



**HAL**  
open science

## **Impact of Immediate Feedback on Students' Learning in a Virtual Reality Environment for Chemical Education**

Valentin Midez, Audrey Serna, Florent Diet, Victor Oliva Torres, Élise Lavoué

### ► **To cite this version:**

Valentin Midez, Audrey Serna, Florent Diet, Victor Oliva Torres, Élise Lavoué. Impact of Immediate Feedback on Students' Learning in a Virtual Reality Environment for Chemical Education. European Conference on Technology Enhanced Learning, Sep 2025, Durham, United Kingdom. pp.337-351, <10.1007/978-3-032-03870-8\_23>. <hal-05273385>

**HAL Id: hal-05273385**

**<https://hal.science/hal-05273385v1>**

Submitted on 22 Sep 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# Impact of Immediate Feedback on Students' Learning in a Virtual Reality Environment for Chemical Education.

Valentin Midez<sup>1</sup>, Audrey Serna<sup>1</sup>[0000-0003-1468-9761], Florent Diet<sup>1</sup>, Victor Oliva Torres<sup>1</sup>, and Élise Lavoué<sup>1,2</sup>[0000-0002-2659-6231]

<sup>1</sup> INSA Lyon, CNRS, Université Claude Bernard Lyon 1, LIRIS, UMR5205, 69621 Villeurbanne, France

`prenom.nom@insa-lyon.fr`

<sup>2</sup> Université Jean Moulin Lyon 3, iaelyon School of Management

`elise.lavoue@univ-lyon3.fr`

**Abstract.** Virtual Reality (VR) offers the opportunity to create immersive environments for safe, practice-based learning, but its intense immersion can limit critical thinking. This study explores the impact of immediate feedback on chemistry learning in an immersive environment. It involved 46 participants divided into two conditions, with and without immediate feedback. We assessed their performance, including correct answers and procedures performed, as well as perceived cognitive load. The results show that immediate feedback does not increase the perceived cognitive load, but its effect on learning depends on the tasks performed. They can be detrimental to performance for a single task, whereas they lead to significant improvements for repeated tasks. These observations highlight their usefulness in repetitive scenarios while underlining the need for appropriate design to avoid adverse effects on learning.

**Keywords:** Virtual reality, Feedback, Reflection, Immersive learning, Chemical education.

## 1 Introduction

Virtual reality (VR) can be used to simulate complex, interactive environments, providing immersive, engaging experiences for learning and training activities [2]. Several studies have shown that VR is particularly effective in supporting active learning and practical application [4]. It provides a safe environment where learners can experiment and develop practical skills without risk while enhancing their engagement. Immersive simulations offer embodied, realistic, and repeatable experiences that help learners explore strategies for managing complex situations and develop their skills [12].

However, some studies have shown that such immersive environments can hinder learners' ability to reflect critically on their actions [11]. To improve the effectiveness of VR environments in educational contexts, Roussou and Slater

[16] emphasize the importance of supporting learners' reflection during task execution. To address this issue, several studies have explored the integration of immediate feedback, reporting positive outcomes on learners' engagement, critical reflection, and performance [14,13,1]. Others, however, point to more mixed effects, particularly regarding learners' perceived cognitive load depending on the complexity of the feedback [21,15]. The design and pedagogical role of such feedback thus remain key challenges to explore.

In this study, we focus on chemistry education, where part of the learning objectives involves performing laboratory-based tasks and procedures, which can be trained in immersive environments. In this context, we investigate whether integrating immediate feedback in a VR learning environment has an effect on the learning experience. We conducted a comparative study with 46 participants, randomly assigned to two conditions: with or without immediate feedback. We analyzed the impact of immediate feedback on both performance (regarding expected answers and completed procedures) and perceived cognitive load in different types of tasks (single vs. repeated). The results show that immediate feedback does not increase perceived cognitive load, but its effect on performance depends on the nature of the task. These findings suggest that integrating immediate feedback into VR environments can be beneficial if carefully designed according to the learning scenario.

## 2 Related Work

### 2.1 Virtual Reality in Education

In education, virtual reality technologies offer immersive and interactive environments that allow learners to practice without physical or ethical risks [6,1]. These environments can enhance learners' attention by fostering a sense of presence, defined as the subjective experience of being in the virtual environment [16]. VR can improve retention of both knowledge and procedural skills, especially when traditional methods such as physical demonstrations or classroom simulations are difficult to implement due to material, time, or safety constraints [1]. It allows learners to experiment and correct their mistakes without real-world consequences freely [8]. In their meta-analysis, Cromley et al. [5] highlight that immersive environments in VR promote knowledge acquisition and learners' engagement, which can limit their ability to reflect on their actions. Jiang and Ahmadpour [11] point out that the intense immersion and flow state experienced in VR reduce self-awareness, thus hindering reflexivity about the actions performed.

According to Elaish et al. [7], the successful integration of VR in education requires an instructional design that aligns immersion with learning objectives, ensuring that virtual environments provide real educational value. Roussou and Slater [16] further recommend combining the interactive features of VR with support for reflective processes. They suggest that immersive and interactive environments, combined with explicit feedback to inform, confirm, or correct

learners' actions, can encourage critical reflection and prevent potential errors while reducing the risk of cognitive overload. These theoretical insights underscore the importance of examining how reflective processes can be effectively supported in VR-based learning environments.

## 2.2 Reflective Processes and Feedback in Virtual Reality

Reflective processes involve reflecting on how the task was carried out, in order to reevaluate the initial objectives and the strategies used to achieve them [24]. Reflecting on their own learning, learners become aware of their learning processes and identify possible alternative strategies [18]. This introspective work enables them to adjust existing practices, develop new ones, or redefine their meaning [23]. Schön [17] expands this notion by distinguishing between reflection-in-action, which supports immediate adjustments, and reflection-on-action, which allows for retrospective and critical analysis. Reflection-in-action can lead to constructing new understandings and real-time adaptation during task execution.

One way to support reflective processes is to provide feedback during the task (reflection-in-action) or after it (reflection-on-action). Hattie and Timperley [10] identify three levels of feedback: task-level feedback (specific information on performance), process-level feedback (guidance on strategies used), and self-regulation feedback (supporting learners' ability to self-assess). This study focuses on feedback related to task execution in immersive VR environments. Recent systematic reviews have emphasized the importance of feedback in immersive learning environments. Lee et al. [14] highlight the role of feedback in serious games using virtual, augmented, or mixed reality to foster reflection, increase engagement, and support the development of complex cognitive skills. Similarly, Lampropoulos et al. [13] show that combining VR with gamification elements such as levels, badges, and scores helps learners engage in critical thinking by providing immediate feedback on their actions.

Several comparative studies have also investigated the impact of immediate feedback in VR-based training. Almousa et al. [1] found that visual and auditory cues helped learners evaluate and adjust their performance in real-time while practicing cardiopulmonary resuscitation. In a study on breastfeeding training, Tang et al. [21] showed that adding a progress bar, badges, scores, and auditory prompts maintained engagement and supported learners' reflection on their performance. Van Ginkel et al. [22] used VR to support the development of public speaking skills, showing that real-time indicators (e.g., prolonged eye contact or fast speech) enabled immediate correction and improved presentation skills.

However, despite these benefits, concerns remain about the cognitive load induced by immediate feedback in VR. Cognitive load refers to the balance between task demands and the cognitive resources available to the learner [20]. Ma et al. [15] found that providing complex or multiple feedback cues can overload cognitive resources and reduce learning effectiveness. Furthermore, Tang et al. [21] reported a significant increase in time pressure and mental effort caused by immediate feedback.

In conclusion, while VR has shown its potential to support learning, the specific role of immediate feedback remains underexplored. Although promising results have been reported, the issue of cognitive load induced by real-time feedback still needs to be addressed to ensure optimal learning conditions.

### 3 Research Questions

This study examines whether adding immediate feedback in a virtual reality learning environment influences learning outcomes. We focus specifically on the acquisition of chemistry-related skills involved in laboratory manipulations. We address the following two research questions:

- **RQ1:** Does integrating immediate feedback into the VR environment affect learners’ performance in learning activities? In terms of performance, we analyze the relevance of the answers provided and the procedures performed by the learners about the target skills defined in the learning scenario.
- **RQ2:** Does integrating immediate feedback into the VR environment impact learners’ perceived cognitive load?

We explore these questions on different types of tasks (single vs. repeated tasks) using a dedicated scenario. In the following sections, we describe the VR environment, the learning scenario and the integration of feedback designed to support learners’ reflexivity.

## 4 Virtual Reality Environment for Learning Chemistry

As part of the Renforce project, we developed a virtual reality (VR) learning environment for chemistry education. Some of the targeted skills relate to the practice of scientific reasoning: proposing hypotheses to answer a scientific question, designing experiments to test them, measuring physical quantities directly or indirectly, interpreting experimental results, drawing conclusions, and communicating them with arguments. Other targeted skills focus on the adoption of ethical and responsible behavior, particularly in relation to safety rules and procedures related to chemical handling. VR environments allow training these skills, especially gestures and practical manipulations (e.g., pouring a liquid from one beaker to another).

The VR environment, developed using Unity, is a virtual laboratory that reproduces all the necessary elements for chemical manipulations (workbench, fume hood, taps, glassware, flasks, and measurement equipment). It also integrates a liquid management system to prepare mixtures (Figure 1).

### 4.1 Learning Scenario

To address our research questions, we designed a learning scenario covering basic chemistry skills: observing, experimenting, and measuring. The scenario focuses



Fig. 1: Virtual chemistry laboratory environment.

on the organization and transformation of matter, specifically on identifying whether a solution is acidic or basic using pH measurement. This topic, introduced in lower secondary school in the French curriculum, offers a simple context that requires no prior knowledge and is accessible to various profiles of learners.

A tutorial video presented to participants introduces the key knowledge and skills covered in the scenario, including identifying the acidic or basic nature of a solution based on the presence of  $H^+$  and  $OH^-$  ions; understanding chemical transformations in different contexts (e.g., daily life, environment); using several methods to measure pH (pH strips, pH meter, pH indicators); identifying reactions between acidic and basic solutions; and applying appropriate procedures during chemical manipulations.

The scenario consists of two manipulations, each designed to assess the activation of chemistry-related knowledge and skills across two dimensions: the relevance of the learners' answers and the proper execution of the expected procedures. To examine the effects of immediate feedback on various types of tasks, each manipulation included either single tasks (manipulation 1) or a sequence of repeated tasks (task 1 followed by task 2 in manipulation 2).

- **Manipulation 1:** Learners have to identify whether five solutions (distilled water, coffee, lemon juice, seawater, and bleach) are acidic, basic, or neutral. They use pH strips and a reference chart to measure and classify each solution. This task involves pipetting the liquid, applying it to a strip, observing the color change, and classifying the solution accordingly. **Targeted skills:** measuring pH using pH strips; classifying solutions as acidic, basic, or neutral. **Expected answers:** 2 acidic (lemon juice, coffee), 1 neutral (distilled water), 2 basic (seawater, bleach). **Required procedures:** use one strip per solution; discard each strip after use; rinse the pipette between measurements.
- **Manipulation 2:** Learners have to repeat the same acid-base reaction task twice. These two tasks aim to reach a specific pH using a combination of acidic and basic solutions and a colored indicator to assess the pH reached. In Task 1, they transform an acidic solution (white vinegar) into a basic solution with a pH of 8. In Task 2, they transform a basic solution (ammonia) into

an acidic solution with a pH of 6. Learners have to select the appropriate titration solution (a base in Task 1, an acid in Task 2) and choose the correct pH indicator based on its color change range. They observe the color during titration and can use pH strips at any time to verify the pH reached. **Targeted skills:** performing a transformation by selecting a titrant and a pH indicator based on its transition range; reaching a target pH. **Expected answers:** pH indicator (Task 1: phenol red; Task 2: bromothymol blue); titrant (Task 1: sodium hydroxide; Task 2: hydrochloric acid); target pH (Task 1: 8; Task 2: 6). **Required procedures:** respect the correct mixing order (indicator first); monitor the color transition; rinse the pipette before use; use a pH strip to confirm the result; accurately reach the pH value; discard the strip after use.

## 4.2 Integration of Feedback into the Environment

The feedback system aims to inform learners about their acquisition of the knowledge and skills targeted in the learning scenario. Feedback relies on the assessed answers and the procedures specified for each experimental task.

Immediate feedback combines visual and auditory cues to provide real-time input as learners act, to support reflection-in-action [17] (Fig. 2). For the design of these elements, we draw inspiration from recent studies that propose immediate, game-inspired audiovisual feedback mechanisms [21,13]. For correct answers, feedback takes the form of colored stars that appear accompanied by a sound symbolizing the twinkling of the star. In the case of an incorrect answer, learners also receive immediate feedback in the form of greyed-out stars. This visual cue indicates failure without corrective guidance to avoid disrupting procedural learning dynamics. For procedures, the feedback is represented by a progress bar (or XP bar). Each time a correct procedure is detected, the bar increases and is accompanied by a sound indicating progress.

To support reflection-on-action [17], a summary of correctly (or incorrectly) performed answers and procedures is displayed at the end of each manipulation. This summary enables learners to review their performance after each task and engage in retrospective reflection. The summary is available to all participants, regardless of their experimental condition.

## 5 Research Protocol

### 5.1 Participants

Participants were recruited via a university mailing list. Individuals with prior training or education in chemistry were excluded from the study to ensure homogeneous baseline knowledge across groups. 46 participants, aged between 19 and 36 ( $M = 23.4$ ,  $SD = 3.92$ ), participated in the experiment, including 17 women and 29 men. Participants were randomly assigned to one of two experimental groups of 23 participants each: one group with immediate feedback (With IF) and one group without immediate feedback (Without IF).

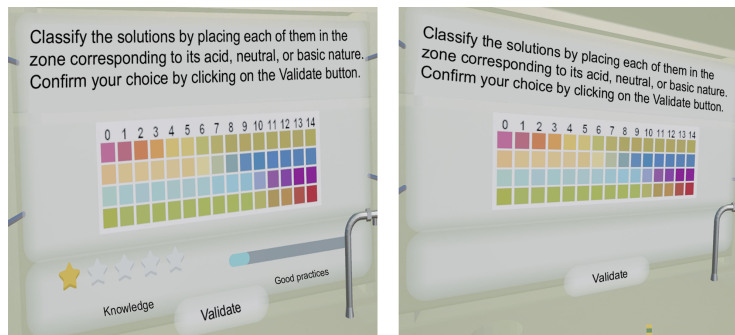


Fig. 2: Instruction panel displayed for each experimental condition for the group with IF on the left (stars and XP bar), and without IF on the right.

We used non-parametric tests to compare the groups due to small sample sizes (Mann–Whitney for quantitative variables and Chi-squared tests for categorical variables). As reported in Table 1, the two groups were well balanced regarding demographic characteristics, with no significant differences in age or gender. We found no significant differences in the number of years of study or participants’ prior experience with VR.

Each participant received a €10 gift card as compensation for their time and participation. The research protocol was approved by the Research Ethics Committee from the University of Lyon.

## 5.2 Experimental Procedure

We set up a dedicated 6×6 meter room to allow participants to move freely and interact with the VR environment designed for the study. We followed a between-subjects study design, therefore participants were divided between the two experimental conditions and followed the same procedure in the same order. The only difference between the groups was the presence or absence of immediate feedback during the learning tasks. Before starting, all participants signed an informed consent form in accordance with ethical guidelines. They then completed a demographic questionnaire, followed by questions about their previous VR experience and virtual reality gaming habits. Participants were guided through a VR tutorial to familiarize them with the headset (Vive Pro Eye) and the Valve Index controllers. The tutorial included instructions on manipulating virtual objects and navigating the environment. The experimental group also received instructions to understand the functioning of immediate feedback. After the tutorial, a video covering basic acid–base chemistry concepts was shown outside of VR to provide the necessary knowledge while minimizing VR exposure time. Participants were then equipped with the headset to complete the learning scenario consisting in manipulation 1 followed by manipulation 2 in the virtual laboratory. At the end of the session, participants completed the NASA-TLX questionnaire to assess their perceived cognitive load during the manipulations.

**Table 1:** Distribution of participants by gender, age, years of study after high school, VR experience frequency, and VR gaming habits between the immediate feedback and no immediate feedback groups. P-values from Mann–Whitney and Chi-squared tests.

Variables	Categories	With IF (n = 23)	Without IF (n = 23)	p-value
<b>Gender</b>	Male	14 (61%)	15 (65%)	1.00
	Female	9 (39%)	8 (35%)	
<b>Age (M ± SD)</b>		23.83 ± 4.93	22.91 ± 2.57	0.80
<b>Years of study after high school (M ± SD)</b>		4.65 ± 1.67	5.35 ± 1.85	0.17
<b>Have you ever worn a VR headset?</b>	Never	10 (61%)	8 (47%)	0.64
	Only once	8 (24%)	11 (33%)	
	Several times	4 (12%)	2 (13%)	
	Regularly	1 (3%)	2 (7%)	
<b>Do you play VR video games?</b>	Never	18 (62%)	19 (58%)	1.00
	Occasionally	5 (38%)	4 (42%)	
	Often	0 (0%)	0 (0%)	

### 5.3 Data Collection and Analysis

All interaction data were collected automatically through simple actions such as “pouring a solution into a beaker” or “grasping a pH strip”. These actions enabled us to analyze participants’ performance during each manipulation by verifying the accuracy of their answers or detecting the completion of expected procedures (e.g., “selecting the correct color indicator” or “respecting the correct mixing order” in Manipulation 2). We identified whether participants provided correct or incorrect answers for each item evaluated in the learning scenario. We thus defined performance as the combination of correct answers and the execution of expected procedures for each manipulation, reflecting the acquisition of procedural skills related to the experimental tasks.

We conducted two types of analyses to address our research questions: an analysis of participants’ performance across the different manipulations of the scenario (RQ1) and an analysis of perceived cognitive load (RQ2). We conducted the analyses primarily using Python (*Pandas*, *SciPy*), with complementary analyses run in Jamovi for the NASA-TLX self-assessment questionnaire [9]. For RQ1, as data were not normally distributed, group performances were compared using the Mann-Whitney test based on scores from correct and incorrect answers and scores for completed procedures in each manipulation. For the repeated tasks in Manipulation 2 (Task 1 and Task 2), we used Wilcoxon tests with and without Benjamini-Hochberg correction to analyze within-group differences. Given

the small sample size and the non-normal distribution of almost all the data, we used non-parametric tests. For RQ2, we assessed perceived cognitive load using the NASA-TLX, which evaluates six specific dimensions: mental demand, physical demand, temporal demand, effort, perceived performance, and frustration. We used these dimensions to compute a global cognitive load score reflecting participants’ subjective experience. We conducted between-group comparisons of these scores using the Mann–Whitney test.

## 6 Results

### 6.1 RQ1a: Performance on Expected Answers

In Manipulation 1, participants spent 5.2 minutes on average in both groups. The group with immediate feedback scored significantly lower on correct answers and higher on incorrect answers than those without immediate feedback (Table 2). The moderate effect size confirms that this difference is noteworthy [3]. We reported both metrics since participants could also choose not to answer.

In Manipulation 2, Task 1 was slightly longer in the group without IF (mean=4.5 mins), compared to the group with IF (mean=4.1 mins). Conversely, Task 2 was shorter in the group without IF (mean=2.4 mins) than in the group with IF (mean=2.9 mins). Table 3 shows the overall results for this manipulation and the detailed results for each task. We observe no significant difference between the two groups, regardless of whether they received immediate feedback. However, we note a slightly higher number of incorrect answers, close to statistical significance, for the group with immediate feedback during Task 1.

**Table 2:** Manipulation 1. Mean and standard deviation of correct and incorrect answers (scores out of 5 points) for the experimental group (With IF) and the control group (Without IF). Mann–Whitney test results: p-value and effect size (Cohen’s d). Statistical significance at  $p < 0.05$  is indicated by an asterisk (\*).

	Experimental condition	Mean (SD)	p-value	Effect size
Correct answers	With IF	4.48 (0.59)	0.03*	0.69
	Without IF	4.83 (0.39)		
Incorrect answers	With IF	0.52 (0.59)	0.03*	0.69
	Without IF	0.17 (0.39)		

In contrast, the intra-group analyses presented in Table 4 reveal interesting findings for the group with immediate feedback. Benjamini–Hochberg corrected post-hoc tests show a significant improvement in the number of correct answers between the two tasks and a significant reduction in errors. We observe no such improvement in the group without immediate feedback.

**Table 3:** Manipulation 2. Mean and standard deviation of correct and incorrect answers (scores out of 6 points, 3 per task) for the experimental group (With IF) and the control group (Without IF). Mann–Whitney test results: p-value and effect size (Cohen’s d).

	Answers	Experimental condition	Mean (SD)	p-value	Effect size
Entire manipulation	Correct	With IF	4.17 (1.37)	0.72	0.13
		Without IF	4.35 (1.34)		
	Incorrect	With IF	1.61 (1.03)	0.12	0.47
		Without IF	0.13 (1.01)		
Task 1	Correct	With IF	1.78 (0.85)	0.25	0.40
		Without IF	2.09 (0.67)		
	Incorrect	With IF	1.09 (0.67)	0.08	0.54
		Without IF	0.74 (0.62)		
Task 2	Correct	With IF	2.39 (0.84)	0.911	0.14
		Without IF	2.26 (1.05)		
	Incorrect	With IF	0.52 (0.67)	0.34	0.19
		Without IF	0.39 (0.72)		

**Table 4:** Manipulation 2. Results of Wilcoxon post-hoc tests between Task 1 and Task 2: raw p-value, corrected p-value (Benjamini–Hochberg), and effect size (Cohen’s d). Statistical significance at  $p < 0.05$  is indicated by an asterisk (\*).

	Experimental condition	Raw p-value	Corrected p-value	Effect size
Correct answers	With IF	0.001*	0.038*	0.72
	Without IF	0.555	1.0	0.20
Incorrect answers	With IF	0.007*	0.027*	0.85
	Without IF	0.078	0.313	0.52

## 6.2 RQ1b: Performance on Procedures

We analyzed the performance on procedures for both manipulations. Without considering tasks separately, we observe no significant differences between the two groups for either the first manipulation or the second manipulation.

Table 5 provides the detailed results for each task of Manipulation 2. Before correction, significant differences were observed, with a medium effect size, for two specific procedures: *Monitor the color transition* during Task 1, and *Use a pH strip* during Task 2, both in favor of the group with immediate feedback. However, Benjamini-Hochberg corrected post-hoc tests no longer show statistically significant results, although they remain close to the significance threshold.

Finally, the intra-group analyses presented in Table 6 highlight a significant improvement in the *Accurately reach the pH value* procedure for the group with immediate feedback between Task 1 and Task 2, with a large effect size. After

**Table 5:** Manipulation 2. Inter-group comparison of scores obtained for the procedures in Tasks 1 and 2 (scale from 0 to 1). Mann–Whitney test results: raw and corrected p-values (Benjamini–Hochberg), and effect sizes (Cohen’s d). Asterisks (\*) indicate statistical significance at  $p < 0.05$ .

Procedure	Task	With IF	Without IF	p-value	p-value	Effect size
		Mean (SD)	Mean (SD)	(raw)	(corrected)	
Respect the mixing order	T 1	0.86 (0.34)	0.92 (0.22)	0.907	1.000	0.23
	T 2	0.88 (0.29)	0.87 (0.29)	0.552	0.663	0.04
Monitor the color transition	T 1	0.65 (0.49)	0.30 (0.47)	0.020*	0.121	0.73
	T 2	0.57 (0.51)	0.57 (0.51)	1.000	1.000	0.00
Rinse the pipette	T 1	0.87 (0.34)	0.78 (0.42)	0.452	0.978	0.23
	T 2	0.74 (0.45)	0.57 (0.51)	0.226	0.452	0.36
Use a pH strip	T 1	0.91 (0.29)	0.91 (0.29)	1.000	1.000	0.00
	T 2	1.00 (0.00)	0.83 (0.39)	0.041*	0.225	0.63
Accurately reach the pH value	T 1	0.17 (0.39)	0.26 (0.45)	0.489	0.978	0.21
	T 2	0.52 (0.51)	0.26 (0.45)	0.075	0.225	0.54
Discard the used pH strip	T 1	0.78 (0.42)	0.78 (0.52)	0.954	1.000	0.00
	T 2	0.74 (0.45)	0.65 (0.49)	0.535	1.000	0.19

Benjamini-Hochberg corrected post-hoc test, this difference remains marginally significant. No such significant evolution was observed for the group without immediate feedback.

### 6.3 RQ2: Cognitive Load

The comparison of the NASA-TLX scores between the two experimental conditions shows no significant differences in any of the measured variables: mental demand, physical demand, temporal demand, effort, frustration, and self-rated performance (see Table 7). The effect sizes, calculated using rank-biserial correlation, were consistently small across all dimensions. These findings confirm that immediate feedback did not impact participants’ perceived cognitive load.

## 7 Discussion and Conclusion

In this study, we investigated the impact of immediate feedback on participants’ performance in a VR learning environment dedicated to chemical education. This work builds upon previous findings on the role of immediate feedback in performance improvement [1,21,22] by distinguishing its effects on expected answers and completed procedures. Our results tend to nuance the positive outcomes reported in earlier studies.

**Table 6:** Manipulation 2. Intra-group comparison of scores obtained for the procedures in Tasks 1 and 2 (scale from 0 to 1). Wilcoxon test results: raw and corrected p-values (Benjamini–Hochberg), and effect sizes (Cohen’s d). Asterisks (\*) indicate statistical significance at  $p < 0.05$ .

Action	Experimental condition	Task 1 Mean (SD)	Task 2 Mean (SD)	p-value (raw)	p-value (corrected)	Effect size
Respect the correct mixing order	With IF	0.86 (0.34)	0.88 (0.29)	0.934	0.934	0.07
	Without IF	0.92 (0.22)	0.87 (0.29)	0.473	0.567	0.21
Monitor the color transition	With IF	0.65 (0.49)	0.57 (0.51)	0.559	0.838	0.17
	Without IF	0.30 (0.47)	0.57 (0.51)	0.080	0.369	0.53
Rinse the pipette	With IF	0.87 (0.34)	0.74 (0.45)	0.277	0.554	0.76
	Without IF	0.78 (0.42)	0.57 (0.51)	0.123	0.369	0.72
Use a pH strip	With IF	0.91 (0.29)	1.00 (0.00)	0.162	0.485	0.43
	Without IF	0.91 (0.29)	0.83 (0.39)	0.397	0.567	0.05
Accurately reach the pH value	With IF	0.17 (0.39)	0.52 (0.51)	0.015*	0.089	0.77
	Without IF	0.26 (0.45)	0.26 (0.45)	1.000	1.000	0.00
Discard the used pH strip	With IF	0.78 (0.42)	0.74 (0.45)	0.744	0.892	0.10
	Without IF	0.78 (0.52)	0.65 (0.49)	0.419	0.567	0.26

**Table 7:** Comparison of NASA-TLX scores between the two experimental conditions. Mann-Whitney test results (p-value). Effect size (rank-biserial correlation)

Variable	Experimental Condition	Mean (SD)	Median	p-value	Effect size
Mental demand	Without IF	11.35 (3.81)	12.00	0.817	0.042
	With IF	11.09 (4.94)	12.00		
Physical demand	Without IF	2.91 (3.88)	2.00	0.771	0.051
	With IF	2.70 (3.44)	1.00		
Temporal demand	Without IF	3.52 (4.07)	3.00	0.273	0.183
	With IF	2.48 (3.74)	0.00		
Performance (self-rated)	Without IF	13.61 (3.66)	14.00	0.799	0.045
	With IF	14.09 (3.29)	14.00		
Effort	Without IF	5.96 (2.75)	6.00	0.383	0.151
	With IF	5.43 (4.26)	5.00		
Frustration	Without IF	14.30 (5.27)	16.00	0.178	0.231
	With IF	12.13 (5.86)	13.00		

The findings suggest that immediate feedback, as designed in this study, could be integrated without causing cognitive overload, even in tasks requiring heightened attention and procedural accuracy. This result contradicts the findings of Tang et al. [21], who reported increased temporal demands and effort. We believe that this difference could be explained by the amount of information provided by the feedback and its visualisation, as in [21] the textual overlays on

the elements added complexity to the visual scene [15]. This aligns with Souchet et al. [19] who emphasize that cognitive load varies according to several factors, including task complexity and level of interaction required to complete the task, and observed increased performance with heightened cognitive load. Future studies should therefore investigate in more depth the relationship between cognitive load, performance and feedback.

Regarding performance, the results show differentiated effects of immediate feedback depending on the type of task. For single tasks, such as in Manipulation 1, immediate feedback resulted in significantly fewer correct answers compared to the group without feedback, even though we did not observe this effect for completed procedures. This suggests that immediate feedback may increase simultaneous information processing and reduce users' attention, affecting performance for this type of task. In contrast, for repeated tasks in Manipulation 2, the group with immediate feedback improved their correct answer rate while reducing the number of errors. In addition, they significantly (before correction, and a near-significantly after correction) improved one specific procedure.

Conversely, we did not observe such improvement in the group without feedback. These results illustrate the role of immediate feedback in supporting the progressive adjustment of actions and the acquisition of procedures through repetition.

In conclusion, while immediate feedback may restrain performance in single tasks, our findings demonstrate its potential to support learning in repeated tasks without increasing cognitive load. These observations highlight the importance of adapting feedback design to the nature of the task and learners' needs to maximize its effectiveness in immersive educational contexts. This finding aligns with Tang et al. [21], who advocate for a cautious approach to feedback integration, as it may sometimes divert attention from essential aspects of learning. The feedback design and the amount of information provided raise issues that should be further investigated to maximize the balance between cognitive load and performance.

Finally, this study presents several limitations. First, the small sample size limits the strength of the results, and the limited number of task repetitions does not allow for a full exploration of the cumulative effects of immediate feedback. Secondly, the participants' profiles and the specific learning context constrain the generalizability of these findings to other domains. In addition, cognitive load was only assessed using a questionnaire, while Souchet et al. [19] recommend mixed subjective and physiological measures.

Nevertheless, this study opens several promising avenues for future research, such as exploring how immediate feedback can be optimally designed with the relevant amount of information and for various types of tasks. Further research should also be conducted on how different types of immediate feedback—such as task-, process-, self-regulation- or self-level cues [10]—affect cognitive load and performance. Finally, we will investigate the impact of task repetition and feedback role on long-term knowledge retention and skills transfer.

**Acknowledgements.** This research was conducted as part of the RENFORCE project (ANR-22-CE38-0008), funded by the French National Research Agency (ANR) and endorsed by the Minalogic competitive cluster. The authors would like to thank all participants for their involvement in the study.

## References

1. Almousa, O., Prates, J., Yeslam, N., Mac Gregor, D., Zhang, J., Phan, V., Nielsen, M., Smith, R., Qayumi, K.: Virtual Reality Simulation Technology for Cardiopulmonary Resuscitation Training: An Innovative Hybrid System With Haptic Feedback. *Simulation & Gaming* **50**(1), 6–22 (Feb 2019). <https://doi.org/10.1177/1046878118820905>
2. Checa, D., Bustillo, A.: A review of immersive virtual reality serious games to enhance learning and training. *Multimedia Tools and Applications* **79**(9), 5501–5527 (Mar 2020). <https://doi.org/10.1007/s11042-019-08348-9>
3. Cohen, J.: Statistical power analysis for the behavioral sciences, Rev. ed. *Statistical power analysis for the behavioral sciences*, Rev. ed., Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US (1977), pages: xv, 474
4. Conrad, M., Kablitz, D., Schumann, S.: Learning effectiveness of immersive virtual reality in education and training: A systematic review of findings. *Computers & Education: X Reality* **4**, 100053 (2024). <https://doi.org/10.1016/j.cexr.2024.100053>
5. Cromley, J.G., Chen, R., Lawrence, L.: Meta-Analysis of STEM Learning Using Virtual Reality: Benefits Across the Board. *Journal of Science Education and Technology* **32**(3), 355–364 (Jun 2023). <https://doi.org/10.1007/s10956-023-10032-5>
6. Di Pasquale, V., Cutolo, P., Esposito, C., Franco, B., Iannone, R., Miranda, S.: Virtual Reality for Training in Assembly and Disassembly Tasks: A Systematic Literature Review. *Machines* **12**(8), 528 (Aug 2024). <https://doi.org/10.3390/machines12080528>
7. Elaish, M.M., Yadegaridehkordi, E., Ho, Y.S.: Publication performance and trends in virtual reality research in education fields: a bibliometric analysis. *Multimedia Tools and Applications* (Apr 2024). <https://doi.org/10.1007/s11042-024-19238-0>
8. Gloy, K., Weyhe, P., Nerenz, E., Kaluschke, M., Uslar, V., Zachmann, G., Weyhe, D.: Immersive Anatomy Atlas: Learning Factual Medical Knowledge in a Virtual Reality Environment. *Anatomical Sciences Education* **15**(2), 360–368 (2022). <https://doi.org/10.1002/ase.2095>
9. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock, P.A., Meshkati, N. (eds.) *Advances in Psychology, Human Mental Workload*, vol. 52, pp. 139–183. North-Holland (Jan 1988). [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
10. Hattie, J., Timperley, H.: The Power of Feedback. *Review of Educational Research* **77**(1), 81–112 (Mar 2007). <https://doi.org/10.3102/003465430298487>
11. Jiang, J., Ahmadpour, N.: Beyond Immersion: Designing for Reflection in Virtual Reality. In: *Proceedings of the 33rd Australian Conference on Human-Computer Interaction*. pp. 208–220. OzCHI '21, Association for Computing Machinery, New York, NY, USA (Sep 2022). <https://doi.org/10.1145/3520495.3520501>

12. Kiegaldie, D., Shaw, L.: Virtual reality simulation for nursing education: effectiveness and feasibility. *BMC Nursing* **22**(1), 488 (Dec 2023). <https://doi.org/10.1186/s12912-023-01639-5>
13. Lampropoulos, G., Kinshuk: Virtual reality and gamification in education: a systematic review. *Educational technology research and development* **72**(3), 1691–1785 (Jun 2024). <https://doi.org/10.1007/s11423-024-10351-3>
14. Lee, L.K., Wei, X., Chui, K.T., Cheung, S.K.S., Wang, F.L., Fung, Y.C., Lu, A., Hui, Y.K., Hao, T., U, L.H., Wu, N.I.: A Systematic Review of the Design of Serious Games for Innovative Learning: Augmented Reality, Virtual Reality, or Mixed Reality? *Electronics* **13**(5), 890 (2024). <https://doi.org/10.3390/electronics13050890>
15. Ma, X., Wijewickrema, S., Zhou, Y., Zhou, S., O’Leary, S., Bailey, J.: Providing Effective Real-Time Feedback in Simulation-Based Surgical Training. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) *Medical Image Computing and Computer-Assisted Intervention MIC-CAI 2017*. pp. 566–574. Springer International Publishing, Cham (2017). [https://doi.org/10.1007/978-3-319-66185-8\\_64](https://doi.org/10.1007/978-3-319-66185-8_64)
16. Roussou, M., Slater, M.: Comparison of the Effect of Interactive versus Passive Virtual Reality Learning Activities in Evoking and Sustaining Conceptual Change. *IEEE Transactions on Emerging Topics in Computing* **8**(1), 233–244 (Jan 2020). <https://doi.org/10.1109/TETC.2017.2737983>
17. Schon, D.A., DeSanctis, V.: The Reflective Practitioner: How Professionals Think in Action. *The Journal of Continuing Higher Education* **34**(3), 29–30 (Jul 1986). <https://doi.org/10.1080/07377366.1986.10401080>
18. Schunk, D., Zimmerman, B.: *Self-regulated Learning: From Teaching to Self-reflective Practice*. London: The Guilford Press, New-York, London (1998)
19. Souchet, A.D., , Stéphanie, P., , Domitile, L., , Leroy, L.: Measuring Visual Fatigue and Cognitive Load via Eye Tracking while Learning with Virtual Reality Head-Mounted Displays: A Review. *International Journal of Human–Computer Interaction* **38**(9), 801–824 (May 2022). <https://doi.org/10.1080/10447318.2021.1976509>
20. Sweller, J.: Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science* **12**(2), 257–285 (1988). [https://doi.org/10.1207/s15516709cog1202\\_4](https://doi.org/10.1207/s15516709cog1202_4)
21. Tang, K., Gerling, K., Vanden Abeele, V., Geurts, L., Aufheimer, M.: Playful Reflection: Impact of Gamification on a Virtual Reality Simulation of Breastfeeding. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. vol. 783, pp. 1–13. CHI ’23, New York, NY, USA (Apr 2023). <https://doi.org/10.1145/3544548.3580751>
22. Van Ginkel, S., Ruiz, D., Mononen, A., Karaman, C., de Keijzer, A., Sitthiworachart, J.: The impact of computer-mediated immediate feedback on developing oral presentation skills: An exploratory study in virtual reality. *Journal of Computer Assisted Learning* **36**(3), 412–422 (2020). <https://doi.org/10.1111/jcal.12424>
23. Verlaine, E.d.: Discovering How to Do Reflexivity and Self-Reflexivity: A Longitudinal Empirical Research Findings. *Advances in Social Sciences Research Journal* **9**(8), 371–394 (Aug 2022). <https://doi.org/10.14738/assrj.98.12931>
24. Zimmerman, B.J.: Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice* **41**(2), 64–70 (May 2002). [https://doi.org/10.1207/s15430421tip4102\\_2](https://doi.org/10.1207/s15430421tip4102_2)