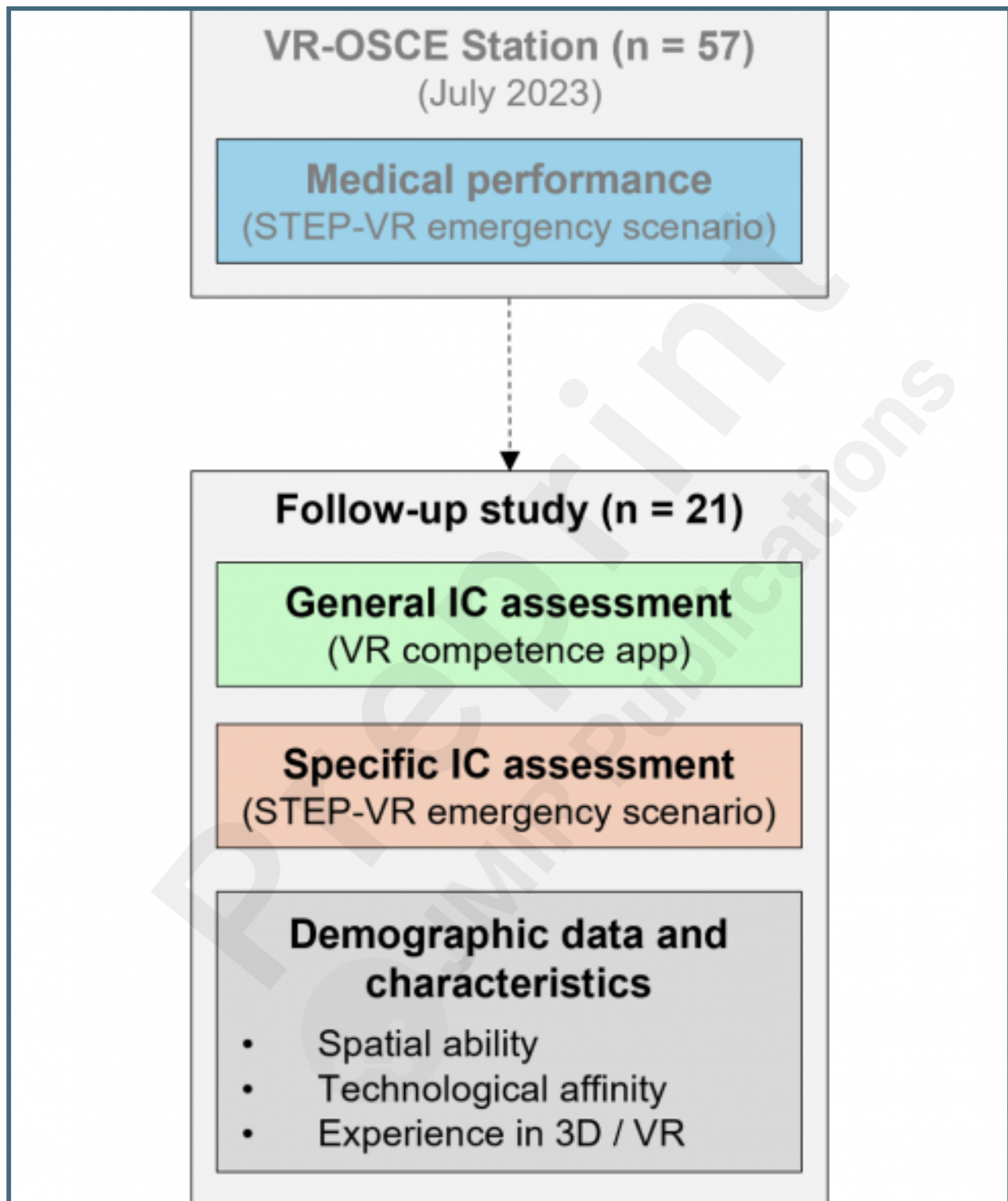
24. Chen M-X, Hu H, Yao R, Qiu L, Li D. A Survey on the Design of Virtual Reality Interaction Interfaces. *Sensors (Basel)* 2024. doi:10.3390/s24196204.
25. van Deursen A, van Dijk J. Internet skills and the digital divide. *New Media & Society*. 2011;13:893–911. doi:10.1177/1461444810386774.
26. Wulf G, Shea CH. Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychon Bull Rev*. 2002;9:185–211. doi:10.3758/BF03196276.
27. Hornbæk K. Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*. 2006;64:79–102. doi:10.1016/j.ijhcs.2005.06.002.
28. Yeşilyurt E, Vezne R. Digital literacy, technological literacy, and internet literacy as predictors of attitude toward applying computer-supported education. *Educ Inf Technol (Dordr)*. 2023;1–27. doi:10.1007/s10639-022-11311-1.
29. Karrer-Gauß K, Roesler E, Siebert FW. Neuauflage des TAEG Fragebogens: Technikaffinität valide und multidimensional mit einer Kurz- oder Langversion erfassen. *Z. Arb. Wiss.* 2024;78:387–406. doi:10.1007/s41449-024-00427-4.
30. Hamstra SJ, Brydges R, Hatala R, Zendejas B, Cook DA. Reconsidering fidelity in simulation-based training. *Acad Med*. 2014;89:387–92. doi:10.1097/ACM.0000000000000130.
31. Motola I, Devine LA, Chung HS, Sullivan JE, Issenberg SB. Simulation in healthcare education: a best evidence practical guide. *AMEE Guide No. 82. Med Teach*. 2013;35:e1511-30. doi:10.3109/0142159X.2013.818632.
32. Mühling T, Späth I, Backhaus J, Milke N, Oberdörfer S, Meining A, et al. Virtual reality in medical emergencies training: benefits, perceived stress, and learning success. *Multimedia Systems*. 2023;29:2239–52. doi:10.1007/s00530-023-01102-0.
33. Zentrum für Testentwicklung und Diagnostik. Test für Medizinische Studiengänge und Eignungstest für das Medizinstudium III. 1st ed. Göttingen: Hogrefe; 2019.
34. Vandenberg SG, Kuse AR. Mental rotations, a group test of three-dimensional spatial visualization. *Percept Mot Skills*. 1978;47:599–604. doi:10.2466/pms.1978.47.2.599.
35. Schönrock-Adema J, Heijne-Penninga M, van Hell EA, Cohen-Schotanus J. Necessary steps in factor analysis: enhancing validation studies of educational instruments. The PHEEM applied to clerks as an example. *Med Teach*. 2009;31:e226-32. doi:10.1080/01421590802516756.
36. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ*. 2011;2:53–5. doi:10.5116/ijme.4dfb.8dfd.
37. Möltner A, Schellberg D, Jünger J. Grundlegende quantitative Analysen medizinischer Prüfungen - Basic quantitative analyses of medical examinations. *GMS Z Med Ausbild* 2006. doi:shtml.
38. Kalantari S, Xu BT, Mostafavi A, Lee AS, Yang Q. Comparing Spatial Navigation

- and Human Environment Interaction in Virtual Reality vs. Identical Real Environments across the Adult Lifespan; 30.08.2023.
39. Liang H-N, Lu F, Shi Y, Nanjappan V, Papangelis K. Evaluating the effects of collaboration and competition in navigation tasks and spatial knowledge acquisition within virtual reality environments. *Future Generation Computer Systems*. 2019;95:855–66. doi:10.1016/j.future.2018.02.029.
 40. Bond M, Marín VI, Dolch C, Bedenlier S, Zawacki-Richter O. Digital transformation in German higher education: student and teacher perceptions and usage of digital media. *Int J Educ Technol High Educ* 2018. doi:10.1186/s41239-018-0130-1.
 41. Gelman A, Stern H. The Difference Between “Significant” and “Not Significant” is not Itself Statistically Significant. *The American Statistician*. 2006;60:328–31. doi:10.1198/000313006X152649.
 42. Sternberg RJ. A triarchic view of intelligence in cross-cultural perspective. In: Irvine SH, Berry JW, editors. *Human Abilities in Cultural Context*. Cambridge: Cambridge University Press; 1988. p. 60–85. doi:10.1017/CBO9780511574603.003.
 43. Case R, Demetriou A, Platsidou M, Kazi S. Integrating concepts and tests of intelligence from the differential and developmental traditions. *Intelligence*. 2001;29:307–36. doi:10.1016/S0160-2896(00)00057-X.
 44. Riggins F, Dewan S. The Digital Divide: Current and Future Research Directions. *JAIS*. 2005;6:298–337. doi:10.17705/1jais.00074.

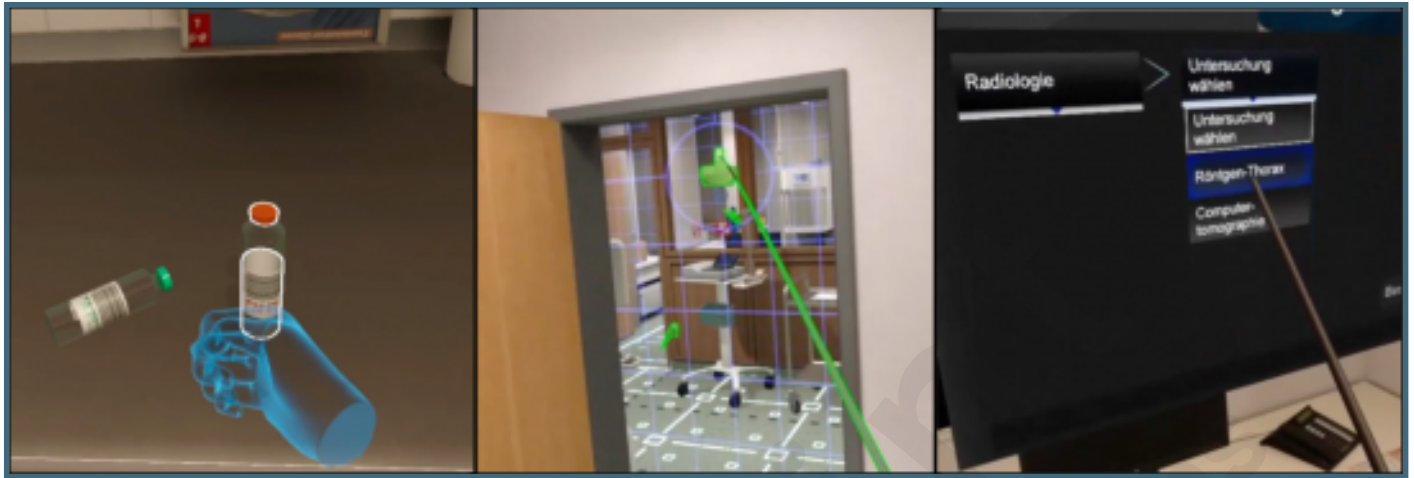
Supplementary Files

Figures

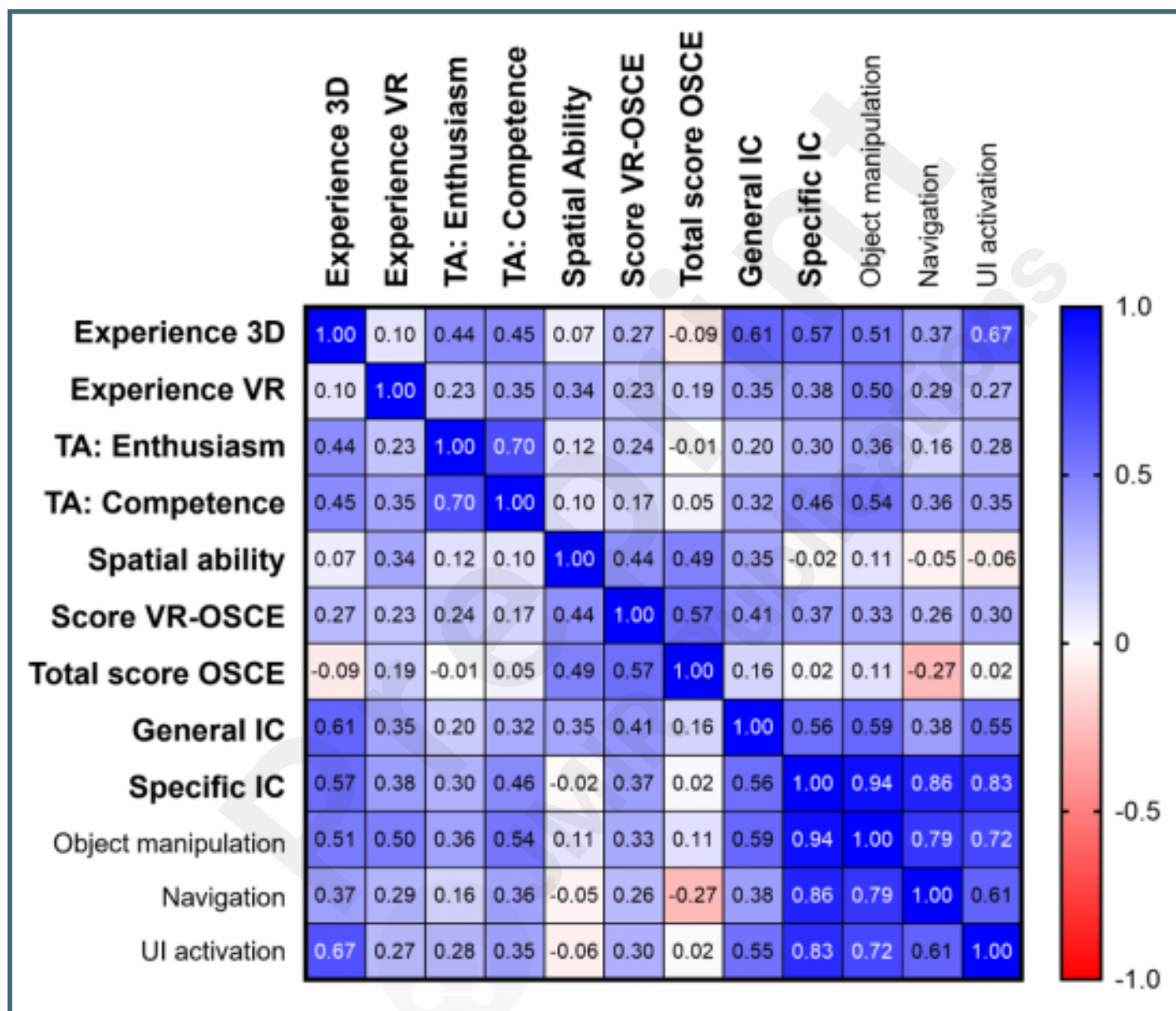
Study Procedure and Data Collection. The present study was conducted with participants who completed the VR-OSCE station in July 2023.



Exemplary interaction steps in the native control logic of STEP-VR. Left: Picking up blood culture bottles (object manipulation). Middle: Teleporting to the nurse station (navigation). Right: Selecting diagnostics and consults in the virtual computer menu (UI activation).



Correlation matrix between prior experience in 3D and VR applications, technical affinity, spatial ability, and students' performance scores of the VR-OSCE station and the total OSCE as well as general and specific IC (the latter subdivided into object manipulation, navigation and UI activation). The heatmap scale (right panel) indicates the color coding for positive (blue) and negative (red) correlations. The more intense the color, the higher the correlation coefficient. Corresponding significance values are listed in Multimedia Appendix 2. IC: immersive competence, OSCE: objective structured clinical examination, TA: technical affinity questionnaire, UI: user interface, VR: virtual reality.



Multimedia Appendixes

Original item set for specific IC.

URL: <http://asset.jmir.pub/assets/f0cf8228dcfdc92e2591c724007da157.xlsx>

Significance values for correlation data.

URL: <http://asset.jmir.pub/assets/1d6720098d47faeae67221464370bf2.xlsx>

