*European Journal of Legal Education*
Vol. 5, No. 1, June 2024, 157–190

# Pleading in the Virtual Courtroom: Exploring Experiential Learning in Law through Virtual-Reality-Based Exercises and Student Feedback

Emanuel van Dongen*

## Abstract

By using virtual-reality-based technology, Dutch law students train their pleading skills in a virtual courtroom. While watching each other's performance, students give each other feedback by means of a specially designed feedback app. Afterwards, they can download the feedback and a recording of their performance. This creates an opportunity to practise their pleading skills and to improve their performance in the courtroom. We studied whether elevated levels of heart rate and electrodermal activity are found in connection with the virtual-reality-based exercises, which could indicate arousal and that proved to be the case, denoting the authenticity of the simulation. We also studied whether or not these virtual-reality-based exercises (positively) influence student experience and student learning by focusing on four mechanisms related to the virtual reality exercises: its value/usefulness, competence perceived by students, students' confidence, and ultimately the role of peer feedback and students' reflective thinking. Significant increases in heart rate and electrodermal activity occurred from rest into pleading. The construct reflective thinking appeared to be correlated with feedback, value and self-confidence. The virtual-reality-based exercises, as an authentic task, seem to create the opportunity, impossible in real settings, to provide extra training in an authentic environment and to make feedback possible. It remains crucial to learn from feedback and to work on self-determined learning goals.

Keywords: Virtual reality, student learning, experiential learning, research into legal education, presentation skills.

* Utrecht University.

ISSN 1684-1360 print/ISSN 1750-4686 online

## Introduction and theoretical framework

*Problem statement*

Oral presentation skills belong to the competences of highly-educated legal professionals.[1] A specific form of oral presenting is pleading, i.e. the skill to present an argument in a court-setting. Pleading is an important skill for law students. Practice will result in an improved quality of pleading, as deliberate practice makes perfect.[2] However, deliberate efforts over a longer period are needed to change particular aspects of performance. To that effect, teachers need to give immediate feedback, designing practice activities individual students can engage in between teacher-student meetings, in order to allow students to work on an individualised set of practice goals designed by their teacher.[3] Furthermore, practice will help students to develop their professional identities, i.e., to conduct themselves in a manner that conforms to customs, values, and mores of the legal profession.[4]

Unfortunately, a frequently heard complaint is a lack of sufficient opportunities to practise pleading in the study program. The large number of students currently enrolling into Law programs at Dutch universities may further complicate this situation (e.g., approx. 900 first-year law students at Utrecht University). Various kinds of presenting assignments are often part of regular university courses in order to train presenting (or: pleading) skills, but not always in the context of an actual Moot Court. Moot Courts, when part of educational programs,[5] are often concluded with pleas at an actual or staged court of law. Presenting in general is, however, often mentioned as causing one

---

[1] L. De Grez, *Optimizing the Instructional Environment to Learn Presentation Skills* (diss. Gent), 2009, p. 1; see also D. Kerby & J. Romine, 'Develop Oral Presentation Skills through Accounting Curriculum Design and Curriculum-Embedded Assessment', *Journal of Education for Business* 85 (2009), 3, p. 172, 176.
[2] K.A. Ericsson, R.T. Krampe & C. Tesch-Römer, 'The Role of Deliberate Practice in the Acquisition of Expert Performance', *Psychological Review* 100 (1993), p. 363–406.
[3] K.A. Ericsson, 'The Differential Influence of Experience, Practice, and Deliberate Practice on the Development of Superior Individual Performance of Experts', in: K.A. Ericsson, R.R. Hoffman, A. Kozbelt, & A.M. Williams (Eds.), *The Cambridge Handbook of Expertise and Expert Performance*, Cambridge University Press 2018, p. 745ff., 755ff.
[4] E.S. Fruehwald, 'Developing Law Students' Professional Identities', *University of La Verne Law Review* 37 (2015), 1.
[5] At Utrecht University, only a select group of honours students, around 100 per year, who follow an extensive program ('Utrecht Law College'), follow an obligatory first-year Moot Court.

of the main fears in social contexts.[6] Developing students' oral presentation competences might reduce this fear. This calls for taking a closer look at pleading. In this contribution, we used an adapted version of the definition of De Grez's definition of oral presentation competence, and define pleading as 'a combination of knowledge, skills and attitudes needed to plea in court in order to inform, self-express, relate and to persuade'.[7]

In a Dutch, university-transcending, innovation project,[8] a virtual, authentic (and thus near-realistic) courtroom has been developed. While wearing actual lawyer gowns, students present part of their plea in a virtual courtroom using a virtual headset (HP Reverb G2 V2).[9] Their pleads are heard and questioned by a legal professional (projected inside the courtroom as a judge). Furthermore, they receive digital feedback from fellow students by means of a peer feedback app. Students can download their feedback, together with a recording of their presentation, and do so directly after their presentation, in class. Thus, they can review it and read peer feedback which is believed to promote self-reflection. Feedback is not considered information provided by an agent such as a teacher or peer,[10] but understood as

[6] C.M. Smith & T.M. Sodano, 'Integrating Lecture Capture as a Teaching Strategy', *Active learning in Higher Education* 12 (2011), 3, esp. p. 153.
[7] De Grez, *Optimizing the Instructional Environment to Learn Presentation Skills*, p. 5 ('the combination of knowledge, skills and attitudes needed to speak in public in order to inform, to self-express, to relate and to persuade').
[8] In this Project 'Leren met Virtual Reality: The Next Level', three universities (University of Groningen, Free University Amsterdam and Utrecht University) work together. The first author of this contribution was work packages leader on Education Research and – not from the start, but soon after – its general project leader. This is a follow-up project on an earlier collaborative innovation project, which yielded promising results, as students were enthusiastic and positive on this authentic exercise.
[9] A computer with specific specifications and software was installed at the university. Three rooms were used: a presentation room, feedback location and a green screen-room for the judge.
[10] J. Hattie & H. Timperley, 'The Power of Feedback', *Review of Educational Research* 77 (2007), 1, p. 81.

> a process in which students interpret information from various sources about their performance and use it to improve the quality of their work or learning strategies.[11]

Therefore, after receiving and reading the feedback ('interpreting', thus giving it meaning), students were given the opportunity to ask clarifying questions, after which a final general conversation was monitored by the teacher who also gave some additional feedback if needed. Before leaving the classroom, students had to take note of some points of improvement, based on feedback received, but also implicitly, namely based on what they learned from the other students' presentations. The students also had to note some points for improvement they would work on the next time they plead. This approach was chosen to spur them into action and to stimulate their own responsibility in their learning process. Students, watching their peers pleading, gave peer feedback across the collaborating universities. Feedback was based on, or started from, a scoring rubric.[12] Thus, the amount[13] and the reliability of peer feedback increased, making it more complete. Furthermore, this approach also gave additional learning opportunities to the peers who provided feedback, deducing insights to improve their presentations.[14] As we made formative use of the rubric we used it as a tool to guide and support learning, an assessment tool for students to review their peers' presentations in a peer assessment setting.[15]

These virtual-reality-based training sessions create an authentic opportunity for students to practise their pleading skills and to improve their performance in the courtroom. It is however unclear i) which elements of our virtual-reality-

---

[11] Cf. S. Vanhoof & G. Speltincx, *Feedback in de klas. Verborgen leerkansen*, Lannoo Campus 2023, p. 19. See also D. Carless & D. Boud, 'The Development of Student Feedback Literacy: Enabling Uptake of Feedback', *Assessment and Evaluation in Higher Education* 43 (2018), 8, p. 1315: 'feedback is defined as a process through which learners make sense of information from various sources and use it to enhance their work or learning strategies'.

[12] Rubrics improve feedback processes that aim to enhance learning outcomes; see, *e.g.*, A. Jonsson & G. Svingby, 'The Use of Scoring Rubrics: Reliability, Validity and Educational Consequences', *Educational Research Review* 2 (2007), p. 130-144; E. Panadero & A. Jonsson, 'The Use of Scoring Rubrics for Formative Assessment Revisited. A Review', *Educational Research Review* 9 (2013), p. 129-144; F.J. Prins, R. de Kleijn & J. van Tartwijk, 'Students' Use of a Rubric for Research Theses', *Assessment and Evaluation in Higher Education* 2015, p. 128-150.

[13] The amount of feedback increases as students are present at more locations, can see the presenter and give feedback.

[14] Vanhoof & Speltincx, *Feedback in de klas*, n. 11, p. 106.

[15] Prins, De Kleijn & Van Tartwijk, 'The Use of Scoring Rubrics for Formative Assessment Revisited', n. 12, p. 129.

based task positively influence student experience and ii) if the exercises in a virtual reality promote student learning. Another related, preliminary, question is if the VR-environment is authentic. If so, one would expect 'arousal', a physiological reaction to stimuli (i.e., increased heart rate and blood pressure),[16] connected with pleading in a virtual reality context since the plea occurs before an expert audience.[17] In the main part of this contribution, I will discuss elements of our virtual-reality-based training sessions that positively influence student experience and whether they promote (various mechanisms underlying) student learning. Pleading in a virtual reality is very promising. Furthermore, the answers to these questions are highly relevant, considering the large number of law students and the restricted possibilities of oral feedback during the law curriculum, provided positive effects on learning experience and learning outcomes can be determined. At present, such research is still lacking.

## The use of Virtual Reality in education: learning outcomes and intervention

### *Virtual Reality: the potential*

Virtual Reality (VR), just as Augmented Reality (AR) and Mixed Reality (MR), is considered to be an enhanced learning technology. It provides opportunities to give students a learning experience that is not possible in a classroom setting. The expansion of developing AR, VR, and MR has stimulated the exploration of the learning potential of these technologies. Virtual Reality is related to the concept of hyperreality, coined by the French sociologist J. Baudrillard.[18] In his theory, this was linked to the idea of simulacrum in which the signs of the real are substituted by simulations of reality. Rubin/Grey defined VR as

> a technology by which computer-aided stimuli create the immersive illusion of being somewhere else – and a topic on

[16] Comp. J. Cusveller, C. Gerrtitsen & J. De Man, 'Evoking and Measuring Arousal in Game Settings', in: S. Göbel & J. Wiemeyer (eds.), *GameDays 2014*, 2014, p. 165.
[17] According to studies in medical education, presentations before content-expert audiences are likely to result in high levels of speech anxiety, see S. van Ginkel *et al*., 'Fostering Oral Presentation Competence through a Virtual Reality-Based Task for Delivering Feedback', *Computers & Education* 134 (2019), p. 85.
[18] J. Baudrillard, 'Simulacra and Simulations', in: J. Baudrillard, *Selected Writings*, ed. M. Poster, Stanford: Stanford University Press 1988, p. 119-148.

> which the middle ground is about as scarce as affordable housing in Silicon Valley.[19]

It is precisely the illusion that is the strength of virtual reality: it allows for acquiring knowledge and skills without distractions of the environment. It provides a personal experience, in a safe environment, less daunting than a classroom.[20] It focuses on immediate and deliberate practice, which contributes to learning. Simulations have also been defined as 'interactive digital learning environments that imitate a real-life process or situation'.[21] Furthermore, virtual-reality-based technology combines crucial elements of memory: comprehending and communicating knowledge, metacognition and motivation, promoting engagement. These elements are most important for effective learning.[22]

In two Dutch commercial studies, a positive effect of virtual-reality-based exercises was found. In a study conducted by PwC, compared to the classroom situation, students appeared to be 275% more confident to apply skills learned after the training.[23] According to a study of Lepaya, virtual reality was 85% more effective to increase skills, compared to face-to-face and virtual classrooms.[24] These studies, however promising, are not fully based on scientific standards and concern virtual reality in avatar mode, which is different from our design (in which we used a 360 degrees recording). In a previous scientific study, the effectiveness of a virtual-reality-based task to practise presentations, facilitating feedback with regard to students'

---

[19] P. Rubin & J. Grey, 'The WIRED Guide to Virtual Reality. Everything you ever wanted to know about VR Headsets, Oculus, Vive, and Simulator Sickness', 8 March 2020, at https://www.wired.com/story/wired-guide-to-virtual-reality/?CNDID=52507010&mbid=nl_091118_daily_list1_p1 (accessed 01 June 2024). See also Lepaya, 'Could Virtual Reality be a L&D solution or is it just a cool gimmick?' at https://www.lepaya.com/guides/vr-whitepaper p. 5 (accessed 01 June 2024).
[20] See also Lepaya, *ibid* p. 8.
[21] Z. Merchant *et al*., 'Effectiveness Virtual Reality-Based Instruction on Students' Learning Outcomes in K-12 and Higher Education: A Meta-Analysis', *Computers & Education* 70 (2014), p. 30.
[22] Lepaya, 'Could Virtual Reality be a L&D solution or is it just a cool gimmick?' (whitepaper), p. 8.
[23] PwC, The Effectiveness of Virtual Reality Soft Skills Training in the Enterprise, public report 25 June 2020 at https://www.pwc.com/us/en/services/consulting/technology/emerging-technology/assets/pwc-understanding-the-effectiveness-of-soft-skills-training-in-the-enterprise-a-study.pdf (accessed 01 June 2024).
[24] The combination of both virtual reality and facilitator-led sessions proved even to be more effective; Lepaya, n. 19, p. 18.

presentation competence, was studied. The results indicated a significant progress in the areas of cognition (i.e. students' knowledge of the main criteria for presentation), presentation behaviour and attitudes towards presenting.[25] Other research demonstrated the effectiveness of virtual reality learning environments on various academic, communication and domain-specific skills and competences.[26]

*Design of learning task*

Ideally, learning takes place in an authentic setting.[27] In our virtual-reality-based exercises students plea in an authentic environment, being immersed in a 360 degrees (virtual) recording of the Court of first instance (*Rechtbank*) in Groningen, while wearing their real-life gowns, being questioned and interrupted by lawyers. Pleading in a moot court is a form of *experiential learning*. According to Kolb's experiential learning theory, experiential learning consist of a fourth-stage learning cycle: concrete experience, reflective observation, abstract conceptualisation and active experimentation.[28] Therefore, the experience itself is not enough, but experiential learning should consists of a contextually rich experience.[29] Furthermore, an affective learning activity or affective environment 'emphasizes the experience of concrete events'.[30] Thus they can imitate professional practice. Although Dorresteijn *et al.* indicate that the feeling of a (physical) courtroom is difficult to imitate online,[31] that is precisely what our virtual, authentic (and thus near-realistic), courtroom provided.

---

[25] Van Ginkel *et al.*, 'Fostering Oral Presentation Competence through a Virtual Reality-Based Task', n. 17, p. 78-97.
[26] See, e.g., O. Noroozi, H. Dehghanzadeh & E. Talaee, 'A Systematic Review on the Impacts of Game-Based Learning on Argumentation Skills', *Entertainment Computing* 35 (2020), p. 1-14; G. Makransky, S. Borre-Gude & R.E. Mayer, 'Motivational and Cognitive Benefits of Training in Immersive Virtual Reality based on Multiple Assessments', *Journal of Computer Assisted Learning* 35 (2019), p. 691-707.
[27] A. Herrington & J. Herrington, *Authentic Learning Environments in Higher Education*, Information Science Publishing 2006.
[28] D. Kolb, Experiential Learning: Experience as the Source of Learning and Development, 2nd ed., Pearson Education 2015.
[29] T.H. Morris, 'Experiential learning – a systematic review and revision of Kolb's model', *Interactive Learning Environments*, 28 (2020), 8, p. 1064-1077.
[30] Kolb, Experiential Learning, n. 28, p. 277.
[31] C. van Dorresteijn, F. Cornelissen & M. Volman, 'De (on)mogelijkheden van online ervaringsgericht onderwijs', *OVO* 52 (2023), 1, p. 12-17.

For the design of the learning task, real-life tasks were used, i.e. cases/dossiers in which lawyers in real-life cases defended their clients. A main characteristic of real-life tasks is that they are ill-structured rather than well-structured, and have, inter alia, multiple acceptable solutions.[32] Using real-life tasks as a basis for learning tasks ensures that learners engage in activities that directly involve them with the constituent skills involved.[33] The medium used to work on learning tasks may be the real task environment or a simulated task environment. In our case, for practical and instrumental reasons, a simulation was chosen. Court pleading practice would otherwise be impossible due to limited availability, and could be harmful for the legal position of clients present. Therefore, a safe and controlled virtual environment was created. In simulations the degree of fidelity, i.e., 'the degree of similarity between the simulated and the real task environment' is important.[34] Our virtual courtroom looks exactly like the Court of first instance in Groningen, because it is recorded at 360 degrees. Thus it has a high physical fidelity but also a high psychological fidelity, carrying out the same actions as in a real-life environment.[35] The downside of high-fidelity environments may be the confrontation of novice learners with an amount of information and work stress that interferes with their learning. Nevertheless these become increasingly important for more experienced learners.[36] When designing learning support in virtual reality, built-in task support and problem-solving guidance is needed.

---

[32] J.G. van Merriënboer & P.A. Kirschner, Ten Steps to Complex Learning: A Systematic Approach to Four-Component Instructional Design, 2017, p. 53-54.
[33] *Ibid*., p. 55-56.
[34] *Ibid*., p. 56.
[35] *Ibid*., p. 58-59. See also R.T. Hays & M.J. Singer, *Simulation Fidelity in Training System Design. Bridging the Gap Between Reality and Training*, New York 1989, who make a distinction between psychological fidelity – stress might also be a relevant element to our study, connected to this kind of fidelity –, functional fidelity and physical fidelity. According to Hays and Singer, *Simulation Fidelity in Training System Design*, p. 50, simulation fidelity is 'the degree of similarity between the training situation and the operational situation which is simulated. It is a two dimensional measurement of this similarity in terms of: (1) the physical characteristics, for example, visual, spatial, kinesthetic, etc.; and (2) the functional characteristics, for example the informational, and stimulus and response options of the training situation.' Our environment might be considered to have a medium degree of functional fidelity.
[36] Van Merriënboer & Kirschner, (n. 32) *Ten Steps to Complex Learning*, p. 59.

For the increase of problem-solving skills, process worksheets[37] and performance constraints[38] can be used - as we do in our virtual reality-based exercises

Students' perceptions towards their learning task and learning environment is, together with cognitive load and intrinsic motivation, a crucial intermediate variable in learning processes. Based on literature, Sichterman *et al.* (2022) distinguish four educational design principles: i. minimize the complexity of the AR environment to reduce extraneous cognitive load; ii. provide an interactive 3D AR environment which enables students to manipulate objects in order to improve visualisation and practice, as well as domain-specific skills; iii. construction of an active and collaborative AR learning environment providing challenging learning activities is emphasized to enhance student's motivation and attitudes for learning; iv. make the AR learning environment easy to use in a way it enhances students' positive perceptions of learning with AR.[39] Although written on AR, the same principles can also be applied to VR. A study of students' perceptions is part of our research design, as will be discussed below.

*Students' learning*

It has been well recognized that authentic learning positively influences learners' success, motivation, attitude, and self-directed learning skills.[40] In this regard, we expect our virtual-reality-based exercise to have a positive effect on student (perceived) learning. We will operationalise 'student experience and

[37] See also Merrïenboer & Kirschner, *Ten Steps to Complex Learning*, p. 77, n. 32, in which a table is included on the phases and rules of thumb for the complex skill 'preparing a plea', taken from R.J. Nadolski, P.A. Kirschner, J.J.G. van Merriënboer & H.G.K. Hummel, 'A Model for Optimising Step Size of Learning Tasks in Competency-Based Multimedia Practicals', *Educational Technology Research and Development* 49 (2001), p. 87-103, esp. p. 89.
[38] Van Merriënboer & Kirschner, *Ten Steps to Complex Learning*, n. 32, p. 78. See also R.J. Nadolski, P.A. Kirschner & J.J.G. van Merriënboer, 'Process Support in Learning Tasks for Acquiring Complex Cognitive Skills in the Domain of Law', *Learning and Instruction* 16 (2006), p. 266-278.
[39] B. Sichterman *et al*., 'Fostering Students' Engineering Competence by Adopting Augmented Reality: a Proposed Randomized Controlled Trial Study', 2022 *IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, p. 261-262.
[40] C. Hursen, 'The Impact of Curriculum developed in line with Authentic Learning on the Teacher Candidates' Success, Attitude and Self-Directed Learning Skills', *Asia Pacific Education Review* 17 (2016), 1, p. 73-86.

learning' by focusing on the following mechanisms: Value/Usefulness (I); (Perceived) Competence (II); Confidence (III); Reflective thinking (IV).

The Value/Usefulness concept (I) is used with the idea that people internalize and become self-regulated learners as to activities they experience as useful or valuable.[41] The (Perceived) Competence concept (II) is theorized to a positive predictor of both self-report and behavioural measures of intrinsic motivation.[42] Confidence (III) is theorized to be a component that influences student motivation to learn.[43] Finally, Reflection (IV) is an important aspect in the self-regulation process and thus for student learning.[44] Especially combined with Feedback, reflection is a powerful strategy to enhance performance.[45] In this regard, peer feedback is an effective feedback strategy in higher education.[46]

Our study starts with an exploration of this authentic environment. Therefore, before discussing the questionnaire study above any further, we will deal with arousal effects an authentic environment may cause, by measuring physiological data while pleading and comparing the data with a rest position. Due to a potential discomfort using a VR-head set, the maximum duration of learning activity was restricted to ten minutes.

---

[41] E.L. Deci, H. Eghrari, B.C. Patrick & D. Leone, 'Facilitating Internalization: The Self-Determination Theory Perspective', *Journal of Personality* 62 (1994), 119-142.
[42] R.M. Ryan, 'Control and Information in the Intrapersonal Sphere: An Extension of Cognitive Evaluation Theory', *Journal of Personality and Social Psychology* 43 (1982), p. 450-461.
[43] J. Keller, 'Development and Use of the ARCS Model of Instructional Design', *Journal of Instructional Development*, 10 (1987), 3, p. 2-10.
[44] P. R. Pintrich, 'The Role of Goal Orientation in Self-Regulated Learning', in: M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation*, San Diego, CA: Academic Press 2000, p. 451–502.
[45] F. Anseel, F. Lievens & E. Schollaert, 'Reflection as a Strategy to enhance Task Performance after Feedback', *Organizational Behavior and Human Decision Processes* 110 (2009), 1, p. 23–3.
[46] N. Winstone & D. Carless, *Designing Effective Feedback Processes in Higher Education: a Learning-Focused Approach*, Routledge 2019; B.J. Zimmerman, 'Attainment of Self-Regulation: A Social Cognitive Perspective', in: M. Boekaerts, P.R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation*, San Diego, CA: Academic Press 2000, p. 13-39.

## The current study

*Research question and hypothesis*

In this research, we were interested in the effect pleading in a virtual-reality-based and authentic task has on student experience and (perceived) learning. We explored whether physiological changes ('stress levels') confirmed the authenticity of our virtual-reality based exercises.[47] Authenticity in our context can be understood as similarity in experience to participating in a real courtroom.[48] Furthermore, we expected our virtual-reality-based exercises to positively influence student experience and to promote student (perceived) learning. Our hypothesis is that practising pleading in a virtual authentic courtroom is a useful and valuable experience for students (H1), which <pleading experience in virtual reality> will also increase perceived competence and promote confidence (H2). Moreover, practising pleading will also enhance critical self-reflective thinking – and will stimulate students to become reflective practitioners –,[49] as well as self-regulation skills when effective peer feedback is provided, taking into consideration that good instructions and learning activities are of central importance in virtual reality (H3).[50] With reflective thinking we mean the extent to which critical self-reflective thinking was occurring – measured by the scale Reflective Thinking (see under *Method*), i.e. the extent to which students have opportunities to reflect on their own learning and thinking.[51]

---

[47] The question whether elevated levels of arousal might negatively influence the competence and confidence could not be answered, due to the small sample size and lack of approval by participants beforehand. We believe that higher levels of arousal might be overcome by confidence, a construct we studied, see below.

[48] M. McDowell, 'Increasing Authenticity in the Classroom. Real-life scenarios enhance lessons and shouldn't require teachers to completely overhaul their practices', available at www.edutopia.org/article/increasing-authenticity-classroom/, defines authenticity as a set of experiences in which teachers and students engage in contexts and content that align with real-world experiences, and students have choice in the experiences.

[49] D. A. Schön, The Reflective Practitioner: How Professionals Think in Action, New York: Basic Books 1983.

[50] Educate-it, Virtual Reality and Augmented Reality in Education: A review, Utrecht: Utrecht University 2020.

[51] Cf. D. Maor, 'A Teacher Professional Development Program on Using a Constructivist Multimedia Learning Environment', *Learning Environments Research* 2 (2000), p. 307–330.

*Course and disciplinary context*

At the Department of Law at Utrecht University,[52] the Moot Court (*Oefenrechtbank*) is an obligatory part of the curriculum offered to honours students at the Utrecht Law College, in the last period of their first academic year where they have been introduced to the main areas of Dutch public, private and criminal law. This honours programme is characterised by a strong sense of community, which may have a positive impact on the quality of feedback and its reception. During the Moot Court, based on a real-life case file, students write an orientation assignment (in groups), follow an introductory lecture by a judge, practise in our virtual reality courtroom, have an instruction (workshop) by a legal practitioner or public prosecutor, write a pleading, practise part of their pleas in virtual reality, receive feedback, and hold a final pleading before a jury. Students receive an individual assessment of their final plea at the final court sessions, from the jury, which consists of teachers, legal practitioners and real-life judges.

## Method

*General information: informed consent and ethical approval*

Informed consent of students was asked once, at the beginning of this study. These consents, in written form and signed by hand, together with paper-based questionnaires were collected and stored safely. Digital data, inserted in Excel and SPSS, were all stored in a safe UU environment (Yoda). The ethical quality of the questionnaire had already been approved by the Faculty's Ethical Review Committee in 2020, and additional approval for the physiological measurements was granted in 2022.

*Four kinds of measures*

*Physiological responses*. Pleading in virtual reality is a form of experiential learning, and part of the affective learning activity is the emphasis on the experience of concrete events. Students' affective experiences have influence on their emotions. Using physiological responses of students may provide insight into the moment-to-moment change in their affective arousal before,

[52] VR exercises are offered at three universities: Rijksuniversiteit Groningen, Vrije Universiteit Amsterdam and Utrecht University. Unfortunately, it appeared impossible to collect sufficient data under students at the two first universities. Therefore, we decided to focus on Utrecht University.

during and after their plea. Increased heart rate can be considered an 'indicator of importance, urgency, or task engagement which only occur in motivated performance situations which a person has judged as personally relevant'.[53] A heart rate increase may be considered an indication of personal relevance, importance and/or involvement.[54] Such engagement would be expected in our authentic learning environment (with its high-fidelity). In such environment, there is similarity in experience with reality – the heightened sense of realism brought by VR known as 'presence', refers to the 'user's authentic experience of being in the virtual world as a result of visual or auditory displays generated by a computer'.[55] In previous research, real-life and VR height exposures using 3D-360° videos were mostly indistinguishable on a psychophysiological level – stress responses did not differ between real and VR environment.[56]

In our study, heart rates and electrodermal activity are measured in a real-life context, within individuals, over time, measured during the whole period of the VR exercise. An E4 wristband (*Empatica E4*) was chosen, due to its relative low degree of invasiveness and cost. Before the start of the actual pleading exercise, students' heart rate and electrodermal activity measured in rest-mode. Heart rate and electrodermal activity is then monitored during the VR-pleading exercise. The wristband also measures the skin conductance level – sweat production increases under stress, resulting in an increased conductance of the

[53] M.H. Donker, In DEPTh: Dynamics of Emotional Processes in Teachers – An Exploration of Teachers' Interpersonal Behavior and Physiological Responses, Utrecht: Utrecht University 2020, p. 73.
[54] D. Scheepers, F. de Wit, N. Ellemers & K. Sassenberg, 'Social Power makes the Heart work more Efficiently: Evidence from Cardiovascular Markers of Challenge and Threat', *Journal of Experimental Social Psychology* 48 (2012), p. 371–374; A. Scholl, F. de Wit, N. Ellemers, A.K. Fetterman, K. Sassenberg & D. Scheepers, 'The Burden of Power: Construing Power as Responsibility (rather than as Opportunity) alters Threat-Challenge Responses', *Personality and Social Psychology Bulletin* 44 (2018), p. 1024–1038; J. Storbeck & G.L. Clore, 'Affective Arousal as Information: How Affective Arousal influences Judgments, Learning, and Memory', *Social and Personality Psychology Compass* 2 (2008), p. 1824–1843. In literature, one has controlled the measured heart rate for the effect of physical activity (cf the Additional Heart Rate method, see M. Myrtek, *Heart and Emotion. Ambulatory Monitoring Studies in Everyday Life*, Göttingen, Germany: Hogrefe & Huber Publishers 2004). In our study we did not perform such control since the physical activity degree was negligible.
[55] R. Lavoie, K. Main, C. King & D. King, 'Virtual Experience, Real Consequences : the Potential Negative Emptional Consequences of Virtual Reality Gameplay', *Virtual Reality* 25 (2021), p. 69.
[56] B. Schöne et al., 'The Reality of Virtual Reality', Front. Psychol. 14 (2023).

skin[57] – and movements. Both averages as lowest/highest points were measured.[58] In order to interpret changes in physiological activation, context is important. For that reason, a student-assistant noted the moments in time when events (such as interventions by the judge and the start of pleading) occur.

The physiological responses of students were measured at the moment of the virtual-reality-based exercises (*n*=27).[59] Before the start of these exercises we tested the wristband (Empatica E4), measured one minute relax time for each student after the wristband was put on their wrist[60] (this moment was fixed by pushing the button), then measured while wearing it while pleading, while under questioning,[61] and then, after the plea was finished, fixed the time with the button and noted the time again. The physiological data were monitored during the plea via an Iphone 10,[62] and afterwards analysed via the Empatica system on a pc, where csv-files were inserted in Excel. The times mentioned in Empatica were then converted into current Dutch times and averages in relax, plea and question intervals were calculated in Excel in order to analyse the

---

[57] N. Sharma & T. Gedeon, 'Objective Measures, Sensors and Computational Techniques for Stress Recognition and Classification: A Survey', *Computer Methods and Programs in Biomedicine* 108 (2012), p. 1287-1301; Donker, n. 53.

[58] For further research, in order to interpret the changes that occurred, it would be valuable to measure whether a person's peaks in heart rate differ from their own mean heart rate (variability, ie deviation for an individual's average), cf M. Wichers, J.T.W. Wigman & I. Myin-Germeys, 'Micro-Level Affect Dynamics in Psychopathology viewed from Complex Dynamical System Theory', *Emotion Review* 7 (2015), p. 362-367.

[59] These measurements were conducted at all three universities that collaborated in this project. These measurements took place on 19 April (VU), 9 May (RUG) and 10 May (UU). As to the question of the importance of stress at the moment of VR exercises and the usefulness of this method, similar amounts were measured at the universities (double the amount at Utrecht University - the number of participating students in the virtual-reality-based exercises was also much larger). The following number of physiological responses were measured: 7 (VU), 6 (RUG) and 14 (UU), so 27 (students) in total.

[60] Three different wristbands (all Empatica E4) were used, for the risk of low batteries or other errors. Most students were right handed, one was left handed. This is relevant to choose the wrist to put the wristband on.

[61] Questions (and/or interventions) were posed during the plea (UU) or after the plea was finished (VU, RUG, UU). Because of this difference, at UU we had to separate the period [question-answer] from the 'normal pleading time' (see below, analysis). These moments in time were all noted by hand in a logbook.

[62] Logbooks were made for each student including information on location, student-number, start time, Empatica number and left/right hand. Furthermore, times were registered and included button markers: begin relax moment, end relax moment, begin plea, end plea, begin question judge, end question judge, and any other special relevancies while recording. For one student his plea and question could not be separated. For that particular case, the same values are mentioned twice in the logbook.

electrodermal activity (EDA) and Heart Rate (HR). Subsequently these data were imported in SPSS for further analysis.

*Questionnaires*. For students, participation to the current study consisted of filling in three questionnaires,[63] and using the heartbeat monitor. Students could also participate in only one of them (or none). For the questionnaire of 24 (or 28) questions[64] – based on existing scales, translated into Dutch and contextualised to our context – a 5-points Likert scale is used ranging from 'rarely or never true' (1) to 'always or almost always true' (5). The constructs are measured by multiple items distributed throughout the questionnaire.

The first construct *Value/Usefulness* has six items, e.g.: I find pleading in Virtual Reality an important activity, the second construct, *Perceived Competence*, also has six items, e.g.: I think I'm good at pleading. The third construct, *Confidence*, has six items as well, e.g.: As I take this course, I think I can successfully complete (the final session of) advocacy if I try hard enough. The perceived competence and value/usefulness constructs are derived from the Intrinsic Motivation Inventory[65] and the confidence construct is derived from the Course Interest Survey.[66] The fourth construct, *Reflective Thinking*, has four items, e.g.: The *PleitVRij* exercise makes me think hard about how I learn, and (as will be shown below in the analysis) is closely connected with the *Feedback* construct, which has six items, e.g.: The feedback via the *PleitVRij FeedbackApp*, included suggestion(s) I can use to improve my pleading skills. The feedback construct is derived from a Master thesis,[67] and

[63] Data points obtained by the questionnaires are aimed at the beginning of the course, immediately after the VR exercise and after the final plea at the Court of first instance. Data from the final evaluation of the presentations ('grades') take place only at the end of the courses, after the final plea.
[64] In the questionnaire directly after the virtual-reality-based exercises four additional items were included to measure the construct reflective thinking.
[65] R. M. Ryan, 'Control and Information in the Intrapersonal Sphere: An Extension of Cognitive Evaluation Theory', *Journal of Personality and Social Psychology* 43 (1982), p. 450-461.
[66] J. Keller, 'Development and Use of the ARCS Model of Instructional Design', *Journal of Instructional Development* 1 (1987), 3, p. 2-10.
[67] E. van Amerongen, *Effect van Instructie op Kwaliteit Peerfeedback* (masterthesis Onderwijskunde), 2009, p. 55.

the reflective thinking construct is taken from the Multimedia Constructivist Learning Environment Survey.[68]

In 2021/2022, a post-VR exercises questionnaire measuring constructs of interest was developed and pilot tested for validation purposes. The constructs of interest included: Value/Usefulness (I); Competence (II); Confidence (III); Reflective thinking (construct integrated with the feedback construct; IV[69] – in the current study (see below), I studied the (items of these) constructs separately. Items on reflective thinking were only added for the second measurement in time (not in the pre-test or post-test)). The following steps were taken for the development of the questionnaire and pilot testing: 1. Literature was searched for validated questionnaires, or clusters of questions to measure the four mechanisms mentioned; 2. clusters of questions were combined, translated and contextualized into a (Dutch) questionnaire. These questions/items were discussed with experts; 3. The assembled questionnaire was pilot tested in our specific teaching context ($n$=57).[70] We conducted a confirmatory factor analysis[71] to test construct validity, as well as Cronbach's alpha to test reliability or internal consistency using R.[72] The scale reliability, in other words the homogeneity of the items of the questionnaire, was calculated by means of Cronbach's Alpha (α). [73] Cronbach's α is a measure of internal consistency, i.e. how closely a set of items is related as a group. The Cronbach's α for the construct perceived competence was .843 (acceptable); for the construct Confidence .084 (not acceptable); for the Value/Usefulness construct .913 (acceptable, one might even say excellent); for the Feedback construct .759 (acceptable); and for the construct Reflective Thinking .817

[68] A.E. Permanasari, E. Swandana, I. Hidayah & E.S. Rahayu, 'Evaluation of Multimedia Usage for supporting Blended Learning', *International Journal of Applied Engineering Research* 11 (2016), p. 6121-6127.
[69] In the final analysis reflective thinking and feedback were not integrated because the reflective thinking was only measured at the post-VR moment, not at the pre- and post-measurement moment. Because of comparability we decided not to integrate these two constructs in the final analysis.
[70] Originally, N-60, but 3 were deleted as they had complete missingness in one construct (feedback construct).
[71] We had to deal with several (52) missing variables – we imputed all of the missing responses using the MICE function in R (which replaces the missing value with a value using multiple imputations).
[72] This validation procedure was conducted by R. Drbohlav Ollerton, at that time student-assistant, now PhD-researcher at the School of Humanities and Digital Sciences, Tilburg University. For the rest of this study, SPSS has been used instead of R.
[73] The scale is from 0 to 1, from totally not to perfectly homogeneous.

(acceptable).[74] By confirmatory factor analysis it was shown that the constructs confidence and Perceived Competence were correlated. Both findings led to rethinking on the items for construct confidence. As a result of the validation procedure two questions were slightly altered and the questionnaire's validity was reassessed within the current study.[75]

For the current study, pre-test measurements (i.e. questionnaires) were taken at the beginning of courses and were filled in together with the informed consent declarations, on 21 and 23 March 2023. VR-measurements were collected directly after the virtual-reality-based exercises on 8-12 May 2023. First, most of the items of the questionnaire were filled in. After the peer feedback was read and discussed, the remaining items were filled in. Post-measurements were collected directly after the final pleas of the course on 30 June 2023. Some of the participants ($n$=87) did not fill in all three questionnaires (see also below). The data on paper were inserted by hand into a separate Excel file and imported into SPSS. We reanalysed the reliability of the constructs. The Perceived competence construct was based on six items, one question had to be reversed coded[76] - with a Cronbach's α of .873, .906 and .811 (acceptable, even excellent). The Confidence construct was based on six items, of which one item had to be reverse coded.[77] The α was already much improved compared to the validation phase. However, it could still be further improved. Two items were deleted, which led to a Cronbach's α of .628, .680 and .583 (still doubtful).[78] The value/usefulness construct was based on six items. It showed a Cronbach's α of .872, .884 and .916 (acceptable, or even excellent). The feedback construct consisted of six items. In order to improve Cronbach's α, I deleted one item, leading to Cronbach's α of .855, .759 and .882

[74] Preliminary results, i.e., from the questionnaire, which was only collected after VR exercises, in the pilot study, showed that students about half the time perceived to be competent (M=3.167), usually felt confident (M=3.991) and valued VR (M=3.848) and about half the time valued their reflective thinking (M=3.539).
[75] Items 7 and 21 were reformulated (old: Ik krijg voldoende feedback op mijn presentatievaardigheden via de PleitVRij FeedbackApp om te weten hoe goed ik het doe; 21. Ik vind het uitdagingsniveau in deze cursus ongeveer goed: niet te gemakkelijk, niet te moeilijk. New: 7. De feedback die ik kreeg op mijn presentatievaardigheden via de PleitVrij FeedbackApp gaf mij inzicht hoe goed ik gepleit had; 21. Ik vind het uitdagingsniveau van de cursus goed: het is niet te moeilijk).
[76] 1Q5 and 3Q5, 2Q6.
[77] 2Q26, 1Q22 and 3Q22.
[78] Instead of .512, .489, .492. Items 6 en 14 were deleted (questionnaires 1 en 3) resp. items 7 en 16 (questionnaire 2).

(acceptable).[79] The reflective thinking construct was only based on 4 items, and only in the measurement we registered directly after the virtual reality-based exercises a Cronbach's α of .804 (acceptable) resulted.

*Focus group meeting*. A focus group meeting with teachers gave insights in students' experiences with virtual reality exercises during our courses. The focus group meeting was held on 26 June 2023 via Microsoft Teams. The discussion was held by an assistant who was not part of the teaching team, based on a written guideline, and was recorded and fully transcribed. The teachers (*n*=5) were all part of the inter-university innovation team. The main researcher responsible for the current research, although also involved in teaching, was not present to make sure teachers felt free to say what they wanted.

*Assessment*. The final pleas on 30 June 2023 were assessed on five elements (vocalisation, posture/gestures, contact with the court, adequately reacting to questions and structuring in reply/rejoinder) on a scale from moderate, sufficient, good (or not applicable) by a court consisting of three legal professionals, consisting of judges, practising lawyers and academics. These results (*n*=78) were coded and imported in SPSS for further analysis.

## Analysis

*Physiological responses*. Three measurements were registered per construct (heart rate (HR) and the electrodermal activity (EDA - used to measure arousal)), namely the moment of relax (i.e. a baseline per individual), and the changes that occurred under activities, namely at moment of plea and questions. For these three constructs, the means, and the highest and lowest values were calculated. From the created photoplethysmography - an uncomplicated and inexpensive optical measurement method that is often used for heart rate monitoring purposes[80] -, we derived the HR and the EDA. A one-

[79] For that Q12 (questionnaire 1 and 3) resp. 14 (questionnaire 2) had to be deleted. Otherwise it would have been .759, .607 and .775.
[80] D. Castaneda et al., 'A review on wearable photoplethysmography sensors and their potential future applications in health care', *Int J Biosens Bioelectron.* 4 (2018), 4, p. 195–202.

way repeated measured analysis of variance (ANOVA) was conducted to study if significant differences occurred.

*Questionnaires and exam results.* A one-way repeated measured analysis of variance (ANOVA) was conducted to evaluate the null hypothesis that there is no change in students' perceived competence, confidence, value of virtual reality-based exercises, and feedback before, during and after virtual-reality based exercises. Further, a Pairwise difference has been calculated to study if there is a significant difference between the three measurement moments (at the beginning of the course, directly after the VR exercises and at the end of the course, after the final pleas). Finally, I calculated the construct reflective thinking at the moment after their virtual reality exercises (measurement moment 'during' the course), and the scores given by legal professionals, on five aspects (vocalisation, posture/gestures, contact with the court, adequately reacting to questions and structuring in reply/rejoinder) at the final pleas. Lastly, I calculated the correlations between the perceptions on the four constructs, reflective thinking resulting from VR (construct reflective thinking) and the final grades (E1-5).[81]

*Focus group*. In the final stage we conducted one focus group with the teaching staff involved.[82] First we planned an online meeting with the academic staff involved in *Pleitvrij* (RUG, VU and UU) and before the meeting, we created a conversation guide.[83] Based on the transcription of the focus group, the researcher thematically summarised the main findings. During the meeting we

[81] We neglected the correlation between various aspects of grading at the final stage (i.e. between E1-5) and between factors among themselves (e.g. PC1 and V1) and/or at various moments (e.g. PC1 and PC2 etc.).
[82] We also planned a focus group with representatives of the student population. Unfortunately, however, at the end this appeared to be unrealistic because of their exams, the summer holidays etc.
[83] In setting-up, conducting and analysing the focus group we profited from the insights given by, *inter alia*, R.A. Krueger & M.A. Casey, 'Designing and Conducting Focus Group Interviews', in: R.A. Krueger *et al*., 'Social Analysis. Selected Tools and Techniques', *Social Development Papers* 36 (2001), p. 4-23; M. Matthijssen, *Gespreksleider focusgroepen*, 5th edn., Rotterdam 2019.

focused on the students' perspective and on the added value of our virtual reality-based exercises.[84]

## Results

*Results: EDA and HR*

Public speaking is a common source of stress, especially when performed before peers. If our VR environment is authentic, a similar form of stress is to be expected. In this study, EDA has been used to measure 'sympathetic nervous system arousal' and to 'derive features related to stress, engagement, and excitement'.[85] Skin conductance consists of a tonic and a phasic component: the tonic component consists of slowly varying activity, also called skin conductance level. It generates a moving baseline per individual[86] which we measured in relax modus. Table 1 shows the means measured of the averages in relax modus, in plea and while under questioning.

*Table 1: Electrodermal Activity and Heart Hate (n=27)*

| | *M* | *SD* |
|---|---|---|
| EDA | | |
| relax | 0.56 | 0.73 |
| plea | 1.34 | 2.4 |
| questions | 1.48 | 2.48 |
| | | |
| HR | | |
| relax | 84.1 | 9.93 |
| plea | 93.8 | 12.9 |
| questions | 90.7 | 17.8 |

[84] Four questions were asked: i. how did our virtual-reality-based exercises contribute to an increase in pleading competences? ii. how did our virtual-reality-based exercises contribute to an increase in self-confidence? iii. In what way did our virtual reality-based exercises contribute to reflective thinking of students? iv. What was the value of our virtual-reality-based exercises for students?

[85] E4 wristband from Empatica. User's manual, p. 4.

[86] A. Horvers, N. Tombeng, T. Bosse, A.W. Lazonder & I. Molenaar, 'Detecting Emotions through Electrodermal Activity in Learning Contexts: A Systematic Review', *Sensors* 2021 Nov 26;21(23):7869.

According to the repeated measures ANOVA, the difference between the means was statistically significant.[87] A follow up analysis, using a Post hoc comparison with a Bonferroni correction, showed that – as could be expected – the mean score of EDA in rest (i.e., the baseline measurement), not only gave the lowest result, but was also substantially different than, on average, in situations of pleading (mean. diff. = -.781; $p$ = .141) and under question (mean diff. -.923; $p$ = .077),[88] pleading being a phasic activity representing a reactive response. Looking at the EDA's increase towards pleading, on average, there is an increase of .78, but with a large variability (*SD*=1.95).[89] This all shows that a physiological reaction actually occurs when pleading in virtual reality.

Also in the case of heart rates, the difference between the means at three moments in time was statistically significant.[90] A follow up analysis, using a post hoc comparison with a Bonferroni correction, showed that the heart rate at relax level – which was, as could be expected the lowest in time – was significantly different from the mean heart rate while pleading, which had a significant higher rate (mean diff. -9.73; $p$ = .014). On average, the mean increase of the heart rate from rest towards pleading was 9.73, it showed a large variability (*SD*=16.3).[91] Although the mean score of heart rates for pleading were higher than under questioning, this difference was not significant.

*Table 2: Differences in moments: EDA and HR (n=27)*

| | *M* | *SD* | Min. | Max. |
|---|---|---|---|---|
| DifferenceRelaxPleading | 0.78 | 1.95 | -0.25 | 9.86 |
| DifferenceRelaxPleadingMax | 1.25 | 3.16 | -0.2 | 16.32 |
| DifferenceRelaxPleadingMin | 0.32 | 0.84 | -0.34 | 3.89 |
| | | | | |

[87] F (1.032; 26840) – 5.002, $p$ = 0.033. As the assumption of Sphericity was violated (Mauchly test; p<.001), I used the Greenhouse-Geisser correction instead, which gave a statistically significant result.

[88] We also see that there is a large variability. It is interesting to know if these physiological measures correlate with the constructs measured in the questionnaire. The physiological measures were however anonymised and the overlapping sample too small. Therefore, we could not study this question in the current study.

[89] Thus, the EDA's increase on average was the highest for one student: 9.86, while the opposite student showed a decrease of .25. The EDA at highest point for one student was 16.32, while at the same moment one student had a decrease of .20.

[90] The assumption of sphericity was met ($p$ = .305); F (2,52) = 4.634, $p$ = .014.

[91] Thus, on average, the heart rate increase was the highest for one student: 45.13, while the opposite showed a decrease of 19.8. The heart rates at the highest point for one student was 66.49, while at the same moment another student had a decrease of 7.40.

| | | | | |
|---|---|---|---|---|
| DifferenceHeartRateAverage | 9.73 | 16.3 | -19.8 | 45.13 |
| DifferenceHeartRateHigh | 27.5 | 18.1 | -7.4 | 66.49 |
| DifferenceHeartRateLow | -10 | 14.5 | -35.6 | 20.06 |

We see that pleading in virtual reality is a situation that significantly affects heart rate and skin conductance responses. However, the degree in which persons responded to pleading in virtual reality and the extent to which a response occurred is difficult to express unequivocally. As a follow-up, further study is needed to categorise and to mediate effects. While the personal experience of the virtual reality exercise varies among students, it lacks the significant real-world consequences associated with failure in a legal context Exactly because of this fact it is valuable. Therefore, and this stands out as a significant conclusion, the task is authentic, also because of the large variability in physiological measures.

*Results: questionnaires*

The results of the repeated measures analysis of variance (ANOVA), on the pre-, VR- and post-test, are presented per construct. See Table 3.

*Table 3: Self-reported values on five constructs on a 5-point Likert scale*

| | pre | | VR | | post | | *n* |
|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | |
| Value/usefulness | 3.719 | .7533 | 3.961 | .7679 | 3.298 | .9648 | 62 |
| Confidence | 3.970 | .5659 | 4.117 | .5842 | 3.934 | .5157 | 66 |
| Perceived confidence | 3.027 | .6877 | 3.029 | .7768 | 3.345 | .5972 | 64 |
| Feedback | 3.736 | .7594 | 3.856 | .5612 | 2.992 | .8377 | 62 |
| Reflective thinking | | | 3.708 | .7540 | | | 75 |

*Value/usefulness*. The values of the participants' reported value/usefulness when measured before, during and after the course indicated a significant time effect.[92] Therefore, there is significant evidence to reject the null hypothesis. A follow up analysis, using a post hoc comparison with a Bonferroni correction, showed that the mean score of the pretest was significantly different from that of the post-test (mean dif. = -.421; $p$ = .003), as was also between the mean score of the VR-measurement and that of the post-test (mean dif. = .663; $p$ < .001), however not between the mean score of the pre-test and that of the VR-test. This means that the usefulness of virtual reality exercises increased in the sense that the self-reported values were higher than expected at the beginning of the course, when asked directly after the virtual reality exercises; however, in the perception of students, when we asked them again at the end of the course, the values reported for the usefulness of the virtual reality exercises decreased.

*Confidence*. The values of the participants' reported confidence when measured before, during and after the course indicated a significant time effect.[93] Therefore, there is significant evidence to reject the null hypothesis. A follow up analysis, using a post hoc comparison with a Bonferroni correction, showed that pairwise difference between the mean score at the pretest and that at the VR-test was significant (mean dif. = -.148; $p$ = .037), as was also the difference between the mean score at the VR-measurement and that of the post-test (mean dif. = .184; $p$ = .014) was significant, not between pre- and post-test. Although an average increase in reported confidence can be seen after the VR exercises, a decrease took place in the period between virtual reality exercises and the final pleas.

*Perceived competence*. The participants' reported perceived competence when measured before, during and after the course indicated a significant time effect.[94] Therefore, there is significant evidence to reject the null hypothesis. A follow up analysis, using a post hoc comparison with a Bonferroni correction, showed a significant difference between the mean score at the pretest and the post-test (mean dif. = -.318; $p$ < .001), as well as between the VR-measurement and the post-test (mean dif. = -.316; $p$ < .001), however not between pre-test and VR-measurement. We can see an increase in perceived competence during

[92] Wilks' Lambda =.536, F (2, 60) = 25.947, $p$ = <.001, η2 = .464.
[93] Wilks' Lambda =.852, F (2, 64) = 5.563, $p$ = .006, η2 = .148.
[94] Wilks' Lambda =.712, F (2, 62) = 12.567, $p$ = <.001, η2 = .288.

the course, which is shown after virtual reality exercises as well as (again) after the final pleas.

*Feedback*. The participants' reported feedback construct when measured before, during and after the course indicated a significant time effect.[95] Therefore, there is significant evidence to reject the null hypothesis. A follow up analysis, using a post hoc comparison with a Bonferroni correction, showed that pairwise difference between the mean score of the pretest and the mean score at the post-test was significant (mean dif. = .744; $p < .001$), as was also the difference between the mean score at the VR-measurement and the post-test (mean dif. = .865; $p < .001$) was significant, however not between pre-test and VR-measurement. Clearly there was a decrease in the value of feedback, when construed and/or compared by students at the moment after their final plea. Still, the value of feedback was rated quite highly (M=3.845).

*Reflective thinking*. Finally, the construct reflective thinking was measured at the moment after their virtual reality exercises (measurement moment 'during' the course) – only at this moment – not at the beginning or at the end of the course – because only during the course they received feedback in the sense of our definition (see above).

*Results: final assessments*

During the final plea students were scored on five aspects. See the results in Table 4.

*Table 4: Final assessment on a 3-points scale*

| | *M* | *SD* | *n* |
|---|---|---|---|
| vocalization | 2.737 | .3841 | 78 |
| posture | 2.699 | .4213 | 78 |
| contact with judges | 2.686 | .4566 | 78 |
| adequate reactions | 2.550 | .5873 | 60 |
| rejoinder/rebuttal | 2.684 | .4565 | 38 |

[95] Wilks' Lambda =.516, F (2, 60) = 28.166, $p = <.001$, $\eta 2 = .484$.

*Results: correlation between constructs*

An important correlating construct with other constructs appeared to be Reflective Thinking, which construct also had a central role in our hypotheses. In H1 (practising pleading in a virtual, authentic, courtroom will be a useful and valuable experience for students), H2 (pleading experience in virtual reality increases perceived competence and promotes confidence) and H3 (practising pleading will also enhance reflective thinking and self-regulation skills when effective peer feedback is provided) , it was Reflective thinking, connected with the virtual reality exercises, that linked these constructs in H1, H2 and H3 together. See Tables 5-7 and discussion below. We measured Reflective Thinking by the four items of the construct Reflective Thinking in our questionnaire (see above, *Method*).

*Table 5: Correlations between reflective thinking and value attributed to pleading in virtual reality*

| | RT2 | V1 | V2 | V3 |
|---|---|---|---|---|
| RT2 | 1 | .395** (n=68) | .645** (n=74) | .543** (n=69) |
| V1 | | 1 | .386** (n=67) | .393** (n=68) |
| V2 | | | 1 | .704** (n=69) |
| V3 | | | | 1 |

*Hypothesis H1*. Table 5[96] shows the Pearson correlations (*r*) between reflective thinking following on the virtual reality exercises (Reflective Thinking measured during the course, after the VR exercises (i.e. the second measurement): RT2), and the value attributed to pleading in virtual reality at the start of the course (V1), after the virtual reality exercises (V2) and at the end of the course (V3). We found significant positive correlations between reflective thinking resulting from the virtual reality exercises and the value attributed to practising in virtual reality at the beginning of the course, after the virtual reality exercises and at the end of the course. These increased after the virtual reality-based exercises but decreased at the end of the course.

*Table 6:Correlations between reflective thinking, confidence and perceived competence*

| | RT2 | C1 | C2 | C3 | PC1 | PC2 | PC3 |
|---|---|---|---|---|---|---|---|

[96] In the correlation tables (i.e., Table 5-7) as well as all results presented on correlations, * shows significant correlation at .05 level, while ** shows correlation at .01 level (2-tailed).

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| RT2 | 1 | -0.192 (*n*=70) | -0.05 (*n*=75) | .252* (*n*=70) | -0.153 (*n*=71) | -0.004 (*n*=73) | -0.001 (*n*=69) |
| C1 | | 1 | .620** (*n*=71) | .484** (*n*=74) | .690** (*n*=79) | .596** (*n*=69) | .603** (*n*=73) |
| C2 | | | 1 | .514** (*n*=71) | .594** (*n*=72) | ,696** (*n*=74) | .597** (*n*=70) |
| C3 | | | | 1 | .293* (*n*=75) | .318** (*n*=69) | .610** (*n*=77) |
| PC1 | | | | | 1 | .740** (*n*=70) | .617** (*n*=74) |
| PC2 | | | | | | 1 | .639** (*n*=68) |
| PC3 | | | | | | | 1 |

*Hypothesis H2.* Table 6 shows the Pearson correlations (*r*) between reflecting thinking following the virtual reality exercises (RT2) and confidence and. perceived competence at the beginning (C1 and PC1), after the virtual reality exercises (C2 and PC2) and at the end of the course (C3 and PC3). Furthermore, it shows the correlations between the constructs of confidence and perceived competence – which became apparent while analysing the data. We found significant positive correlation between reflective thinking due to the virtual reality exercises and confidence at the end of the course, not at beginning of the course or after the virtual reality exercises. No correlation was established between reflective thinking due to the virtual reality exercises and perceived competence at the beginning, after the virtual reality exercises or at end of the course. Nevertheless, both constructs confidence and perceived competence are significantly correlated at all moments, to all moments.

*Table 7: Correlations (r) between reflective thinking (RT) and feedback (FC)*

| | RT2 | FC1 | FC2 | FC3 |
|---|---|---|---|---|
| RT2 | 1 | .398** (*n*=72) | .399** (*n*=66) | .478** (*n*=69) |

| | | | |
|---|---|---|---|
| FC1 | 1 | .405** (*n*=64) | .136 (*n*=71) |
| FC2 | | 1 | .282* (*n*=64) |
| FC3 | | | 1 |

*Hypothesis H3.* Table 7 shows the Pearson correlations (*r*) between reflecting thinking following the virtual realities (RT2) and feedback construct at the beginning (FC1), after the virtual reality exercises (FC2) and at the end of the course (FC3). We found significant positive correlation between reflective thinking due to the virtual reality exercises the peer feedback received on their pleading in virtual reality exercises and the peer feedback received on their pleading in virtual reality when judged at the beginning of the course, after the virtual reality exercises as well as at the end of the course.[97]

*Other results.* We also studied the relationship between the self-reported values (of value/usefulness of the virtual reality-based exercises, of confidence and of perceived competence) as filled in by students and the assessments provided by legal experts (not included due to its size). We found significant positive correlations between confidence measured directly after the virtual reality exercises and adequately reacting at the final pleas (.331*), between perceived competence directly after the virtual reality exercises and contact with judges at the final plea (.237*). We found significant positive correlations between perceived competence at the beginning of the course resp. directly after the virtual reality-based exercises, with the grade on vocalisation at the final plea (.238* resp. .346**), but surprisingly, significant negative correlation between the value attributed to virtual-reality-based exercises at the beginning of the course and vocalisation at the final plea (-.266*).

*Results: focus group*

The results of the focus group can be summarised as follows.

[97] Furthermore, a significant negative correlation was found between reflective thinking resulting from the virtual reality exercises and adequately reacting at the final plea (-.288**).

i. *It was difficult to assess the added value of our virtual reality based exercises*

Teachers received many positive responses from students. Teachers thought that the virtual-reality-based exercises only contributed to students' competences if they were given points to work on. They however reported that extra practice helps, but that for a genuine contribution to students' competences, investment in feedback is necessary. The role of the virtual-reality-based exercises also depend on the period remaining before the final pleading session, for its role (e.g., general repetition), but also of extra stress afterwards in the event students forgot their lines.

ii. *The contribution of the virtual-reality-based exercises to stress and/or self-confidence varied*

The teachers' impression was that a positive contribution to self-confidence could be witnessed in the situation in which they, together with students, gave feedback and the feedback was discussed with them. The virtual-reality-based exercises can be considered an extra practice moment, which contributes to self-confidence. Most students left the room with a positive vibe, some of them however felt nervous. Teachers could not conclude if students had greater confidence after performing virtual-reality-based exercises.

iii. *For teachers, the influence and role of reflection was unclear*

The role of reflection depended on how seriously students gave feedback. If students knew each other well they were often very positive and less critical.

iv. *Added value: mainly extra practice time with feedback*

Teachers have the impression that added value resulted particularly after extra practice time with feedback, not so much just by using virtual reality. According to one of the teachers, if students had learning points written down beforehand, and added notes during the session, feedback could be given on these points.[98] The added value of the virtual-reality-based exercises is an extra practice moment and the possibility to look into the room. If you let students plead longer, e.g. ten minutes instead of five, they will be involved more deeply which provides added value. One group of students had the possibility to

[98] One of the teachers was very critical, and thought the same result could have been achieved without virtual glasses. In order to study this statement a different research design is needed.

practise again after the feedback, which was very useful, also taking into account that the first time they wear virtual reality glasses they might feel uncomfortable. After the first time they were used to them and less distracted. Upon further questioning, it became apparent that teachers can say little about the effects of our virtual-reality-based exercises. They emphasize that it is crucial to focus on feedback and learning points.

## Discussion

VR (or, broader: XR) might be considered an exciting new way to enhance learning. It has great potential due to its possibilities of enhancing learning through immersion, through an authentic learning experience and environment. For law students, a recent Dutch innovation created a unique opportunity for students to practise their pleading skills and to improve their performance in the courtroom. The current contribution studied whether the virtual reality-based exercises (positively) influence student experience and student learning. In this final part, we will answer the research questions, summarise the results and their implications, discuss the answers to the hypothesis, the limitations of the current study, and offer some suggestions for further study.

In this study, we were interested in the effect pleading in a virtual-reality-based and authentic task has on student experience and (perceived) learning. First, we explored whether physiological changes ('stress levels') confirmed the authenticity of our virtual-reality based exercises. Our simulation, i.e., a virtual courtroom, can be considered an authentic environment since it looks exactly like the Court of first instance in Groningen, being filmed at 360 degrees. Therefore, it has high physical, as well as high psychological fidelity, as more or less the same actions are performed as in a real-life environment.[99] Stress is relevant for pleading – a form of presentation often considered to be stressful – and is related to psychological fidelity.[100] Our results show that pleading in a virtual reality courtroom environment significantly affects heart rate and skin conductance responses, which indicates the realism of the task/simulation. Reacting to questions is less an issue, showing fewer changes (the EDA slightly increased, the HR even decreased). Mainly the step between rest and status of pleading drove up the heartbeat and the electrodermal activity. On the other hand, the degree in which the students responded to pleading in virtual reality and the extent to which a response occurs is difficult to express unequivocally.

[99] Van Merriënboer & Kirschner, *Ten Steps to Complex Learning*, n. 32, p. 58-59.
[100] Cf Hays & Singer, Simulation Fidelity in Training System Design. n. 35.

As a follow-up, further study is needed to categorize and to mediate effects. While the virtual reality exercise offers a personalized 'immersive' experience for each student based on a real case file, it lacks the major consequences failure may have in legal reality. Exactly because of this fact it is valuable. The task's authenticity is particularly evident due to the large variability in physiological measures. In order to better understand the degree of simulation fidelity[101] a comparative-physiological study is needed between the simulation virtual court room and pleading in an actual court room. That might be difficult, but is not impossible to realise.

In order to answer the research question, we started from the well-known idea that authentic learning positively influences learners' success, motivation, attitude, and self-directed learning skills. We operationalised student learning in our 'authentic' virtual-reality-based environment and student experiences connected to it, by focusing on four mechanisms: the value/usefulness of the virtual reality exercises, students' perceived competence in pleading, students' confidence in pleading, and, last but not least, the role of peer feedback and students' reflective thinking in pleading. The results show an increase of the usefulness of virtual reality exercises and an increase in reported confidence, in the sense that the increase was higher than expected at the beginning of the course, when asked after the virtual reality exercises, but decreased in the perception of the students at the end of the course. We see an increase in perceived competence during the course, as shown after the virtual reality exercises as well as (again) after the final pleas. There was a clear decrease in the value of feedback, when construed and/or compared by students at the moment after their final plea. We can conclude that in the perception of students pleading in our virtual reality court room increased their learning levels. Furthermore, their experience was even better than expected at the outset of the course. However, some perceptions changed or were reinterpreted, after they faced a 'real' experience in front of real practitioners, but not their perceived competence which even continued to increase. Also, the feedback from professionals, at the end of the course, let them re-evaluate peer feedback after our virtual reality exercises in a different (less positive) light. Still, the value of feedback was evaluated as quite good, which is positive, certainly considering the following correlations.

[101] Comp. ibid., p. 50.

Our first hypothesis was that practising pleading in a virtual, authentic, courtroom is a useful and valuable experience for students. This proved to be the case: reflective thinking related to/following after the virtual reality exercises was significantly positively correlated with the value attributed to pleading in virtual reality at the start of the course, after the virtual reality exercises and at the end of the course. Our second hypothesis was that pleading experience in virtual reality increases perceived competence and promotes confidence. We found significant positive correlation between reflective thinking due to the virtual reality exercises and confidence at the end of the course, not at beginning of the course or after the virtual reality exercises. No correlation was found between reflective thinking due to the virtual reality exercises and perceived competence at the beginning, after the virtual reality exercises or at end of the course. Nevertheless, both constructs confidence and perceived competence were found to be significantly correlated at all moments, to all moments. The third hypothesis was that practising pleading also enhances reflective thinking and self-regulation skills provided effective peer feedback is given. We found significant positive correlations between reflective thinking due to the virtual reality exercises, peer feedback received on their pleading in virtual reality exercises, and peer feedback received on their pleading in virtual reality when judged both at the beginning of the course, after the virtual reality exercises as well as at the end of the course. In fact, it was reflective thinking connected with virtual reality exercises that linked these constructs together in H1, H2 and H3.

From correlations between variables mentioned above and the final assessments, we can deduce that confidence measured directly after the virtual reality exercises was positively correlated with adequately reacting at the final pleas, between perceived competence directly after the virtual reality exercises and contact with judges at the final plea and with the grade on vocalisation at the final plea. Confidence in pleading, promoted by our virtual reality exercises, was thus correlated with good objective results. Surprisingly, we found significant negative correlation between the value attributed to virtual-reality-based exercises at the beginning of the course, and vocalisation at the final plea. Possibly, though this is only one possible interpretation, these honours students were quite negative due to their already high-level of vocal performance.

In the focus group, teachers suggested that it was mainly feedback and reflective thinking (‘which points to work on’) that determined the added value

of virtual-reality-based exercises. The increase in self-confidence, as indicated in the questionnaires, finds support in teachers' responses ('if feedback was given and discussed with them'). This further underscores the pivotal role of reflective thinking. Of course, the quality of feedback 'in the traditional sense' is crucial. Where feedback is considered in our definition, it also involves and/or stimulates reflective thinking. A contradiction that could be deduced from the focus group meeting is that it is about either using virtual reality or (extra time with) giving feedback. Without an authentic environment, necessarily in a simulation setting (for the reasons given above), feedback on a plea cannot be given. In that sense virtual reality provides the context that allows feedback. So, there is no contradiction. Exactly the illusion or realism part of VR, provides added value. Further research therefore is needed to be able to use the opportunities virtual reality can provide us. Moreover, it is crucial to get the most out of feedback and to work on (self-determined) learning goals, also in the event of exercises in virtual reality. For experiential learning, besides personal experiences, self-evaluation of experiences and individual experimenting, and reflection, the influence of others is important (teachers, other students); social context influences learning. Therefore, interaction remains important.[102] Furthermore, the profits from (critical) reflection could be improved by letting students, after their experience that is the catalyst that ignites the reflection process, find meaning by connecting to theories and to expectations.[103] Besides, critical reflection could be improved by additional, regular, visits to courts.[104]

In this study, we measured heart rates and electrodermal activity. Elevated stress levels can however impede performance on tasks that require divided attention, working memory, retrieval of information from memory, and decision making.[105] Stress might therefore be connected with the virtual-

[102] See V. Schutjens, *Experiential Education in Geography. Meer leren buiten je comfortzone* (inaugural lecture Utrecht University), Utrecht 2023, p. 20-21, with reference to L.S. Vygotsky & M. Cole, *Mind in Society: Development in Higher Psychological Processes*, Cambridge (MA) 1978.

[103] The framework provided by M. Fraser, A. Wotring, C.A. Green & M.J. Eady, 'Designing a framework to improve critical reflection writing in teacher education using action research', *Educational Action Research* 2022, could be a useful framework for students – which will then have to be, written - reflections.

[104] There, students can not only observe cases but also witness how lawyers and judges behave and act. This enables them to engage in critical reflection based on theory, experience, and expectations.

[105] V.R. LeBlanc, 'The Effects of Acute Stress on Performance: Implications for Health Professions', *Education, Academic Medicine* 84 (2009), 10, p. 25-33.

reality-based exercises, bringing negative effects on self-confidence. This, too, will have to be studied further. Unfortunately, that was not possible in this study. Physiological data could not be related to the other data. Furthermore, study of heart rate variability when pleading could show when students are more and less stressed. However, physiological responses and changes in general are of interest, because if arousal can be reduced, it can lead to improved performance. Also, extra practice might lead to enhanced self-competence and a better performance. That would fit a previous study which concluded that stress management workshops had a positive effect and can lead to significant short-term improvement in stress and burnout test scores.[106] Supporting students' stress resilience when pleading in court is not only important for better (academic) performance and results but also for students' well-being and resilience to stress. This offers good starting points for further study.[107]

## Acknowledgments

This study is financed by the Dutch Ministry of Health, Welfare and Sport Under Grant No. OO22-04 ('Leren met Virtual Reality: The Next Level'). The author thanks the work packages leaders from this project, who gave feedback on an earlier draft of this version, and, in general, the colleagues from this university-transcending innovation project with whom he pleasantly worked together. The pre-study validation was made possible by a SOTL-Grant (2021-2022) by Utrecht University. He also thanks dr. Monika Donker for her guidance in using physiological instruments, dr. Frans Prins for his educational guidance, Sharon Padt MSc for her research assistance, and the anonymous peer reviewers for their comments.

[106] J.D. McCue & C.L. Sachs, 'A Stress Management Workshop improves Residents' Coping Skills', *Archives of Internal Medicine* 151 (1991), p. 2273–2277.
[107] This contribution did not touch upon the role of gender dimension. A final suggestion for further research is to study the importance of gender, i.e. the (possible) differences of gender for both the issue of stress and for pleading (skills) in virtual reality.