



**HAL**  
open science

## Virtual reality: A new method to investigate cognitive load during navigation

A. Armougum, E. Orriols, A. Gaston-Bellegarde, C. Joie-La Marle, P. Piolino

### ► To cite this version:

A. Armougum, E. Orriols, A. Gaston-Bellegarde, C. Joie-La Marle, P. Piolino. Virtual reality: A new method to investigate cognitive load during navigation. *Journal of Environmental Psychology*, 2019, 65, pp.101338 -. 10.1016/j.jenvp.2019.101338 . hal-03487259

**HAL Id: hal-03487259**

**<https://hal.science/hal-03487259>**

Submitted on 20 Jul 2022

**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 Attribution - NonCommercial 4.0 International License

## **VIRTUAL REALITY: A NEW METHOD TO INVESTIGATE COGNITIVE LOAD DURING NAVIGATION**

A.ARMOUNGUM<sup>1,2,4</sup>, E.ORRIOLS<sup>1,2</sup>, A.GASTON-BELLEGRARDE<sup>1,2</sup>, C.JOIE-LA MARLE<sup>4</sup> AND P.PIOLINO<sup>1,2,3</sup>

<sup>1</sup>*Memory, Brain and Cognition Lab, Institute of Psychology, Paris Descartes University, Sorbonne Paris Cité, Boulogne-Billancourt, France;*

<sup>2</sup>*INSERM, UMR S894, Centre for Psychiatry and Neurosciences, Paris, France*

<sup>3</sup>*Institut Universitaire de France (IUF), Paris, France*

<sup>4</sup>*SNCF Mobilités, Transilien, SID, Lab Mass transit, Paris, France*

1 **ABSTRACT** *Cognitive load has for long been studied in relation with learning processes. In our*  
2 *study, we investigated the impact of cognitive load in real-life situation taking the example of train*  
3 *travelers looking for relevant information in a train station. For this purpose, we created a virtual*  
4 *reality model of the tested train station from which we conducted a real-life study. Our aim was to*  
5 *compare travelers' cognitive load impact in real-life environmental condition versus virtual reality*  
6 *simulation of the same environment. Regular and occasional travelers were recruited and were*  
7 *assumed as experts and novices, respectively. Investigation of cognitive load was based on*  
8 *physiological, subjective and behavioral aspects. These were measured using electrodermal activity,*  
9 *NASA-Task Load Index and recognition of relevant factual and contextual information seen by*  
10 *travelers in the train station, respectively. These three indicators used in both real-life and virtual*  
11 *reality, were expected to follow the same trend, irrespective of the environmental condition, but were*  
12 *expected to vary with respect to expertise level. A higher cognitive load was forecasted for novice*  
13 *travelers than for expert travelers. The findings revealed no difference on the three indicators*  
14 *between virtual and real-life conditions. However, novice travelers showed higher cognitive load*  
15 *responses than expert travelers, in both environmental conditions. Our results suggest virtual reality*  
16 *as a promising technique for cognitive load analysis during navigation, and an effective method for*  
17 *neurocognitive assessments in daily life situations.*

18 Keywords: Cognitive load; virtual reality; memory; electrodermal activity; train travelers;  
19 expertise

## 20 **1. Introduction**

21 “*I move, therefore I am*” said the Japanese writer Haruki Murakami (2009). Human beings  
22 are in constant movement for living, working and discovering the world. We rely on  
23 information all around us to understand and control our environment. This is the case,  
24 whenever we are in a train station seeking for the right information to reach our destination on  
25 time. During this process of information acquisition, cognitive load is generated. This  
26 phenomenon corresponds to the total amount of resources invested in learning process  
27 (Sweller, 1988). From this definition, cognitive load can be hypothetically represented as a  
28 reservoir of cognitive resources available to perform multiple tasks. It is thus necessary to  
29 optimize its level, while avoiding saturation or overload of our cognitive resources (Kirsh,  
30 2000). This is usually the case, for instance, by the end of a big day of work, when we are  
31 sometimes unable to process a cognitively low-demanding task. To be able to use strategically  
32 our limited cognitive resources, we first need to understand the underlying theory: Cognitive  
33 Load Theory (CLT) (Sweller, 1988; Sweller, Van Merriënboer & Paas, 1998).

34 CLT does not have its own model *per se*, but instead is constructed from another validated  
35 model, working memory model (Baddeley, 1986). CLT is completed with studies conducted  
36 on attentional resources to account for limitations in cognitive resources and individual  
37 differences (Chanquoy, Tricot & Sweller, 2007). Working memory resources are available in  
38 three interdepending pools: *intrinsic cognitive load* (Sweller, 1994) which depends on the  
39 structure of the task itself, *extraneous cognitive load* (Sweller and Chandler, 1991) which  
40 arises whenever cognitive resources are allocated to working memory for operations that are  
41 irrelevant to the task, and *Germane cognitive load* (Sweller et al., 1998) characterized as the  
42 learning process itself. It is related to schema creation, automatization and procedural thinking  
43 and action (Chanquoy et al., 2007). Therefore, to avoid cognitive overload, we should  
44 optimize intrinsic cognitive load, reduce extraneous cognitive load and maximize Germane  
45 cognitive load. Several authors have worked on cognitive load optimization, especially in the  
46 field of educational psychology (for examples in visual representations in science education:  
47 Cook, 2006; in statistics: Paas, 1992; in problem solving: Sweller, 1988; in health education:  
48 Van Merriënboer & Sweller, 2010). These studies have put in light other non-cognitive  
49 factors affecting our cognitive resources that need to be considered when conducting studies  
50 on cognitive load. A few examples of these factors are: emotional state, fatigue level,  
51 vigilance level, expertise level, to name a few (see Chanquoy et al., 2007 for review). These  
52 non-cognitive factors together with cognitive factors, such as memory and attention, form the  
53 mental load (Jex, 1998). These phenomena suggest the use of different types of measure for  
54 quantifying cognitive load. The combination of three main types of cognitive load  
55 measurement techniques are mostly used in cognitive load investigation (Cain, 2007; Cegarra  
56 & Chevalier, 2008; Kramer, 1991). The first one corresponds to *physiological measures*  
57 (Paas, Renkl & Sweller, 2003; Sweller, Ayres & Kalyuga, 2011) such as  
58 electroencephalography (Gevins et al., 1998; Ke et al., 2014), Heart rate: Paas & van  
59 Merriënboer, 1994), and more recently electrodermal activity (EDA) (Shi, Ruiz, Taib, Choi,  
60 & Chen, 2007). It measures skin conductance response (former *galvanic skin response*),  
61 which is known to rise during cognitive effort, providing an indirect measure of cognitive  
62 load (Chen et al., 2016). EDA has also been investigated in relation to other physiological  
63 factors that may interfere with cognitive load, such as stress level (Setz, Arnrich, Schumm, La  
64 Marca, Tröster, & Ehlert, 2009). EDA reflects generalized changes in the state of arousal.  
65 These can be caused by emotional, cognitive or physical stimulation (Picard, 2015). EDA  
66 levels have also been shown to vary with different cognitive load levels (e.g. in real-world  
67 driving tasks: Healey and Picard, 2005). A second type of cognitive load measure is related to

68 *subjective measures* (Vidulich & Tsang, 2012). Two main subjective scales used in the  
69 literature are the NASA-Task Load Index (Hart & Staveland, 1988) and the Paas scale (Paas,  
70 1992). A third type of measure recommended in experiment conducted in cognitive load  
71 theory is *behavioral measures* through performance analysis (Paas & Van Merriënboer, 1994;  
72 Brunken, Plass, & Leutner 2003). The measuring techniques are usually chosen according to  
73 the experimental context. So far, most of these studies have investigated cognitive load in a  
74 static context, such as performing a cognitive task on a computer. This static configuration,  
75 however, does not reflect our everyday life activities, where we learn about our environment  
76 by moving around and exploring it. This raises the question on how our cognitive load varies  
77 during navigation in a non-static ecological context.

78 We are obviously limited in available techniques to investigate cognitive load in a non-static  
79 condition. This is due to the fact that we cannot create the same scenario with same external  
80 real-life factors, on and on again for a statistically reliable sample size. One promising  
81 method, which has proven its efficacy for reproducing real-life environments, is virtual reality  
82 (VR) technology (Koutsoudis, Arnaoutoglou & Chamzas, 2007; Van Veen, Distler & Braun,  
83 1998; Whyte, Bouchlaghem & Thorpe, 2000).

84 VR technology has bloomed since the first appearance of 3D representations in 1838 with the  
85 use of stereoscope. Today numerous virtual headsets capable of displaying high 3D quality  
86 images have been developed (e.g. HTC vive, Oculus rift, Tobii) and are used in neuroscience  
87 and psychology research, to investigate healthy adults (e.g. virtual visit of museum:  
88 Carrozzino & Bergamasco, 2010; evolution of episodic memory in normal ageing: Plancher,  
89 Gyselinck, Nicolas & Piolino, 2010; influence of action on episodic memory: Plancher, Barra,  
90 Orriols & Piolino, 2013; enactment and episodic memory among adults: Jebara, Orriols,  
91 Zaoui, Berthoz & Piolino, 2014; working memory and episodic memory: Plancher, Gyselinck  
92 & Piolino 2018) as well as patients (e.g. mild cognitive impaired patients and Alzheimer  
93 patients: Plancher, Tirard, Gyselinck, Nicolas & Piolino, 2012; psychotherapy: Riva, 2005).  
94 In these fields of research, VR is commonly used to help researchers design new fictional and  
95 ecological scenarios that allow them to control random variables that are usually difficult or  
96 impossible to manipulate in real-life. VR also helps to create complex real-life environments  
97 or structures that represent an effective method for neurocognitive assessments of cognitive  
98 load in daily life situations. This technology is also compatible with other interfaces, for  
99 instance with a crowd simulator, to observe how an individual's reaction depend on different

100 crowd densities, if immersed in a virtual environment (VE) (Musse & Thalmann, 2001;  
101 Ulciny & Thalmann, 2001).

102 One important aspect of using VR technology is to generate self-experience in individuals,  
103 bodily immersion and sensation, which were believed to arise solely in real-life condition  
104 (Murray & Gordon, 2001). Studies that succeeded to recreate a VE closed to real-life  
105 condition, managed to generate same self-experience in individuals. For instance, in a study  
106 conducted in VR in the subway of London which aimed to show that healthy participants  
107 experienced unfounded paranoid thoughts, had put in light that the same degree of paranoid  
108 reactions were observed in VR environment as in real-life situations. This proved VR to be a  
109 potent tool for the understanding of paranoid reactions (Freeman et al., 2008). In another  
110 study, social phobia therapy had been investigated in VR. The efficacy of this technology was  
111 shown to be significantly closed to classical cognitive behavior therapy (Klinger et al., 2005).  
112 More recently, in a study on mindfulness practice and meditation, psychological and  
113 physiological responses to stressful situations had been studied in an immersive VR  
114 environment. An increase in mindfulness and anxiety levels had been shown through the use  
115 of realistic life situations designed in a virtual reality context (Crescentini, Chittaro, Capurso,  
116 Sioni & Fabbro, 2016). This immersive level is closely related to the optimization of the VE  
117 and is measured through subjective sense of presence of the participant in the VE (Schubert,  
118 Friedmann & Regenbrecht, 2001). If the VE is sufficiently vivid, extensive and inclusive  
119 (Slater & Wilbur, 1997), and as close as possible to the real world (Barfield & Hendrix, 1995;  
120 Bystrom, Barfield, & Hendrix, 1999), the higher will be the sense of presence (Schubert, et  
121 al., 2001).

122 Studies conducted in navigation and more precisely wayfinding in the field of spatial  
123 cognition, have greatly benefits from the use of VR (Chen & Stanney, 1999; Darken &  
124 Peterson, 2014; Darken & Sibert, 1996; Lee & Wong, 2015; Steinicke, Visell, Campos &  
125 Lécuyer, 2013; Viaud-Delmon, Ivanenko, Berthoz, & Jouvent, 1998). Besides, VR has been  
126 used to help reduce cognitive load associated to complex surgical procedures in  
127 otorhinolaryngology, by providing virtual training sessions to surgeons (Andersen, Konge &  
128 Sørensen, 2018). To this date, not much has been done on the relationship between navigation  
129 and cognitive load analysis in virtual reality. This implies that before investigating complex  
130 situations in VR, we first need to validate its efficacy for cognitive load analysis during  
131 navigation.

132 In the present study, we aimed to set-up two experiments. The first one was conducted in real-  
133 life condition, in an actual train station, and the second one was carried out in the VR version  
134 of the same tested train station. To reduce the effects of any covarying factors, we tested for  
135 known effects on cognitive load in our experiment. We evaluated the difference in expertise  
136 level of travelers: regular travelers as experts and occasional travelers as novices. This  
137 assumption was based on the fact that expertise improves mnemonic performance (for review:  
138 expertise of taxi drivers for street names, Kalakoski & Saariluoma, 2001; Memory of drawing  
139 experts, Perdreau & Cavanagh, 2015; Chess masters' performance compared to novices',  
140 Reingold, Charness, Pomplun & Stampe, 2001). We predicted occasional travelers to show  
141 higher physiological activity than regular travelers, during information processing. Higher  
142 self-perceived mental load were also expected for occasional travelers. We also hypothesized  
143 regular travelers to demonstrate a higher mnemonic performance for relevant information on  
144 signage boards and information display screens, than occasional travelers, for factual  
145 information and their corresponding contextual specifications. These known cognitive  
146 differences between experts and novices should be statistically the same for both real-life and  
147 virtual reality experiments.

148

## 149 **2. Method**

150 The real-life experiment was performed in a Parisian train station, Saint-Michel Notre  
151 Dame, train C. The virtual reality experiment was conducted in our University laboratory, in a  
152 fully equipped room.

### 153 2.1 Participants

154 112 participants of age 26( $\pm$ 5) years old, constituting a stratified sample of 50% male,  
155 were selected for the study, from 121 participants recruited in total. 56 participants were  
156 randomly assigned to the real-life condition, and the other 56 participants randomly to the VR  
157 condition. Half of the participants in each of these two conditions were female. The  
158 recruitment for this study was made among undergraduate students from Paris Descartes  
159 University, from the Faculty of Psychology. These participants were rewarded additional  
160 marks for one course module. Other undergraduate participants were recruited online from a  
161 site dedicated to cognitive science studies and they were rewarded too. A second group of  
162 participants was recruited directly from the investigated train station.

163 Participants' use of video games was monitored, since virtual reality has a common interface  
 164 to that of video games, which could be subjected to expertise for regular gamers. Our sample  
 165 consisted of participants who played video games at most once a month.

166 2.1.1 Expertise level

167 Participants' selection was also based on their expertise level of train station and  
 168 particularly for the investigated one, using a survey. The latter informed us about the  
 169 frequency of visits of Saint-Michel Notre Dame train station. This category contained three  
 170 questions: question one investigated the frequency of utilization of Saint-Michel Notre Dame  
 171 for line C, ranging from *every day* to *one to two times a year*, on a five segmented Likert  
 172 scale. The second question dealt with the frequency of utilization of Saint-Michel Notre Dame  
 173 globally (for all trains' lines except for line C), using the same scale as for the previous  
 174 question. Information such as exit names, train's map, visual and spoken information, to name  
 175 a few, were common among the different lines in the investigated train station. Thus,  
 176 passengers could develop indirect knowledge of the investigated line that needed to be  
 177 accounted for. The third question was based on utilization of SNCF train station around Paris  
 178 and suburb (apart from Saint-Michel Notre Dame), with the same five Likert scale. Some  
 179 features could also be common among different train stations from the same train operator,  
 180 such as the charter used on information displays and signage boards, but also the instructional  
 181 design used to present information. A threshold score was defined from these three questions  
 182 to distinguish regular travelers (experts) from occasional travelers (novices). The scale used  
 183 and the threshold estimated for participants' categorization in different expertise groups, are  
 184 illustrated in table 1 below.

185

5	4	3	2	1	0
Everyday	At least 3 times a week	Between one and two times a week	Once or twice a month	Once or twice a year	Not at all

186

187 *Table.1. Categorization of travelers' sample*

188

189 *Threshold expert: Participant needs to obtain a score of 5, 4 or 3 in at least two of the three*  
 190 *questions described above. One of which should be question one. The sum*  
 191 *of answers for the three questions should be between 7 and 15 inclusive)*

192

193 *Threshold novice: Sum of answers for the three questions should be between 3 and 6 inclusive*

194



195 We chose these questions for our expertise survey since we expected to evaluate memory as  
 196 an indicator of performance in cognitive load. Besides, according to our experimental  
 197 protocol, participants had to find particular landmark and region of the train station that  
 198 should be assessed rapidly if they had a prior knowledge of the train station, or schemas from  
 199 other train stations of the same train operator. In our experimental sample, regular travelers  
 200 were assimilated as *experts*, since they take suburban train every day to go to work or  
 201 University, in Paris central. Besides, most of them have been living for their whole life in the  
 202 same region of the suburb, and all of them have been taking the same train for more than 10  
 203 years, practically every day. This period of time corresponds to the estimated time required  
 204 for a chess player beginner to reach expert level (Simon and Gilmarin, 1973). On the other  
 205 hand, our novice sample consisted mostly of people living in Paris center and taking the  
 206 investigated train station for at most once or twice a month, though most of them claimed to  
 207 take the train between once or twice a year.

208 Once the participants filled in our short survey for evaluation of their expertise level, they  
 209 were assigned to 1 of 4 different groups of investigated profile (male novice, male expert,  
 210 female novice, female expert). We continued our recruitment until we obtained 112  
 211 participants, with 28 participants per group. We assumed that the three items of the survey  
 212 were sufficient for an expertise assessment of our 112 randomly distributed in the condition  
 213 factor (*real-life vs virtual reality*). In order to test statistically if these items were good  
 214 predictors of expertise, we conducted a multiple linear regression test. We used the following  
 215 model for our regression and results of this analysis are presented in table 2 below.

$$216 \quad Expertise_i = \beta_0 + \beta_1 * Investigation\ train\ line_i + \beta_2 * Investigation\ train\ station_i + \beta_3 * Train$$

$$217 \quad stations\ of\ the\ same\ train\ operator_i + \epsilon_i$$

Variable	Mean ±SD for Novice	Mean ±SD for Expert	Estimate	Std Error	t- value	p	CI 2.5%	CI 97.5%
Utilization of investigated line	0.91±0.84	3.98±0.80	.101	.015	6.89	<.001***	.072	.130
Utilization of whole train station	0.95±0.82	4.07±0.81	.091	.015	5.91	<.001***	.060	.121
Utilization of other train station of the same train operator	1.05±0.77	4.07±0.81	.106	.015	7.03	<.001***	.076	.136

218 *Table.2. Mean scores and Regression analyses of final 112 participants*

219

220 This gave the following regression model:

$$221 \quad Expertise_i = -0.746 + .101 * Investigation\ train\ line_i + .091 * Investigation\ train\ station_i + \\ 222 \quad .106 * Train\ stations\ of\ the\ same\ train\ operator_i + .15$$

223 Results showed that the three investigated items are responsible for the prediction of  
224 expertise,  $F(3,108) = 378.8$ ,  $p < .001$ , and account for 91.1% of the prediction. The associated  
225 residual error is .15. Analyses of t-values indicate that each variable contributed significantly  
226 to the prediction of expertise ( $p < .001$ ).

### 227 2.1.2 Cognitive neurological assessments

228 To complete the inclusion process, we conducted a survey to make sure that our  
229 112 participants were not under medication. We also made sure that they were not suffering  
230 from any particular neurological or psychiatric problems, and presented normal basic  
231 cognitive skills, by completing a set of neuropsychological tests at the beginning of the  
232 experiment. These tests comprised the short version of the Beck Depression Inventory (BDI),  
233 used to evaluate depression (Collet & Cottraux, 1986; Beck, Steer & Brown 1996), the  
234 revised Piper Fatigue scale, used to evaluate fatigue (Piper, Dibble, Dodd, & Weiss 1988),  
235 State-Trait Anxiety Inventory (STAI Y-A) used to evaluate state anxiety level (Spielberger,  
236 1983), the Prospective and Retrospective Memory Questionnaire (PRMQ) used to evaluate  
237 subjective prospective and retrospective memory skills (PM and RM respectively) in daily life  
238 (Smith, Del Sala, Logie & Maylor, 2000) and the ZOO map test, used to evaluate planning  
239 skills (Wilson, Evans, Emslie, Alderman & Burgess, 1998). These neurological tests served as  
240 inclusion criteria for participants with no depression, low fatigue, low anxiety, normal mental  
241 planning level and no apparent subjective complaints of prospective and retrospective  
242 memory. Six participants were not kept for the final experiment after these analyses. We thus  
243 replaced them after investigation of 15 additional profiles. A 2-way MANOVA (context vs  
244 profile) conducted on these different variables gave a non-significant effect on *context* (VR vs  
245 real-life),  $F(3,108) = 1.58$ ,  $p = .15$ ,  $\eta^2 = .10$ , and no significant profile effect (regular vs  
246 occasional),  $F(3,108) = 0.87$ ,  $p = .06$ ,  $\eta^2 = .06$ . No interaction was noted between context and  
247 profile factors,  $F(3,108) = 1.58$ ,  $p = .15$ ,  $\eta^2 = .10$ . These non-significant differences between  
248 contexts and profiles allowed us to check that no bias could intervene as secondary factor  
249 before investigating cognitive load. Table 3 below illustrates the mean score, the lowest and  
250 the highest scores of our sample.

	<b>BDI</b>	<b>Fatigue scale</b>	<b>STAI Y-A</b>	<b>STAI Y-B</b>	<b>ZOO map test</b>	<b>PM</b>	<b>RM</b>
<b>Mean score</b>	0.99	2.54	33.7	38.6	161	18.4	8.49
<b>Lowest value</b>	0	0.75	26	34	0	11	9
<b>Highest value</b>	4	4.33	45	45	338	30	27
<b>Threshold values</b>	No depression : 0-4	Low fatigue level : 0-4.5	Low anxiety : < 39	Low anxiety : < 39	±2 standard deviations from mean value		

Table.3. Table of scores to neuropsychological tests of final 112 participants

## 2.2 Experimental environments

### 2.2.1 Real-life environment

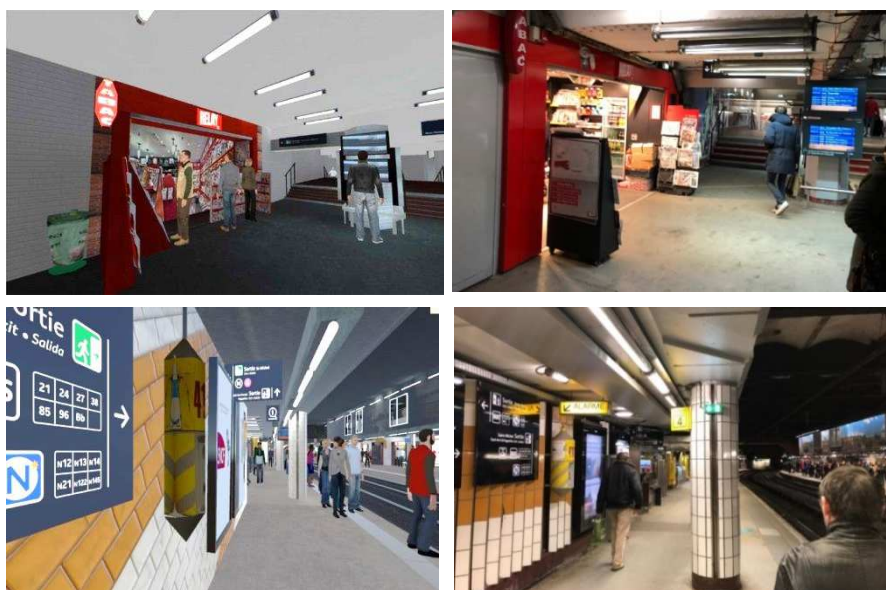
The selected environment was Saint-Michel Notre Dame, train station. This train station accommodates nearly 100K travelers each day. This phenomenon represents a real case of *mass transit*, characterized by the continuous movement of thousands of people in space and time, which in this case is the limited space of the train station. The highest level in passengers' flow can be reached during peak hours, usually at 8 a.m. and 5 p.m.; before and after office hours.

### 2.2.2 Virtual reality environment

In the laboratory, a VR multimodal environment was designed to simulate different everyday life scenarios at Saint-Michel Notre Dame train station, based on observations made in the actual train station. In order to create a virtual environment that could be used as a replacement of a real-life condition, according to Fuchs and Moreau (2006), several indications had to be followed: immersion, interaction, and interface. The immersive level of the VE puts in action the participant's senses, and mainly sight, hearing and touch. This was done by using the Head Mounted Display of HTC Vive. The interactive aspect depends on the relation with other avatar(s) and objects in the virtual train station. This represents a key procedure in the development of an ecological VE. The interface used, should be easily understood and ergonomic (Fuchs, Moreau, Guitton, 2011). Pictures taken from the actual train station, together with textures available on Unity® asset store, that matched objects from the train station, were used to create a realistic interface. Participants navigated in the VE using a controller in each hand, by moving them up and down. During the conception of the

275 VE, multiple tests were conducted to make sure that physical effort invested by participants  
276 while moving the controllers, were proportional to the displacement of the participant's avatar  
277 in the VE. Other prior tests were conducted on angular displacement of the headset with slow  
278 and quick movement, so as to avoid a mismatch between the visual characteristics of the  
279 virtual scenes and what it would have been in the real world. These tests allowed us to give an  
280 ecological aspect to our VE and to reduce the risk of cybersickness and more precisely motion  
281 sickness (Bruck & Watters, 2011). Though we optimized as much as possible the VE, the risk  
282 of cybersickness could not be totally removed. This side effect of virtual technologies,  
283 characterized by nausea and head ache, was explained to the participants and the latter were  
284 asked to let the experimenter know if any of cybersickness symptoms arose, in order to stop  
285 the experiment. During our experiment, 3 participants suffered slightly from cybersickness  
286 and their session was cancelled instantaneously. Necessary assistance was provided to these  
287 participants and we made sure that they felt healthy again before leaving. To balance our  
288 sample three other participants with same profile were later recruited.

289 For the purpose of our experiment we created a VE from the architectural blueprints of Saint-  
290 Michel Notre Dame train C. We used the 3D software Unity3D® to design the train station.  
291 We applied the rules of interaction, immersion and interface (Fuchs & Moreau, 2006) to  
292 design our 3D environment. The detailed procedures for the construction of our virtual train  
293 station are illustrated in the Appendix section of this paper. Some pictures obtained from the  
294 virtual train station (left side) are compared to pictures taken from the same region in the real  
295 train station (right side), as shown in figure 1 below.



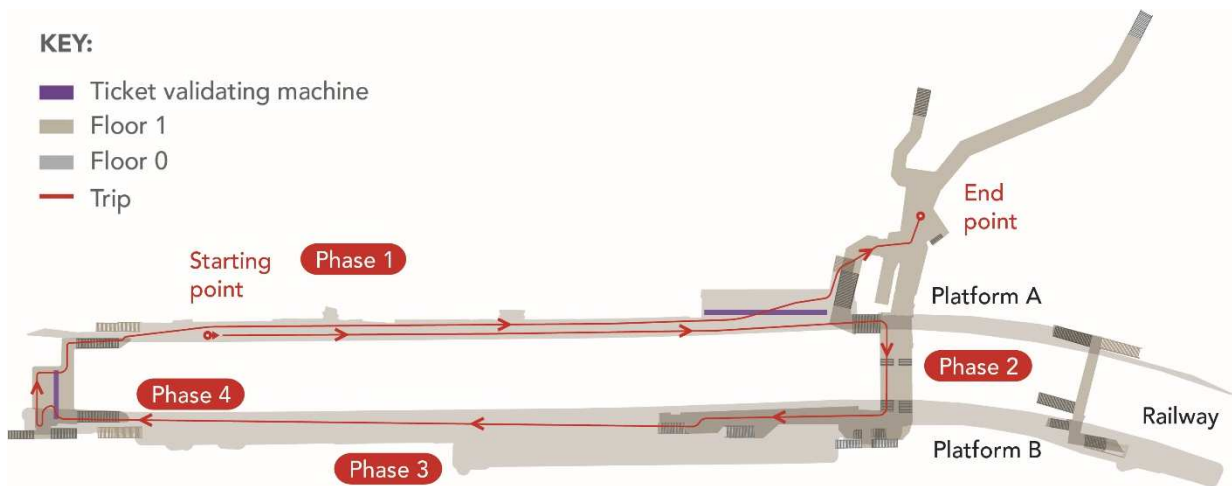
303 *Figure.1. Virtual vs real life train station of Saint-Michel Notre Dame, train C*

304 The overall purpose of the experiment was clearly explained to the participants, without  
305 giving information that could bias the experiment. The participants agreed to the terms by  
306 signing an informed consent form, explaining all possible health related effects due to our  
307 experimental setup and material. Ethical clearance was granted before running the  
308 experiment. Every possible risk associated to the use of virtual reality was explained, such as  
309 the limit of movement in the experimental room while wearing the virtual reality headset, to  
310 avoid knocking against wall was prevented using a virtual limitation shown in the virtual  
311 environment itself. The participants were volunteers and could stop the experiment at any  
312 moment.

### 313 2.3 Navigation path

314 During the study, participants taking part in the virtual experiment were required to  
315 navigate in the train station using a head mounted VR headset HTC Vive to be immersed in  
316 the VE. Walking action was simulated using HTC controllers, by moving them up and down.  
317 The path was the same for both experimental conditions (real-life and virtual). The trip of the  
318 participant is illustrated in figure 2 below.

319



320

320 *Figure.2. Navigation path of travelers*

321

322 The starting point was on platform A and the participants had to move to the mezzanine on  
323 the upper floor, cross the bridge and find the stairs leading to platform B. Afterwards, they  
324 had to move on platform B to a specific exit, *Petit Pont Notre Dame*. From there, they needed  
325 to meet a virtual avatar who handed them a book. Eventually, the participants had to go back

326 to platform A, using the stairs and walked till the information display screen next to the ticket  
327 validating machines on platform A. During the experiment there was an expertise effect on  
328 the time taken by travelers to visit each section of the train station. The mean time taken by  
329 occasional travelers ( $M = 4.57 \pm .17$  min) was significantly higher than that of regular travelers  
330 ( $M = 3.74 \pm .04$  min) for each experimental phase ( $p < .001$ ). There was no environmental  
331 effect (real-life x virtual reality) ( $p > .18$ ).

## 332 2.4. Indication

333 Before the experiment, indications were given to the participants in a story telling  
334 format (for e.g. Bucher, 2017; Ibanez & Aylett, 2003; Shin, 2018), where they had to find the  
335 next move for their journey in the train station. There was no time limit to end the experiment,  
336 but the participants were nevertheless encouraged to perform as quickly as possible, as it  
337 would be the case in real-life. The indication was chosen so that the participants cover over  
338 90% of the train station, allowing us to evaluate the use of information in almost all the train  
339 station.

340 The indication was as follows: “*You are heading to Orsay museum where a group of friends*  
341 *are waiting for you. You have decided to take train C at Saint-Michel Notre Dame. You*  
342 *arrived at the train station by a connection leading to platform A. Before going to find your*  
343 *train to Orsay museum, you decided to buy a magazine at the Relay shop in the train station,*  
344 *before heading to the exit Petit Pont Notre Dame. There, a friend is waiting for you to hand*  
345 *you a book. Your task is to find the path to your final destination (Orsay museum), while*  
346 *accomplishing all the previous tasks.*”

## 347 2.5 Cognitive load measures and materials

348 Cognitive load was evaluated during and after the experiment. Three main methods  
349 were used: *physiological, subjective and behavioral.*

### 350 2.5.1 Physiological measure

351 Electrodermal activity (EDA) was chosen as a *physiological measure* during the  
352 experiment. Endosomatic EDA was measured from the inner wrist skin of the participants  
353 using a sensor which quantified this activity for later analyses. In both virtual and real-life  
354 conditions of our experiment, an E4 wristband sensor Empatica® was used to monitor the  
355 EDA level. This wristband contained two electrodes on the sensor’s band for EDA monitoring

356 and has a permanent internal memory of 60 hours, with a synchronization resolution of 5  
357 seconds. The data collected on the device can be streamed via bluetooth® on a smartphone.  
358 The sensor measured blood volume at a frequency of 64 Hz, with an electrodermal activity of  
359 4Hz, at an acceleration of 32Hz and a skin temperature measured every 4 seconds. EDA data  
360 were measured in micro Siemens ( $\mu\text{s}$ ) at a frequency of 4Hz every 0.25 second.

361 Variables known to impact EDA recordings were supervised according to recommendations  
362 from previous studies (Boucsein, 2012). Variations in temperature were measured during the  
363 experiment, since EDA level is affected by it. Room temperature for virtual reality experiment  
364 was monitored and maintained at 20°C using an air-conditioner, and the participants were  
365 asked to wear sport clothing, to be more comfortable and to prevent drastic rise in body  
366 temperature during the experiment due to sweating. The same condition was applied for real-  
367 life experimental condition, with the exception of the air-conditioned temperature. However,  
368 the average temperature of the air in both environments was  $20.3\pm 0.2^\circ\text{C}$  with a non-  
369 significant difference between the two environmental conditions ( $p = .76$ ). Data from  
370 electrodermal activity were collected at four different check points in the virtual environment.  
371 These were platform A, upper platform, platform B and the exit (Petit Pont Notre Dame). We  
372 expected regular travelers to show lower EDA score than occasional travelers. We expected  
373 no statistical difference in EDA score between the two investigated environmental conditions.

#### 374 2.5.2 Subjective measure

375 NASA- Task Load Index (NASA-TLX) (Hart & Staveland, 1988) was chosen to  
376 evaluate *subjective measure* during our experiment. This test gave an indication for self-  
377 perceived mental load by the participants. This measure is operationalized using weighted  
378 mean scores for 6 dimensions: mental demand, physical demand, temporal demand,  
379 performance, effort and frustration, perceived by the individual with respect to the task.

380 We expected occasional travelers to express higher subjective mental load than regular  
381 travelers. No significant difference was predicted between real-life and virtual experiments.

#### 382 2.5.3 Behavioral measure

383 In addition to the first two measures, a *behavioral measure* was conducted on the  
384 participants' performance by analyzing factual and contextual memory using a memory test,  
385 designed for this purpose. Our aim was to assess factual and contextual memory encoding

386 performance, i.e. featuring binding. Binding refers to link a stimulus to its contextual  
387 information (Sperduti, Armougum, Makowski, Blondé & Piolino, 2017). This test gave an  
388 indication of contextual information memorized by participants, linked to episodic memory  
389 (Johnson and Raye, 1981; Tulving, 2002).

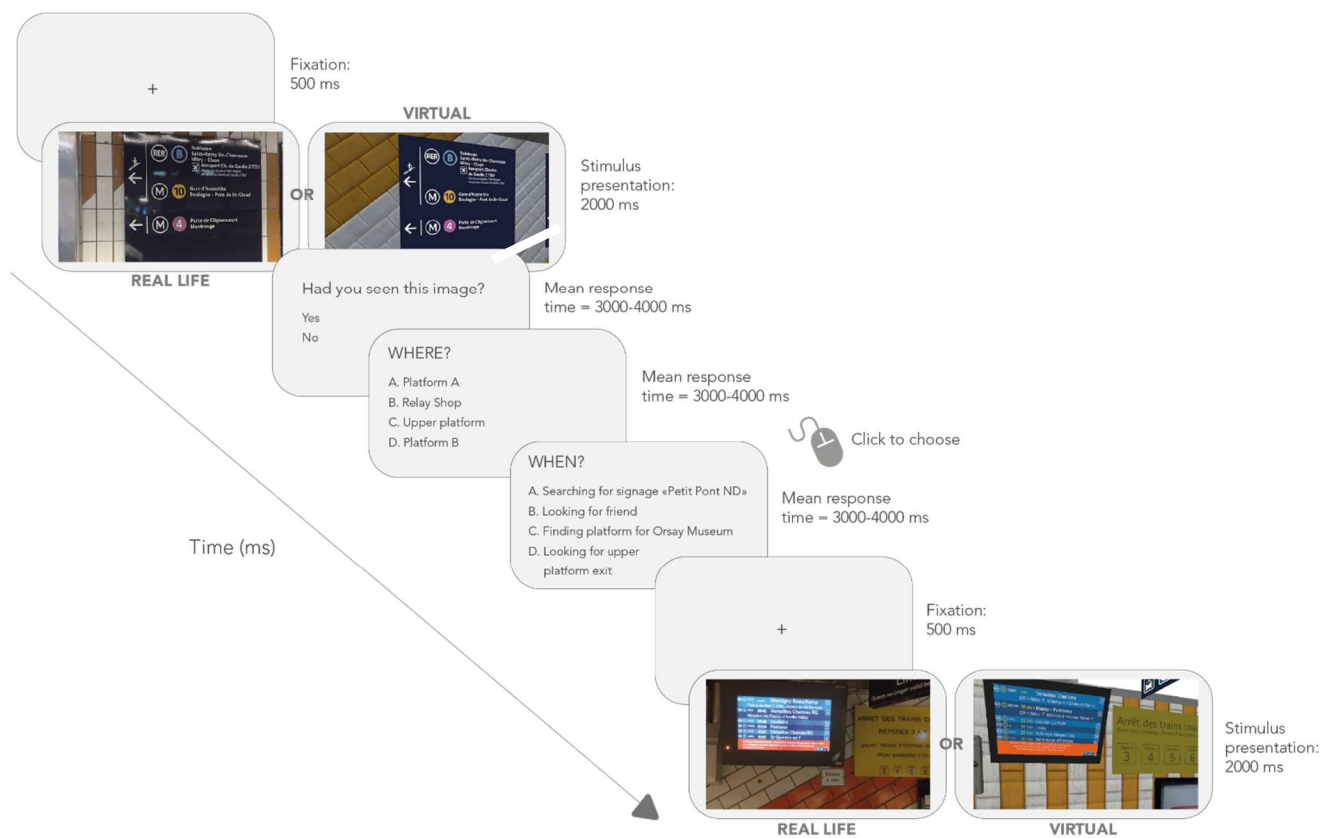
390 The memory test was designed on OpenSesame Experiment and presented on a Hewlett-  
391 Packard HP EliteBook 820 G2, processor Intel® Core™ i5-5200U CPU @ 2.20 GHz, after  
392 the experiment in the train station (virtual or real). During this test, participants who took part  
393 in the virtual version of the experiment, were presented a series of pictures (N=30 hits and  
394 n=15 distractors) taken from the virtual train station. Participants who took part in the real-life  
395 experiment were presented 30 other pictures (hits) and 15 distractors, taken from the actual  
396 train station. For both environmental conditions, the chosen pictures were those of signage  
397 boards and information displays with the train station's background as it was seen by the  
398 participants during the experiment. Names of exits, directions for different platforms and  
399 trains' destinations were modified for distractors. For each session, the 45 images were  
400 randomized to avoid any effect related to order of presentation or participant's fatigue. The  
401 questions asked during the test investigated the "what" of the object, corresponding to the  
402 recognition of the object and the ability to identify information among fake ones. The  
403 participants had to choose between *yes* or *no*, to say if the image has been presented during  
404 navigation or not. The "where" was also measured, which was related to the spatial  
405 localization of the stimulus in the train station, viewed during navigation. For this question,  
406 the participant had to choose among platform A, Relay shop, upper platform and platform B.  
407 Finally, the "when" was recorded, which informed us on the temporal localization of the  
408 stimulus in the environment. Participants had to choose among: *when looking for Petit Pont*  
409 *Notre Dame*, *when looking for friend*, *when finding platform for Orsay museum* and *when*  
410 *looking for upper platform exit*. Combining these three scores (*what*, *where* and *when*), gave  
411 an indication of the binding score. The higher the binding score, the stronger is the link  
412 between the stimulus and the contextual information (for e.g.: Makowski, Sperduti, Nicolas &  
413 Piolino, 2017; Meiser, Sattler & Weißer, 2008).

414 In order to have an adequate timing for stimuli presentation in our memory test, we pre-tested  
415 its different phases with 15 participants who were not part of our final experimental cohort.  
416 During the pre-test, the image identification difficulty was assessed according to the number  
417 of correct answers of the pre-tested participants. This allowed us to obtain a measure of the  
418 mean difficulty level of a list of pictures. For the final test, each image was presented for



419 2000ms after which the participant had to choose if the image was presented or not, earlier in  
 420 the train station (virtual or real-life condition). There was no time limit for the participants to  
 421 decide: if the image was initially presented or not, in which temporal and spatial contexts it  
 422 was presented. The participants were however encouraged to choose an answer as quick as  
 423 possible. The mean time taken by the participants to answer each question was 3000 ms. The  
 424 different steps for our memory test are illustrated in figure 3 below.

425 Higher binding scores (*what, when, where, total binding*) were expected for regular travelers  
 426 than occasional travelers, with no environmental condition effect.



427

428

Figure.3. Behavioral cognitive load measure - Memory test

429

430 2.6 Procedure

431 The experiment started with a set of tests consisting of six neuropsychological tests  
 432 (BDI, STAI-YA, Fatigue scale, PRMQ and Zoo map test) assigned to each participant. For  
 433 the next steps, the procedures varied according to the experimental environment: real-life and  
 434 VR. The real-life experiment took place between 10 to 12 a.m. at off-peak hours. The VE was  
 435 designed to take place during this time interval. Only one participant took part in the

436 experiment at a time. Participants were equipped with an E4 wristband sensor and the  
437 recordings of EDA signals were triggered. They were then asked to move around the train  
438 station for 10 minutes, checking their environment in a designated area, indicated by the  
439 experimenter, excluding the experimental perimeter of the train station. This was done to  
440 obtain a baseline of EDA signals for the participants in a non-experimental context. The data  
441 were measured online, with a feedback on our smartphone that allowed us to monitor  
442 variations in EDA level. The recorded EDA scores were accessible offline after the  
443 experiment. After the 10 minutes of baseline encoding, written and oral indications for the  
444 actual experiment were given to the participants in a story-telling format, as seen in section  
445 2.4. Once the experiment started, another trigger was activated on the EDA wristband to mark  
446 the beginning of the experiment. During the study, the experimenter followed the participant  
447 at a distance of two meters, while avoiding any contact with the latter.

448 For the *virtual reality experiment*, 56 participants were also recruited. 28 of them were regular  
449 and 28 were occasional, with 50% of women in each category. Here also, one participant took  
450 part in the experiment at a time. Before starting the actual experiment in the virtual condition,  
451 the participants were able to navigate for 10 minutes in an empty version of the train station,  
452 without furniture, object, train, avatars and signage boards. There was only the building, the  
453 walls' texture and lights. This was done to help the participants accommodate to virtual reality  
454 and be at ease with this system. The participants tested all the settings of the joystick: walk  
455 slowly, walk quickly, accelerate, decelerate, walk upstairs quickly and slowly. Afterwards, the  
456 participants were equipped with an E4 wristband sensor, and the EDA baseline was recorded  
457 for 10 minutes in the empty virtual train station. Once this phase ended, written and oral  
458 indications for the experiment were given to the participants in a story-telling format, as seen  
459 in section 2.4. The virtual train station was then switched on and a trigger was activated on the  
460 EDA wristband to mark the start of the experiment, with a timer integrated in the app on  
461 which the data could be monitored online.

462 During both the virtual and real-life experiments, any unexpected behavior or reaction was  
463 noted by the experimenter. The participants were not supposed to ask any question to the  
464 experimenter. The experiment ended when the participants accomplished all the tasks  
465 mentioned in section 2.3. An additional neuropsychological test was then assigned to the  
466 participants: STAI Y-B to investigate anxiety level after the experiment. To evaluate the  
467 subjective mental load measure related to the experiment, the participants had to fill up the  
468 NASA-TLX test. A questionnaire assessing the sense of presence for an indirect subjective

469 measure of immersion in virtual reality was filled up by participants who took part in the  
470 virtual train station (Schubert et al., 2001).

471 Participants were then asked to take part in a memory test designed especially for this study as  
472 shown in figure.3. The purpose of the experiment was finally explained to the participants and  
473 their feedbacks were noted in order to explain possible deviant results and for potential  
474 upgrades of the study.

### 475 **3. Data processing**

476 During this experiment, we evaluated cognitive load of participants through EDA  
477 analyses for physiological dimension, NASA-TLX for subjective dimension and memory  
478 binding score for behavioral dimension

479 Data collected were analyzed for two factors: travelers' profile (regular or occasional) and  
480 environmental situation (virtual reality or real-life) using the statistical computing and  
481 graphics software, R. Descriptive analyses were first performed using the package *tidyverse*  
482 (Wickham, 2017). The mean values and standard deviations for all collected data were hence  
483 obtained from these descriptive analyses. Normality and homoscedasticity tests were  
484 conducted using *Car* package (Fox et al., 2012). ANOVA was then calculated using the same  
485 package. *Lsr* (Navarro, 2015) for effect size calculation. Posthoc pairwise analyses were  
486 performed, using *emmeans* package (Estimated Marginal Means aka Least-Squares Means)  
487 (Lenth, 2016). *LmerTest* (Kuznetsova, Brockhoff & Christensen, 2017) and *Psycho*  
488 (Makowski, 2018) for Mixed Effects Models.

#### 489 *3.1 Electrodermal activity*

490 EDA data were uploaded in Matlab-based software Ledalab, designed to analyze skin  
491 conductance data (Benedek & Kaernbach, 2010). The format of EDA E4 Empatica  
492 corresponds to "Text Type I". The sampling frequency was 4Hz and was down-sampled to  
493 1Hz, with threshold amplitude of 0.01 $\mu$ s. The signal was first conditioned before storage. The  
494 uploaded data were preprocessed using low-pass Butterworth filter, down-sampled, cut,  
495 smoothed, and artifact-corrected. The data were then optimized by clicking on the  
496 corresponding button.

497 A Continuous Decomposition Analysis (CDA) was then performed by decomposing the raw  
498 Skin Conductance (SC) data into continuous tonic and phasic activities providing an unbiased  
499 quantification of SCR amplitudes and number of SCR peaks.

500

### 501 *3.2 NASA-Task Load Index*

502 Scores for NASA-TLX were calculated using a weighted mean score. This was done by  
503 comparing pairwise the 6 investigated dimensions, namely mental demand, physical demand,  
504 temporal demand, performance, effort and frustration level. The task of the participant was to  
505 choose which dimension between each pair, was more important with respect to the demand  
506 of the task. For each comparison, a point was granted to the selected dimension. After each  
507 two by two comparison, the weighted mean score was calculated by multiplying score  
508 assigned to each dimension out of 20, by the number of times that each dimension was chosen  
509 during pairwise comparisons.

### 510 *3.3 Binding score*

511 Data obtained from Opensesame after memory test were automatically saved in a  
512 spreadsheet, obtained after each participant has taken part in the memory test. Each  
513 participant's spreadsheet contained the percentage of correct answers, wrong answers, spatial  
514 binding temporal binding and total binding for each stimulus presented. The time taken in  
515 milliseconds to respond to each question was also recorded. Additional data concerning  
516 keyboard letters pressed were also saved, to check if the participants' wrong answers were not  
517 due to wrong typing.

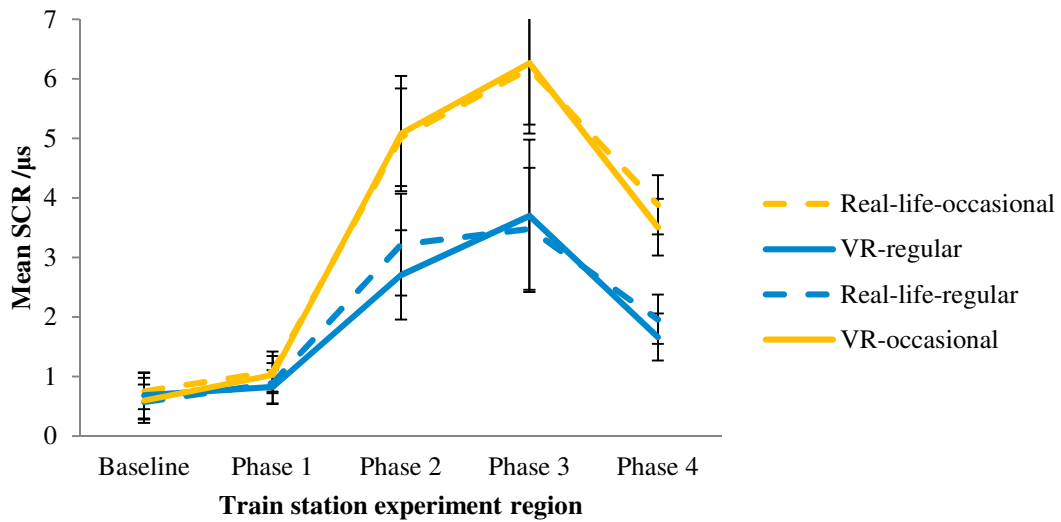
518

## 519 **4. Results**

### 520 4.1 Electrodermal activity

521 Data obtained by continuous deconvolution analysis showed a mean SCR score of  
522 0.66 $\mu$ s at baseline for the four groups of travelers (occasional travelers in real-life condition,  
523 occasional travelers in virtual reality, regular travelers in real-life condition and regular  
524 travelers in virtual reality condition). This score increased exponentially during phase 1 to  
525 phase 3 and decreases from phase 3 to phase 4. The mean SCR scores of occasional travelers

526 increased more than those of regular travelers at any moment. The results obtained are  
 527 illustrated in figure 4 below.



528 Figure. 4. EDA of regular and occasional travelers in virtual reality and real-life conditions,  
 529 through mean SCR (CDA)

530 Since EDA was measured repeatedly throughout the experiment for the same participant, the  
 531 independence of data was no more satisfied. A linear mixed model analysis was thus  
 532 conducted on our two experimental factors (profile and condition), to test for random and  
 533 fixed effects to account for spacing between each time points (Krueger & Tian, 2004; Jaeger,  
 534 2008), while taking time point (EDA\_Phase) as error. Our model was as follows:

$$535 \text{EDA\_value} = \text{condition} * \text{profile} + (1 | \text{EDA\_Phase})$$

536  
 537 *Where EDA\_value was skin conductance responses*  
 538 *EDA\_phase was the order of EDA measurements*  
 539

540 Analysis of fixed model gave no condition effect between *virtual reality* and *real-life* contexts  
 541  $F(3,120) = .07, p = .79$ . A profile effect was however detected between *regular* and  
 542 *occasional* travelers,  $F(3,120) = 111, p < .001$ . No interaction was observed between  
 543 condition and profile factors,  $F(3,120) = .02, p = .89$ . Analysis of the main effects of the  
 544 linear mixed model is presented in table 4 below.

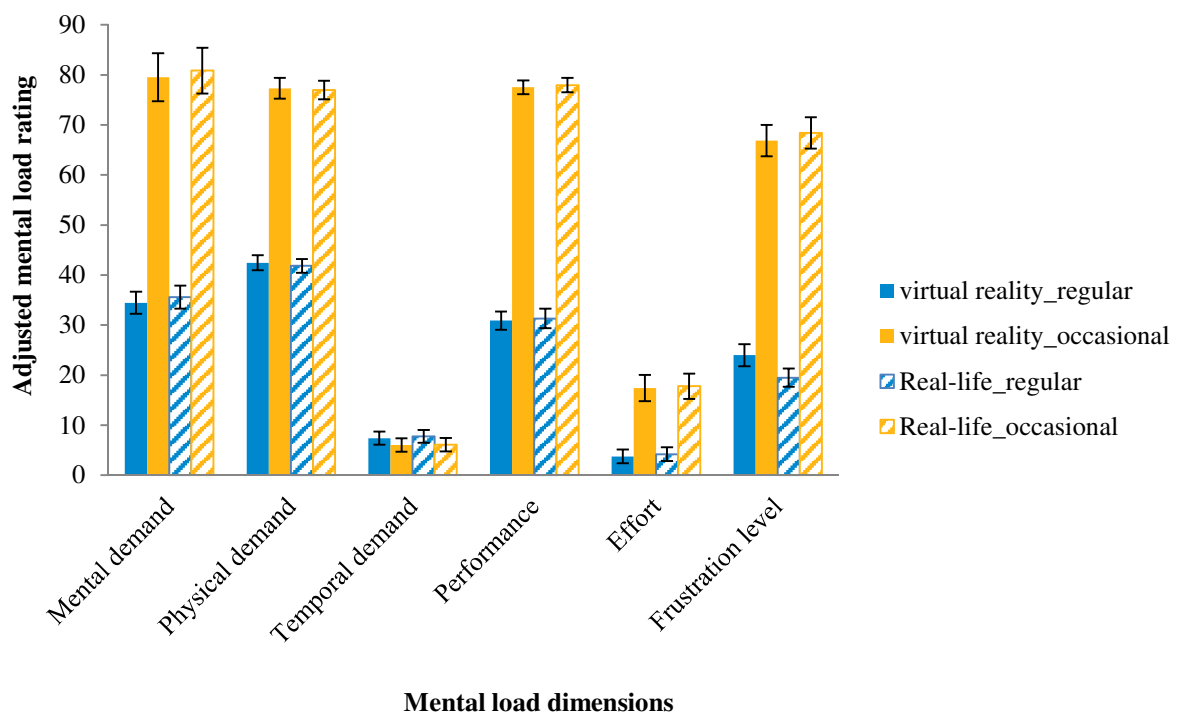
Variable	Coef	SE	t	df	p	Coef_std	SE_std	Effect_Size	CI lower	CI higher
(Intercept)	1.75	0.75	2.31	4.04	.08	-0.25	0.44	small	0.12	3.37
condition	-0.02	0.09	-0.27	552	.79	-0.01	0.05	Very small	-0.19	0.15
profile	0.91	0.09	10.5	552	<.001***	0.53	0.05	medium	0.74	1.08
condition * profile	-0.02	0.12	-0.13	552	.89	-0.01	0.07	Very small	-0.26	0.22

Table.4. Linear mixed model for EDA of travelers in different train station contexts

545 The overall model predicting (EDA\_value ~ condition\* profile+ (1 | EDA\_Phase)) has a total  
 546 explanatory power (conditional R<sup>2</sup>) of 85.25%, in which the fixed effects explain 5.76% of the  
 547 variance (marginal R<sup>2</sup>). The model's intercept is at 1.75 (SE = 0.75, 95% CI [0.12, 3.37]).  
 548 Within this model, *condition* factor was not significant ( $\beta = -0.02$ , SE = 0.09, 95% CI [-0.19,  
 549 0.15],  $t(552) = -0.27$ ,  $p > .1$ ) and can be considered as very small (std.  $\beta = -0.01$ , std. SE =  
 550 0.05). *Profile* factor, on the other hand, proved to be significant ( $\beta = 0.91$ , SE = 0.09, 95% CI  
 551 [0.74, 1.08],  $t(552) = 10.54$ ,  $p < .001$ ) and can be considered as medium (std.  $\beta = 0.53$ , std. SE  
 552 = 0.05). No significant interaction was found between *condition* and *profile* factor ( $\beta = -0.02$ ,  
 553 SE = 0.12, 95% CI [-0.26, 0.22],  $t(552) = -0.13$ ,  $p > .1$ ) and can be considered as very small  
 554 (std.  $\beta = -0.01$ , std. SE = 0.07).

#### 555 4.2 NASA-Task Load Index

556 Cognitive load measured by the participant using NASA-TLX subjective measure of  
 557 mental load were collected and illustrated in figure 5 below.



558 **Mental load dimensions**  
 559 *Figure.5. NASA-TLX Mental load results distribution*

560 Globally, mental load dimensions were higher for occasional travelers than for regular  
 561 travelers (except for temporal demand dimension). 2X2 ANOVA was conducted on traveler's  
 562 profile (regular vs occasional) and environmental condition (real-life vs virtual reality). The  
 563 analysis of mental demand showed significant effect for profile,  $F(3,108) = 4201$ ,  $p < .001$ ,

564  $\eta_p^2 = .97$ , higher for occasional travelers than for regular travelers. Analysis of environmental  
 565 condition showed,  $F(3,108) = 3.13$ ,  $p = .08$ ,  $\eta_p^2 = 0.03$ , indicating no significant difference in  
 566 virtual reality and real-life conditions. The interaction between environmental condition and  
 567 profile was also non-significant:  $F(3,108) = .02$ ,  $p = .90$ ,  $\eta_p^2 = .00$ . These findings indicate a  
 568 higher mental load level for occasional travelers compared to regular travelers. This  
 569 difference is maintained in real-life and virtual environments. Similar analyses were  
 570 conducted on physical demand, temporal demand, performance, effort and frustration level,  
 571 and the results are resumed in table 5 below.

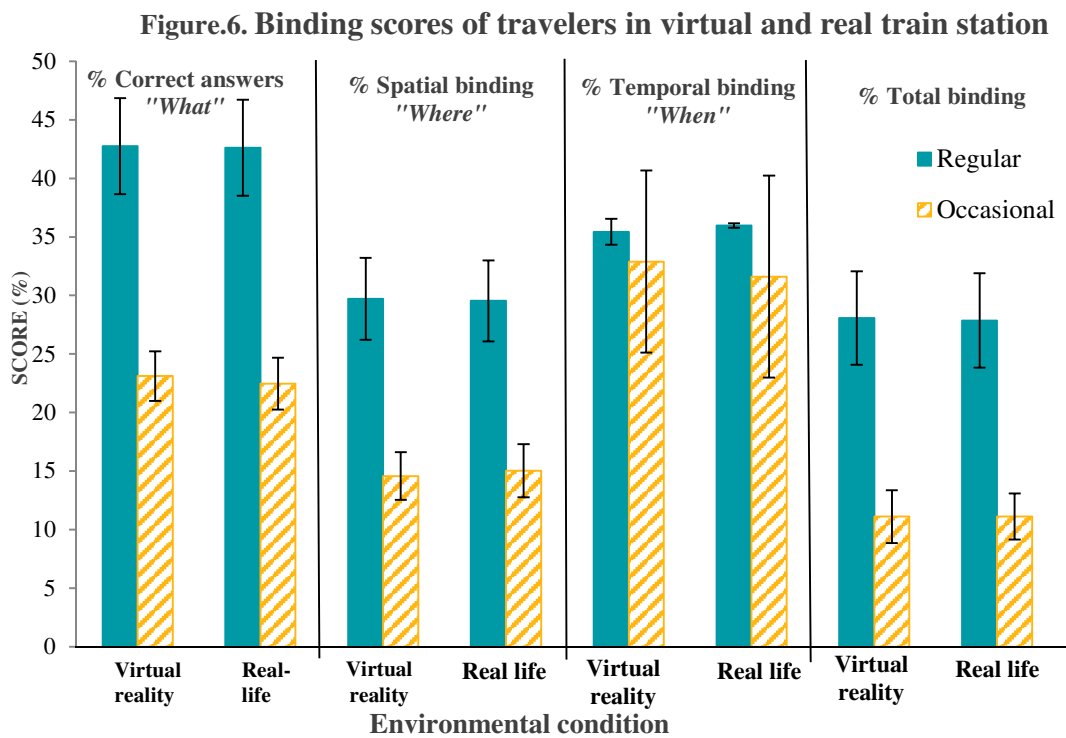
572 For frustration level, the mean adjusted load ratings were higher for occasional travelers than  
 573 for regular travelers. This difference was significant,  $F(3,108) = 8495$ ,  $p < .001$ ,  $\eta_p^2 = .99$ ,  
 574 indicating a Travelers' profile effect. Adjusted load ratings were also higher for real-life  
 575 condition than for VR condition, for occasional travelers. The data were the other way round  
 576 for regular travelers (i.e. higher for VR condition than for real-life condition). This effect was  
 577 significant,  $F(3,108) = 8.87$ ,  $p < .001$ ,  $\eta_p^2 = .08$ . Analysis of interaction between  
 578 Environmental condition X Travelers' profile, gave,  $F(3,108) = 37$ ,  $p < .001$ ,  $\eta_p^2 = .25$ ,  
 579 indicating an interaction between our two factors. We thus performed a post hoc Tukey HSD  
 580 test to verify pairwise interactions between each pair of data. The travelers' profile effect  
 581 persisted ( $p < 0.001$ ) when environmental condition was kept constant and when only the  
 582 traveler's profile was modified (e.g. occasional travelers in real-life condition vs regular  
 583 travelers in real-life). However, the environmental condition effect persisted only for regular  
 584 travelers ( $p < 0.001$ ), when comparing regular travelers in real-life vs regular travelers in VR.  
 585 No environmental condition was observed when comparing occasional travelers in real-life  
 586 condition to occasional travelers in VR ( $p = 0.13$ ). These results suggest that frustration level  
 587 varies significantly in regular travelers with respect to the environmental condition.

Subjective cognitive score	Environmental condition	Travelers' profile	
		Travelers' profile	Environmental condition*
<b>Mental demand</b>	$F(3,108)= 3.13$ , $p=.08$ , $\eta_p^2=.03$	$F(3,108)=4201$ , $p<.001$ , $\eta_p^2=.97$	$F(3,108)=.02$ , $p=.90$ , $\eta_p^2=.00$
<b>Physical demand</b>	$F(3,108)=2.02$ , $p=.16$ , $\eta_p^2=.02$	$F(3,108)=11486$ , $p<.001$ , $\eta_p^2=.99$	$F(3,108)=.19$ , $p=.66$ , $\eta_p^2=.00$
<b>Temporal demand</b>	$F(3,108)=.86$ , $p=.36$ , $\eta_p^2=.01$	$F(3,108)=37$ , $p<.001$ , $\eta_p^2=.25$	$F(3,108)=.41$ , $p=.52$ , $\eta_p^2=.17$
<b>Performance</b>	$F(3,108)=1.84$ , $p=.18$ , $\eta_p^2=.02$	$F(3,108)=21733$ , $p<.001$ , $\eta_p^2=.99$	$F(3,108)= .00$ , $p=1.00$ , $\eta_p^2=.00$
<b>Effort</b>	$F(3,108)=1.09$ , $p=.30$ , $\eta_p^2=.01$	$F(3,108)= 1201$ , $p<.001$ , $\eta_p^2=.92$	$F(3,108)= .02$ , $p=.89$ , $\eta_p^2=.43$
<b>Frustration level</b>	$F(3,108)=8.87$ , $p<.001$ , $\eta_p^2=.08$	$F(3,108)= 8495$ , $p<.001$ , $\eta_p^2=.99$	$F(3,108)= 37$ , $p<.001$ , $\eta_p^2=.25$

Table.5. Analyses for subjective cognitive load (adjusted load rating) for environmental situation factor (real life vs virtual reality) and travelers' profile factor (regular vs occasional)

588 4.3 Binding scores

589 Comparison of mean scores for correct answers, spatial binding, temporal binding and total  
 590 binding are illustrated in figure 6 below. In each of these four dimensions, regular travelers showed a  
 591 better performance than occasional travelers, reaching however a highest score of 43% in *correct*  
 592 *answers* dimension.



593

594 A 2x2 ANOVA was conducted on travelers' profile (regular x occasional) and the  
 595 environmental situation (virtual reality x real-life) on the four measured dimensions (*what*,  
 596 *when, where and total binding*).

597 For analysis of *correct answers*, we obtained a significant difference for travelers' profile:  
 598  $F(3,108) = 1032, p < .001, \eta_p^2 = .905$ . This showed a significant difference in correct answers,  
 599 based on travelers' profile, higher for regular travelers than for occasional travelers. In  
 600 environmental condition, however, the analysis was not significant:  $F(3,108) = .40, p = .53,$   
 601  $\eta_p^2 = .004$ . Analysis of interaction between environmental condition and travelers' profile  
 602 indicated a non-significant interaction:  $F(3,108) = .16, p = .69, \eta_p^2 = .002$ . These findings  
 603 showed that regular travelers expressed higher contextual and factual memories than  
 604 occasional travelers, with similar results in real-life and virtual environments.



605 The same analyses performed on the other three dimensions showed the same significant p-  
 606 values for different travelers' profile. No significant difference was obtained for  
 607 environmental condition, for any memory performance. The results are summarized in table 6  
 608 below.

Score	Environmental condition	Travelers' profile	Environmental condition*
			Travelers' profile
<b>Correct answers</b>	F(3,108)=.40, p=.53, $\eta^2$ =.004	F(3,108)=1032, p<.001, $\eta^2$ =.905	F(3,108)=.16, p=.69, $\eta^2$ =.002
<b>Spatial binding</b>	F(3,108)=.07, p=.80, $\eta^2$ =.001	F(3,108)=733, p<.001, $\eta^2$ =.872	F(3,108)=.34, p=.56, $\eta^2$ =.003
<b>Temporal binding</b>	F(3,108)=.12, p=.73, $\eta^2$ =.001	F(3,108)=9.77, p<.001, $\eta^2$ =.083	F(3,108)=.68, p=.41, $\eta^2$ =.006
<b>Total binding</b>	F(3,108)=.55, p=.46, $\eta^2$ =.000	F(3,108)=278, p<.001, $\eta^2$ =.878	F(3,108)=.82, p=.37, $\eta^2$ =.000
<b>Response time</b>	F(3,108)=.55, p=.46, $\eta^2$ =.005	F(3,108)=278, p<.001, $\eta^2$ =.721	F(3,108)=.82, p=.37, $\eta^2$ =.008

609 Table.6. Analyses of binding scores for environmental situation factor (real life vs  
 610 virtual reality) and travelers' profile factor (regular vs occasional)

#### 611 4.4 Additional measures

612 Evaluation of travelers' subjective immersion level from virtual reality session was  
 613 assessed using the *presence questionnaire* (Schubert et al., 2001). 4 dimensions were  
 614 evaluated from 14 items, on a 7 divided, ranging from -3 to 3. These dimensions were *general*  
 615 *experience*, *spatial presence*, *personal involvement* and *relevant realism*. Immersion level was  
 616 high in each of the four dimensions and significantly different from the mean value of 0  
 617 (corresponding to a moderate sense of presence in the VE); general presence,  $t(55)=16.3$ ,  $p <$   
 618  $.001$ ; spatial presence,  $t(55)=46.5$ ,  $p < .001$ ; involvement,  $t(55)=46.4$ ,  $p < .001$  and  
 619 experienced realism,  $t(55)=51.9$ ,  $p < .001$ .

620  
 621 When comparing the two levels of profile in a 1-way MANOVA, we found that regular  
 622 travelers expressed no significant difference compared to occasional travelers,  $F(1,54) =$   
 623  $1.23$ ;  $p=.311$ ,  $\eta^2= .09$ . A tendential significance was however noted in spatial presence  
 624 dimension, indicating a difference between the two investigated profiles of travelers, being  
 625 higher for regular travelers,  $F(1,54)=4.15$ ,  $p=.05$ ,  $\eta^2$ =.07. Results of descriptive analysis and  
 626 full one-way MANOVA on each of the dimensions are presented in table 7 below.

627

Profile	General_Presence	Spatial_Presence	Involvement	Experienced_Realism
Occasional	2.11±.92	2.16±.34	2.21±.38	2.14±.34
Regular	1.93±.94	2.34±.37	2.29±.35	2.09±.27
MANOVA	F(1,54)=.52, p=.47, $\eta^2=.01$	F(1,54)=4.15, p=.05, $\eta^2=.07$	F(1,54)=.53, p=.47, $\eta^2=.01$	F(1,54)=.22, p=.64, $\eta^2=.00$

628

*Table.7. Descriptive analysis and MANOVA for immersion dimensions in VR*

629

## 630 5. Discussion

631 Information processing generates cognitive load that has a direct impact on  
632 information acquisition and learning (Sweller, 1988) and the amount of information recalled  
633 (Kirsh, 2000). In our experiment, we took the example of train travelers who are exposed to  
634 numerous amounts of information that help them get to a desired point in a train station. Our  
635 experiment was conducted in real-life and in virtual reality (condition factor), with occasional  
636 and regular travelers (expertise factor), comparing their cognitive load via physiological,  
637 subjective and behavioral responses. The findings showed no environmental condition effect  
638 between virtual and real-life conditions. EDA, NASA-TLX (except for subjective level of  
639 frustration in regular travelers), and Binding memory scores were not statistically different for  
640 a given level of expertise, when compared between two experimental conditions – expert  
641 travelers had same cognitive load in virtual and real-life conditions, and so are the results for  
642 novice travelers. In our study, we test for expertise effect as a confirmatory effect in order to  
643 evaluate variations in cognitive load, when contrasting real-life and virtual environments.  
644 Results showed higher cognitive load for novice travelers than for expert travelers, in both  
645 experimental conditions.

### 646 Modulations of cognitive load components

647 Cognitive load varies in real-life activities, due to modulations in cognitive resources  
648 (Chanquoy et al., 2007). These modulations can be analyzed through different pools of  
649 resources associated to intrinsic (Sweller, 1994), extraneous (Sweller and Chandler, 1991) and  
650 Germane cognitive load (Sweller et al., 1998). In this study we investigated cognitive load as  
651 a whole, while considering the three different components of cognitive load during the  
652 experimental set-up.

653 Intrinsic cognitive load in our study was related to the degree of knowledge of the train  
654 station's charter, necessary to understand information conveyed to travelers. This allows

655 travelers to encode information and use it to find their way through the train station. Travelers  
656 with knowledge of the investigated train station or those who had visited another train station  
657 using same charter in terms of information display, have a greater expertise level than those  
658 less exposed to these stimuli.

659 On the other hand, extrinsic cognitive load depends on the nomenclature of information itself.  
660 Since in our experiment we used actual train station's displays without any modification, this  
661 secondary factor was thus controlled. Extrinsic cognitive load depends on the expertise and  
662 knowledge for the investigated train station, of our participants.

663 Germane cognitive load corresponds to the relevant information recalled by the travelers  
664 related to their journey in the train station and varies oppositely to extrinsic cognitive load.  
665 For this purpose, we investigated the behavioral aspect of cognitive load through performance  
666 analysis, using binding score. The use of binding score as a behavioral measure in cognitive  
667 load analysis has never been used, as it is the case in the present study. One main advantage  
668 of using this behavioral measure was to take into account interindividual differences in  
669 information processing. This is due to the fact that participants, irrespective of their expertise  
670 level, have a unique interpretation of information and pay attention to different details in their  
671 surroundings. This would be a great advantage for the analysis of cognitive load in complex  
672 situations, where interindividual differences are much important. Binding score proved to be a  
673 convenient and promising behavioral measure for cognitive load.

#### 674 **Experimental condition factor: real-life or virtual reality environment**

675 The absence of experimental condition effect suggests that travelers can experience  
676 same cognitive load level, irrespective of the format of presentation of information (virtual  
677 reality and real-life). This observation is in line with findings obtained from previous studies  
678 on 3D modeling of virtual environment to investigate cognitive processes, initially assessed in  
679 real-life condition (Andersen et al., 2018; Chen & Stanney, 1999; Crescentini et al., 2016;  
680 Darken & Sibert, 1996; Darken & Peterson, 2014; Freeman et al., 2008; Klinger et al., 2005;  
681 Steinicke et al., 2013; Viaud-Delmon et al., 1998). We extended here the findings by  
682 revealing similar cognitive load measurement based on physiological (Gevins et al., 1998;  
683 Paas et al., 2003; Sweller et al., 2011), subjective (Vidulich & Tsang, 2012) and behavioral  
684 (Brunken et al., 2003; Paas & Van Merriënboer, 1994) responses. Several reasons have been  
685 spotted for the success of the validation of virtual reality for cognitive load analysis. Two

686 main reasons could be the design of the virtual environment itself and the use of storytelling  
687 format for giving indications to participants.

688 Storytelling has proven its efficacy in highly rated video games experience, where players felt  
689 more invested and immersed in a virtual universe. Storytelling also provides a mnemonic  
690 technique to help participants understand and learn indications, through the use of a story  
691 rather than listing a set of tasks for the participant to perform. Our study is in line with former  
692 studies, which have proven the rise in mnemonic performance using storytelling format  
693 (Bucher, 2017; Ibanez & Aylett, 2003; Shin, 2018). It would be interesting to investigate the  
694 use of storytelling, mostly in experiment with elaborate indications.

### 695 **Expertise effect: a cognitive load indicator for virtual environment validation**

696 Our findings confirmed that expertise effect, where experts have higher performances  
697 than novices, has an important impact on cognitive load (Kalakoski & Saariluoma, 2001;  
698 Perdreau & Cavanagh, 2015; Reingold et al., 2001). Moreover, whenever all secondary  
699 factors were kept constant, experts showed same performance in both environmental contexts.  
700 This phenomenon was used as an indicator to validate our virtual environment.

701 When traveling in big cities, with a mass transit context such as Paris, it is essential to think  
702 about all the complex situations that may arise each day in a train station, and especially how  
703 these situations may impact on the travelers' cognitive load. Interestingly, we found only one  
704 difference regarding subjective measures, since the experts felt more frustration in virtual  
705 environment than in the real environment. It suggests that experts could feel annoyed as a  
706 result of being unable to achieve their navigation as they used to do in the real-life, i.e. alone  
707 and without being observed by an experimenter. Accordingly, the experts seem more attached  
708 and sensitive to their habits than the novices and seem to be more vulnerable to social  
709 pressure. It would thus be interesting to investigate cognitive load in a more complex and  
710 unpredicted situations, in order to see how expert effect would vary in these conditions.  
711 Previous studies had shown the existence of expertise reversal effect in non-optimal  
712 situations, where experts showed significantly lower performance than novices (Kalyuga,  
713 2009; Sweller et al., 2003). For examples, delayed or cancelled trains, engineering work on  
714 railway, traveler's accident and strike of employees, to name a few. Results from cognitive  
715 load analyses of travelers in these complex situations could help us optimize information  
716 given to travelers, in terms of instructional design. These concern the format (redundancy

717 effect: Chandler & Sweller, 1991), the modality (Mayer, 2002) and the attentional levels (split  
718 attention effect: Chandler & Sweller, 1992; Mayer & Moreno, 1998) of information.

### 719 **Strength, limitations, and future avenue of research**

720 In our study we assessed young adults, with a mean age of 26( $\pm$ 5) years old. This is  
721 however not representative of all travelers in Saint-Michel Notre Dame train station. It would  
722 be interesting to extend our sample to higher age groups and evaluate cognitive load of elder  
723 travelers. For this to be efficient, it would be necessary to adapt our training session for our  
724 VE, by extending the training session and by creating additional ones. The participant would  
725 for instance perform some particular exercises in a virtual environment such as walking up  
726 and down stairs, moving sideways to avoid obstacles, picking up objects. These activities,  
727 however, can contribute to amplify one main side effect of VR: cybersickness. This main  
728 limitation in virtual reality technologies requires further studies to find practical solutions to  
729 facilitate its reduction. Besides from optimization of the virtual environment, one main  
730 method that proved its efficacy against cybersickness, is to provide repeated training sessions  
731 with a gradual increase in difficulty.

732 Besides extending and diversifying training sessions, it would also be important to think about  
733 a common threshold to define expertise, since expert travelers from the elder population  
734 would have a greater time exposure than younger experts, in the tested train station (e.g.  
735 expertise among chess players: Chase & Simon, 1973). It would be interesting to investigate  
736 expert memory models (Gobet, 1998) in relation with cognitive load theory.

737 When choosing the train station for our experiment, we look for different criteria such as:  
738 localization, passengers' flow, travelers' profiles, size and volume of travelers per area of  
739 train station (mass transit characteristics) and number of anomalies reported by travelers to  
740 local agents. This investigation led us to the conclusion that train stations, though located in  
741 the same city, are significantly different from each other. We can thus ask ourselves if our  
742 results obtained from Saint-Michel Notre Dame could be applied to other mass transit train  
743 stations in Paris city. It would be interesting to design different train stations and perform the  
744 same analyses on cognitive load, to contrast our results. These data obtained on cognitive load  
745 could be implemented in a supervised learning function of machine learning for diagnostics  
746 classification. We would hence be able to predict cognitive load of travelers in a given train  
747 station, using data collected from other train stations. This perspective is in line with past  
748 studies in machine learning prediction (for cancer prognosis and prediction: Kourou,

749 Exarchos, Exarchos, Karamouzis & Fotiadis 2015; classification of mental tasks from EEG:  
750 Liang, Saratchandran, Huang & Sundararajan, 2006).

751

## 752 **6. Conclusion**

753 This study validates the use of virtual reality as a technique for investigating cognitive  
754 load in daily life situations, using expert effect as an indicator of learning process. Different  
755 techniques that accounted for cognitive load in our study proved to be practical and  
756 nevertheless promising for studies involving navigation. A whole new field of work in non-  
757 static cognitive load investigation, involving non-optimal environmental scenarios can now be  
758 tested. These modulations in environmental conditions could give rise to learning effects,  
759 explained by the Cognitive Load Theory, and would represent a great opportunity to study  
760 these effects in applied experiments.

761

## 762 **REFERENCES**

- 763 **Andersen, S. A. W., Konge, L., & Sørensen, M. S. (2018).** The effect of distributed virtual  
764 reality simulation training on cognitive load during subsequent dissection training.  
765 *Medical teacher*, 1-6.
- 766 **Baddeley, A., Logie, R., Bressi, S., Sala, S. D., & Spinnler, H. (1986).** Dementia and working  
767 memory. *The Quarterly Journal of Experimental Psychology Section A*, 38(4), 603-618.
- 768 **Barfield, W., & Hendrix, C. (1995).** The effect of update rate on the sense of presence within  
769 virtual environments. *Virtual Reality*, 1(1), 3–15. DOI:10.1007/BF02009709
- 770 **Beck, A. T., Steer, R. A., & Brown, G. K. (1996).** Beck depression inventory-II. *San Antonio*,  
771 78(2), 490-8.
- 772 **Benedek, M., & Kaernbach, C. (2010).** A continuous measure of phasic electrodermal  
773 activity. *Journal of neuroscience methods*, 190(1), 80-91.
- 774 **Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014).** Measuring  
775 neurophysiological signals in aircraft pilots and car drivers for the assessment of mental  
776 workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, 44, 58-75.
- 777 **Boucsein, W. (2012).** *Electrodermal activity*. Wuppertal, Germany: Springer Science &  
778 Business Media.
- 779 **Bruck, S., & Watters, P. A. (2011).** The factor structure of cybersickness. *Displays*, 32(4),  
780 153–158. DOI: 10.1016/J.DISPLA.2011.07.002
- 781 **Brunken, R., Plass, J. L., & Leutner, D. (2003).** Direct measurement of cognitive load in  
782 multimedia learning. *Educational psychologist*, 38(1), 53-61.
- 783 **Bucher, J. (2017).** *Storytelling for virtual reality: methods and principles for crafting*  
784 *immersive narratives*. New York & London: Focal press.
- 785 **Bystrom, K.-E., Barfield, W., & Hendrix, C. (1999).** A Conceptual Model of the Sense of  
786 Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*,  
787 8(2), 241–244. <https://doi.org/10.1162/105474699566107>

788 **Cain, B.** (2007). *A review of the mental workload literature*. Toronto, Canada: Defence  
789 Research And Development.

790 **Carrozzino, M., & Bergamasco, M.** (2010). Beyond virtual museums: Experiencing  
791 immersive virtual reality in real museums. *Journal of Cultural Heritage*, 11(4), 452-458.

792 **Cegarra, J., & Chevalier, A.** (2008). The use of Tholos software for combining measures of  
793 mental workload: Toward theoretical and methodological improvements. *Behavior*  
794 *Research Methods*, 40(4), 988-1000.

795 **Cegarra, J., & Morgado, N.** (2009, September). Étude des propriétés de la version  
796 francophone du NASATLX. In Communication présentée à la cinquième édition du  
797 colloque de psychologie ergonomique (Epique) (pp. 28-30).

798 **Chandler, P., & Sweller, J.** (1991). Cognitive load theory and the format of instruction.  
799 *Cognition and instruction*, 8(4), 293-332.

800 **Chandler, P., & Sweller, J.** (1992). The split-attention effect as a factor in the design of  
801 instruction. *British Journal of Educational Psychology*, 62(2), 233-246.

802 **Chanquoy, L., Tricot, A., & Sweller, J.** (2007). *La charge cognitive: Théorie et applications*.  
803 Paris: France: Armand Colin

804 **Chase, W. G., & Simon, H. A.** (1973). Perception in chess. *Cognitive psychology*, 4(1), 55-  
805 81.

806 **Chen, J. L., & Stanney, K. M.** (1999). A theoretical model of wayfinding in virtual  
807 environments: Proposed strategies for navigational aiding. *Presence*, 8(6), 671-685.

808 **Chen, F., Zhou, J., Wang, Y., Yu, K., Arshad, S. Z., Khawaji, A., & Conway, D.** (2016).  
809 Galvanic Skin Response-Based Measures. In *Robust Multimodal Cognitive Load*  
810 *Measurement* (pp. 87-99). Cham: Springer International Publishing. DOI: 10.1007/978-  
811 3-319-31700-7\_5

812 **Collet, L., & Cottraux, J.** (1986). The shortened Beck depression inventory (13 items). Study  
813 of the concurrent validity with the Hamilton scale and Widlöcher's retardation scale.  
814 *L'Encephale*, 12(2), 77-79.

815 **Cook, M. P.** (2006). Visual representations in science education: The influence of prior  
816 knowledge and cognitive load theory on instructional design principles. *Science*  
817 *education*, 90(6), 1073-1091.

818 **Crescentini, C., Chittaro, L., Capurso, V., Sioni, R., & Fabbro, F.** (2016). Psychological and  
819 physiological responses to stressful situations in immersive virtual reality: Differences  
820 between users who practice mindfulness meditation and controls. *Computers in*  
821 *Human Behavior*, 59, 304-316.

822 **Darken, R. P., & Sibert, J. L.** (1996). Navigating large virtual spaces. *International Journal*  
823 *of Human-Computer Interaction*, 8(1), 49-71.

824 **Darken, R. P., & Peterson, B.** (2014). *Spatial orientation, wayfinding, and representation:*  
825 *Handbook of Virtual Environment Technology*. Monterey, California: Stanney, K. Ed

826 **Freeman, D., Pugh, K., Antley, A., Slater, M., Bebbington, P., Gittins, M., & Garety, P.**  
827 (2008). Virtual reality study of paranoid thinking in the general population. *The British*  
828 *Journal of Psychiatry*, 192(4), 258-263.

829 **Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., ... & Heiberger, R.**  
830 (2012). Package 'car'. *Vienna: R Foundation for Statistical Computing*.

831 **Fuchs, P., & Moreau, G.** (2006). *Le traité de la réalité virtuelle (Vol. 2)*. Paris, France :  
832 Presses des MINES.

833 **Fuchs, P., Moreau, G., & Guitton, P.** (2011). *Virtual reality: concepts and technologies*.  
834 London, UK CRC Press.

835 **Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., & Rush, G.** (1998).  
836 Monitoring working memory load during computer-based tasks with EEG pattern  
837 recognition methods. *Human factors*, 40(1), 79-91.

838 **Gobet**, F. (1998). Expert memory: A comparison of four theories. *Cognition*, 66(2), 115-152.

839 **Gopher**, D., & Donchin, E. (1986). Workload: An examination of the concept. In K. R. Boff,  
840 L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance*,  
841 Vol. 2. Cognitive processes and performance (pp. 1-49). Oxford, England: John Wiley  
842 & Sons.

843 **Hart**, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index):  
844 Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp.  
845 139-183). North-Holland.

846 **Healey**, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks  
847 using physiological sensors. *IEEE Transactions on intelligent transportation systems*,  
848 6(2), 156-166.

849 **Ibanez**, J., Aylett, R., & Ruiz-Rodarte, R. (2003). Storytelling in virtual environments from a  
850 virtual guide perspective. *Virtual Reality*, 7(1), 30-42.

851 **Jebara**, N., Orriols, E., Zaoui, M., Berthoz, A., & Piolino, P. (2014). Effects of enactment in  
852 episodic memory: a pilot virtual reality study with young and elderly adults. *Frontiers*  
853 *in aging neuroscience*, 6, 338.

854 **Jex**, S. M. (1998). *Stress and job performance: Theory, research, and implications for*  
855 *managerial practice*. Sage Publications Ltd.

856 **Johnson**, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological review*, 88(1), 67.  
857 DOI: 10.1037/0033-295X.88.1.67

858 **Kalakoski**, V., & Saariluoma, P. (2001). Taxi drivers' exceptional memory of street names.  
859 *Memory & Cognition*, 29(4), 634-638.

860 **Kalyuga**, S. (2009). The expertise reversal effect. In *Managing Cognitive Load in Adaptive*  
861 *Multimedia Learning* (pp. 58-80). IGI Global.

862 **Ke**, Y., Qi, H., He, F., Liu, S., Zhao, X., Zhou, P. & Ming, D. (2014). An EEG-based  
863 mental workload estimator trained on working memory task can work well under  
864 simulated multi-attribute task. *Frontiers in human neuroscience*, 8, 703.

865 **Kirsh**, D. (2000). A few thoughts on cognitive overload. *Intellectica*. 1 (30):19-51

866 **Klinger**, E., Bouchard, S., Légeron, P., Roy, S., Lauer, F., Chemin, I., & Nugues, P. (2005).  
867 Virtual reality therapy versus cognitive behavior therapy for social phobia: A  
868 preliminary controlled study. *Cyberpsychology & behavior*, 8(1), 76-88.

869 **Kourou**, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015).  
870 Machine learning applications in cancer prognosis and prediction. *Computational and*  
871 *structural biotechnology journal*, 13, 8-17.

872 **Koutsoudis**, A., Arnaoutoglou, F., & Chamzas, C. (2007). On 3D reconstruction of the old  
873 city of Xanthi. A minimum budget approach to virtual touring based on  
874 photogrammetry. *Journal of Cultural Heritage*, 8(1), 26-31.

875 **Kramer**, A. F. (1991). Physiological metrics of mental workload: A review of recent  
876 progress. *Psychophysiological metrics of mental workload*, 279(12)

877 **Krueger**, C., & Tian, L. (2004). A comparison of the general linear mixed model and  
878 repeated measures ANOVA using a dataset with multiple missing data points.  
879 *Biological research for nursing*, 6(2), 151-157. DOI: 10.1177/1099800404267682

880 **Kuznetsova**, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests  
881 in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13).  
882 DOI:10.18637/jss.v082.i13

883 **Lee**, E. A. L., & Wong, K. W. (2014). Learning with desktop virtual reality: Low spatial  
884 ability learners are more positively affected. *Computers & Education*, 79, 49-58.

885 **Lenth**, R. V. (2016). Least-squares means: the R package lsmeans. *Journal of statistical*  
886 *software*, 69(1), 1-33.progress. *Multiple-task performance*, 279-328.



- 887 **Liang**, N. Y., Saratchandran, P., Huang, G. B., & Sundararajan, N. (2006). Classification of  
888 mental tasks from EEG signals using extreme learning machine. *International journal*  
889 *of neural systems*, 16(01), 29-38.
- 890 **Makowski**, D., Sperduti, M., Nicolas, S., & Piolino, P. (2017). “Being there” and  
891 remembering it: Presence improves memory encoding. *Consciousness and cognition*,  
892 53, 194-202.
- 893 **Makowski**, (2018). The psycho Package: an Efficient and Publishing-Oriented Workflow for  
894 Psychological Science. *Journal of Open Source Software*, 3(22), 470.  
895 DOI:10.21105/joss.00470
- 896 **Mayer**, R. E. (2002). Multimedia learning. In *Psychology of learning and motivation (Vol.*  
897 *41, pp. 85-139)*. Academic Press.
- 898 **Mayer**, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence  
899 for dual processing systems in working memory. *Journal of educational psychology*,  
900 90(2), 312.
- 901 **Meiser**, T., Sattler, C., & Weißer, K. (2008). Binding of multidimensional context  
902 information as a distinctive characteristic of remember judgments. *Journal of*  
903 *Experimental Psychology: Learning, Memory, and Cognition*, 34(1), 32.
- 904 **Murakami**, H. (2009). *IQ84*. Japan, Kyoto: Shinchosha.
- 905 **Murray**, C. D., & Gordon, M. S. (2001). Changes in bodily awareness induced by immersive  
906 virtual reality. *CyberPsychology & Behavior*, 4(3), 365-371.
- 907 **Musse**, S. R., & Thalmann, D. (2001). Hierarchical model for real time simulation of virtual  
908 human crowds. *IEEE Transactions on Visualization and Computer Graphics*, 7(2), 152-  
909 164.
- 910 **Navarro**, D. (2015). *Learning statistics with R*. Adelaide, South Australia: lulu.
- 911 **Paas**, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in  
912 statistics: A cognitive-load approach. *Journal of educational psychology*, 84(4), 429.
- 913 **Paas**, F. G., & Van Merriënboer, J. J. (1994). Variability of worked examples and  
914 transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of*  
915 *educational psychology*, 86(1), 122.
- 916 **Paas**, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional  
917 design: Recent developments. *Educational psychologist*, 38(1), 1-4.
- 918 **Perdreau**, F., & Cavanagh, P. (2015). Drawing experts have better visual memory while  
919 drawing. *Journal of vision*, 15(5), 5-5.
- 920 **Picard**, R. W. (2015, March). Recognizing stress, engagement, and positive emotion. In  
921 *Proceedings of the 20th International Conference on Intelligent User Interfaces* (pp. 3-  
922 4). ACM.
- 923 **Piper**, B. F., Dibble, S. L., Dodd, M. J., Weiss, M. C., Slaughter, R. E., & Paul, S. M.  
924 (1998). The revised Piper Fatigue Scale: psychometric evaluation in women with breast  
925 cancer. In *Oncology nursing forum*. Oncology Nursing Society.
- 926 **Plancher**, G., Gyselinck, V., Nicolas, S., & Piolino, P. (2010). Age effect on components of  
927 episodic memory and feature binding: A virtual reality study. *Neuropsychology*, 24(3),  
928 379.
- 929 **Plancher**, G., Tirard, A., Gyselinck, V., Nicolas, S., & Piolino, P. (2012). Using virtual  
930 reality to characterize episodic memory profiles in amnesic mild cognitive impairment  
931 and Alzheimer's disease: influence of active and passive encoding. *Neuropsychologia*,  
932 50(5), 592-602.
- 933 **Plancher**, G., Barra, J., Orriols, E., & Piolino, P. (2013). The influence of action on episodic  
934 memory: a virtual reality study. *The Quarterly Journal of Experimental Psychology*,  
935 66(5), 895-909.
- 936 **Plancher**, G., Gyselinck, V., & Piolino, P. (2018). The Integration of Realistic Episodic

- 937 Memories Relies on Different Working Memory Processes: Evidence from Virtual  
 938 Navigation. *Frontiers in psychology*, 9, 47.
- 939 **Reingold**, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual span in expert  
 940 chess players: Evidence from eye movements. *Psychological Science*, 12(1), 48-55.
- 941 **Riva**, G. (2005). Virtual reality in psychotherapy. *Cyberpsychology & behavior*, 8(3), 220-  
 942 230.
- 943 **Schubert**, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor  
 944 analytic insights. *Presence: Teleoperators & Virtual Environments*, 10(3), 266-281.
- 945 **Setz**, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., & Ehlert, U. (2009).  
 946 Discriminating stress from cognitive load using a wearable EDA device. *IEEE*  
 947 *Transactions on information technology in biomedicine*, 14(2), 410-417.  
 948 DOI:10.1109/TITB.2009.2036164
- 949 **Shi**, Y., Ruiz, N., Taib, R., Choi, E., & Chen, F. (2007, April). Galvanic skin response (GSR)  
 950 as an index of cognitive load. *CHI'07 extended abstracts on Human factors in*  
 951 *computing systems*, 2651-2656. DOI: 10.1145/1240866.1241057
- 952 **Shin**, D. (2018). Empathy and embodied experience in virtual environment: To what extent  
 953 can virtual reality stimulate empathy and embodied experience? *Computers in Human*  
 954 *Behavior*, 78, 64-73.
- 955 **Simon**, H. A., & Gilmarin, K. (1973). A simulation of memory for chess positions. *Cognitive*  
 956 *psychology*, 5(1), 29-46. DOI:10.1016/0010-0285(73)90024-8
- 957 **Slater**, M., & Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE):  
 958 Speculations on the Role of Presence in Virtual Environments. *Presence:*  
 959 *Teleoperators and Virtual Environments*, 6(6), 603-616.  
 960 DOI:10.1162/pres.1997.6.6.603
- 961 **Smith**, G., Del Sala, S., Logie, R. H., & Maylor, E. A. (2000). Prospective and  
 962 retrospective memory in normal ageing and dementia: A questionnaire study. *Memory*,  
 963 8(5), 311-321.
- 964 **Sperduti**, M., Armougum, A., Makowski, D., Blondé, P., & Piolino, P. (2017). Interaction  
 965 between attentional systems and episodic memory encoding: the impact of conflict on  
 966 binding of information. *Experimental brain research*, 235(12), 3553-3560.
- 967 **Spielberger**, C. D. (1983). *STAI State-trait Anxiety Inventory for Adults Form Y: Review*  
 968 *Set; Manual, Test, Scoring Key*. Mind Garden.
- 969 **Steinicke**, F., Visell, Y., Campos, J., & Léculyer, A. (2013). *Human walking in virtual*  
 970 *environments* (pp. 199-219). New York, NY, USA: Springer.
- 971 **Sweller**, J. (1988). Cognitive load during problem solving: Effects on learning.  
 972 *Cognitive Science*, 12(2), 257-285.
- 973 **Sweller**, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and*  
 974 *instruction*, 8(4), 351-362.
- 975 **Sweller**, J. (1994). Cognitive load theory, learning difficulty, and instructional design.  
 976 *Learning and instruction*, 4(4), 295-312.
- 977 **Sweller**, J., Van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and  
 978 instructional design. *Educational psychology review*, 10(3), 251-296.
- 979 **Sweller**, J., Ayres, P. L., Kalyuga, S., & Chandler, P. (2003). The expertise reversal effect.
- 980 **Sweller**, J., Ayres, P., & Kalyuga, S. (2011). Measuring cognitive load. In *Cognitive load*  
 981 *theory* (pp. 71-85). Springer, New York, NY.
- 982 **Tulving**, E. (2002). Episodic Memory: From Mind to Brain. *Annual Review of Psychology*,  
 983 53(1), 1-25. DOI: 10.1146/annurev.psych.53.100901.135114
- 984 **Ulicny**, B., & Thalmann, D. (2001). Crowd simulation for interactive virtual

985 environments and VR training systems. In *Computer Animation and Simulation 2001*  
 986 (pp. 163-170). Springer, Vienna.

987 **Van Merriënboer, J. J., & Sweller, J. (2010).** Cognitive load theory in health professional  
 988 education: design principles and strategies. *Medical education*, 44(1), 85-93.

989 **Van Veen, H. A., Distler, H. K., Braun, S. J., & Bülthoff, H. H. (1998).** Navigating through a  
 990 virtual city: Using virtual reality technology to study human action and perception.  
 991 *Future Generation Computer Systems*, 14(3-4), 231-242.

992 **Viaud-Delmon, I., Ivanenko, Y. P., Berthoz, A., & Jouvant, R. (1998).** Sex, lies and virtual  
 993 reality. *Nature Neuroscience*, 1(1), 15.

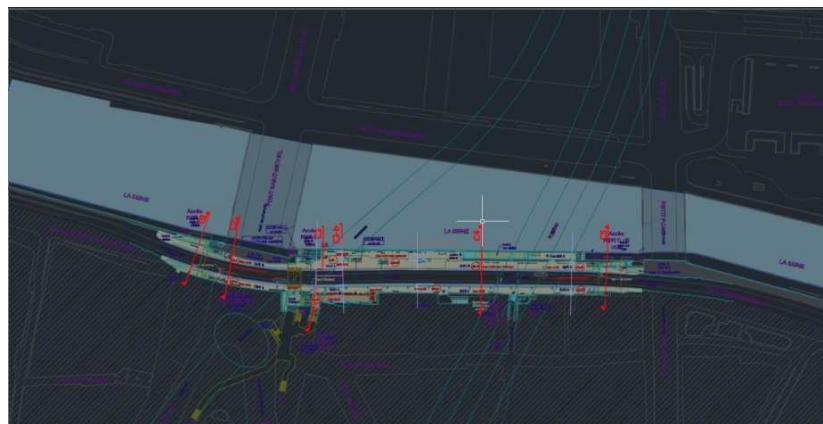
994 **Vidulich, M. A., & Tsang, P. S. (2012).** Mental workload and situation awareness. *Handbook*  
 995 *of human factors and ergonomics*, 4, 243-273.

996 **Whyte, J., Bouchlaghem, N., Thorpe, A., & McCaffer, R. (2000).** From CAD to virtual  
 997 reality: modelling approaches, data exchange and interactive 3D building design tools.  
 998 *Automation in construction*, 10(1), 43-55.

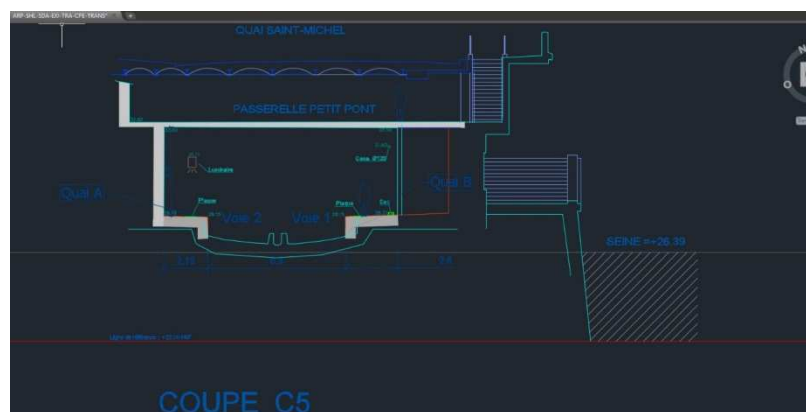
999 **Wickham, H. (2017).** The tidyverse. *R package ver. 1.1. 1.*

1000 **Wilson, B. A., Evans, J. J., Emslie, H., Alderman, N., & Burgess, P. (1998).** The  
 1001 development of an ecologically valid test for assessing patients with a dysexecutive  
 1002 syndrome. *Neuropsychological rehabilitation*, 8(3), 213-228.

1003  
 1004  
 1005 **Appendix**  
 1006



1008 *Figure 7. Top view of Saint-Michel Notre Dame RERC train blueprints*

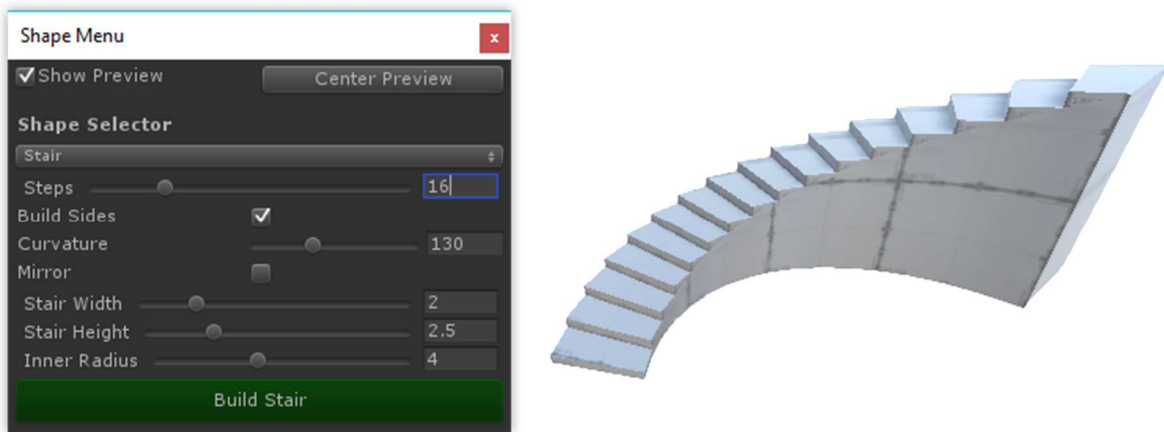


*Figure 8. Side view of Saint-Michel RERC train station blueprints*

1010

1011 We chose to use the multiplatform game interface Unity 3D® for the development of  
1012 the train station and for integration of 3D elements in the environment. For 3D construction of  
1013 the walls and buildings in VR, we used the integrated Unity plug-in Pro-Builder. This  
1014 function allowed us to easily create buildings as in figure 9 below.

1015

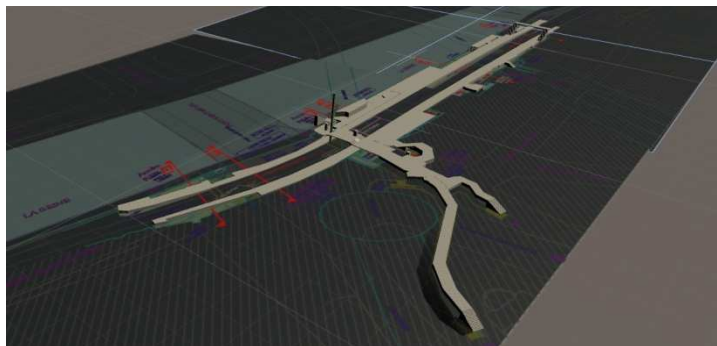


1016

1017 *Figure 9. Creation of stairs in Pro-Builder*

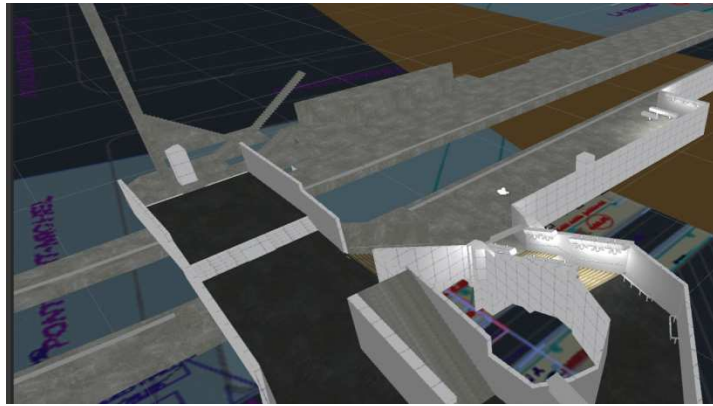
1018 We started by designing the train station floor with the designated metrics from the  
1019 architecture plans before implementing the walls with the proportioned heights and altitudes.  
1020 From pictures and videos made directly in the train station we have spotted the position of the  
1021 light system and had integrated a complex light system with adjusted brightness and contrast  
1022 as shown in figures 10i and 10ii below.

1023



1024

*Figure.10i. Creation of train station's floor from the architecture plans*



1025

Figure.10ii. Creation of train station's walls and lights from the architecture plans

1026

1027 The different objects and furniture in the train station were designed using videos,  
 1028 pictures and observations made directly in the train station. Some objects were taken from the  
 1029 assets store of Unity and from Sketchup (Sample of RER train). Some objects' textures were  
 1030 modified using 3DSMAX and animation of characters and objects were made using integrated  
 1031 functions from unity3D. The acoustic environment was designed in the virtual environment  
 1032 using recordings made in the actual train station. These recordings were, the sound of  
 1033 travelers walking in the train station, a train arriving and departing from the train station,  
 1034 conversations heard in the train station and the acoustic recording for each region in the train  
 1035 station.

1036 The contents of signage board and display screens designed in the virtual environment  
 1037 were obtained from passengers' information department of SNCF. We synchronized the time  
 1038 displayed on dynamic screens throughout the train station for a better immersion of  
 1039 participants in the train station. Passengers' flow was designed using integrated Unity plug-in  
 1040 for flow management, compatible with the avatars created to represent the travelers in the  
 1041 train station as shown in figure 11 below.

1042

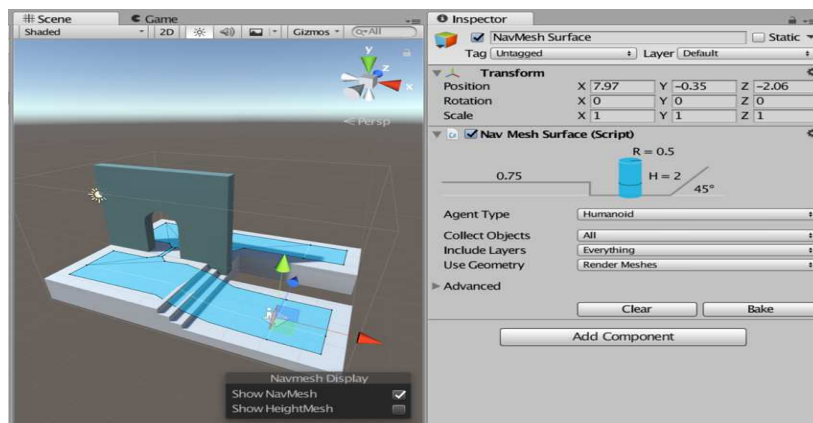


Figure.11. Navigation Mesh surface for crowd simulation

1043           Some scenarios were created for the crowd movement strategy, based on observation  
1044 made in the actual train station. Particular scenarios at strategic points in the train stations  
1045 were triggered by the experimenter using the keyboard. This was for instance, activating  
1046 passengers' flow to another train's line after an announcement was made about a cancelled  
1047 train. Modifications were made on the virtual environment until it was closed to the actual  
1048 train station.