This paper must be cited as:


The final publication is available at

https://doi.org/10.1089/cyber.2020.0560

Copyright  Mary Ann Liebert
Abstract

Alcohol use disorder (AUD) is a major global problem. Neuropsychological studies have shown that AUD causes deficits in executive functions (EF), a set of higher-order cognitive skills that govern individual behavior in every-day situations. Many standardized neuropsychological tests are used to evaluate EF. These are reliable and valid, but have limitations in predicting real-life performance. To address this, we present a preliminary study to test the Virtual Cooking Task (VCT) as an alternative to standardized neuropsychological tests. The VCT includes four subtasks developed to assess attentional, planning, and cognitive shifting abilities; it was tested in an immersive 3D environment.

To evaluate the VCT performance and standardized neuropsychological tests, data were gathered from a sample of healthy subjects (CG; n=23) and AUD patients (n=18). The standardized neuropsychological measures used consisted of questionnaires (attentional control scale, Barratt impulsiveness scale, and cognitive flexibility scale), and specific tests (Dot-probe task, Go/No-go test, Stroop test, the trail making test, and Tower of London test).

The results showed significant higher correlations for AUD patients than for the CG for the VCT, questionnaires, and specific tests, mainly related to planning and cognitive shifting abilities. Furthermore, comparative analyses of the VCT performance showed that the AUD patients made more errors and had higher latency times than the control group.

The present study provides initial evidence that a more ecologically valid assessment can be a useful tool to detect cognitive impairments in many neuropsychological and mental disorders, affecting daily activities.

Keywords: Alcohol use disorder, executive functions, virtual reality, task performance, neuropsychological assessment.
Introduction

Alcohol use disorder (AUD) is one of the main health and social problems affecting individual health and well-being; it is considered the most prevalent addiction in economically developed countries. Many previous studies into AUD showed negative effects on brain structures, leading to impaired functioning mainly in the pre-frontal and frontal areas. These areas are particularly responsible for the set of basic and higher-order executive functions (EF) that encompass the abilities to pay attention, shift or switch attention, remember, plan, inhibit behaviors, control interference, and solve problems.

Current EFs assessment include standardized global neuropsychological batteries, such as the Mini-mental state examination (MMSE), and specific tests, as the trailing making test for assessing set shifting and attention abilities, the dot-probe task, Go/No-go, and the Stroop task, used to assess attention, inhibition abilities and control interference, and the Tower of London test, to assess planning and problem-solving abilities. These tests present evidence of reliability and validity but have some limitations in terms of social desirability response bias, subjective interpretations, and ecological validity. Social desirability response bias refers to the individual’s tendency to respond to a self-report scale by presenting a favorable image of him/herself, but which may not reflect reality. Furthermore, the outcomes of paper-and-pencil tests depend on the subjective interpretations of experts that could affect the objectivity of the results. Finally, ecological validity refers to the ability of a test to predict the individual’s real-life performance; the standardized measures of EF are considered too abstract, decontextualized and incapable of capturing the real dynamic and complex performance of daily activities.
Several studies have shown that low scores in traditional measures are not associated with impaired executive behaviors in real life, and vice versa.

Virtual reality (VR) use has increased substantially over the last decade, allowing making more ecological measurements and collecting more objective data. VR is an advanced interactive computer technology able to generate non- or immersive real-simulated environments. Factors such as the number of senses stimulated, the interaction and the ability of the system to isolate the user from external stimuli contribute to the sense of immersion provided by a VR system. Non-immersive VR systems use conventional computer desktops and the interaction is via a mouse or keyboard. An immersive system displays the visual environment via a head mounted display (HMD) device and the interaction is provided by controllers or gloves; this allows the user to navigate in a simulated world and interact with the artificial objects there as if (s)he was in the real world.

Various VR applications have been developed for neuropsychological assessment – such as virtual classrooms and shopping centers - and tested on different clinical populations. For example, Cipresso et al., (2014) developed a virtual supermarket in which Parkinson’s patients, with and without cognitive impairments and a control group, have to select and buy products. The results showed that VR could discriminate between patients with and without cognitive impairments and control groups.

Other every-day activity is the act of cooking. Cooking requires the ability to plan, pay attention, remember, and the shift abilities between one task and another. Although there are few related studies, previous research has shown that virtual cooking tasks could be considered an ecologically and construct valid test to assess EF, for patient impairments, elderly adults evaluations along with traditional neuropsychological tests. For example, Craik and Lockheart (2006) developed a non-immersive cooking task
VIRTUAL COOKING TASK FOR EXECUTIVE FUNCTIONS

(CT) to test planning ability in the elderly, and showed that it was sensitive to the influence of age. Giovanetti et al.\textsuperscript{48} tested a virtual touch-screen versus a real breakfast and lunch task on younger and older adults, showing that older adults made more mistakes than younger in both conditions. Similarly, Tanguay et al.\textsuperscript{50}, using the non-immersive CT developed by Craik and Lockheart (2006)\textsuperscript{47}, compared patients with acquired brain injury with healthy subjects showing that patients presented significant difficulty to execute the CT compared to the healthy subjects.

Regarding the construct validity of the CT, Doherty et al. (2015)\textsuperscript{49} developed a similar non-immersive CT to Craik and Lockheart (2006)\textsuperscript{47}, to which they added a dual task that is, setting a table during the cooking process, and further levels of difficulty, which they tested on healthy subjects, and showed that it was able to discriminate among the EFs standardized measures. Finally, Chicchi Giglioli et al. (2019)\textsuperscript{46} developed and compared the feasibility and the sense of presence of an immersive virtual CT (VCT) versus an augmented reality CT, showing that the VCT produced a greater usability and feasibility, as well as a higher sense of presence than the augmented CT.

To our knowledge no previous studies have developed an immersive virtual CT (VCT) and tested it on AUD patients, the main aim of this study was to compare the performance data of AUD, and healthy subjects derived from both traditional EF assessments and the VCT.

Material and Methods

Subjects

The experimental sample consisted of 18 AUD patients (AUD) (7 males and 11 females; M=45.4, SD=9.83; age range: 27-62) and a control group (CG) of 23 healthy
participants (9 males and 14 females; M=44.7, SD=9.72; age range: 33-62). The AUD were recruited from the inpatient unit of a public hospital in Alicante (Spain) and the CG were recruited through local advertisements between college students and employees of the university.

The inclusion criteria for the AUD group were: (a) having an AUD diagnosis (DSM-5), (b) they had drunk alcohol within 12 months of the time of the study. The inclusion criteria for the CG were: (a) a cut-off score > 24 on the MMSE and a score ≤ 8 on the questionnaire Alcohol Use Disorders Identification Test (AUDIT).

Before participating in the study, participants of both centers received written information about the study and they were required to give written consent for the inclusion in the investigation. The study received the ethical approval of the Ethical Committees of both centers.

**Psychological assessment**

First, the MMSE was administered. MMSE is a short and validated paper-and-pencil test that measures performance in various cognitive abilities, such as orientation, attention, short-term memory, verbal fluency, and constructional apraxia. A score of > 24 is the standardized cut-off score, indicating the absence of cognitive impairments.

Second, the AUDIT was administered to the CG to assess their normal drinking behaviors. AUDIT is a 10-item screening tool able to assess alcohol consumption, drinking behaviors, and alcohol-related problems. A score of 8 or more indicates strong likelihood of hazardous or harmful alcohol consumption. Patients’ group presented AUD diagnosis administered by the public hospital in Alicante.

Third, the following questionnaires were administered to each participant:
• Attentional Control Scale (ACS)\textsuperscript{51}: evaluates individual’s attentional control through 20 questions with 4 possible answers (1 = almost never; 4 = always). Higher scores show a greater ability to maintain voluntarily attention towards a task, while low values suggest attention deficits.

• Barratt Impulsiveness Scale (BIS-11)\textsuperscript{52-53}: measures impulsiveness through 30 questions with 4 possible answers (1 = rarely or never, 4 = always or almost always). A score of 72 or more means that the individual is highly impulsive. Scores between 52 and 71 are considered within the normal range. A score below 52 suggests the subject is excessively controlled.

• Cognitive Flexibility Scale (CFS)\textsuperscript{54}: this consists of 12 questions that are scored on a 6-point scale where 1 means “totally disagree” and 6 means “totally agree”; a score of 60 or more indicates that the individual has high cognitive flexibility.

Fourth, the participants completed a total of 5 standardized tasks (ST): to assess attentional and inhibition control abilities, the dot-probe task (DOT)\textsuperscript{55}, the Go/No-go task\textsuperscript{56}, and the Stroop test\textsuperscript{57}; to assess set shifting ability, the trail making task (TMTA-B)\textsuperscript{18} was used; and the Tower of London - Drexler test (TOLDX)\textsuperscript{21} was used to evaluate planning ability. The outcomes for each ST included total and latency times and correct answers/errors. For the Tower of London, the outcomes also included execution time, excess movements, and total score.

The VCT

This virtual system was developed using Unity 5.5.1f1 software, applying c# programming language using the Visual Studio tool.
The VCT is a kitchen-based scenario consisting of four increasingly difficult subtasks (Table 1). Before tasks, a tutorial consisted of an introductory cooking task, in order to learn the main body movements and hands’ interactions using two controllers, was performed by participants. Participant could train for as long as needed and when he/she felt confident with the virtual movements and interaction, he/she pulsed a button to start the experimental tasks.

The four subtasks were based on cooking a series of foods within a set time, while avoiding burning them or allowing them to cool. The subtasks are made progressively more difficult by the introduction of various additional activities (Fig.1). Before each subtask, the system explains to participants the specific activity they should carry out, the total cooking time available, the cooking times for each food, and reminded them not to burn or let the food cool down (Fig.2). The first subtask consisted of cooking three foods on one burner for 2 minutes; the second subtask consisted of cooking 5 foods on 2 burners for a total time of 3 minutes. The third and fourth subtasks each has two tasks: in the third, the participants had to cook 5 foods on 2 burners and add the appropriate ingredient (such as salt, pepper, cinnamon, vanilla, etc.) to each food. In the fourth subtask, the participants had to cook 5 foods on 2 burners while setting also the table.

Each subtask had a total time, continuously displayed in the virtual environment and two cooking time countdowns (one graphic and one numeric) for each food appeared all the time over the pan where the food was introduced. When the countdown time finished the graphic and numeric elements over the pan appeared green indicating that the food was cooked and ready to move to the dish.

Participants passed on to the following subtask when they had completed the previous subtask. The subjects did not interrupt an activity if they did not finish it within the predefined time. The clock continued to run, capturing time taken for each task and any overruns. The
time frames for each subtask were based on the reference literature and adjusted to laboratory
pre-tests in accordance with the VR scenario and system\textsuperscript{46,47}.

The virtual system gathered: (a) the total time taken to perform the tasks; (b) the
cooking time, that was, the time that the participants took to cook each food; (c) burning
time, that was, the time that they allowed food to burn by not taking it out of the pan, or
turning the burner off, after the allotted cooking time; (d) cooling time, that was the amount
of time they left the food in the pan to cool down after it was cooked; and (e) the order in
which the foods were cooked. In addition, in the third subtask the system also recorded
whether the selected ingredient was appropriate and, in the fourth, the moment when the
participants set the table (before, during or after cooking).

**Experimental procedure**

After the subjects had given written informed consent for their participation, they
were first assessed with MMSE\textsuperscript{17}. The CG also completed the AUDIT\textsuperscript{49} questionnaire to
evaluate if their alcohol drinking behaviors were in the normal ranges. Second, the
participants completed, using personal computers, the questionnaires and the standardized
tests, randomly presented. After the neuropsychological assessment, the participants carried
out the VCT in a real kitchen, wearing an HMD device.

The VCT begins with tutorial explaining main actions (displacements and interaction
with the virtual elements). The training time of the tutorial varied according to the
participant’s confidence with the system. The total time taken to complete the experiment
was around 20 minutes for each participant.
Statistical analyses and experimental design

The analyses were performed using SPSS version 22.0 (Statistical Package for the Social Sciences for Windows, Chicago, IL) for Windows. We first verified the assumptions of normality by applying the Kolmogorov Smirnov test; the internal consistency of the scales was assessed using Cronbach’s alpha. Second, we verified the cognitive functioning of both groups using the three questionnaires. Third, the Pearson correlations were computed between the psychological questionnaires, standard task and performance in the VCT. Finally, two variance analyses (ANOVA) were performed to discover if the traditional neuropsychological assessment methods highlighted differences between the CG and AUD groups. The level of significance was set at $\alpha = 0.05$.

Results

The normality assumption was confirmed (Kolmogorov Smirnov $p > .05$) as was the internal consistency of the self-report scales (Cronbach’s alpha $\alpha_{ACS} = .839$, $\alpha_{BIS} = .816$, $\alpha_{CFS} = .757$; bootstrap 95%).

Regarding cognitive functioning (Table 2), both groups showed to be within the normal limits on maintain attentional control ($ACS: AUD= 46.8; CG= 59.6$), impulsivity, ($BIS: AUD= 67.5; CG= 58.5; normal range 52-71$) and cognitive flexibility ($CFS: AUD= 37.4; CG= 49.7$). Although, the cognitive functioning of AUD patients resulted in the normal limits, they showed lower results in maintaining attentional control and cognitive flexibility, and higher impulsivity than CG. Table 2 also reports the descriptive data on the standardized tasks. They are based on the mean scores, standard deviation, and the range values for each group.
Pearson correlations calculated for each questionnaire, standard task and the VCT performance showed significant relationships among variables (see Table 3 and 4). Specifically, AUD patients’ responses to questionnaires showed higher relationships to the VCT performance than CG. Regarding correlations between standard tasks and the VCT performance, the performance in the first subtask showed main relationships with the standard tasks in both groups. In addition, AUD patients’ performance in the second and third subtasks of the VCT showed high relationships with attention, cognitive flexibility and planning abilities.

On behavioral data, two ANOVAs were performed to discover if the traditional neuropsychological assessment methods highlighted statistically significant differences between groups. First, in the analysis of the questionnaires, the differences between the means of the groups were significant. This can be observed in more detail in Table 5.

Second, the analysis of the standardized tests showed significant differences between the groups. The dot-probe and the Go/No-go tasks did not show significant differences, while the other tasks showed one or more variables able to differentiate between the groups. In Table 6 we see that in the TMT task the total time variable was significant; it shows that the AUD group took longer than the CG; in the Tower of London test the execution and total time variables indicate differences between the groups. The Stroop task showed differences between the groups in latency and total time.

In the VR task, the groups statistically differed in all subtasks. In more detail, the total time taken to complete the 4 subtasks showed significant differences between the groups \( F = 8.565, p <0.01 \), and in the mean time of the 4 levels \( F = 10.957, p <0.01 \). We found similar results in the total times taken to complete each subtask, and in the order that the food was cooked at levels 2 and 4. Table 7 shows the significant results.
Discussion

EF impairments are common in AUD and are linked to significant daily-life
dysfunctions. Traditional measures showed some limitations in predicting real-life
performance, and the main aim of this study was to examine the potential of a VCT for EF
assessment.

Our results showed that, although both groups presented a cognitive functioning
within the normal limits, AUD patients showed a lower functioning than CG, and the
correlation results showed moderate to high relationships between standardized
neuropsychological tests and the VCT. Regarding questionnaires, higher relationships were
found in the AUD group in attention control, impulsiveness and cognitive flexibility than in
the CG. More specifically, attention control and cognitive flexibility were inversely related
to the attention paid to food that was burning while the other foods were being cooked.
Significant relationships were found between greater impulsiveness and a higher likelihood
of burning food and spending more time to finalize the tasks. Similar relationships were
found between the traditional attention control tasks (DOT and STROOP), planning (TOL
task) and cognitive flexibility (TMT-AB task) and total and burning times for the AUD
group. In accordance with this result, the previous literature on alcohol-dependence showed
an affectation of various cognitive processes, such as attention, cognitive flexibility,
problem-solving, planning abilities, as well as impairments in inhibitory control of
impulsivity.

Regarding the variance analyses in the traditional questionnaires, the findings showed
that, the AUD group, had lower attention control and cognitive flexibility scores but higher
impulsiveness scores than CG. Furthermore, the AUD lower scores were greater than CG in
tests requiring planning, attention control, and cognitive flexibility. These results are coherent with the previous literature that has demonstrated behavioral impairments on multiple EFs with a particular focus on impulsivity as the main factor in the ability to control and inhibit responses goal-directed\(^9\). The no cognitive control involves a variety of behaviors, as acting without planning, difficulty to pay attention, and not considering all information to execute a task. Behavioral impairments have been demonstrated to depend on a dysfunctional integrity of brain areas involved in cognitive control, including, among others, the dorsolateral prefrontal cortex, lateral orbitofrontal cortex and the anterior cingulate cortex, regulated goal-directed behaviors\(^5\text{-}^7\).

Finally, regarding performance comparison, our findings showed that VCT can discriminate between CG and AUD. Specifically, the time factor variables significantly differed between CG and AUD group, both between time taken to complete each level and by the total time to perform the task, as well as by the burning time. These results seem to suggest, as demonstrated by similar previous studies\(^50\text{-}^51\), that patients with lower executive functioning completed the VCT tasks slower than individuals with normal executive functioning.

Our study also evaluated how subjects planned to cook the foods, since a certain cooking order allowed the subjects to carry out the tasks in less time. Our results for levels 2 and 4 showed that AUD patients had less tendency to plan the optimal cooking order and simply prepared the food according to the order of its appearance on the table. A further difference in planning and cognitive flexibility was observed in subtask 4: AUD patients set the table before, or after, cooking all the food, whereas the CG set the table while the food was cooking, thereby reducing the overall time. Similar results have been found by Tanguay et al. (2014)\(^50\) on a similar task performed by brain injury patients.
As to attention and cognitive flexibility abilities, in subtask 3 the AUD group had longer cooking and total times than the CG. These results seem to suggest that the introduction of new activities can affect the ability to plan a sequence of actions goal-directed of the AUD group more than the performance of the CG.

The present study, according to a clinical perspective and implication, attempted to explain the relation between the VCT measures and basic and higher-order cognitive processes such as attention, control inhibition, cognitive flexibility and planning on the other. The VCT involved rules and sequential steps that reproduce those required in real life meal preparations, providing ecological validity and the possibility to be applied to a wide range of populations and sensitive to various neuropsychological impairments. The results can provide guidelines for the assessment of these processes in ecologically valid settings, as well as an enjoyable and engaging evaluation of everyday behaviors. Furthermore, the possibility to use the VCT for various neuropsychological impairments could reduce costs and waiting lists, enhancing the functional recovery of these processes.

While the findings of this study are interesting and valuable, it has some limitations. First, the small sample size, as well as gender imbalance, might limit the generalizability of the results. Second, it is important also to assess the individual’s perception of the usability of the VCT (e.g., difficulty using controllers, picking up the foods, cooking, and learning to move within the environment). A last limitation relates to the final scoring. Indeed, traditional neuropsychological tests are corrected for age and the VCT results should be viewed in accordance with these corrections. To address these limitations, future studies are needed to: (a) explore the relative impact of age, gender, and education on VCT performance; (b) evaluate test-retest reliability and temporal stability.
Conclusions

VR is overcoming limitations of traditional measures by facilitating the development of more ecological performance-based environments and generating more accurate control stimuli and data than traditional methods. Indeed, the VCT was contextualized in a real situation that people face every day and subjects were evaluated stealthily and more objectively than traditional assessment, reducing also social desirability bias. In conclusion, this study offers initial evidence that more ecologically valid assessments can be also useful, alongside standard assessments, for detecting functional cognitive impairments with respect to daily activities in AUD patients.

References

5. Harris GJ, Jaffin SK, Hodge SM, Kennedy D, Caviness VS, Marinkovic K, ... & Oscar-Berman M. Frontal white matter and cingulum diffusion tensor imaging


33. Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philosophical Transactions of the Royal Society B: Biological Sciences 2009; 364(1535): 3549-57.


