[Page de garde]

HEC MONTRÉAL

Vers un design optimal pour l'amélioration des capacités et résultats d'apprentissage de simulateurs pour l'entrainement de professionnels en médecine et en aviation

par

Laurie-Jade Rochon

Pierre-Majorique Léger, Ph.D., Constantinos K. Coursaris, Ph.D., Sylvain Sénécal, Ph.D.
HEC Montréal
Codirecteurs de recherche

Sciences de la gestion
Spécialisation expérience utilisateur dans un contexte d'affaires

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences en gestion
(M. Sc.)

Août 2022

Résumé

Les principes de design de Mayer issus de la théorie cognitive de l'apprentissage multimédia[1] (TCAM) ont précédemment été élaborés afin de fournir des lignes directrices pour la conception d'environnements multimédia adaptés pour l'apprentissage. Les simulateurs d'entrainement destinés à la formation de professionnels comportent les caractéristiques des environnements multimédia. En outre, ils présentent des interfaces utilisateur (IU) élaborés desquels émanent des interactions complexes. Ces interfaces doivent être conçues dans le but de favoriser un apprentissage optimal.

L'objectif de cette recherche était de répondre à la question suivante : Dans quelle(s) modalité(s) sensorielle(s) le matériel de formation d'une simulation sur ordinateur doit-il être présenté pour permettre d'améliorer significativement les états cognitifs d'apprentissage et la performance en contexte de formation professionnelle ?

Ce mémoire présente une étude réalisée en deux phases, au cours desquelles des évaluations utilisateurs des simulateurs d'entrainement dans deux secteurs distincts furent effectuées : une simulation pour la formation de médecins anesthésistes, ainsi qu'un simulateur de vol sur ordinateur pour la formation de pilotes de l'air.

La Phase I de ce projet était une étude de faisabilité de type exploratoire réalisée à distance qui consistait à évaluer les expériences explicite et implicite de 10 étudiants en médecine alors qu'ils suivaient le didacticiel d'un simulateur en anesthésie. Précisément, nous évaluions comment le design multimédia de l'interface impactait ses capacités d'apprentissage[2] et en conséquence, l'expérience des sujets. Le didacticiel était suivi d'une série de tâches de rappel qui furent sous-catégorisées pour tenter de déceler des corrélations intéressantes (e.g., tâche informationnelle versus tâche transactionnelle). Les résultats de la Phase I suggèrent la pertinence d'utiliser une approche multidimensionnelle pour mesurer l'expérience utilisateur de participants pendant l'encodage (i.e., le didacticiel) et pendant le rappel (i.e., tâches de rappel). Une corrélation négative fut identifiée entre les résultats d'auto-efficacité et de charge cognitive perçues. Les résultats

[1] Traduction de « Cognitive Theory of Multimedia Learning (CTML) »
[2] Traduction de « Learnability »

d'une analyse qualitative des points de frictions de valence émotionnelle s'appuyant sur la TCAM suggèrent que les modalités sensorielles sélectionnées pour présenter du matériel de formation dans un simulateur impacte l'expérience d'apprentissage de professionnels.

Au cours de la Phase II réalisée en laboratoire avec 30 étudiants pilotes d'avion, nous avons manipulé la modalité sensorielle de présentation des éléments d'interface d'un simulateur de vol dans trois groupes expérimentaux. Au cours de deux scénarios de vol suivants les « Règles de Vol à Vue »[3] (RVV), des directives de vol étaient transmises via *audio et texte* (i.e., bimodal), *audio* (i.e., mode auditif), ou *texte* (i.e., mode visuel) aux participants. Les états cognitifs d'apprentissage évalués pendant un vol d'entrainement et pendant un vol d'évaluation des acquis incluent la charge cognitive, l'attention visuelle, la motivation et l'état affectif des pilotes. Les résultats de la Phase II suggèrent que l'utilisation d'instructions de vol sous forme de *texte* est plus appropriée que des instructions de vol *audio* ou *audio et texte* dans un simulateur de vol lorsque l'attention des sujets est divisée entre les tâches d'opération de l'aéronef et entre la réception et le traitement des instructions de vol.

Ce mémoire contribue à la littérature sur le design de didacticiel et de matériel de formation pour des simulateurs d'entrainement destinés à des professionnels. Il contribue à spécifier les limites de la CTML. Finalement, nous proposons un cadre méthodologique multidimensionnel pour mesurer l'expérience d'utilisateurs dans un environnement d'interactions complexes qui puisse être répliqué à distance et lors de tests utilisateurs.

Mots clés : Apprentissage Multimédia • Capacité d'Apprentissage • Expérience Utilisateur • Reconnaissance Automatique d'Expressions Faciales • Entropie de Transition du Regard • Attention Focale/Ambiante • Ingénierie pédagogique • Simulation • Formation

[3] Traduction de « Visual Flight Rules, (VFR) »

Abstract

Mayer's design principles from the Cognitive Theory of Multimedia Learning (CTML) have been developed to provide guidelines for the design of multimedia environment suitable for learning. Training simulators for professionals include the characteristics of multimedia environment. They may include elaborate user interface (UI) from which emerge complex interactions. These interfaces must be designed to promote optimal learning. The purpose of this research is to answer the following question: What sensory modality(s) should the instructional material of computer-based simulation be displayed to enable a significant improvement in cognitive learning states and learning performance in a professional training context?

This thesis presents a two-phase study during which user evaluations of training simulators were conducted in two distinct areas: a computer-based simulator for anesthesiologist's training, and a computer-based flight simulator (FS) for airline pilot's training.

The Phase I of this project was a remote exploratory/feasibility study that aimed at evaluating the explicit and implicit experiences of 10 medical students as they followed the tutorial of an anesthesia simulator. Specifically, we assessed how the multimedia interface design impacted its learnability capabilities, and consequently, the overall subjects' experience. The tutorial was followed by recall tasks that were subcategorized in an attempt to detect interesting correlations (e.g., informational versus transactional task). The Phase I results suggested the relevance of using a multidimensional approach to measure participant user experience, both during encoding (i.e., the tutorial) and during recall (i.e., recall tasks). In addition, a negative correlation was identified between self-efficacy and perceived cognitive load scores. A qualitative analysis of valence pain points (VPPs) realized through the lens of the CTML suggests that sensory modalities selected to provide training material in a simulator impacts the learning experience of professionals.

During the Phase II conducted in the laboratory with 30 student pilots, we manipulated the sensory modality of various FS interface elements, in three experimental groups. In two flight scenarios following the Visual Flight Rules (VFR) set of regulations, flight instructions were transmitted via *audio and text* (i.e., bimodal), *audio* (i.e., auditory mode), or *text* (i.e., visual mode) to students. The cognitive learning states of subjects assessed included cognitive load, visual attention, motivation, and affect. Results of Phase II suggest that providing pilots with *text*-based flight instructions is more effective than *audio* or *audio and text*-based flight instructions, if a pilot's attention is split between aircraft operating tasks and the processing of flight instructions.

This work contributes to the literature on the design of tutorial and instructional material for professional training simulators. It contributes to the specification of the boundaries of the CTML. Finally, we propose a multidimensional methodological framework to assess the experience of users in a highly interactive environment that can be replicated remotely and during user testing.

Keywords : Multimedia Learning • Learnability • User Experience • Automated Facial Expression Analysis • Gaze Transition Entropy • Ambient/Focal Attention • Instructional Design • Simulation • Training.

Table des matières

Liste des figures et des tableaux

La numérotation des figures et tableaux a été modifiée dans les articles pour faciliter la lecture de ce mémoire.

Liste des figures

Liste des tableaux

Liste des formules

Liste des abréviations

Abréviation	Signification
3D	Tridimensionnel (Three-dimensional)
AD	Avancé (Advanced)
ANOVA	Mesures répétées d'analyses de variance (Repeated measures ANalysis Of VAriance)
AOI	Aire d'intérêt (Area of Interest)
APA	Agent pédagogique animé (Animated Pedagocical Agent)
ATC	Contrôleur de trafic aérien (Air Traffic Controller)
EV/VE	Environnement virtuel (Virtual Environment)
FPM	Mode de jeu libre (Free Play Mode)
FS	Simulateur de vol (Flight Simulator)
GTE	Transition de l'entropie du regard (Gaze Transition Entropy)
IEQ	Questionnaire d'expérience immersive (Immersive Experience Questionnaire)

INT	Intermédiaire (Intermediate)
IU/UI	Interface utilisateur (User Interface)
L-SES	Échelle d’auto-efficacité perçue d’apprentissage (Learning Self-Efficacy Scale)
MFS	Microsoft Flight Simulator (Microsoft Flight Simulator)
NASA-TLX	Indice de charge des tâches de la NASA (NASA-Task Load Index)
NOV	Novice (Novice)
OTW	Hors de la fenêtre (Outside The Window)
PCPD	Pourcentage de changement du diamètre de la pupille (Percentage Change in Pupil Diameter)
PSSUQ	Questionnaire d’utilisabilité post-étude du système (Post-Study System Usability Questionnaire)
RV/VR	Réalité virtuelle (Virtual Reality)
RVV/VFR	Règles de vol à vue (Visual Flight Rules)
SCR	Stimulus – Traitement central – Réponse (Stimulus - Central processing – Response)
SIMS	Échelle de motivation situationnelle (Situational Motivation Scale)
TCAAM/ CATLM	Théorie cognitive-affective de l’apprentissage avec les médias (Cognitive-Affective Theory of Learning with Media)
TCAM/ CTML	Théorie cognitive de l’apprentissage multimédia (Cognitive Theory of Multimedia Learning)
TCC/CLT	Théorie de la charge cognitive (Cognitive Load Theory)
TFH	Heures totales de vol (Total Flight Hours)
UX	Expérience utilisateur (User eXperience)
VPP	Points de friction de valence (Valence Pain Points)

Avant-propos

Avec l'autorisation de la direction administrative du programme de la Maîtrise ès Science en Gestion, ce mémoire en expérience utilisateur dans un contexte d'affaires prend la forme de deux articles portant sur l'évaluation du design de simulateurs à haute validité écologique pour la formation et l'entrainement de professionnels.

Les articles ont été ajoutés à ce mémoire avec le consentement signé des coauteurs. Le premier article est une évaluation des capacités d'apprentissage du didacticiel d'introduction d'un simulateur sur ordinateur pour la formation de médecins anesthésistes. Cet article utilise une approche méthodologique multidimensionnelle pour mesurer l'expérience utilisateur d'étudiants en médecine, et ce, à distance, pour informer le design d'interface en se basant sur la TCAM. Cette étude préliminaire a permis d'identifier le principe de modalité tiré de la TCAM tel qu'un principe potentiellement influent sur l'expérience d'apprentissage de professionnels dans un simulateur d'entrainement. Le second article évalue comment l'inclusion d'instructions de vol de diverses modalités sensorielles dans le syllabus d'un simulateur de vol destiné au public influence les états cognitifs d'apprentissage et la performance d'apprentissage de pilotes de l'air à l'entrainement, tel que décrit par leur charge cognitive, leurs états attentionnel, motivationnel et affectif.

Pour chaque article, une introduction relative aux concepts de psychologie de l'apprenant et de l'expérience utilisateur est effectuée. L'approche méthodologique et son implémentation sont détaillées; les résultats obtenus sont analysés et mis en relations avec des concepts théoriques complémentaires. Finalement, les implications pour les concepteurs de simulateurs sont discutées.

Remerciements

J'aimerais remercier quelques personnes qui ont contribué au succès de ce projet de recherche.

Pierre-Majorique Léger, Constantinos K. Coursaris et Sylvain Sénécal, merci de m'avoir fait confiance dès mes débuts au Tech3Lab en m'attribuant un projet stimulant. Je suis reconnaissante d'avoir pu faire partie d'une équipe à la fois créative, professionnelle et complice. Vous m'avez appris le travail ardu, mais efficace, dans un environnement de recherche rigoureux. Vous êtes des modèles de professeurs, chercheurs, gestionnaires et collègues sans précédents. Il fut sensationnel de pouvoir apprendre à vos côtés au quotidien.

Alexander Karran, merci d'avoir été pour moi un mentor tout au long du processus d'idéation et de rédaction. Votre habileté à continuellement remettre en perspective le statut quo fait de vous un chercheur hors-pair.

Frédérique Bouvier, merci de m'avoir permise de développer ma confiance professionnelle. Ton calme, ta prestance et ton empathie sont à la fois rassurants et inspirants. Les opportunités que tu m'as offertes ont fait de mon passage au Tech3Lab, une expérience hautement formative. J'en ressors définitivement grandie.

Merci aux collaborateurs du Tech3Lab, Shang Lin Chen, François Courtemanche, Salima Tazi, Emma Rucco, David Brieugne et aux assistants de recherche, sans qui la réalisation de ce projet d'envergure n'aurait été possible. Un merci particulier à Jean-François Delisle chez CAE Inc. pour la confiance. Je remercie également l'école de pilotage Cargair Ltée pour la précieuse collaboration. Merci à mes amis et coéquipiers de classes virtuelles – Shady, Tanguy et Xavier. Je remercie mes proches pour leurs aide et soutien : mes parents – Nancy et Patrick – et mes autres amis du quotidien – Philippe, Daphné, Laurie-Délia et Xavier. Finalement, je tiens à remercier le CRSNG ainsi que la chaire de recherche industrielle NSERC-Prompt en UX pour le soutien financier tout au long de ma maitrise.

Chapter 1. Introduction

Les experts prévoient une augmentation de la taille du marché de l’entrainement et de la simulation de réalité virtuelle (RV) avec un taux de croissance annuel composé de 13,30%, passant d’une évaluation de 336,8 millions USD en 2021 pour atteindre 1036,2 millions USD d’ici 2030 dans les industries de la santé, de la défense et des forces armées, du transport et du sport (Verified Market Research, 2022). Cet engouement s’explique par les avancées technologiques récentes dans le domaine, par l’optimisation des outils d'apprentissage en ligne et par l’amélioration de notre compréhension des processus d’apprentissage (Verified Market Research, 2022). Les plus récents simulateurs de formation en RV incluent des technologies qui monitorent et classifient les actions humaines pour mettre à la disposition de l'utilisateur un environnement interactif et immersif.

Outre l’amélioration continue de la puissance des simulateurs, plusieurs motifs expliquent l’augmentation en popularité de la technologie pour la formation de personnel qualifié. D’abord, les simulations permettent un apprentissage dit « expérientiel » qui permet l’acquisition rapide et le maintien prolongé de connaissances dans la mémoire des utilisateurs grâce à la participation, à la réflexion et à l’intégration multisensorielle (Fowler, 2008). Dans plusieurs domaines d’expertise, les tâches à effectuer au travail sont procédurales motrices. Elles peuvent être apprises et peaufinées grâce à la mémoire musculaire. La formation par la simulation est également engageante, motivante et divertissante (Makransky et Petersen, 2019). Ainsi, éliminer les formats traditionnels fastidieux peut être une stratégie efficace pour augmenter le temps passé par les professionnels à s’entrainer de manière autonome et pour augmenter l’efficacité de l’entrainement (e.g., Egle *et al.* (2015)).

Ensuite, les simulations pour l’entrainement permettent de pratiquer des scénarios rares, risqués, complexes ou coûteux (Riener et Harders, 2012). Il peut arriver que les apprenants ne soient pas encouragés à expérimenter leur rôle professionnel dans le monde réel, car ils pourraient nuire à l’image de l’entreprise, à l’équipement ou à des individus. Dans un environnement simulé, une rétroaction instantanée permet la correction des erreurs, facilitant l'apprentissage. De surcroit, en vivant les conséquences simulées de leurs actions, les apprenants peuvent mieux comprendre pourquoi la bonne application des procédés au travail est fondamentale.

Finalement, pour les organisations, les simulations d'apprentissage sont des outils efficaces pour changer et standardiser le comportement des apprenants dans un environnement hautement contrôlé. En outre, il s'agit d'une opportunité pour que les entreprises collectent des données pour l'analyse de la performance. Les entreprises qui comptent un grand nombre de professionnels à former peuvent trouver coûteux de mettre en œuvre des programmes de formation traditionnels à l'échelle organisationnelle. En outre, les montants reliés à divers facteurs (e.g., disponibilité de formateurs, de l'équipement technique ou des lieux de formation, disposition géographique, etc.) peuvent s'élever au-delà du plafond de dépenses engendré par l'achat d'équipements, de licences ou pour l'implémentation de programmes pour la formation à l'aide de simulations d'entrainement. Néanmoins, les comités décisionnels d'organisations devraient entreprendre des analyses de coût-efficacité pour assurer le retour sur investissements des programmes d'entrainement de simulation en évaluant les besoins en ressources, les coûts associés et les résultats ultérieurs (Hippe *et al.*, 2020; Isaranuwatchai *et al.*, 2014).

La question de l'expertise

L'utilisation de simulations en RV a un grand potentiel pour l'entrainement de professionnels pour deux raisons : l'entrainement de professionnels est dispendieux et l'expertise des professionnels est indispensable pour la société. L'utilisation de simulations d'entrainement peut aider à réduire les coûts et le temps nécessaire pour l'entrainement et la formation, bénéficiant à la société.

Les professionnels ont besoin d'entrainements théoriques, pratiques et techniques sophistiquées. Leurs connaissances et compétences leur sont enseignées par des spécialistes séniors. La rareté de ces experts, leur demande et les coûts qu'ils engendrent créent ainsi une demande croissante pour le développement de matériel de formation sophistiqué. L'utilisation de telles simulations est une opportunité pour que les professionnels en formation se familiarisent avec leur environnement de travail et son équipement; pour qu'ils augmentent les heures de pratique et la variété des scénarios expérimentés. Ce faisant, les professionnels en formation gagnent de la confiance avant de passer dans un environnement de travail ou la pression et le stress sont souvent élevés, ayant un impact positif sur leur hygiène de santé mentale (Martin, 1986).

Afin de rendre bénéfique l'utilisation de tels simulations d'entrainement pour la formation de professionnels, les systèmes doivent être conçus de manière à favoriser un apprentissage maximal tout en misant sur l'efficacité (Hippe *et al.*, 2020). L'efficacité se réfère à la capacité du système à permettre aux utilisateurs d'atteindre leurs buts en ne dépensant qu'un minimum de ressources, mesuré tel que les ressources dépensées par l'utilisateur en relation à l'exactitude et à la complétude des buts atteints (ISO, 2018). Pour la formation de professionnels en demande, la demande temporelle est un enjeu important. Les simulations et les leçons qu'elles contiennent doivent permettre l'atteinte de résultats d'apprentissage élevés dans un temps restreint. Pour ce faire, les concepteurs peuvent manipuler différentes variables, fonctions et composantes d'interface tout en se conformant aux caractéristiques du système humain du traitement de l'information.

Pour de prendre des décisions de design éclairées, les concepteurs s'inspirent des théories de la psychologie de l'apprenant et de l'ingénierie pédagogique en interactions homme-machine[4]. Parmi celles-ci, nous retrouvons la « Théorie de la charge cognitive[5] (TCC) » (Sweller, 2011), la « Théorie cognitive de l'apprentissage multimédia[6] » (Richard Mayer et Mayer, 2005), et la « Théorie cognitive-affective de l'apprentissage avec le média (TCAAM) » (Moreno, 2005, 2006, 2007, 2009) sur lesquelles s'appuient ce travail.

La TCC décrit les implications de la conception pédagogique d'un modèle d'architecture cognitif humain fondé sur un bagage de connaissances permanentes (i.e., la mémoire à long terme) et un processeur conscient et temporaire de l'information (i.e., la mémoire de travail) (Sweller, 2011). Les caractéristiques essentielles de la mémoire de travail sont sa

---

[4] Traduction de « Human-Computer Interactions (HCI) »
[5] Traduction de « Cognitive Load Theory (CLT) »
[6] Traduction de « Cognitive Theory of Multimedia Learning (CTML) »

capacité et sa durée limitées. Si ces limites sont dépassées, la mémoire de travail devient surchargée et l'apprentissage est inhibé (Baddeley, 1992).

La TCAM soutient que le multimédia est compatible avec la manière dont le cerveau humain traite l’information et apprend. Les individus apprennent plus en profondeur à partir de mots et de photos qu’à partir de mots seulement (i.e., principe de multimédia) (Butcher, 2014). Le multimédia est défini tel qu’une combinaison de texte et images ou les « mots » peuvent être écrits ou parlés et les « images » peuvent consister d’illustrations, de photos, d’animations, de vidéos, etc. La théorie suggère que l’apprentissage multimédia survient lorsque des représentations mentales des mots et des images sont créées et intégrées ensemble (Richard Mayer et Mayer, 2005).

La TCAAM reprend les concepts de cognition de la TCAM et les élargit pour prendre en considération le rôle des facteurs motivationnels et affectifs lors de l’apprentissage (Moreno, 2006, 2007). La motivation se réfère à « l’état interne qui initie, maintient et énergise les efforts de l’apprenant à s’engager dans des processus d’apprentissage » (Richard E Mayer, 2014).

Alors que les TCAM et TCAAM furent largement étudiées pour décrire l’apprentissage faisant intervenir plus d’un médium (Richard Mayer et Mayer, 2005; Moreno, 2005, 2006, 2007), aucune théorie ne fut spécifiquement conçue pour décrire l’apprentissage très caractéristique qui survient dans un simulateur sur ordinateur avec du contenu complexe destiné aux professionnels. Les simulateurs d’entrainement en RV ont un haut niveau d’interactivité incluant divers éléments et acteurs. L’information à traiter provient de diverses sources (e.g., environnements réel et virtuel, équipements réel et virtuel, instructeurs réel et virtuel) et médiums (e.g., texte, image, contenu interactif), sollicitant tous les sens humains. Pour la formation d’experts, l’opération de l’équipement spécialisé et la prise de décision requièrent un haut niveau de compétences résultant d’années d’apprentissage.

Questions de recherche

L'utilisation des simulateurs en RV pour l'entrainement de professionnels en médecine et en aviation comporte plusieurs avantages, notamment la réduction des coûts et l'accélération du processus d'entrainement et de formation pouvant faire bénéficier la société (Ortiz, 1994).

Les théories de l'apprentissage sont le fondement d'une expérience éducative efficace. La complexité des interactions entre un expert et un simulateur et le manque d'une théorie unifiée pour l'apprentissage lors de l'entrainement dans un simulateur en RV oblige les designers en expérience utilisateur (UX) et IU à se référer à diverses théories pour concevoir des environnements virtuels immersifs, fidèles et qui facilitent l'apprentissage. Les simulations en RV comportent les caractéristiques des environnements multimédias. Dans ce travail, nous examinons les principes de design de l'apprentissage multimédia issus de la TCAM et de la TCAAM (Richard Mayer et Mayer, 2005; Moreno, 2005, 2006, 2007) appliqués à des simulations en RV destinées aux professionnels. Pour ce faire, nous réalisons deux évaluations d'utilisabilité employant une recherche de méthodes mixtes incluant des mesures psychométriques et psychophysiologiques.

L'Article 1 présente une étude préliminaire nous ayant permis d'évaluer les principes d'ingénierie pédagogiques des simulateurs de formation et d'entrainement pour professionnels. Dans cet article, nous analysons l'expérience d'étudiants en médecine introduits à un nouveau simulateur d'entrainement pour anesthésistes, sous l'angle des principes de design de l'apprentissage multimédia de Mayer. Nous tentons de répondre à la question de recherche préliminaire suivante :

QR préliminaire : Quelle caractéristique du design d'une simulation en RV tridimensionnelle (3D) sur ordinateur peut-elle influencer significativement l'expérience utilisateur de professionnels et les capacités d'apprentissage d'un simulateur?

Dans la seconde étude présentée dans ce mémoire, la caractéristique d'interface sélectionnée lors de l'étude 1 est manipulée. Cette caractéristique d'intérêt est la modalité sensorielle des instructions de vol incluses dans le scénario de tâche de la simulation. L'expérience de pilotes d'avion est évaluée au cours d'un scénario de tâche d'entrainement en RVV. Nous tentons de répondre à la question de recherche principale suivante :

QR principale : Dans quelle(s) modalité(s) sensorielle(s) le matériel de formation d'une simulation sur ordinateur doit-il être présenté pour permettre d'améliorer significativement les états cognitifs d'apprentissage et la performance d'apprentissage de professionnels?

Objectifs de l'étude

Ce mémoire par article a pour objectif préliminaire l'évaluation d'un simulateur pour l'entrainement de professionnels à la lumière des principes de design de l'apprentissage multimédia de Richard Mayer et Mayer (2005). L'objectif principal de cette recherche est de déterminer dans quelle(s) modalité(s) sensorielle(s) le matériel de formation d'une simulation sur ordinateur doit être présenté pour permettre d'améliorer significativement les états cognitifs d'apprentissage et la performance en contexte de formation professionnelle. Faisant ainsi, nous espérons définir et/ou confirmer les conditions d'application des concepts de la TCAM aux simulations en RV 3D sur ordinateur au travers d'un angle théorique élargi reprenant divers concepts de la psychologie de l'apprenant, de l'ingénierie pédagogique et des interactions homme-machine.

Ce projet utilisera une recherche de méthodes mixtes pour faire l'évaluation des interactions entre un expert et une simulation en RV 3D sur ordinateur. Nous y développons une démarche robuste pour effectuer une sélection de mesures permettant la captation d'un spectre multidimensionnel et élargi d'états cognitifs d'apprentissage (i.e., des processus cognitifs, attentionnels, motivationnels et affectifs). En utilisant cette approche méthodologique, nous espérons contourner les limites du seul recours aux méthodes psychométriques ou physiologiques. De plus, nous y dépeignons un portrait de l'expérience des utilisateurs qui sache indiquer aux concepteurs UX/IU comment optimiser les simulations en RV 3D sur ordinateur pour l'entrainement et la formation.

Structure du mémoire

Ce travail prend la forme d'un mémoire par articles. Pour écrire ce travail, deux projets de recherche furent menés à terme successivement. Suite à la complétion de chaque étude, un article fut rédigé afin d'être soumis à une revue scientifique. Les deux projets de recherche testent l'expérience de professionnels en contexte de formation et d'entrainement alors qu'ils utilisent un simulateur de vol en RV 3D sur ordinateur. Les deux études consistent de recherches de méthodes mixtes telles que détaillées par O'Brien et Lebow (2013). Les méthodes mixtes sélectionnées incluent des mesures subjectives (auto-rapportées) et objectives (comportementales et physiologiques) afin de couvrir les aspects pragmatiques et hédoniques de l'expérience des sujets (O'Brien et Lebow, 2013).

Dans l'introduction, nous définissons la problématique et son contexte. Nous énonçons les questions ayant guidé les deux études réalisées et spécifions des objectifs de recherche clairs. Les cadres théoriques ayant guidé les deux études (i.e., TCC, TCAM and TCAAM) sont introduits succinctement dans l'introduction, puis ils sont détaillés dans les deux articles subséquents. Les travaux de recherche sont précédés d'une brève section d'information mettant en contexte les étapes de recherche et de rédaction pour l'Article 1 et pour l'Article 2.

Le premier article évalue l'expérience utilisateur de professionnels de la santé alors qu'ils sont introduits au didacticiel d'un nouveau simulateur d'entrainement. Nous y générons des recommandations de design pour améliorer les capacités d'apprentissage du système. Le second article évalue l'expérience d'entrainement et d'apprentissage de professionnels du domaine de l'aviation alors qu'ils testent pour la première fois un module de formation de RVV d'un simulateur d'entrainement au vol. Nous formulons des recommandations de design pour améliorer les états cognitifs et résultats d'apprentissage des pilotes.

Nous concluons ce mémoire en rappelant nos questions de recherche, puis nous mettons en commun les principaux résultats des deux études. Nous poursuivons en étayant nos contributions théoriques, pratiques et méthodologiques, puis terminons en soulevant quelques limites et pistes de recherche futures.

Méthodologie de recensement des écrits

Pour la réalisation de ce mémoire, la méthode employée pour recenser les écrits était dépendante de l'étape de recherche en cours. Une revue de la littérature initiale a servi à identifier et spécifier les sujets et questions à investiguer. En utilisant le moteur de recherche librement accessible Google Scholar et la bibliothèque en ligne d'HEC Montréal, des mots-clés tels que « training », « virtual reality », « simulation » et « learning + simulation » furent utilisés. Cette méthode d'identification d'articles principaux a permis la création d'une banque d'articles exhaustive par effet boule de neige : les travaux importants cités dans un article étaient consultés, puis sélectionnés ou éliminés lorsque jugés pertinents ou non. Des critères de sélection primordiaux incluaient la date de publication (i.e., 2010 ou plus récent), le nombre de citations (i.e., 50 ou plus), la disponibilité de l'article en accès libre ou via HEC Montréal, la complétude de la revue de la littérature ou la robustesse de l'expérimentation, etc. En outre, nous avons recensé quelques travaux de revue sur divers sujets, notamment sur des théories d'intérêt (Richard Mayer et Mayer, 2005; Moreno, 2005, 2006, 2007; Sweller, 2011) ou sur la méthodologie à employer dans divers contextes de recherche (e.g., Glaholt (2014)). Faisant ainsi, nous avons pu développer une connaissance détaillée du domaine des simulateurs en RV 3D sur ordinateur pour l'entrainement et la formation. Les contextes des deux études furent ainsi définis.

Une deuxième vague de recensement des écrits fut effectuée afin de déterminer dans quelle mesure des tendances pertinentes et interprétables étaient présentes pour divers domaines d'études spécifiques. Puisqu'aucune théorie unifiée de l'apprentissage en simulateur de RV 3D sur ordinateur pour la formation et l'entrainement de professionnels n'existe, nous avons identifié des articles issus de divers domaines connexes comme l'apprentissage multimédia, le design de jeux en RV, les processus cognitifs en aviation, le design de didacticiels, l'ingénierie pédagogique en RV, etc. Ces articles supplémentaires ont permis l'interprétation et la discussion des résultats à la lumière de données probantes empiriques et de théories existantes.

Cette méthodologie fut employée itérativement et cycliquement pour combler les manques et assurer la qualité de notre démarche.

Informations sur l'article 1

Le premier article fut soumis et présenté à la conférence Human Computer Interactions International qui s'est tenue virtuellement à Wasington D.C. en juillet 2021. L'article fut publié en libre accès dans le livret « Late Breaking Paper – Multimodality, eXtended Reality, and Artificial Intelligence » (Rochon *et al.*, 2021; Stephanidis *et al.*, 2021). Le but de la première étude était l'évaluation du design d'un nouveau simulateur en RV 3D sur ordinateur destiné à la formation de médecins anesthésistes.

L'objectif global de ce mémoire était l'amélioration de l'apprentissage permis par le design d'un simulateur en RV 3D sur ordinateur pour la formation de professionnels. Dans ce contexte, l'étude de Phase I était exploratoire et permettait la définition des paramètres pour l'étude suivante (i.e., Phase II). Nous désirions premièrement identifier un ou plusieurs cadres théoriques pertinents pour guider l'évaluation du design de l'interface. Nous voulions deuxièmement affiner les mesures à exploiter pour évaluer en profondeur l'expérience des participants.

De surcroit, il s'agissait d'une étude de faisabilité puisque nous avions accès à un échantillon de commodité[7] et que l'étude fut implémentée à distance vues les mesures sanitaires imposées par la Covid-19.

En effet, la première contrainte de l'étude était l'évaluation d'un simulateur ayant pour but la mise en pratique de connaissances sommatives en anesthésie en utilisant un échantillon qui consistait presqu'exclusivement d'étudiants de 3^{e} à 5^{e} année d'un programme universitaire de médecine. Nous devions vérifier si les sujets possédaient les connaissances préalables nécessaires pour pouvoir compléter les tâches de l'étude tout en nous permettant d'obtenir des résultats valides pour les construits d'intérêt. Pour assurer le succès de complétion des tâches par les participants lors de la Phase I de ce travail, nous

[7] Traduction de « convenience sample »

avons évalué diverses mesures de l'expérience des participants alors qu'ils suivaient le didacticiel d'introduction au simulateur. Ce didacticiel était interactif. Il était constitué de segments d'explications et d'instructions. Les participants ont effectué des tâches qui nous ont permis d'évaluer les acquis de l'entrainement d'introduction au système. Ainsi, cette étude de faisabilité évaluait les capacités d'apprentissage du simulateur via son didacticiel introductif, plutôt que les acquis de l'entrainement en anesthésie. Nous avons recherché dans la littérature pour trouver une manière de mesurer si l'expérience des participants était influencée par des éléments de design du simulateur ou, à l'inverse, si le manque de connaissances en anesthésie avait pu prévaloir sur les données collectées. L'auto-efficacité, définie telle que la perception ou le jugement d'être en mesure d'atteindre un objectif spécifique (Zulkosky, 2009), est un terme omniprésent en psychologie et ses domaines connexes. Il est connu que l'auto-efficacité affecte la manière dont les individus réfléchissent, agissent, ressentent et se motivent (Bandura, 1997). Dans cette étude de faisabilité, nous avons mesuré l'auto-efficacité perçue, puis, nous avons corrélé cette mesure avec la charge cognitive perçue par les sujets de l'étude. Nous discutons des résultats et implications dans l'Article 1 ci-bas.

La deuxième contrainte de l'étude concernait l'implémentation d'une méthode rigoureuse pour assurer la validité des résultats malgré les limites imposées par les environnements non-contrôlés hors du laboratoire. Nous désirions tester quels outils et méthodes il était possible d'utiliser dans un tel contexte pour mesurer des interactions complexes entre un participant et le système mis à sa disposition. Le contexte de pandémie mondiale a permis à notre laboratoire de recherche de développer la méthode décrite par Giroux *et al.* (2021) pour la collecte simultanée de données objectives (i.e., psychophysiologiques et comportementales) et subjectives (i.e., psychométriques). Les outils sélectionnés pour réaliser cette étude à distance étaient, notamment, Lookback.io (Montréal, Québec, Ca) et Noldus FaceReader v8.0 (Noldus, Wageningen, NL). La méthodologie employée pour la réalisation de la Phase I de cette étude est décrite plus amplement dans l'Article 1 ci-bas.

Informations sur l'article 2

Le second article présenté dans ce mémoire est en cours de soumission au journal Frontiers in Psychology. Cet Article 2 décrit une étude dans laquelle les caractéristiques du matériel de formation du scénario d'un simulateur d'entrainement destiné à des experts furent manipulées. Plus précisément, en nous basant sur le « principe de modalité » de la TCAM, nous manipulons les modalités sensorielles du matériel pédagogique en comparant des instructions de vol *bimodales – texte et audio*, *unimodales – texte*, et *unimodales – audio*.

Certains paramètres de la première étude furent conservés (i.e., le cadre théorique) et d'autres furent modifiés (i.e., le lieu de collecte, la nature du simulateur, l'expertise évaluée, les outils de mesure) pour la réalisation de la deuxième étude. Les concepts importants pour la compréhension des interactions homme-machine et de l'apprentissage cognitif lors de l'entrainement et de la formation contextualisés au milieu de l'aviation sont présentés dans l'Article 2. Nous y détaillons également la méthodologie et outils sélectionnés, puis nous terminons en discutant les résultats obtenus et leurs implications.

Pour la réalisation de l'étude 2, il fut décidé d'utiliser un simulateur de vol destiné à la formation de pilotes plutôt que de poursuivre en exploitant un simulateur de vol spécialisé pour la formation de médecins anesthésistes. La corrélation négative détectée entre les résultats d'auto-efficacité perçue et de charge cognitive perçue pendant l'étude 1 suggérait qu'en poursuivant avec un échantillon de commodité pour l'étude 2, nous obtiendrions des résultats d'apprentissage suboptimaux par rapport à ceux attendus. Il fut décidé d'utiliser un simulateur de vol en RV 3D à haute validité écologique conçu pour être utilisé indépendamment d'un instructeur. Notre échantillon était constitué de 30 pilotes commerciaux en formation. Tous les sujets avaient en main leur licence privée, ce qui assurait qu'ils possédaient les compétences nécessaires pour compléter les tâches expérimentales d'entrainement.

Vu l'assouplissement des mesures sanitaires imposées par la Covid-19, il fut décidé de mener l'étude 2 en laboratoire. En effet, réaliser l'étude à distance dans un environnement non-contrôlé limitait le choix des outils pour mesurer l'expérience des participants ainsi que leurs états cognitifs d'apprentissage. En recherche de méthodes mixtes, la triangulation des mesures à une unité de temps donnée permet de mieux comprendre l'expérience d'un participant lors de son interaction avec une technologie (Léger *et al.*, 2019). Spécifiquement en entrainement et en formation en aviation, plusieurs travaux antérieurs ont mis en valeur l'utilisation de mesures oculométriques pour décrire les états cognitifs des pilotes (pour une revue, voir Glaholt (2014)). L'oculométrie est un outil sensible dont l'utilisation nécessite un environnement hautement contrôlé. Plusieurs mesures oculométriques furent sélectionnées pour évaluer les états cognitifs et attentionnels des pilotes lors de l'étude 2. Ces mesures ainsi que les méthodes employées sont présentées dans l'Article 2.

Contributions et responsabilités personnelles

Activité	**Contribution**
Définition des requis	Définition de la question de recherche et de la problématique – 60% • Problématique existante à l'initiation du projet • Contextualisation de la problématique élaborée en collaboration avec un partenaire d'entreprise • L'équipe et le partenaire d'entreprise ont participé à la définition des questions de recherche et à l'approche adoptée
Revue de la littérature	Revue de la littérature sur l'apprentissage multimédia – 100% Revue de la littérature sur les mesures neurophysiologiques – 100% Revue de la littérature sur les simulateurs de vol – 100% Justification des choix de design et développement – 100%
Conception du design expérimental	Élaboration de la demande au CER et des demandes de changement – 60% • Demande de CER soutenue par l'équipe opérationnelle • Développement des formulaires de consentement et de compensation à partir de modèles

	• L'équipe opérationnelle s'est occupée de mettre à jour et de renouveler le CER Protocole expérimental – 90% • Développé en me basant sur le modèle de protocole expérimental du laboratoire Installation de la salle de collecte – 80% • Tests techniques aidés par un instructeur de vol externe pour la Phase II du projet Familiarisation avec les simulateurs (simulation anesthésie, simulateur de vol), tests et sélection des scénarios de tâches – 100% Conception du design expérimental – 90%
Recrutement	Recrutement des participants – 100% Suivi recrutement – 75% • Attribution des créneaux horaire par moi-même • L'équipe opérationnelle a effectué un suivi pour assurer la présence des participants pour la collecte Gestion des compensations – 50% • Les assistants de recherche et moi-même nous sommes occupés de remplir le cartable de compensation à la fin de chaque collecte expérimentale • Présence lors de 100% du processus de collecte Phase I et 75% du processus de collecte Phase II • Compensations Phase I et tirage Phase II effectués par l'équipe opérationnelle
Prétests et collecte	Chargée des opérations lors des prétests – 100% Chargée des opérations lors de la collecte – 75% • Présence lors de 100% du processus de collecte Phase I et 75% du processus de collecte Phase II • Assistée d'assistants de recherche pour la Phase II (ajouts de marqueurs et calibration des outils)

Analyse des données	Traitement des données – 80%Analyses statistiques des données – 80%
Rédaction	Écriture des articles – 100% • Des modifications ont été suggérées par les co-auteurs

Chapter 2. Article 1

Improving Learnability Capabilities in Desktop VR Medical Applications

Laurie-Jade Rochon, Alexander J. Karran, Frédérique Bouvier, Constantinos K. Coursaris , Sylvain Sénécal , Jean-François Delisle , and Pierre-Majorique Léger

Abstract. The main objective of this study was to evaluate the implicit and explicit learning experiences of two distinct training segments, a tutorial and a Free Play Mode (FPM), of a desktop-based virtual reality (VR) medical operations simulator to assess aspects of learnability for a first-time user. Our goal was to evaluate the tutorial simulator and UI design by interpreting results through the lens of Mayer's principles of multimedia learning. The experiment was conducted remotely and the study sample comprised of ten upper-year medical students. The video recording from the participant's desktop camera was retrieved to determine their affective responses by analyzing facial micro-expressions and infer VPPs. Participants performed the simulation's tutorial followed by the FPM tasks, partitioned into two types: twelve retention tasks designed to verify how well users learned UI elements through the tutorial, and an exploration task to observe how the user explored the interface when few instructions were given. Results showed that the explicit user experience did not differ between the tutorial and the retention tasks. In contrast, users reported significantly higher cognitive load and lower system usability during the exploration task than during the tutorial. A negative correlation was found between perceived self-efficacy and perceived cognitive load. Results pertaining to VPPs indicated that FPM tasks were associated with more negative affective responses when compared to the tutorial. The manuscript concludes with methodological guidelines to assess the learnability of complex, ecologically valid simulations while reinforcing the need to use complementary methods to assess the users' experience.

Keywords: Multimedia learning • Learnability • Virtual reality • VR • User experience • Valence pain points • Automated facial expression analysis

Introduction

Virtual environments (VE) are becoming an essential learning tool for various industrial domains such as aerospace, medicine and engineering (Hays *et al.*, 1992; Izard *et al.*, 2018; Wang *et al.*, 2018). Learning in a VR settings has been shown to have beneficial effects on knowledge acquisition, retention, motivation and enjoyment (Chittaro et Buttussi, 2015; Makransky, Borre-Gude et Mayer, 2019; Richard E Mayer, 2014), providing users with a robust environment to facilitate the comprehension of complex conceptual knowledge and procedures through observation, imitation and participation (Goodyear et Retalis, 2010).

It has been widely accepted that applying the principles of multimedia learning for the design of multimedia environments generates deeper, better learning experiences (Richard Mayer et Mayer, 2005). However, to our knowledge, no work has investigated the use of these principles to generate heuristic recommendations for the improvement of VE in a learning context. Therefore, it becomes necessary to develop a method of evaluating user training that accounts for the complexity of such ecologically valid simulations.

The main objective of the present study is to evaluate the implicit and explicit learning experience of two distinct segments – a guided tutorial and an FPM – of a desktop VR medical operations simulator to assess elements of learnability for a first-time user. Specifically, our goal is to evaluate the simulator's UI design by interpreting results through the lens of Mayer's design principles of multimedia learning (Richard Mayer et Mayer, 2005). To form the basis of our study, we posed the following research question: To what extent can the evaluation of the user training experience through implicit and explicit measures inform the design of learning experiences within desktop VR simulations?

Literature Review

VR for Medical Training

VR technology and VE's have been defined by Schroeder (Schroeder, 2008) as "a computer-generated display that allows or compels the user (or users) to have a sense of being present in an environment other than the one they are actually in, and to interact with that environment". Immersion, interactivity, and presence are particular features of VR that distinguish it from other representational technology (Mandal, 2013). VE's may be displayed on a computer screen, a head-mounted display, or projection screen, whereas in desktop VR, interaction usually occurs through a computer monitor using a keyboard, a touch screen, mouse, or joystick (Lee et Wong, 2014). VR offers several advantages, such as real-world familiarity, allowing learners to rehearse sequences of procedures with high physical and psychological fidelity, enabling immediate feedback in a controlled environment (Feng *et al.*, 2018; McComas, MacKay et Pivik, 2002; Rose *et al.*, 2000; Smith et Ericson, 2009).

A recent meta-analysis of eight studies of inpatient deaths put the number of preventable deaths at just over 22,000 a year in the United States (Rodwin *et al.*, 2020). Therefore, refining medical training procedures is pivotal in saving thousands of lives. However, a lack of training personnel and equipment scarcity constrain medical practitioner's training (Makled *et al.*, 2019). VR training provides a realistic environment and is an attractive hands-on training method to address this need as it provides practical know-how while simultaneously minimizing the risks to patients. Moreover, VR-based simulators can offer a large number of diverse use cases, allowing the training of rare but dangerous complications which a trainee might not otherwise experience during a residency program (Riener et Harders, 2012). Finally, VR is relatively inexpensive and widely available, reducing costs associated with booking operating rooms and accelerating the training of medical personnel (Riener et Harders, 2012). To train effectively using VR, one must determine the skills to be learned, distinguishing between basic manipulative skills and procedural skills (Riener et Harders, 2012). The present study focuses on procedural skills aligned with cognitive processes such as problem identification and selecting an appropriate response by the trainee (Riener et Harders, 2012).

Principles for Multimedia Learning

The effectiveness of training depends on multiple factors, namely the training method and the learning processes involved. Learning is viewed as "a process of model transformation, a progression through increasingly sophisticated mental models where each reflects an adequate understanding of the target software" (Bostrom, Olfman et Sein). A novice user can form a system mental model in three different ways independently or simultaneously: through using it (mapping via usage), by drawing analogies from similar systems that are familiar to them (mapping via analogy), and through training (mapping via training) which is the focus of this study (Bostrom, Olfman et Sein). Learning outcomes can be measured through retention tests that emphasize remembering and transfer tests that emphasize understanding (Council *et al.*, 2001), and this study focuses on retention as a learnability outcome during user training.

The CTML has shed light on the notable effects of learning support on learning outcomes (Richard Mayer et Mayer, 2005). The theory proposes three main assumptions. Firstly, there are two separate channels for processing information, the auditory and visual channels. Secondly, each channel has a finite capacity. Thirdly, learning is an active process that requires filtering, selecting, organizing, and integrating information based upon prior knowledge to generate meaningful learning (Richard Mayer et Mayer, 2005). By a series of research that was tested through more than 200 experimental tests, Mayer articulated 15 principles of multimedia learning which are separated into three categories (See Table 1): principles meant to reduce extraneous processing – cognitive processing that does not serve the instructional goal; principles for managing essential processing – cognitive processing needed to acquire the essential information; and principles to foster generative processing – cognitive processing aimed at making sense of the material (Richard Mayer et Mayer, 2005). These principles were translated into design recommendations, making it easy for designers to build high-quality multimedia material (Richard Mayer et Mayer, 2005).

Representative instructional principle	Goal
1. Multimedia	Basis for the theory
2. Coherence	Reduce extraneous processing
3. Signaling	Reduce extraneous processing
4. Redundancy	Reduce extraneous processing
5. Spatial contiguity	Reduce extraneous processing
6. Temporal contiguity	Reduce extraneous processing
7. Segmenting	Manage essential processing
8. Pre-training	Manage essential processing
9. Modality	Manage essential processing
10. Personalization	Foster generative processing
11. Voice	Foster generative processing
12. Image	Foster generative processing
13. Embodiment	Foster generative processing
14. Immersion	Foster generative processing
15. Generative activity principle	Foster generative processing

Table 1. *Principles of multimedia learning and their goal (retrieved and adapted from Richard Mayer et Mayer (2005).*

Implicit and Explicit Measurement in a Learning Context

User experience (UX) is defined by the ISO 9241-210 standard as a "person's perceptions and responses resulting from the use and anticipated use of a product, system or service" (ISO et STANDARD, 2010), including emotions, beliefs, preferences, and physiological responses among others. Most UX research utilizes self-reported qualitative (e.g., interviews, observations, open-ended questionnaire questions) and quantitative (e.g., closed questionnaire questions) measures to capture these responses. Self-reported questionnaires allow the measurement of different aspects of UX, for instance, perceptions of usability (Lewis, 1995), cognitive load (Hart et Staveland, 1988), and perceived efficacy (Kang *et al.*, 2019), among others. However, it is difficult for users to report on their own experience. The evaluation of perceptual emotional reactions at the end of an interaction is associated with a significant loss of moment-to-moment user reactions throughout a task, potentially resulting in important differences between what a user felt during an experience and how it was recalled afterwards (Eich *et al.*, 2000). This type of evaluation corresponds to a global perspective that has shown to be impacted by multiple biases such as the peak effect and the peak-end rule (Cockburn, Quinn et Gutwin, 2015).

Collecting quantitative psychophysiological data to measure UX in real-time has shown to provide a deeper understanding of UX while reducing methodological biases associated with the sole use of perceptual measures (de Guinea, Titah et Léger, 2014). Combining complementary methods through the measurement of implicit antecedents in parallel with explicit measurements via self-report renders a more complete and reliable portrait of UX (de Guinea, Titah et Léger, 2014). The recognition of facial expressions is a psychophysiological measure used to calculate emotional valence, i.e., the emotional spectrum ranging from unpleasant (negative valence) to pleasant (positive valence) (Ekman et Friesen, 1978). The method of identifying psychophysiological pain points (PPPs) developed by Giroux-Huppé *et al.* (2019) generates a better representation of a user journey as it includes automatic, often unconscious, negative physiological manifestations of the user. Shown in **Table 2** is a summary of the constructs investigated in the present study along with their definitions.

Construct	Definition
Valence Pain Points *Implicit*	Frequency of moments of negative emotions during the execution of training task, defined by low valence (lowest 10%)
Cognitive workload *Explicit*	Participant's perceptions of workload during the training task and VR application used
Perceived usability *Explicit*	Participant's perceptions of the VR application used
Self-perceived Efficacy *Explicit*	Participant's perceptions of self-efficacy during the training task

Table 2. *Constructs of interest.*

Hypothesis Development

Our objective was to evaluate the implicit and explicit learning experience of two distinct segments of a desktop VR medical operations simulator to assess elements of learnability for a first-time user. The broader aim of this research is for the results to inform the design of learning experiences within desktop VR simulations.

We hypothesized that an effective tutorial should allow a user to complete retention tasks successfully without frustration. Conversely, an ineffective tutorial will evoke frustration, as indicated by an increased number of VPPs and difficulty completing the task. This expected outcome is due to much of the tutorial affording participants the opportunity to passively process information, whereas users are actively engaged in interaction during FPM tasks. Accordingly, we expect users to experience more moments of frustration when struggling to complete a task than when passively following a tutorial. Therefore, to test the implicit training experience of users, we propose the hypothesis that (H1) a user will experience a significantly higher number of VPPs per minute during an FPM task - for both the retention tasks (H1a) and the exploration task (H1b) - than during a tutorial, translating to a more negative affective experience.

Instructional design aims to control the cognitive load while also stimulating learners to use their available cognitive capacity for better learning (Paas *et al.*, 2005). When completing FPM tasks, users must constantly interact with the VE and various interface elements, which has been shown to have beneficial effects on engagement and motivation (Makransky, Borre-Gude et Mayer, 2019). When motivated, learners are more likely to exert more cognitive effort to schema construction and automation to improve their cognitive task performance (Paas *et al.*, 2005). Previous studies also showed task procedural complexity to be positively associated with intrinsic cognitive load (Sewell *et al.*, 2017; Van Merriënboer et Sweller, 2010). Essential processing is analogous to intrinsic cognitive load in cognitive load theory (CLT) (Sweller, 2011). An FPM task complexity is higher than that of a tutorial in that a user must remember the exact steps of a procedure and replicate them to complete it. Hence, we hypothesize that a user's perceived cognitive load will be significantly higher during an FPM task when compared with a tutorial, for both the retention tasks (H2a) and the exploration task (H2b), as a user interacts with the interface. During their interaction with the interface, the chance of encountering new pain points increases, which can impact the perceived usability of a system.

The work of de Guinea, Titah et Léger (2014) has shown that when frustration is high, neurophysiological memory load has a negative impact on perceived system usability, while conversely, a low level of frustration positively influences a user's perceived system usability. Hence, we posit that effects of frustration and cognitive load experienced when interacting with the medical desktop VR simulator will lead to a lower perceived system usability for both the retention tasks (H3a) and the exploration task (H3b) relative to the tutorial.

Previous studies have also identified a negative correlation between the levels of self-perceived efficacy and cognitive load, where high self-perceived efficacy led to lower cognitive load than low self-efficacy (Redifer, Bae et Zhao, 2021; Vasile *et al.*, 2011). In that sense, a participant's low perceived efficacy for clinical skills will result in more difficulty completing a task and increased cognitive load. We thus hypothesize that (H4) the level of self-perceived efficacy will be negatively related to a user's perceived cognitive load.

Method

Due to COVID-19 restrictions, the experiment was conducted remotely in the presence of a moderator as part of the deployment of a new desktop VR medical platform designed to reproduce clinical environments and train anesthetists to treat patients. A user had to complete a continuously guided tutorial that showed the basic functionality of the system. After completion of the tutorial, a built-in FPM allowed a user to freely explore the simulation environment. This FPM simulator mode allowed participants to reproduce various tasks learned through the tutorial and perform an exploration task. Hence, a one-factor within-subject experimental design was used, where the tutorial task and the FPM tasks measures were compared for each participant.

Participants

Ten medical students were recruited by word-of-mouth to participate in this study. Selected participants were required to have completed a minimum of two years in a university medicine program and speak at an advanced English level. Four participants were 3rd-year students, three participants were 4th-year students, two students were 5th-year students, and one student was a postgraduate student completing a residency training in anesthesia. Participants included five men and five women, ranging from 22 to 25 years old, with a mean of 23.4 years (SD = 1.07). Nine participants spoke French as a first language and reported being bilingual, and the remaining participant spoke English.

Before the experiment, participants had to read and sign a consent form; they received compensation in the form of a $30 gift card at the end of the experiment. Our institution's ethics committee approved this project (2021-4305).

Procedure

The experimental procedure is shown in Fig. 1. The training course (stimulus) for this study consisted of two segments, a guided tutorial and an FPM. First, participants had to follow the interactive guided tutorial of a desktop-based VR medical operations simulator designed to teach them the basic functionalities of the interface. The tutorial included segments of information presented in text boxes. Users were asked to turn the simulation's sound on to hear the information read by a narrative voiceover. At multiple times, the participant was required to step into the tutorial to try various interface features. A participant would complete the tutorial steps at their own pace, but without the possibility of performing backtracking, nor exploring features other than those presented in real-time during the tutorial. When the tutorial was complete and before the FPM task started, the user switched to a questionnaire tab on their computer to complete an online questionnaire.

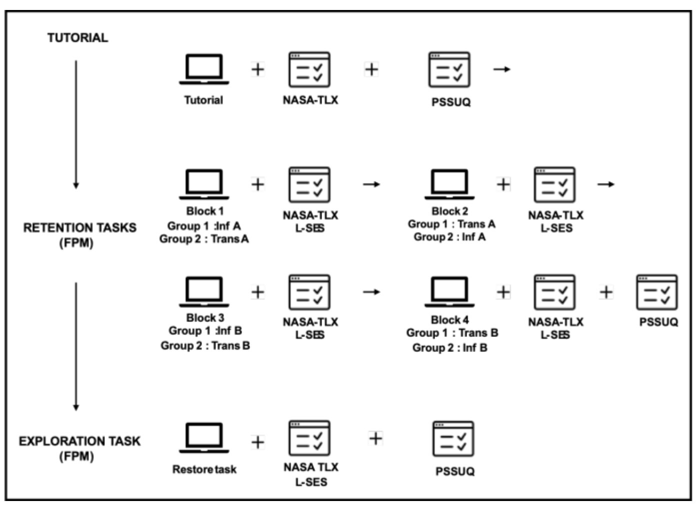


Figure 1. *Experimental procedure.*

Next, a user had to complete twelve retention tasks and one exploration task in a linear sequence using the simulator's FPM. The individual task descriptions were presented in the questionnaire window; a user would hence navigate back and forth between the questionnaire and simulator windows.

The retention tasks aimed to verify how first-time users learned UI elements through the tutorial, defined as "executing the exact steps in a training task that were previously shown" (Council *et al.*, 2001). Six tasks were "informational" where the intent was to acquire information (e.g., Identify the patient's heart rate), and six tasks were "transactional" where the intent was to perform an activity (e.g., Perform five manual ventilations) (Broder, 2002). Empty fields in the questionnaire tab allowed participants to write down their answers to informational questions. Retention tasks were separated into four blocks; after each block, the participant filled out questionnaires. The order of presentation of tasks within a block was randomized, and the presentation order of the blocks was counterbalanced in two equal groups of participants who were randomly assigned to a group.

The exploration task aimed to investigate how users explored the interface when few instructions were provided. We call this task "restore", which consisted of restoring the patient's condition to the best of the participant's ability. The participant would collect information about the patient in the simulated operating room and perform various procedures whose effects could be observed in real-time. Participants were allocated five minutes to complete the task but could tell the experiment moderator they were done before so, which immediately ended the task. This was followed by the completion of a series of questionnaires by switching to the computer's questionnaire tab.

Participant (explicit) cognitive load was assessed using the NASA Task Load Index (NASA-TLX)[8], a widely used subjective multidimensional scale designed to obtain workload estimates that can be used during task completion or immediately afterwards. The scale's six items are Mental, Physical, and Temporal demand, Frustration, Effort, and Performance, each rated on a 100-point range with 5-point intervals, which are then averaged (as no weights were applied to each item) to give the task load index (Hart et Staveland, 1988). The user's experience with the simulator was assessed using the Post-

[8] Voir Annexe 1.1 – Questionnaires : NASA-TLX

Study System Usability Questionnaire (PSSUQ)[9], a 16-item instrument that allows participants to provide an overall evaluation of the system they used (Lewis, 1995). The items are 7-point Likert Scales ranging from "Strongly disagree" (1) to "Strongly agree" (7) and a "Not applicable" (N/A) point outside the scal. The scale's psychometric evaluation revealed three factors, i.e., System Usefulness, Information Quality and Interface Quality (Lewis, 1995), which were also assessed in this study. The perceived self-efficacy for the clinical skills of medical students was assessed using the cognitive domain of the Learning Self-Efficacy Scale (L-SES)[10] (Kang *et al.*, 2019), which includes 4 items rated on 5-point Likert Scales ranging from "Disagree" (1) to "Agree" (5). The L-SES was used to distinguish any difficulties encountered in tasks by users because of a lack of prior knowledge of clinical skills.

To capture emotional valence from participant video recordings, we utilized FaceReader v8.0 (Noldus, Wageningen, NL) and CUBE HX was used to synchronize the data (Léger *et al.*, 2019; Léger *et al.*, 2014). This software uses the Facial Action Coding System (FACS) developed by Ekman et Friesen (1978) to provide inference for six basic emotions, happiness, sadness, anger, disgust, fear, and surprise, with the addition of the neutral emotion (Loijens et Krips, 2018). Video acquisition and analysis was acquired following guidelines for collecting automatic facial expression detection data synchronized with a dynamic stimulus in remote moderated user tests from Giroux *et al.* (2021). These inferences are provided as a value between 0 and 1, from negative to positive (Loijens et Krips, 2018). VPPs were derived based on the participant's peak intensity of negative emotions, which crossed the tenth percentile of emotional valence threshold (Giroux-Huppé *et al.*, 2019; Lamontagne *et al.*, 2021).

Performance was assessed for the retention tasks using task completion time, success rate and partial success rate as proxies. A task was marked as failed if a participant required help from a moderator or if they did not complete a task and moved on to the next. A partial success was counted if a participant navigated to the correct interface element but

[9] Voir Annexe 1.2 – Questionnaires : PSSUQ
[10] Voir Annexe 1.3 – Questionnaires : L-SES

indicated an incorrect final response (informational task) or incorrectly adjusted a parameter (transactional task) (Dargent *et al.*).

Analysis

Using SAS 9.4 (SAS, Cary, USA), the NASA-TLX, PSSUQ and L-SES results were tested with a linear regression with random intercept model (Holm-Bonferroni corrected) at the task level as they were assessed directly after the tasks. To test the difference in the mean number of VPPs between the tasks, a 2-tailed Wilcoxon signed rank test was used.

Results

One informational task was excluded from all analyses because its formulation led to confusion in 6 participants who failed or skipped the task.

Descriptive Statistics

Prior to answering the research question, we investigated whether task characteristics influenced the results. The eleven retention tasks were divided into two groups of tasks, five informational tasks and six transactional tasks (Broder, 2002). **Table 3** shows the mean ratings (and standard deviation) of task completion time, success and partial success rates, perceived cognitive workload, and numbers of VPPs per minute for informational and transactional task types.

Measure	Informational task	Transactional task	P-value
1. Performance			
Completion time	37.75 (32.8)	144.2 (97.9)	$p < .001$
Success rate	0.95 (0.122)	0.87 (0.199)	$p = .153$
Partial success rate	0.05 (0.122)	0.08 (0.183)	$p = .638$
2. Perceived cognitive load	24.72 (17.4)	31.4 (17.3)	$p = .011$
3. VPPs per minute	27.81 (15.5)	21.79 (12.9)	$p = .322$

Table 3. *Mean and standard deviation of six dependent measures for informational and transactional tasks.*

The analysis performed indicate that transactional tasks took significantly more time to perform when compared to informational tasks (B = (−6.0063), SE = 1.4678 (df = 29, t =

–4.09), $p < 0.001$). Overall, the tasks were successfully completed; no significant difference for success and partial success rates were found between informational and transactional tasks. Results indicate that there is an effect associated with the nature of tasks for perceived cognitive load, which was perceived significantly higher by participants when performing transactional tasks when compared to informational tasks ($B = (-0.3065)$, $SE = 0.1131$ ($df = 29$, $t = -2.71$) $p = 0.0112$). The affective experience of users, however, did not differ significantly across both task types.

Hypothesis 1: A User Will Experience a Significantly Higher Number of VPPs Per Minute During an FPM Task - for Both the Retention Tasks (H1a) and the Exploration Task (H1b) - Than During a Tutorial

The first prediction is that the mean number of VPPs will be higher during the FPM tasks as compared with the tutorial. To examine the results from the analysis of VPPs, we produced a "valence journey map" relating participant affective responses with tasks performed within the simulation. The results indicated that both the retention tasks ($Z = 17.5$, $p = 0.0391$, $r = 0.78$) and the exploration task ($Z = 20.5$, $p = 0.0117$, $r = 0.91$) were associated with more negative affective responses when compared to the tutorial task. These results support Hypothesis 1a and 1b that the FPM tasks induced a significantly higher number of VPPs per minute than the tutorial.

Hypothesis 2a and 2b: User's Perceived Cognitive Load Will Be Significantly Higher During an FPM Task - for Both the Retention Tasks (H2a) and the Exploration Task (H2b) - When Compared with a Tutorial

The mean ratings (and standard deviation) of perceived cognitive workload, perceived system usability, and numbers of VPPs per minute for the three tasks performed (i.e., the tutorial, the retention tasks and the exploration task) are shown in **Table 4**.

Hypothesis measure	Task Tutorial task	Retention task	P-value Tut-Ret	Exploration task	P-value Tut-Exp
Perceived cognitive load	22.82 (16.5)	28.06 (17.5)	p =.221	53.00 (23.1)	p <.001
Perceived system usability	6.15 (0.487)	5.993 (0.59)	p =.185	5.706 (0.80)	p =.017
VPPs per minute	11.15 (7.66)	23.95 (10.9)	p =.039	20.16 (7.45)	p =.012

Table 4. *Mean and standard deviation of four dependent measures for the tutorial, retention and exploration tasks.*

The second hypothesis posited that participants will report higher levels of cognitive load during the FPM tasks compared with the tutorial. The results from this analysis indicate that there is an effect associated with perceived cognitive load, in this case between the exploration task and the tutorial (B = (0.8847), SE = 0.2065 (df = 47, t = 4.28), p = < 0.001), where the perceived cognitive workload was perceived higher during the exploration task. Our analysis for the NASA-TLX individual items revealed a significant difference for the levels of temporal demand between the tutorial and the exploration task (B = (4.44), SE = 1.7528 (df = 48, t = 2.82), p = 0.0208); Participants perceived the pace of the exploration task to be more rushed or hurried. Moreover, results indicate a significant difference in the levels of perceived effort between the tutorial and the exploration task (B = (3.6111), SE = 1.3829 (df = 48, t = 2.61), p = 0.036); Participants felt like they had to work harder to accomplish their level of performance during the exploration task. These results do not support Hypothesis 2a, i.e. that the perceived cognitive load induced by a retention task would be significantly higher than that of a tutorial. However, results support Hypothesis 2b, as the perceived cognitive load was reportedly higher during the exploration task than during the tutorial.

Hypothesis 3a and 3b: Users Will Perceive a Lower System Usability for Both the Retention Tasks (H3a) and the Exploration Task (H3b) Relative to the Tutorial

The third hypothesis posited that participants will report lower levels of system usability during the FPM tasks as compared with the tutorial. The perceived system usability was rated significantly higher by users during the tutorial than during the exploration task (B = (−0.4432), SE = 0.1526 (df = 47, t = −2.9), p = < 0.05). We observed the same tendency for reported results of perceived information quality ($B = (-0.6646)$, $SE = 0.1679$ ($df = 47$, $t = -3.96$), $p = < 0.001$), indicating that the tutorial task provided a better-perceived level of information quality when compared to the exploration task; and system usefulness ($B = (-0.5428)$, $SE = 0.1936$ ($df = 47$, $t = -2.8$), $p = < 0.05$), indicating in this case that participants perceived a higher level of system usefulness during the tutorial than during the exploration task. No significant differences were found when comparing the perceived interface quality of the tutorial with that of the retention tasks and the exploration task. These results do not offer support for Hypothesis 3a, i.e. that the users would report significantly lower system usability during retention tasks when compared to a tutorial as no significant difference was observed across tasks. However, results offer support for Hypothesis 3b, as system usability was perceived to be lower for the exploration task than for the tutorial.

Hypothesis 4: Self-efficacy Will Be Negatively Related to a user's Perceived Cognitive Load

The last hypothesis posited that the level of self-efficacy will negatively affect a user's perceived cognitive load. Our results confirmed a negative correlation between the levels of perceived cognitive load and those of perceived self-efficacy ($B = (-1.6483)$, SE = ($df = 38$, $t = -10.13$), $p < 0.0001$) for all the FPM tasks combined. These results provide support for Hypothesis 4 and are consistent with the idea that self-efficacy can positively impact learning (Schunk et DiBenedetto, 2016).

Discussion

The objective of this study was to evaluate the implicit and explicit learning experience of two distinct segments of a desktop VR medical operations simulator to assess elements of learnability for a first-time user. Overall, results suggest that users' training experience was more positive during the tutorial than during the retention and exploration tasks. The results of explicit and implicit measures contribute differently but jointly to explain the training experience of users.

We first discuss the implicit experience of users. The number of VPPs per minute during FPM tasks was significantly higher than during the tutorial. We interpret this result using four situations for illustration. Firstly, a tutorial presents interface functionality adequately, and because of this, the user manages to perform an associated FPM task easily, this inducing a non-significant amount of VPP's during both training segments. Secondly, the presentation of interface functionality during the tutorial is misunderstood by a user, which induces a VPP. While in FPM, a user has difficulty performing the task which induces a new VPP. The number of VPPs in this situation is equal between the two segments. Thirdly, the tutorial shows an interface function and immediately requires a user to test it. The user struggles, and a VPP occurs; however, when a user performs the same task in FPM, they understand the task and no VPP is recorded. The number of VPPs is, therefore, higher during the tutorial. Lastly, the tutorial explains a feature that the participant believes they understand. They perform it for the first time during an FPM task, and they struggle, which induces a VPP. The number of VPPs is now higher during FPM than during the tutorial. It is possible that the last situation presented manages to adequately explain why the number of VPPs was higher during the FPM tasks. This also suggests that designers would benefit from including more interactive segments during a tutorial, which supports the design principle of generative activity, which states that individuals learn better when guided in carrying out activities such as summarizing, mapping, or like in our case, self-testing, by fostering generative processing (Richard Mayer et Mayer, 2005). These results can additionally be explained by the fact that users are more passive during a tutorial than during FPM tasks. Indeed, a portion of the tutorial time is dedicated to listening and reading, whereas users continually interact with the interface when performing tasks such as in FPM; thus, the opportunities for a pain point to arise are more numerous in the latter case.

We now discuss the explicit experience of users. Results highlight that the explicit training experience of users was more positive during the tutorial than during the exploration task, as indicated by lower measures of perceived cognitive load, temporal demand, effort, and higher perceived system usability, information quality and system usefulness. However, the explicit experience did not differ when comparing the tutorial with retention tasks; thus, Hypothesis 1b and 1c were not supported. Indeed, we found no significant change in the levels of perceived cognitive load nor perceived system usability for these tasks. In addition, our results indicate that the levels of perceived cognitive workload are negatively correlated with measures of self-perceived efficacy.

We hypothesized that the perceived cognitive load would be higher during retention tasks when compared to the tutorial; this was justified through an appraisal of the cognitive processes involved during the execution of retention tasks. Previous work that studied the limitations of subjective cognitive load in simulation-based procedural training supported essential load as synonymous with the NASA-TLX mental demand item, while extraneous and generative cognitive loads were not reflected in the questionnaire (Naismith *et al.*, 2015). The fact that instrinsic load is positively correlated with the complexity of the performed task, and negatively correlated with the learner's expertise (Sewell *et al.*, 2017; Van Merriënboer et Sweller, 2010), suggests that retention task complexity was not high enough to impact a user's perceived cognitive load significantly. The failure rate for the retention tasks was 2.73% for all participants, indicating that tasks were completed successfully by participants and this suggests that participants were able to recall the steps shown during the tutorial without involving significantly greater cognitive resources, which together seem to indicate that the tutorial was effective in achieving its goal overall.

Users perceived a significantly higher cognitive load during the exploration (or restore) task when compared to the tutorial. The restore task required users to refer to external knowledge and to understand interface features regardless of their context of use, which characteristics correspond to those of a transfer test, i.e. by verifying if learners can apply what they have learned in various situations (Council *et al.*, 2001). Engaging in generative processes proved to lead to better performance outcomes in transfer tests (Richard Mayer et Mayer, 2005). Since the cognitive load is negatively correlated with the learner's expertise (Sewell *et al.*, 2017; Van Merriënboer et Sweller, 2010), results related to self-efficacy – which indicate a negative correlation between self-efficacy and perceived cognitive load – can inform our interpretation of the NASA-TLX results. In this case, the low level of self-efficacy during the exploration task caused an increase in users perceived cognitive load due to the lack of mastery of the external knowledge required to perform the medical procedures. We know from the retention task results that users remembered the various interface features. However, the restore task results do not allow us to infer whether they were understood.

Interestingly, during the tutorial, several VPPs occurred when principles that foster generative processes were not respected. Notably, the tutorial narration was performed by a robotic non-human voice using formal dialogue. This contradicts the personalization and voice principles, which are based on the premise that human to human communication fosters the creation of a conversational bond. Without this bond, a user must try harder to make sense of what the author is saying (Richard Mayer et Mayer, 2005). The increase in perceived effort can be explained by the users' need to recover prior knowledge activated from long-term memory to select the VE's important signals, then organize this information into a coherent structure mentally, and integrate it (Richard Mayer et Mayer, 2005). The increase in perceived temporal demand can be explained by two different factors. First, immersion is a defining characteristic of VE. When immersion is high, users tend to lose track of time (Brown et Cairns, 2004). When task difficulty increases at a fast rate, immersion decreases (Qin, Rau et Salvendy, 2010). It is therefore possible that a lower level of immersion during the exploration task led to higher perceived temporal demand. Second, users had five minutes to complete the exploration task whereas they could take the time that they needed to complete the tutorial, which could have created pressure for participants during the exploration task.

In line with the work of de Guinea, Titah et Léger (2014) and Lamontagne *et al.* (2021) which showed that when frustration is high, cognitive load negatively impacts perceived system usability, we hypothesized that as a result of an anticipated higher number of VPPs per minute and higher perceived cognitive load, the perceived system usability would be lower during retention tasks. In our case, negative affective responses were more frequent during retention tasks, but perceived cognitive load remained constant as did the perceived system usability. Usability problems are caused by the combination of user interface design factors and factors of usage context (Manakhov et Ivanov, 2016). As all the participants had access to the same VE and functionalities to complete the tutorial and retention tasks and the context of use was the same, participants could have encountered the same usability problems.

Participants rated the usability of the system significantly lower during the exploration task. In this case, the VE was the same as during the tutorial, but the context of use was different, as users had not previously been shown the exact steps to complete the task. We observed a change in the participants' strategies to complete the task. It is therefore possible that the interface was less suitable for this type of interaction. For example, participants would cease to make an intense use of the quick access buttons menu and they would instead directly look for clues within the simulation's VE by clicking on medical objects and tools. In this case, results support our hypothesis and the work of de Guinea, Titah et Léger (2014), i.e., the highest number of VPPs per minute occurred during the exploration task, when the perceived cognitive load was accordingly higher and this influenced negatively perceived system usability. Results also pointed towards lower perceived information quality and lower perceived usefulness, whereas no significant difference in perceived interface quality was observed. This suggests that it was likely the information given to participants to perform the task rather than the interface itself that led to this perceived decline in usability. Once again, the self-efficacy reported by users is aligned with these results, as users felt they did not have the knowledge at their disposal to complete the task. It is interesting to note that some multimedia principles that were put forward during the tutorial were not applied during the FPM tasks. For example, while the tutorial interface elements being presented were highlighted in yellow, referring to the signaling principle (Richard E Mayer et Fiorella, 2014), this was not the case during FPM tasks. Emphasizing elements of simulation during FPM tasks could redirect novice users' attention and thus guide them to the right solution (Richard E Mayer et Fiorella, 2014).

The combination of the explicit and implicit experience measures clearly informs us that the experience of users was suboptimal during the exploration task of the FPM when compared with the tutorial. However, the sole use of perceptual measures would have had no significant difference between the experience of participants when following the tutorial and when completing a retention task. Additionally, the method of identification of VPPs allows to distinguish principles of multimedia learning that seem to impact the training experience of users the most. For example, in a situation where VPPs occurred 50% times more frequently when the robotic voiceover was playing, one could infer that the voice principle should be revisited in priority. Taken together, these results are in line with extant literature that suggests that the combination of explicit and implicit methods provides a deeper understanding of UX (de Guinea, Titah et Léger, 2014; Lamontagne *et al.*, 2021).

This study has limitations that need to be considered for future studies. First, the sample size used for this study respected the minimum number of nine participants required to find more than 80% of VPPs, as indicated by Lamontagne *et al.* (2019). Nevertheless, follow-up studies should be conducted with a larger sample and with other VR training applications to ensure and reinforce the results' validity. Second, the sample used in this study included upper-level medical students ranging from their third undergraduate year to postgraduate training, nine students of which frequented the same university at the graduate level. The characteristics of these participants were therefore necessarily not identical to those of practicing physicians. Being at a lower level of training than the end-users could have negatively impacted the results. Moreover, the attendance of the same school by a vast majority of participants could also have reduced the generalizability of the results, as these students were exposed to the same learning and training methods in their academic environment. Third, this study prioritized ecological validity by using the orientation module of a real medical simulator. The use of such technology meant that it was not possible to control all the stimuli present in the simulation, as it would have been the case for a simulator specifically built for the purpose of the study. Moreover, participants would complete a block of three short tasks followed by the completion of the self-efficacy questionnaire. We did this because retention tasks had to be the same to those previously shown during the tutorial. Administering the questionnaire after each task would have taken too much time. In a controlled VE however, it would be beneficial to create longer tasks to administer the self-efficacy questionnaire once before each task. This would allow to have a better idea of which specific skill was less mastered by participants; this would ensure removing any peak-end effect; and above all it would ensure that the perceived performance of participants with the task did not influence questionnaire response regarding self-efficacy. Fourth, this study only used automated facial expression analysis to infer emotional valence and thus measure the psychophysiological state of participants using remote physiological recording (Vasseur *et al.*, 2021). Previous research has shown however that at least two physiological measures should be assessed at the same time to avoid extraneous noise and to give a richer comprehension of the affective and cognitive state of the user (Charles et Nixon, 2019; Ganglbauer *et al.*, 2009; Standardization, 2010). The psychophysiological pain points (PPPs) identification method used by Giroux-Huppé *et al.* (2019) which derives valence-arousal PPPs and valence-cognitive PPPs from measures of electrodermal activity, pupillometry and user facial expressions generates a deeper representation of a user's journey and should thus be used in follow-up studies (Dawson, Schell et Filion, 2017; Sweller, Ayres et Kalyuga, 2011). In lab data collection could also allow for a better understanding of the learner cognitive engagement via the use of electroencephalography (Lackmann *et al.*, 2021).

Conclusion

This article contributes to the multimedia literature by evaluating the training experience of two distinct segments of a desktop VR medical operations simulator to assess elements of learnability for a first-time user. This study presents a first effort in understanding how the combination of implicit and explicit measures can render a deep comprehension of a user's VR medical training experience to inform heuristic recommendations based on the design principles of multimedia learning (Richard Mayer et Mayer, 2005). Our results show that whereas the number of negative affective responses was significantly higher during the retention and exploration tasks of the FPM as compared with the tutorial, only the exploration task explicit results (i.e., perceived cognitive load and perceived system usability) were different than those of the tutorial. The results also support previous work that showed a negative correlation between the levels of self-efficacy and those of perceived cognitive load (Redifer, Bae et Zhao, 2021; Vasile *et al.*, 2011), and underline the importance of paying attention to task characteristics in VR training assessment. We finally propose a method to inform designer teams of usability problems' severity and UI elements optimization to improve learnability outcomes in complex medical desktop VE.

Acknowledgements. The authors would like to thank CAE Inc for its collaboration and funding as well as the NSERC-PROMPT Industrial Research Chair in UX.

References

1. Bostrom, Robert P., Lorne Olfman et Maung K. Sein (1990). « The importance of learning style in end-user training », *MIS Quarterly*, vol. 14, no 1, p. 101-119.
2. Broder, Andrei (2002). « A taxonomy of web search », *SIGIR Forum*, vol. 36, no 2.
3. Charles, Rebecca L. et Jim Nixon (2019). « Measuring mental workload using physiological measures: A systematic review », *Applied Ergonomics*, vol. 74, p. 221-232.
4. Chittaro, Luca et Fabio Buttussi (2015). « Assessing knowledge retention of an immersive serious game vs. A traditional education method in aviation safety », *IEEE transactions on visualization and computer graphics*, vol. 21, no 4, p. 529-538.
5. Cockburn, Andy, Philip Quinn et Carl Gutwin (2015). « Examining the peak-end effects of subjective experience », communication présentée au *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Seoul, Republic of Korea. Récupéré de https://doi.org/10.1145/2702123.2702139
6. Dargent, Tanguy, Alexander Karran, Pierre-Majorique Léger, Constantinos K Coursaris et Sylvain Sénécal « The influence of task types on user experience after a web interface update ».
7. Dawson, Michael E, Anne M Schell et Diane L Filion (2017). « The electrodermal system ».
8. de Guinea, Ana Ortiz, Ryad Titah et Pierre-Majorique Léger (2014). « Explicit and implicit antecedents of users' behavioral beliefs in information systems: A neuropsychological investigation », *Journal of Management Information Systems*, vol. 30, no 4, p. 179-210.

9. DIS, ISO (2010). « 9241-210: 2010. Ergonomics of human system interaction-part 210: Human-centred design for interactive systems (formerly known as 13407) », *International Standardization Organization (ISO). Switzerland*.
10. Eich, Eric, John F Kihlstrom, Gordon H Bower, Joseph P Forgas et Paula M Niedenthal (2000). *Cognition and emotion*, Oxford University Press on Demand.
11. Ekman, Paul et Wallace V Friesen (1978). *Manual for the facial action coding system*, Consulting Psychologists Press.
12. Feng, Zhenan, Vicente A González, Robert Amor, Ruggiero Lovreglio et Guillermo Cabrera-Guerrero (2018). « Immersive virtual reality serious games for evacuation training and research: A systematic literature review », *Computers & Education*, vol. 127, p. 252-266.
13. Giroux, Félix, Pierre-Majorique Léger, David Brieugne, François Courtemanche, Frédérique Bouvier, Shang Lin Chen, *et al.* (Forthcoming). « Guidelines for collecting automatic facial expression detection data synchronized with a dynamic stimulus in remote moderated user tests », *International Conference on Human-Computer Interaction*.
14. Goodyear, Peter et Symeon Retalis (2010). « Learning, technology and design », dans *Technology-enhanced learning*, Brill, p. 1-27.
15. Hart, Sandra G. et Lowell E. Staveland (1988). « Development of nasa-tlx (task load index): Results of empirical and theoretical research », dans Peter A. Hancock et Najmedin Meshkati (dir.), *Advances in psychology*, vol 52, North-Holland, p. 139-183.
16. Hays, Robert T., John W. Jacobs, Carolyn Prince et Eduardo Salas (1992). « Flight simulator training effectiveness: A meta-analysis », *Military Psychology*, vol. 4, no 2, p. 63-74.
17. ISO, BSEN et BRITISH STANDARD (2010). « Ergonomics of human-system interaction ».
18. Izard, Santiago González, Juan A. Juanes, Francisco J. García Peñalvo, Jesús Mª Gonçalvez Estella, Mª José Sánchez Ledesma et Pablo Ruisoto (2018). « Virtual reality as an educational and training tool for medicine », *Journal of Medical Systems*, vol. 42, no 3, p. 50.

19. Kang, Y. N., C. H. Chang, C. C. Kao, C. Y. Chen et C. C. Wu (2019). « Development of a short and universal learning self-efficacy scale for clinical skills », *PLoS One*, vol. 14, no 1, p. e0209155.
20. Lackmann, Sergej, Pierre-Majorique Léger, Patrick Charland, Caroline Aubé et Jean Talbot (2021). « The influence of video format on engagement and performance in online learning », *Brain Sciences*, vol. 11, no 2, p. 128.
21. Lamontagne, Charles, Sylvain Sénécal, Marc Fredette, Élise Labonté-LeMoyne et Pierre-Majorique Léger (2021). « The effect of the segmentation of video tutorials on user's training experience and performance », *Computers in Human Behavior Reports*, vol. 3, p. 100071.
22. Lee, Elinda Ai-Lim et Kok Wai Wong (2014). « Learning with desktop virtual reality: Low spatial ability learners are more positively affected », *Computers & Education*, vol. 79, p. 49-58.
23. Léger, Pierre-Majorique, Francois Courtemanche, Marc Fredette et Sylvain Sénécal (2019). « A cloud-based lab management and analytics software for triangulated human-centered research », dans *Information systems and neuroscience*, Springer, p. 93-99.
24. Lewis, James R. (1995). « Ibm computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use », *International Journal of Human–Computer Interaction*, vol. 7, no 1, p. 57-78.
25. Loijens, Leanne et Olga Krips (2018). « Facereader methodology note », *A white paper by Noldus Information Technology, Noldus. Zugriff am*, vol. 16, p. 2018.
26. Makransky, Guido, Stefan Borre-Gude et Richard E. Mayer (2019). « Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments », *Journal of Computer Assisted Learning*, vol. 35, no 6, p. 691-707.
27. Mandal, Sharmistha (2013). « Brief introduction of virtual reality & its challenges », *International Journal of Scientific & Engineering Research*, vol. 4, no 4, p. 304-309.
28. Mayer, Richard E (2005). « Cognitive theory of multimedia learning », *The Cambridge handbook of multimedia learning*, vol. 41, p. 31-48.

29. Mayer, Richard E (2014). « Incorporating motivation into multimedia learning », *Learning and Instruction*, vol. 29, p. 171-173.
30. Mayer, Richard E. (2020). *Multimedia learning*, 3e éd., Cambridge, Cambridge University Press.
31. Mayer, Richard E. et Logan Fiorella (2014). « Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles », dans Richard E. Mayer (dir.), *The cambridge handbook of multimedia learning*, 2e éd, Cambridge, Cambridge University Press, coll. Cambridge handbooks in psychology, p. 279-315.
32. McComas, Joan, Morag MacKay et Jayne Pivik (2002). « Effectiveness of virtual reality for teaching pedestrian safety », *CyberPsychology & Behavior*, vol. 5, no 3, p. 185-190.
33. Naismith, Laura M., Jeffrey J. H. Cheung, Charlotte Ringsted et Rodrigo B. Cavalcanti (2015). « Limitations of subjective cognitive load measures in simulation-based procedural training », *Medical Education*, vol. 49, no 8, p. 805-814.
34. Paas, Fred, Juhani E Tuovinen, Jeroen JG Van Merrienboer et A Aubteen Darabi (2005). « A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction », *Educational Technology Research and Development*, vol. 53, no 3, p. 25-34.
35. Pellegrino, James W, Naomi Chudowsky et Robert Glaser (2001). *Knowing what students know: The science and design of educational assessment*, ERIC.
36. Qin, Hua, Pei-Luen Patrick Rau et Gavriel Salvendy (2010). « Effects of different scenarios of game difficulty on player immersion », *Interacting with computers*, vol. 22, no 3, p. 230-239.
37. Redifer, Jenni L., Christine L. Bae et Qin Zhao (2021). « Self-efficacy and performance feedback: Impacts on cognitive load during creative thinking », *Learning and Instruction*, vol. 71, p. 101395.
38. Riener, Robert et Matthias Harders (2012). « Vr for medical training », dans *Virtual reality in medicine*, Springer, p. 181-210.

39. Rodwin, B. A., V. P. Bilan, N. B. Merchant, C. G. Steffens, A. A. Grimshaw, L. A. Bastian, *et al.* (2020). « Rate of preventable mortality in hospitalized patients: A systematic review and meta-analysis », *J Gen Intern Med*, vol. 35, no 7, p. 2099-2106.
40. Rose, F David, Elizabeth A Attree, Bennett M Brooks, David M Parslow et Paul R Penn (2000). « Training in virtual environments: Transfer to real world tasks and equivalence to real task training », *Ergonomics*, vol. 43, no 4, p. 494-511.
41. Schroeder, Ralph (2008). « Defining virtual worlds and virtual environments », *Journal For Virtual Worlds Research*, vol. 1, no 1.
42. Schunk, Dale H et Maria K Dibenedetto (2016). « Self-efficacy theory in education », *Handbook of motivation at school*, vol. 2, p. 34-54.
43. Sewell, Justin L., Christy K. Boscardin, John Q. Young, Olle ten Cate et Patricia S. O'Sullivan (2017). « Learner, patient, and supervisor features are associated with different types of cognitive load during procedural skills training: Implications for teaching and instructional design », *Academic Medicine*, vol. 92, no 11, p. 1622-1631.
44. Smith, Shana et Emily Ericson (2009). « Using immersive game-based virtual reality to teach fire-safety skills to children », *Virtual reality*, vol. 13, no 2, p. 87-99.
45. Sweller, John (2011). « Cognitive load theory », dans *Psychology of learning and motivation*, vol 55, Elsevier, p. 37-76.
46. Sweller, John, Paul Ayres et Slava Kalyuga (2011). « Measuring cognitive load », dans *Cognitive load theory*, Springer, p. 71-85.
47. Van Merriënboer, Jeroen J. G. et John Sweller (2010). « Cognitive load theory in health professional education: Design principles and strategies », *Medical Education*, vol. 44, no 1, p. 85-93.
48. Vasile, Cristian, Ana-Maria Marhan, Florence Mihaela Singer et Daniela Stoicescu (2011). « Academic self-efficacy and cognitive load in students », *Procedia - Social and Behavioral Sciences*, vol. 12, p. 478-482.
49. Vasseur, Aurélie, Pierre-Majorique Léger, François Courtemanche, Élise Labonté-Lemoyne, Vanessa Georges, Audrey Valiquette, *et al.* (Forthcoming). «

Distributed remote psychophysiological data collection for ux evaluation : A pilot project », *International Conference on Human-Computer Interaction*

50. Wang, Peng, Peng Wu, Jun Wang, Hung-Lin Chi et Xiangyu Wang (2018). « A critical review of the use of virtual reality in construction engineering education and training », *International Journal of Environmental Research and Public Health*, vol. 15, no 6.

Chapter 3. Article 2

On the Assessment of Flight Simulator Instructional Design: The Effects of Flight Instruction Modality on Pilots' Cognitive Learning States and Performance

Laurie-Jade Rochon, Alexander J. Karran, Frédérique Bouvier, François Courtemanche, Constantinos K. Coursaris, Sylvain Sénécal, Jean-François Delisle, and Pierre-Majorique Léger

Abstract. Previous research has shown that the flight instructor is the most influential factor in a pilot's training progress. Since flight simulators (FS) deployed to the general public may now include training syllabi and flight instructions to complement or replace live instructions, it appears necessary to evaluate how to integrate flight instructions into lessons to ensure that apprentices are in optimal states for learning to happen. This manuscript presents research which evaluates how different modalities of flight instruction (*unimodal audio*, *unimodal text* and *bimodal audio and text*) influence the cognitive load, visual attention strategies, and motivational and affective states of student pilots when engaged in simulated flight tasks using Microsoft Flight Simulator 2020 (MFS) under VFR. We utilized a between-subject study design in which N=30 flight-school students performed two flight tasks. Task 1) involved an instructional flight task, following instructor guidelines, whereas 2) an evaluation flight task to evaluate the previous task training outcomes, in which the students performed a solo return flight without flight instruction. We found that participants in the *unimodal text* (F(1, 14) = [33.39], p <.0001) and *bimodal* (F(1, 15) = [13.01], p = 0.0052) flight instruction groups had significantly higher visual scanning efficiency (using Gaze Transition Entropy as a proxy, GTE) during the evaluation flight task, whereas participants in the *unimodal text* (F(1, 15) = [7.33], p = 0.0486) flight instruction group had significantly higher emotional valence. Additionally, we observed a non-significant trend across various self-perceived scales (lower implicit cognitive load, higher perceived immersion, and higher perceived motivation), indicating that the *unimodal text* condition is best suited to provide flight instructions to pilots. Our results suggest that the optimal flight instruction modality may vary depending on the flight segment in progress and its inherent tasks. This study adds to previous findings that indicated that performance-only metrics were insufficient to

assess the complex interactions between a pilot and their environment. This study and its results are of benefit to the aerospace training domain by providing a rigorous multidimensional assessment framework to evaluate pilots' cognitive learning states, including cognitive load, attentional strategies, motivational and affective states in a FS. Furthermore, it contributes to the literature related to FS instruction design by characterizing how various sensory modalities of flight instruction impact cognitive processes and learning as a consequence. Finally, it confirms boundary conditions of the CTML.

Keywords: Multimedia Learning • Flight Instruction • Professional Training • Pilot Training • Simulator • Flight Simulator · Ambient/Focal K Coefficient · Gaze Transition Entropy

Introduction

The use of training simulators in various fields presents numerous advantages, including the reduction of costs and risks generated by on-site training and technical equipment (Aragon et Hearst, 2005). In the aerospace industry, Transport-Canada Approved FS and flight training devices are used throughout flight school and airline training of aircraft operators. Moreover, pilots can benefit from training with Transport-Canada uncertified personal FS (Callender *et al.*, 2009) by learning about frequently used airports, familiarizing themselves with avionics and building muscle memory, practicing procedures, maneuvers, and preparing for all eventual flight scenarios. Recent FS deployed to the public are increasingly inexpensive, powerful, versatile, and designed to be used independently of a flight instructor, rivaling Transport-Canada approved systems (Callender *et al.*, 2009). Notably, MFS contains hundreds of task scenarios as well as a complex VE with unprecedented graphics quality that allows its users to immerse themselves in real-world map representations from around the globe. It can be played at home or in more advanced FS setups on various platforms including Windows, Classic Mac OS, and Xbox One.

A simulator training process is made of three parts: a simulator, a training syllabus and associated objectives, and an instructor (Myers III, Starr et Mullins, 2018). Until recently, these three parts were separate; the simulator itself did not train but was a tool used in the training process by both the trainee and the instructor (Myers III, Starr et Mullins, 2018). However, recently developed FS now include training syllabi and flight instructions, which can take the form of a virtual instructor or user interface (UI) display (e.g., flight objectives, graphical representations, and more). Previous work has shown that a flight instructor influences a student's progress significantly more than syllabi variations or the simulator environment (Myers III, Starr et Mullins, 2018). Thus, assessing how FS instructions should be conveyed to the users appears necessary. Moreover, researchers should investigate how the inclusion of FS instructions interacts with other FS characteristics (e.g., the level of fidelity and efficiency) and the user's learning experience (e.g., the user learning state and the user training transfer) (Myers III, Starr et Mullins, 2018).

Therefore, a major challenge for designers of instructional content used as training within FS is to create instructional material both sensitive and congruent with the characteristics of the human information processing system while at the same time reducing any interference with aircraft operating tasks. In this context, researchers and designers must consider different design aspects, including the amount and type of visuospatial and auditory information to convey to the user during a lesson to increase understanding and knowledge assimilation. Indeed, from a CLT perspective, presenting all instructional material through on-screen animations and text imposes a high load in the visual working memory system of pilots, since both types of information are processed in the same system (Baddeley, 1992). Conversely, providing auditory instructions in a FS may induce a lower load in the visual working memory because auditory and visual instructions are processed in their respective systems. This has been known as the "modality effect" (Mayer et Pilegard, 2005).

Sensory modalities, often used interchangeably with « senses », refers to what is perceived after a stimulus (LibreTexts, 2020). A sense corresponds to a group of cells that reacts to a physical phenomenon or signal that can be received and interpreted in specific brain regions (LibreTexts, 2020). Mammalians have the capability to combine different inputs of the sensory system, thus enabling multimodal perception (LibreTexts, 2020). We characterize the integration of two sensory modalities as "bimodal". For instance, visual inputs can include light, text, animations, pictures, whereas auditory inputs include sounds, speech, music, etc. A bimodal stimulus could consist of an on-screen animation combined with narrated speech.

The focus of the research presented in this manuscript is to evaluate the impact of different instructional material delivered using a variety of sensory modalities upon the cognitive learning states and learning performance of pilots over the course of a simulated flight task. To form the basis of this study, the following research questions were developed.

> **RQ1:** *To what extent does the sensory modality of FS instructions affect pilots' cognitive learning states?*
>
> **RQ2:** *To what extent does the sensory modality of FS instructions affect pilots' learning performance?*

To address these questions, we conducted a lab experiment with 30 flight-school pilots who completed two flights in the MFS: one flight that allowed us to assess pilots' cognitive learning states during a VFR flight with a virtual instructor and one VFR solo flight (i.e., without instruction) that allowed us to assess both the pilots' cognitive learning states and learning performance. The participants were divided into three experimental groups where the sensory modality of flight instructions was manipulated: one-third of the participants were presented with *bimodal – audio and text* – flight instructions, a third of the participants were presented with *unimodal audio* flight instructions only, and the remaining participants were presented with *unimodal auditory* flight instructions. To our knowledge, this study is the first to assess how the sensory modality of FS instructions affects pilots' cognitive learning states and learning performance using an educational, ecologically valid and widely used FS and training context.

Background

In the present section, we firstly present an introduction to the theoretical framework that serves as the basis for this study. We secondly provide a summary of the previous literature on the manipulation of navigation instruction sensory modalities in an aircraft/FS. Most studies listed in that field were "data link" studies. By definition, data links are telecommunications over which aircraft and flight information – as digital data – are transmitted between an aircraft operator and an air traffic controller (ATC) (Latorella, 1998). They can be used in various situations, for instance, for crossing the oceans when an aircraft is too far from the ATC to establish a radio communication or make a radar observation possible. Data link studies have longed to evaluate the benefits of transmitting telecommunications (i.e., text) rather than voice communications (i.e., audio) to provide ATC communications to pilots. Data link task paradigms involve concurrent tasks like operating the aircraft, listening/reading and processing navigational instructions similar to those happening in the current study. Consequently, we expect similar learning mechanisms to take place in this research. We thirdly present a synthesis of the literature on the assessment of pilots' cognitive learning states.

Multimedia learning : the « Modality Principle »

Richard Mayer et Mayer (2005) first described the CTML as an attempt to understand how to increase learning during a lesson, based on the basic premise that the arrangement of text and images enables deeper learning than text alone (Butcher, 2014). The vast body of literature derived from the CTML has since demonstrated that many factors (e.g., modalities, segmentation, pre-training, and more) can influence the cognitive load and learning achieved during a multimedia lesson, making clear that not all uses of multimedia are equally effective for the learner (Noetel *et al.*, 2022). These factors were translated into design principles, one of them stating that people learn more deeply from a multimedia message when the words are spoken rather than printed (Richard Mayer et Mayer, 2005). This "modality principle" would allow off-loading the verbal processing from the visual channel to the auditory channel by presenting the words as narration rather than on-screen text, substantially increasing learning (Richard E Mayer et Pilegard, 2005).

The brain processes information through multiple channels, such as the visuospatial and auditory channels, each activating different brain neural substrates (Baldwin *et al.*, 2012). Mayer's CTML outlines three central assumptions on how cognitive processing occurs: humans possess separate systems for processing pictorial and verbal material (i.e., dual-channel assumption) (Baddeley, 1992; Paivio, 1990), meaningful learning involves information processing, including building connections between pictorial and verbal representations to construct coherent mental models (i.e., active-processing assumption) (Richard E Mayer, 2005; Wittrock, 1989), and there is a limit to the amount of information that can be processed simultaneously along each channel (i.e., limited-capacity assumption) (Baddeley, 1992). The construction of both verbal and pictorial models can be influenced by prior knowledge and experiences retrieved from long-term memory (Richard E Mayer, 2003).

Moreno's work on the Cognitive-Affective Theory of Learning with Media (CATLM) (Moreno, 2005, 2006, 2007, 2009) expanded the CTML by incorporating aspects of motivation and meta-cognition to the framework, translated into two additional assumptions: motivational factors mediate learning by increasing or decreasing cognitive engagement (i.e., affective mediation assumption) (Gottfried, 1990; Moreno *et al.*, 2001; Park *et al.*, 2011) and metacognitive factors mediate learning by regulating cognitive and affective processes (i.e., meta-cognitive mediation assumption) (McGuinness, 1990; Morris, 1990). According to the CATLM, affective features of an instructional message can influence the level of learner engagement in cognitive processing during learning (Moreno, 2006). Instructional design techniques that manage to enhance motivation and positive affect, that ensures that the cognitive capacity is not continually overloaded and that the learner can learn the essential material may prove beneficial to learning.

To this day, the modality principle is arguably the most studied of all CTML principles. It was widely studied in cognitive psychology and instructional design. Studies that yielded positive results involved diverse content like geometry, biology and environmental science (Jeung, Chandler et Sweller, 1997; Moreno et Mayer, 2000; Mousavi, Low et Sweller, 1995; Pirovano *et al.*, 2020), and media, such as computer-based lessons, animations, and VR (Makransky, Terkildsen et Mayer, 2019; Moreno, 2006). Moreno (2006) wanted to evaluate the extent to which the modality principle applies to different media, and the extent to which specific media characteristics affect learning. We use the term media to describe the "physical system or vehicle used to deliver the lesson" (Moreno, 2006). In two 2X2-factor design studies, she evaluated the modality principle in low-immersion (desktop VR) and high-immersion (VR using a head-mounted display) conditions, and in the absence or presence of an animated pedagogical agent

(APA). In both studies, while increased immersion and the presence of an APA did not affect learning outcomes, students performed better on tests of retention and problem-solving transfer when words were presented as speech rather than on-screen text, therefore producing a modality effect. Hence, support was garnered for the "media-affects-learning" hypothesis, regardless of the instructional method used (Moreno, 2006).

Since the early foundational studies (Jeung, Chandler et Sweller, 1997; Kalyuga, Chandler et Sweller, 1999; Richard E Mayer et Moreno, 1998; O'Neil *et al.*, 2000; Tindall-Ford, Chandler et Sweller, 1997), several researchers have also worked toward testing the boundaries of the modality principle. For instance, a review performed by Reinwein (2012) showed moderate evidence for the modality principle, particularly when the presentation was system-paced rather than self-paced, when the graphical displays were dynamic rather than static, and when transfer measures were used rather than retention measures. In addition, Richard Mayer et Mayer (2005) outlined that the effects of the modality principle were maximized when there was no split-attention or redundancy of information across modalities.

Research on the modality principle in a FS context is scarce. A FS training has high-element interactivity. That is, understanding requires that all elements be maintained in the learners' working memory and manipulated simultaneously, which can result in highly complex tasks (Pociask et Morrison, 2004). Indeed, FS training involves additional learning mechanisms than those that derive from a usual multimedia lesson. Users are not only faced with instructional content, but also need to be attentive to environmental cues to operate the aircraft accurately. Hence, there are important attentional and psychomotor components and the cost of these cognitive tasks could interact with the treatment of the instructional material.

Manipulating the Modality of Navigational Instructions in the Aircraft: The Case of Data Link

No study have previously compared the impacts on learning of using different sensory modalities to provide instructions to pilots during FS training. Previous research that has investigated the modality principle in a FS were "data link" studies (John R Helleberg et Wickens, 2003; Lancaster et Casali, 2008; Latorella, 1998; McGann *et al.*, 1998;

Rehmann, 1997). These studies aimed at evaluating which modality was better suited for carrying out ATC communications between the ATC tower and an aircraft operator. Text communications, also called data links, were compared with audio communications; some studies added a redundant condition in which data links were displayed while equivalent audio messages were played. We have a particular interest in these study results as they tested how various sensory modalities affected performance metrics outcomes using the same media used in the current study, i.e., an aircraft or a FS. We report below the main results for this field of research, as well as possible explanations that account for the cognitive and attentional demands and processes of aircraft operating tasks.

Both auditory and textual conditions generated positive performance outcomes. Better performance results for the textual condition were justified by the permanent nature of the instructions as well as by a possibility to confirm the accuracy of the perceived information (Rehmann, 1997). An additional explanation referred to the stimulus-central processing-response (SCR) compatibility model proposed by Wickens *et al.* (2021), stipulating that there is compatibility between modes of input (visual or auditory), central processing modes (spatial or verbal) and output modes (speech or manual). This compatibility exists when a spatial task is perceived visually and responded to manually through the visuomotor network) (Fitts et Deininger, 1954; Fitts et Seeger, 1953; Morin et Grant, 1955; Wickens *et al.*, 2021). Thus, navigational instruction perceived on a simulator screen could be understood more easily than an equivalent audio instruction. On the other hand, better performance results were justified by an auditory pre-emption effect (John R Helleberg et Wickens, 2003) and an inherent increased sense of urgency to respond when a signal was perceived auditorily (Latorella, 1998). It was also justified that the code for the acoustic material lasts longer than the visual code, allowing better retention of the perceived information and leading to better performance. Lancaster et Casali (2008) also found that the visual modality induced an increased time to respond and increased ratings of workload relative to the auditory condition, while the auditory mode permitted better navigation message clarification. A possible explanation for this is that since navigational messages are visual, the auditory modality confers a bimodality to information processing, which reduces cognitive load (Mousavi, Low et Sweller, 1995). Additionally, previous research suggested that there may be a conflict when translating textual messages into spatial relations; however, this effect was not observed when messages were heard rather than read (Brooks, 1967). On the contrary, in another study, worst performance results in pilot monitoring and read-back tasks in both auditory and redundant conditions were justified by the disruptive nature of auditory signals, as they would interfere too much with aircraft operating tasks (Helleberg et Wickens, 2003). Previous research has shown that an interruptive auditory task (e.g., an audio navigational instruction) will capture and demand attention, to the detriment of an ongoing visual task (e.g., monitoring signals to operate the aircraft), at a degree to which a visual task will not (e.g., a data link) (Latorella, 1998; Wickens et Liu, 1988). Indeed, auditory channel has attention-demanding properties on the emergence of the stimulus; humans also desire to keep their attention focused longer on an auditory interruptive task of some complexity, since to do otherwise would risk the loss of working memory information (Wickens, Dixon and Seppelt, 2005).

The body of literature on the manipulation of sensory modalities to establish communications in the aircraft offers great insights as to why auditory, text or both types of messages could benefit or hinder learning in a FS training scenario. However, data link studies place great importance on negligible factors in the current context of research, namely, the need to react quickly to, or not to misunderstand an ATC communication. We must also consider that the nature of flight instructions transmitted to pilots during FS training is not strictly navigational. Besides, from these results, it is interesting to recognize that learning processes other than cognitive load (highlighted in CTML), namely attentional and affective processes (highlighted in CATLM), may also have a great impact on results during complex interactions between a user and a FS.

Monitoring pilots cognitive learning states

Based on the CTML (Richard Mayer et Mayer, 2005) and the CATLM (Moreno, 2005, 2006, 2007, 2009), this study proposes a framework (Figure 2) to evaluate the subjects' learning experience that enhances the cognitive perspective by taking attentional, motivational and affective aspects into account. We provide background for various measures selected for this study.

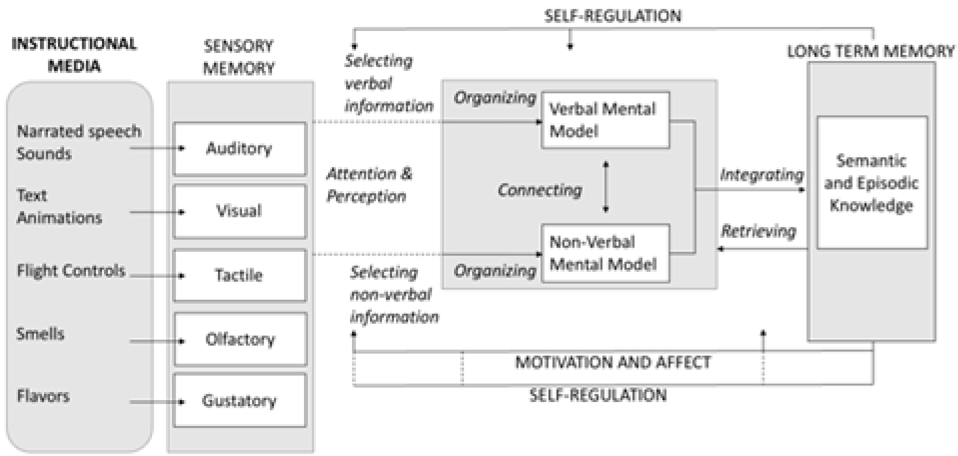


Figure 2. *Cognitive-affective theory of learning with media, reproduced from Moreno (2005).*

Cognitive load in pilot training

In CLT, cognitive load refers to the used amount of working memory resources (Kalyuga, 2008). First, to operate an aircraft efficiently and safely, pilots must continually pay attention to critical instrument panel cues. Processing these signals and producing adequate psychomotor responses involve a significant cognitive load. Second, in multimedia learning, a challenge is created when learners need to allocate extra effort to make sense of the information presented through different sensory modalities. Depending on the instructional design, much of the effort can be aimed at mentally integrating different sources of information rather than directed towards learning the relevant material. Multimedia learning design principles aim at reducing extraneous cognitive processing by creating effective instructional design (Richard E Mayer, 2014). Cognitive load can be operationally defined using perceptual indices such as verbal reports, or through objective behavioral and psychophysiological indices (Charlton, 2002).

The most widespread index used in the literature provides an inference for the level of cognitive load involved during a task, reflecting either task demands or load, or the effort exerted in response to these demands (van der Wel et van Steenbergen, 2018). When individuals face demanding cognitive tasks, their pupils dilate. Different methods have been employed to transform pupillary responses in study data analyses, notably the unadjusted pupil diameter, the Percentage Change in Pupil Diameter (PCPD), the Index of Pupillary Activity, and the Patented Index of Cognitive Activity (J. Attard-Johnson, C et Bindemann, 2019; Beatty, 1982; Weber *et al.*, 2021). Early studies commonly performed analysis on raw pupil diameter data (Hamel, 1974; Nunnally *et al.*, 1967; Scott, Wells et Wood, 1967), however, several have demonstrated that using raw pupil diameter data to estimate cognitive load was not comparable between participants as individuals generally differ in pupil size (Janice Attard-Johnson, Ó Ciardha et Bindemann, 2019). Authors have proposed the use of PCPD instead. PCPD can be measured by calculating the difference between the pupil diameter measured during a task and a pre-stimulus baseline level, divided by the pre-stimulus baseline level (Janice Attard-Johnson, Ó Ciardha et Bindemann, 2019). This baseline typically corresponds to an average value over a few seconds period of pupil diameter data measured before the experiment (J. Attard-Johnson, C et Bindemann, 2019). Pupillary responses reflect a broad range of higher cognitive processes. Changes evoked by pupillary light reflexes can be described as large (i.e., up to several millimeters), whereas those evoked by cognitive activity can be described as relatively small (normally between 0 ,1 and 0.5 mm) (Beatty, 1982). Using pupil diameter as an index to measure various characteristics of a user state therefore requires the control of environmental factors, namely, light or the angle of the camera relative to the user (Duchowski *et al.*, 2018). These external factors controlled for, pupillary responses represent a window into the cognitive processes involved in activities such as learning and memory.

Visual attention in pilot training

Researchers have taken an interest in eye tracking measurements since they represent the possibility of a window on the cognitive state of a pilot, allowing them to deepen their understanding of the behaviors and information processing performed by aircraft operators (for a review, see Glaholt (2014)). For instance, a novice pilot might not detect an aircraft in the surrounding airfield because their attention is focused on processing the signals displayed on the instrument panel. If an expert pilot detects an aircraft whose trajectory represents no danger, both the novice and expert pilots will maintain their respective course and no discernible difference between the performance data will be recorded. In this case however, the eye tracking derived data could allow to differentiate the two individuals.

Pilot training research experimental paradigms have used diverse methods, namely simple eye movement metrics including fixation and saccade data, and sequence analyses approaches using a combination of fixation, saccade and Areas of Interest (AOIs) data to evaluate the cognitive, attentional states of pilots (Glaholt, 2014). We propose utilizing a set of visual scanning measures to render a complete portrait of a pilot's learning experience during FS training.

GTE is an information theory measure (Shiferaw, Downey et Crewther, 2019) that provides an estimation for the level of complexity or randomness in the pattern of visual scanning relative to the overall spatial dispersion of gaze. In other words, it describes the amount of information required to describe the visual strategies of a user, using the formula:

$$H(x) = -\sum_{i=1}^{n} p_i \sum_{j=1}^{n} p(i,j) log_2 p(i,j) \quad (1)$$

where i represents the "from" AOI and j represents the "to" AOI, p_i represents the stationary distribution and $p_{(I,j)}$ represents the probability of transitioning from i to j. Higher GTE denotes more randomness and more frequent switching between AOIs.

GTE is influenced by task complexity, icstress and pilot expertise. Indeed, higher task cognitive load (Ephrath *et al.*, 1980; Van De Merwe, Van Dijk et Zon, 2012; van Dijk, van de Merwe et Zon, 2011) and levels of stress (Allsop et Gray, 2014) correlated with higher GTE. In another study by Diaz-Piedra *et al.* (2019), the complexity of task caused a decrease in pilot's visual transition entropy, whereas C. Lounis, Peysakhovich et Causse (2021) found that expert pilots had more elaborate visual patterns.

Learning to navigate under VFR in a FS involves performing several tasks concomitantly: processing flight instructions, applying corrections, monitoring flight instruments, inspecting the environment outside the window (OTW), operating the aircraft, and more. These tasks all have an attentional cost. As they train, pilots learn to allocate an appropriate proportion of their attentional resources to operate the aircraft safely and efficiently. Hence, it appears appropriate to evaluate how flight instructions of various modalities can capture a user's attention, and how this attention demand affects a user's cognitive learning states. A metric that allows the evaluation of the dispersion of visual attention is the ambient-focal K coefficient introduced by Krejtz *et al.* (2016).

The "K coefficient" allows to investigate the dynamics of changes in ocular exploration and inspection behaviors. Negative and positive ordinates of K indicate ambient viewing (governing initial scene exploration) and focal viewing (common during scene inspection), respectively (Krejtz *et al.*, 2016). As for the abscissa, it indicates time, so that K acts as a dynamic indicator of fluctuation between ambient and focal visual scan behavior. *K* is derived as the mean difference between standardized values (z-scores) of each saccade amplitude (a_{i+1}) and its preceding i^{th} fixation duration (d_i):

$$K_i = \frac{d_i - \mu_d}{\sigma_d} - \frac{a_{i+1} - \mu_a}{\sigma_a} \qquad (2)$$

where μ_d, μ_a are the mean fixation duration and saccade amplitude, respectively, and σ_d, σ_a are the fixation duration and saccade amplitude standard deviations, respectively, computed over all *n* fixations and hence *n Ki* coefficients.

Values of K_i close to zero indicate relative similarity between fixation durations and saccade amplitudes. Positive values of K_i show relatively long fixations followed by short saccade amplitudes, which indicate focal attention. Conversely, negative values of K_i point towards relatively short fixations followed by relatively long saccades, suggesting ambient or diffuse attention (Unema *et al.*, 2005). In the former case (i.e., in focused mode), attention is concentrated at a few areas of interest, specified by a central or peripheral cue. In the latter case (i.e., in diffuse mode), visual attention is allocated to all regions of the visual field in quite equal proportion (Heitz et Engle, 2007).

C. Lounis, Peysakhovich et Causse (2021) found that the attentional dispersion of gaze was more focal in a novice pilot group as compared with an expert group, which suggests that expert pilots have a greater spatial distribution of their visual attention than novices. They also found that the K coefficient showed greater sensitivity to the task difficulty, where difficult tasks led to a switch from focal to ambient processing for both novice and expert groups.

Motivational and affective processes in pilot training

Immersion has been defined as "a state of deep mental involvement in which the individual may experience disassociation from the awareness of the physical world due to a shift in their attentional state" (Agrawal *et al.*, 2020). Immersion is based on the extent to which visual displays support an illusion of reality that is inclusive (denoting the extent to which physical reality is shut out), extensive (the range of sensory modalities accommodated), surrounding (the size of the field of view), and vivid (the display resolution, richness, and quality) (Slater et Wilbur, 1997). Previous studies have demonstrated that high immersion increased user motivation and engagement (Bailenson *et al.*, 2008; Dalgarno et Lee, 2010; Dede, 2009; Dejian Liu *et al.*, 2017). Moreover, designers have longed to design FS that provide the most training transfer (Myers III, Starr et Mullins, 2018). Positive training transfer happens when performance in the aircraft is better than if there was no simulator training provided, as opposed to negative training transfer that happens when performance in the aircraft is poorer than if there was no pre-training at all (Lintern, 1991). Among other factors (e.g., simulator fidelity, presence, operator buy-in), increased immersion has shown to drive positive training transfer (Alexander *et al.*, 2005).

Positive affect has shown to facilitate thinking, problem-solving, motivation and learning (Isen, 2004). Indeed, as per the CATLM, multimedia components that can enhance learners' affective and motivation states can positively mediate their cognitive engagement. Specifically, for self-administered pilot training, motivational and hedonic user states should not be overlooked as they directly affect usage intentions and the overall training efficiency (Wu et Holsapple, 2014).

Hypothesis development

Extensive previous literature has demonstrated that in multimedia learning, presenting instructional material in more than one modality promotes deeper learning, leading to increased performance in retention and transfer tests (Ginns, 2005; Richard E Mayer et

Moreno, 1998; Moreno et Mayer, 1999). However, no study has specifically tested the modality principle in a FS training context.

Performance data derived from aircraft parameters can provide information about a pilot's learning state. These measurements include deviation in heading, altitude, and speed with respect to the flight plan. The phrase "cognitive state" has been defined as the state of human cognitive processes and resources, such as perception, attention, cognitive effort, engagement, working memory, arousal, stress and fatigue (Dirican et Göktürk, 2011). It is possible that an impaired cognitive state during the learning process does not manifest as a significant change of learning performance. However, thoroughly evaluating a user cognitive learning states – i.e., cognitive load, visual attention, motivation and affect – as he trains and at the end of the training process should allow researchers identify an optimal sensory modality to provide FS instructions to student pilots. Hence, we firstly posit that **H1** bimodal (auditory and textual), unimodal auditory and unimodal textual flight instructions will influence pilots' cognitive learning states at different levels.

As per the CTML and the CATL, we distinguish and evaluate three relevant aspects of a pilot's cognitive learning process – **H1A** cognitive load, **H1B** visual attention and **H1C** motivation and affect – therefore, underlying three sub-hypotheses:

H1A. Bimodal (auditory and textual), unimodal auditory and unimodal textual flight instructions will generate different levels of objective and subjective cognitive load. In this case, the FS instruction modality conferring an optimal cognitive state during learning will allow the reduction of extraneous processing, resulting in lower cognitive load during an evaluation flight task.

H1B. Bimodal (auditory and textual), unimodal auditory and unimodal textual flight instructions will influence pilots' attentional strategies, as indicated by various levels of visual transition entropy and visual ambient-focal attention dispersion. In this case, the FS instruction modality that facilitates an optimal cognitive learning state will allow the adoption of appropriate attentional visual strategies, resulting in expanded visual attention dispersion (i.e., low ambient-focal K coefficient) with gaze patterns which are organized rather than random (i.e., low GTE/H_{max}) during an evaluation task.

H1C. Bimodal (auditory and textual), unimodal auditory and unimodal textual flight instructions will influence pilots' motivation and affect, as indicated by various levels of objective emotional valence, subjective motivation and immersion. In this case, the FS instructional modality that places the subject in an optimal cognitive learning state will result in increased emotional valence during an evaluation flight task, as well as increased self-ratings of motivation and immersion after an evaluation task.

The data link body of literature shows that textual and bimodal instructions lead to increased performance, as measured by accuracy in the execution of navigational instructions (John R Helleberg et Wickens, 2003), whereas audio instructions allow better performance when using response times as a proxy (Lancaster et Casali, 2008). The sensory modality of flight instructions that allows for deeper learning in a FS should lead to increased execution of flight objectives. Based on the inconsistency of the aforementioned results, we secondly posit that :

> **H2** Bimodal (auditory and textual), unimodal auditory and unimodal textual flight instructions will generate various levels of pilot learning performance.

Methods

Ethics statement

This research project was approved by our institutional ethics committee (code N°2020-3559). Participants signed an informed consent form prior to beginning the study. They were informed of their right to stop the experiment at any point if they wished to.

Participants

Thirty pilot students were recruited at a local flight school to participate in this study. Selected subjects were older than 18 years old and understood advanced spoken and written French or English. Candidates who had laser vision correction or astigmatism, had a neurological or psychiatric diagnosis, and suffered from epilepsy were excluded. In a recruitment questionnaire, candidates indicated their previous use of various FSs, their Total Flight Hours (TFH), and their flight qualification (aircraft type). Participants were assigned as to control the level of expertise of each sensory modality experimental group. Novices (NOV) had TFH ranging from 25 to 100; intermediates' (INT) TFH ranged from 101 to 200; and advanced' (AD) TFH were over 200. Group 1 included 11 students (4 NOV, 5 INT, 2 AD; 11 males; mean age 24,6 ± 3,6 years), Group 2 included 10 students (3 NOV, 5 INT, 2 AD; 9 males; mean age 24,1 ± 3,7 years), and Group 3 consisted of 9 students (3 NOV, 4 INT, 2 AD; 7 males; mean age 24,3 ± 6,0 years). All subjects had prior theoretical knowledge and a good comprehension about the various information displayed on the cockpit instrument panel; they had flight notions on how to manually interact with the aircraft; and they had already flown a Cessna-152. Finally, all subjects had used a FS, but were first-time users of the VFR module of MFS 2020.

Experimental design and procedure

Three experimental conditions were tested using a between-subject study design, where the sensory modality of flight instructions provided to a user during an instructional flight task was manipulated. The *bimodal* condition consisted of MFS default parameters. It included a synthesized speech virtual instructor as well as a textual flight objectives display. The *unimodal audio* condition included a synthesized speech virtual instructor, however no additional flight objectives were displayed on screen. In the *unimodal text* condition, both textual virtual instructor guidelines and flight objectives were displayed on screen.

The 90-minutes experiment (Figure 3) was conducted at the laboratory. After being guided through completing a consent form and a 7-point eye-tracking calibration, a subject read a task description (i.e., a flight plan) on a computer tab web browser, then the simulation appeared on their screen. The simulation screen displayed a navigation log including the check points, the times in route and the headings to reach Airports A and B. A subjective cognitive load questionnaire (i.e., NASA-TLX) was filled out by participants after each flight task segment, whereas additional psychometric questionnaires (i.e., Immersion, IEQ; Motivation, SIMS) were filled out at the end of the experimental tasks. The session ended with an interview to gain information on subjects' perceptions of their learning experience and about the system's strengths and weaknesses. The collection tools were removed. The participant filled out a form to be compensated $CAD 30 and one participation in a draw to win a prize worth $ 600, i.e., MSF 2020 and an Xbox Series S.

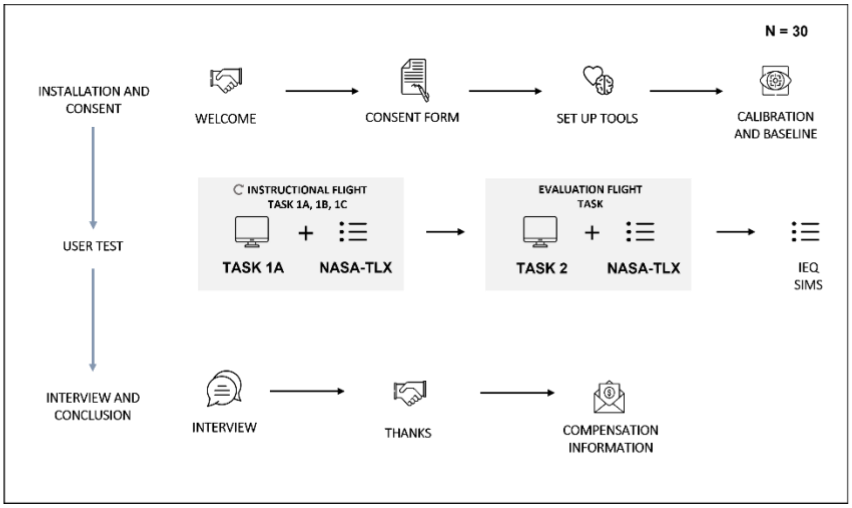


Figure 3. *Experimental procedure.*

Apparatus and simulated scenario

Apparatus

The MFS 2020 software was used for this experiment. The simulation was presented on a 27-inch HP EliteDisplay E2273q (2560 x 1440 px) computer screen. The subjects controlled the aircraft with a yoke, a side-stick, two thrust levers, and a rudder (Figure 4). They could use a joystick on the yoke to change view and gain better visibility OTW in the VE. The participants' screen was recorded and their flight performance was assessed by an experienced pilot after the experimental sessions using the session recordings. The aircraft flown was a Cessna-152. It was depicted accordingly in high definition in the simulation.

Eye movements were recorded at a sampling frequency of 60 Hz using the Tobii Pro Nano (Tobii, Stockholm, Sweden) eye tracker, which uses near infrared diodes to identify the position of each eyeball in the three-dimensional space and to calculate the gaze point on the screen (Tobii Pro, 2021).

The cockpit was split into 8 to 10 AOIs corresponding to the flight instruments and instruction displays necessary for successful task-completion (Figure 5). AOIs included a flight deck, a navigation log and an external view (i.e., OTW). Two condition-specific AOIs were also analyzed: a flight objectives display (*bimodal*, *unimodal text*) and a textual flight instructor (*unimodal text*).

Figure 4. *Apparatus.*

Figure 5. *Overview of the ten different AOIs: 1. Flight deck (including AOIs 4–9), 2. Navigation Log, 3. OTW view, 4. Airspeed Indicator, 5. Attitude Indicator, 6. Altimeter Indicator, 7. Heading Indicator, 8. Vertical Speed Indicator, 9. Power Indicator, 10. Flight Objectives Display.*

Simulated scenario

Following a series of tests carried out with a flight instructor from the local flight school, the VFR module developed by Asobo Studio (Asobo, Bordeaux, France) was selected for this study. The training module presents moderate task difficulty, moderate task length, a familiar aircraft type, various instruction modalities; flying in VFR requires pilots allocating a portion of their visual spatial attention outside the aircraft to locate landmarks, thus creates competition for attention when presenting other instructional material.

The experimentation was separated into two main tasks. In a first “instructional flight task”, a participant flew from Sedona Airport to Flagstaff-Pulliam Airport with the help of their virtual instructor. In a second “evaluation flight task”, a participant flew from the Flagstaff-Pulliam Airport back to Sedona Airport during a solo flight without a virtual instructor. During this second evaluation flight task, no instructions were provided to pilots, thus, the task was identical for across experimental conditions. Subjects were informed that the first task’s flight objectives would be evaluated during the evaluation flight task. Hence, the second task aimed at assessing how the modality of flight instructions during an instructional flight task led to training outcomes during an evaluation flight task. The description of the flight scenario is presented below (Table 5).

Flight Task	Flight task segment	Description
Instructional flight task – From Airport A (Sedona) to Airport B (Flagstaff-Pulliam) — 30 mins	Departure	Via instructions presented throughout the task, a participant learns how to take off from Airport A, how to climb to a prescribed altitude, and how to maintain a prescribed heading.
	Dead Reckoning	Via instructions presented throughout the task, a participant learns how to find a *checkpoint* during a flight, and how to start the stopwatch to calculate the duration of a flight segment to compare it with the *Navigation Log*.
	Arrival	Via instructions presented throughout the task, a participant learns how to *integrate a circuit* when arriving at Airport B, and how to land on a specified *Runway*.
Evaluation flight task – From Airport B (Flagstaff-Pulliam) to Airport A (Sedona) — 30 mins	N/A	In a first solo flight, a participant is required to take off from Airport B, to climb to a cruising altitude and heading; to use diverse methods of navigation (*Landmark Navigation* and *Dead Reckoning*) to navigate to Airport A; to arrive at Airport A, integrate its circuit and land.

Table 5. *Task description.*

The instructions provided to users throughout the session took two forms. First, a synthesized or textual speech virtual instructor informally provided instructions to pilots. The synthesized speech instructor could be heard through the computer speakers, whereas the textual speech instructor could be read directly on-screen. The virtual instructor was responsive to participant behaviors and would therefore repeat instructions, bring back a user to previous flight objectives, or explain participant mistakes if needed. Second, a flight objective display appeared in the upper-right corner of the UI simulator screen and would summarize concise flight objectives in real-time. Flight objectives that had to be met and maintained (e.g., “Maintain 8000 ft.”) would dynamically appear/dissapear on the screen signaled in green when correctly performed by participants, whereas flight objectives that had to be met but not maintained (e.g., “Reach 8000 ft.”) would be successively displayed.

All experimental tasks were performed linearly by participants to reproduce a real-world flight setting (i.e., departure to landing; Airport A to B and back). The learning performance was assessed at the task level (i.e., instructional flight task, evaluation flight task) and at the flight segment level (i.e., departure, navigation, arrival) for the instructional flight task, where each flight objectives displayed in the simulator UI window was evaluated. Performance dependent variables included the exactitude in speed, altitude, heading, power, the deviation from the flight course and “pass or fail” flight objectives.

Operationalization of Measures

To measure a subject’s cognitive learning states during their interaction with MFS, we described participant visual attention, cognitive load, motivation and affect using multiple measures. We also measured the cognitive states’ effects on the learning performance of pilots. The operationalization of the measures is described in the **Table 6**.

The visual attention of pilots was evaluated through measures of visual transition entropy and visual attention dispersion. GTE/H_{max} was used as a measure to quantify pilots’ visual transition entropy using the method previously described in 3.2.1. Pilots’ ambient-focal K coefficient was assessed using the method described in 3.2.2.

To assess participant self-perceived cognitive load, we utilized the NASA-TLX[11], a well-established multi-dimensional rating scale (Hart et Staveland, 1988). The NASA-TLX includes six items and factors: mental demand, physical demand, temporal demand, overall performance, effort, and frustration. The objective cognitive load of pilots was measured using the PCPD using the last 10 seconds of each task as a baseline. This baseline was used to ensure that the screen lighting condition was that of the simulator, and because the simulator took this amount of time before switching view after the last task flight objective was completed by a participant.

To assess the motivational and affective states of pilots, we measured participant emotional valence, subjective motivation and subjective immersion. We assessed

[11] Voir Annexe 1.1 – Questionnaires : NASA-TLX

participant emotions during the task using a facial emotion recognition software (see 4.7.2 below). The participants' subjective motivation was measured using the Situation Motivational Scale (SIMS)[12], a 16-items scale developed by Guay, Vallerand et Blanchard (2000) that includes constructs of intrinsic motivation, identified regulation, external regulation and amotivation. Pilots' perceived immersion was measured using the Immersive Experience Questionnaire (IEQ)[13], a 31-items scale developed by Jennett *et al.* (2008) that includes affective, cognitive, real-world dissociation, challenge and control components while playing a game.

Measure	Psychophysiological State (Response to Stimuli)	Operationalization
Visual attention	· Visual attention dispersion · Visual transition entropy	· Ambient-focal K coefficients (Krejtz *et al.*, 2016) · GTE/H_{max} (Shiferaw, Downey et Crewther, 2019)
Cognitive load	· Cognitive load	· Subjective cognitive load: NASA-TLX (Hart et Staveland, 1988), objective cognitive load: PCPD from baseline (J. Attard-Johnson, C et Bindemann, 2019)
Motivation and affect	· Valence · Motivation · Immersion	· Emotional valence using FaceReader (Loijens et Krips, 2018) · SIMS (Guay, Vallerand et Blanchard, 2000) · IEQ (Jennett *et al.*, 2008)
Learning performance		· Subjective "overall performance": NASA-TLX (Hart et Staveland, 1988) · Observed in-flight performance

Table 6. *Operationalization of the measures.*

[12] Voir Annexe 1.4 – Questionnaires : SIMS
[13] Voir Annexe 1.5 – Questionnaires : IEQ

Data processing

All data were analyzed using the statistical software SAS 9.4 (SAS, Cary, USA). Time markers corresponding to the stimuli, task beginning and end were sent by Tobii Pro Lab v.161 (Tobii, Stockholm, Sweden) to the Observer XT software (Noldus, Wageningen, Netherlands). *Post hoc* synchronization of the physiological data was run via the Cobalt Photobooth software (Courtemanche *et al.*, 2022).

Flight simulator data

An experienced pilot watched the participant screen recordings using the Tobii Pro Lab video replay function to assess flight performance. Each flight objective was marked as « 0 » (i.e., failure), « 1 » (i.e., partial success) or « 2 » (i.e., success). When the FS made the user start over at a previous flight objective, the unsuccessful objective was marked as failed. In this case, we kept the score of the first trial for each flight objective performed twice and started scoring normally when the participant was past the objective that led to the backtracking of the simulation. For the "Maintain altitude/heading/speed" objective types, « 0 » was assigned if the flight objective in the simulator window appeared green less than 25% of the time, « 1 » was assigned if it appeared green 25-75% of the time, and « 2 » if it appeared green more than 75% of the time. If a participant was not able to finish a task, each flight objective not performed was marked as a failure. Weights were applied to flight objectives as to fit the score computed by MFS 2020. During the instructional flight task, each flight segment (i.e., departure, navigation, and landing) gave a total score on 20 Pts. The instructional flight task and the evaluation flight task gave total scores on 60 Pts.

Eye tracking data

The eye tracking data was pre-processed in Tobii Pro Lab v.161 (Tobii, Stockholm, Sweden). As participants could switch views in the aircraft, AOIs were coded manually after data collection. Event markers were positioned at the start and end of each experimental task for each participant. Task duration varied consequently to participant actions. The AOIs data were extracted from the raw data. We utilized the Tobii Pro Lab Tobii I-VT fixation filter which is based on the work of Salvucci et Goldberg (2000) and Komogortsev *et al.* (2010). Fixations inferior to 60 ms were discarded and a velocity threshold of 30 degrees/second was used. To compute the GTE and K coefficients, home-built scripts were coded following the methodology described in Shiferaw, Downey et Crewther (2019) and Krejtz *et al.* (2016), respectively.

Emotional valence

The facial video stream of participants recorded with a webcam was retrieved and analyzed in FaceReader v6.0 (Noldus, Wageningen, Netherlands) to record and model emotional valence in real time. FaceReader analyzes participant facial movements to detect six emotions. It then calculates emotional valence as the intensity of positive emotion minus the intensity of negative emotions, which renders a score between -1 (negative) to 1 (positive) (Ekman et Friesen, 1978; Loijens et Krips, 2018).

Statistical analysis

All analyses were either performed at the flight task level (i.e., "learning" and "evaluation") or at the flight segment level (i.e., "departure", "navigation" and "arrival" segments) of the instructional flight task. The statistical tests are based on data aggregated (i.e., one data point) per participant and task for all analysis performed at the flight task and flight segment levels. Using SAS 9.4 (SAS, Cary, USA), the IEQ and SIMS were assessed once after the instructional flight task and tested using a linear regression with random intercept model. P-values were adjusted for multiple comparison using the Holm-Bonferroni method. A Kruskal Wallis Test was used to evaluate if the performance differed by condition at the flight task and flight segment levels. We performed a repeated measures analysis of variance (ANOVA) for each of the following dependent variables (DVs) to assess the effects of the sensory modalities at the flight task and flight segment levels (Holm-Bonferroni corrected): PCPD from participant baseline, emotional valence, GTE/H_{max} and ambient-focal K coefficient. We conducted a Kruskal-Wallis Test to test the NASA-TLX results at the flight task level, and a linear regression with random intercept model (Holm-Bonferroni corrected) was performed to assess if the DV differed by condition at the flight segment level.

Results

For each hypothesis, we first report results at the *task level* to assess the effect of the modalities on the DVs during an instructional flight task to measure pilots' cognitive learning states, and during an evaluation flight task to measure cognitive learning states and learning performance. For each hypothesis, the results are secondly reported at the *flight segment level* to evaluate the effects of the modalities on the DVs during the departure, navigation, and landing parts of the instructional flight task. Only the answers to the psychometric questionnaires evaluating subjective immersion and motivation were collected only once following the evaluation flight task to minimize the negative effects of a lengthy experimental session (e.g., boredom, fatigue) and to prevent participants' responses from being affected by the redundancy of questions.

Pilot cognitive learning states

Visual Transition Entropy (GTE/H_{max})

The visual transition entropy results are presented in the **Figure 6** below. A two-way type III ANOVA did not reveal that there was a statistically significant difference of visual transition entropy scores between at least two sensory modality conditions during the instructional flight task ($F(2, 24) = [0.23]$, $p = 0.7977$). Results however revealed a statistically significant difference of GTE/H_{max} scores between at least two sensory modality conditions during an evaluation flight task ($F(2, 21) = [12.07]$, $p = 0.0003$).

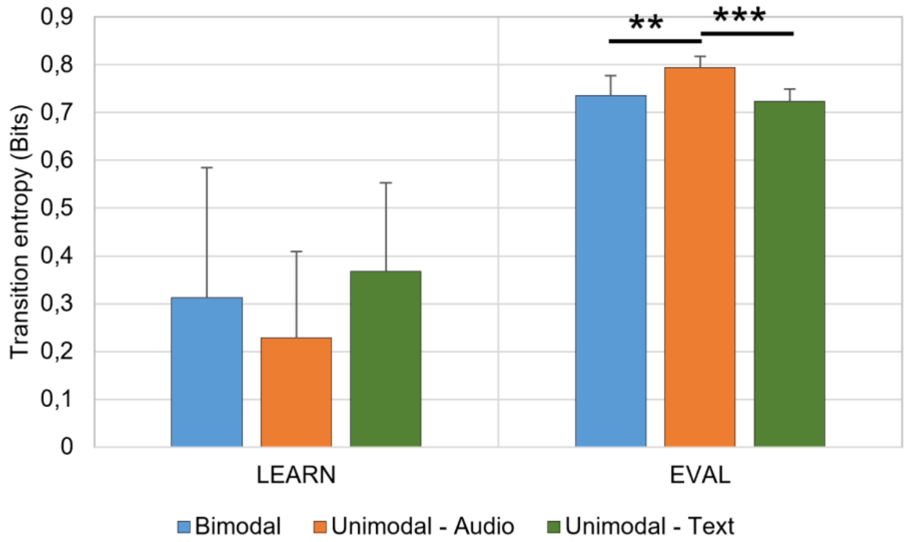


Figure 6. Mean GTE/H_{max} scores per modality group for the instructional flight task (left) and the evaluation flight task (right). *Error bars represent SD, * indicates main effects p < 0.05, ** indicates main effects p < 0.01 and *** indicates main effects p < 0.001.*

The pairwise comparisons indicate that the mean value of GTE/H_{max} was significantly different between the bimodal condition and the audio condition (F(1, 15) = [13.01], p = 0.0052) and between the audio condition and the text condition (F(1, 14) = [33.39], p <.0001). During the solo evaluation flight task, the gaze patterns were more organized for the bimodal and text unimodal conditions as shown by lower entropy ratios, whereas the entropy ratio was slightly higher for the audio unimodal condition, indicating more gaze pattern randomness.

A two-way type III ANOVA did not reveal that there was a statistically significant difference of visual scanning efficiency scores between at least two sensory modality conditions during the departure flight segment (F(2, 24) = [0.65], p = 0.5328), the navigation flight segment (F(2, 21) = [0.52], p = 0.6048) and the arrival flight segment (F(2, 21) = [0.89], p = 0.4261) of the instructional flight task.

Ambient-Focal K coefficient

Two-tailed analysis showed (**Figure 7**) that there was no statistically significant difference of ambient-focal K coefficient results between the three experimental conditions during the instructional flight task ($F(2, 29) = [2.06]$, $p = 0.1454$) and the evaluation flight task ($F(2, 26) = [2.92]$, $p = 0.0719$).

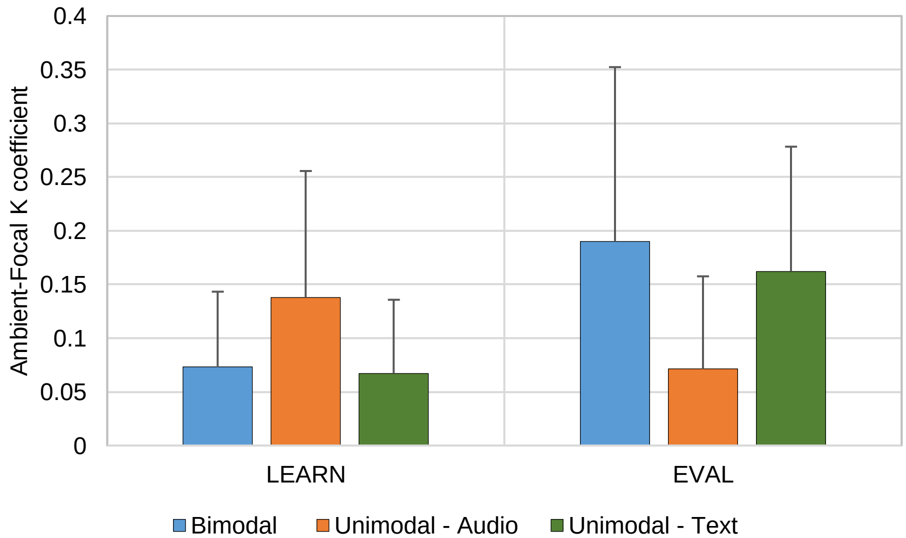


Figure 7. Mean ambient-focal K coefficient per modality group during the instructional flight task (left) and the evaluation flight task (right). *K > 0 indicates a focal visual attention, whereas k < 0 indicates an ambient visual attention. Error bars represent SD.*

The mean ratings show that all results across tasks and conditions were above zero, indicating that fixations were relatively long and followed by short saccade amplitudes, which refers to a focal type of visual attention. During the instructional flight task, results show that the audio unimodal condition participant visual attention was more focal, as shown by a higher positive K coefficient, whereas it was less focal for the bimodal and text unimodal conditions, as shown by lower positive K coefficients. During the evaluation flight task, the mode of visual attention of participants in the bimodal and unimodal text group was rather focal, compared to participants in the audio condition group, whose values of K coefficient were lower. Indeed, the results of visual scanning dispersion show that for the two conditions where more visual cues were displayed on the screen, the scanning strategy shifted from less focal during learning, to more focal during the evaluation flight task. Opposing results were found for the audio condition group for whom fewer visual cues were displayed on-screen. For the latter group, the scanning strategy shifted from more focal during learning, to less focal during an evaluation flight task.

Two-tailed analysis revealed no significant difference of ambient-focal K coefficient results between the three experimental conditions for the departure ($F(2, 26) = [0.66]$, $p = 0.5268$) and arrival ($F(2, 26) = [2.9]$, $p = 0.073$) flight segments (Figure 8). A two-way type III ANOVA revealed a main effect of the flight instruction modality on the arrival flight segment K coefficient results ($F(2, 26) = [3.61]$, $p = 0.0413$). However, the pairwise comparisons did not reveal any statistically significant differences between the bimodal and unimodal audio conditions ($F(1, 19) = [3.41]$, $p = 0.1612$), the bimodal and unimodal text conditions ($F(1, 17) = [1.24]$, $p = 0.281$), and the unimodal audio and unimodal text conditions ($F(1, 16) = [5.76]$, $p = 0.0867$).

The results show that for the unimodal text condition, the visual attention dispersion strategy became more ambient as the flight progressed. For the bimodal and the audio condition groups, the visual attention dispersion strategy became more focal during the navigation part of the flight, and it shifted to an ambient strategy during the arrival segment of the flight.

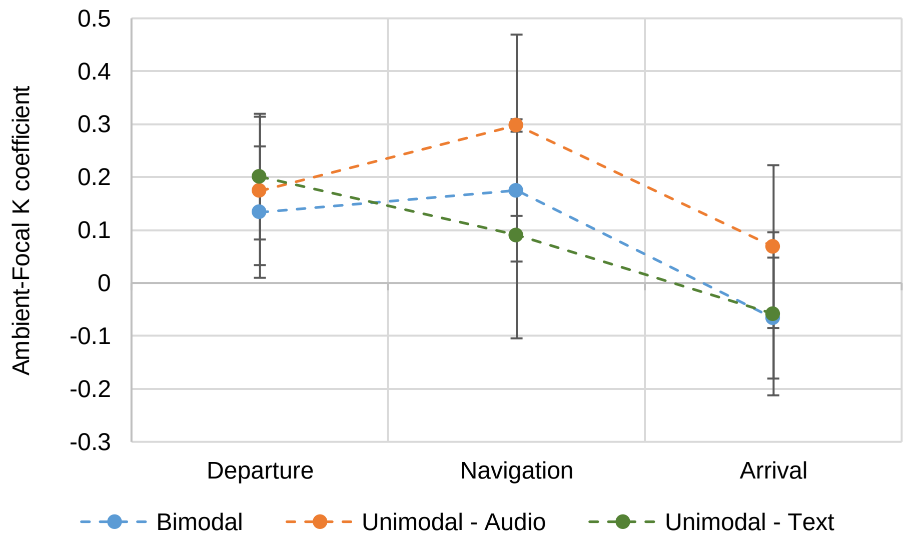


Figure 8. Mean ambient-focal K coefficient for the departure, navigation and arrival flight segments of the instructional flight task for each modality group. *K > 0 indicates a focal visual attention, whereas k < 0 indicates an ambient visual attention. Error bars represent SD.*

Cognitive load

A two-tailed Kruskal-Wallis Test revealed that the NASA-TLX global score, nor the NASA-TLX individual items results (i.e., mental demand, physical demand, temporal demand, overall performance, effort and frustration) significantly differed across sensory modality conditions during the instructional flight task and during an evaluation flight task. The mean ratings shown in the **Figure 9** below indicate that the subjective cognitive load was lower for the bimodal group during learning. During the solo flight evaluation task, the extrinsic cognitive load was equally lower for the bimodal group, and the unimodal text group.

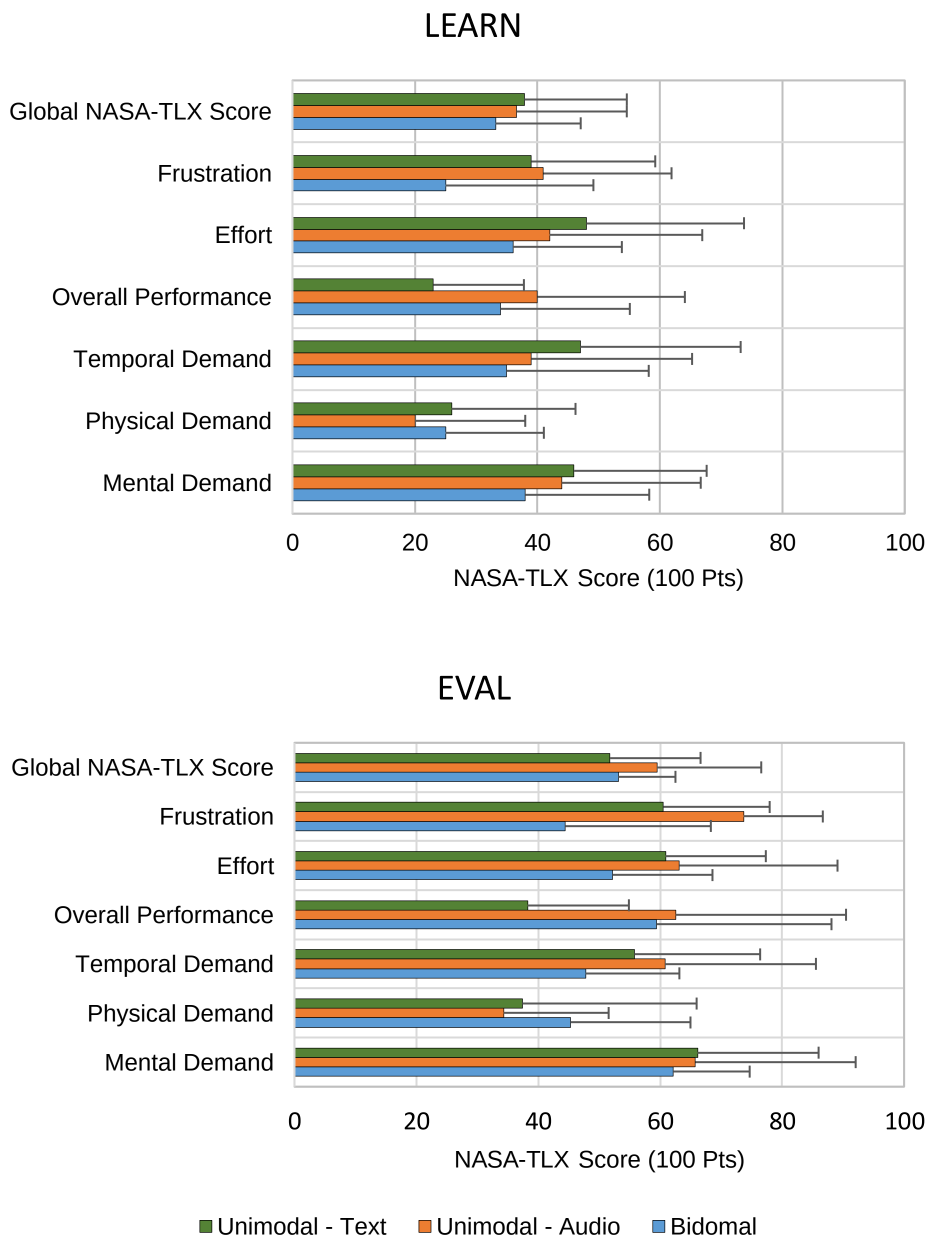


Figure 9. Mean perceived cognitive load (NASA-TLX) per modality group for the instructional flight task (up) and the evaluation flight task (down). *Error bars represent SD.*

We found no statistically significant difference in intrinsic cognitive load between the three modality groups when we assessed the departure ($F(2, 24) = [1.73]$, $p = 0.1991$), navigation ($F(2, 24) = [0.72]$, $p = 0.4981$) and arrival ($F(2, 24) = [1.51]$, $p = 0.2404$) flight segments individually. Similarly, a two-way type III ANOVA did not show any statistically significant difference of extrinsic cognitive load between the three modality conditions for the departure (F(2, 23) = [0.6], p = 0.5582), navigation (F(2, 23) = [0.15], p = 0.858) and arrival (F(2, 23) = [0.21], p = 0.8159) flight segments.

We measured the cognitive load physiologically using the PCPD relative to each participant's baseline state (taken from the last 10 s of each task). The two-way Type III ANOVA results from the PCPD do not indicate that there is a modality specific effect associated with intrinsic cognitive load during the instructional flight task (F(2, 27) = [0.89], p = 0.4208) and the evaluation flight task (F(2, 24) = [1.65], p = 0.2126). The mean ratings shown in the **Figure 10** below indicate that the intrinsic cognitive load was lower for the unimodal text condition during the instructional flight task, and this was reflected during the solo evaluation flight task, as shown by more negative values. Indeed, the pupil was less dilated relative to baseline for the unimodal text condition as compared with the two other modality groups. Negative results (i.e., < 0) indicate that the pupil was less dilated during the experimental task than during the baseline task.

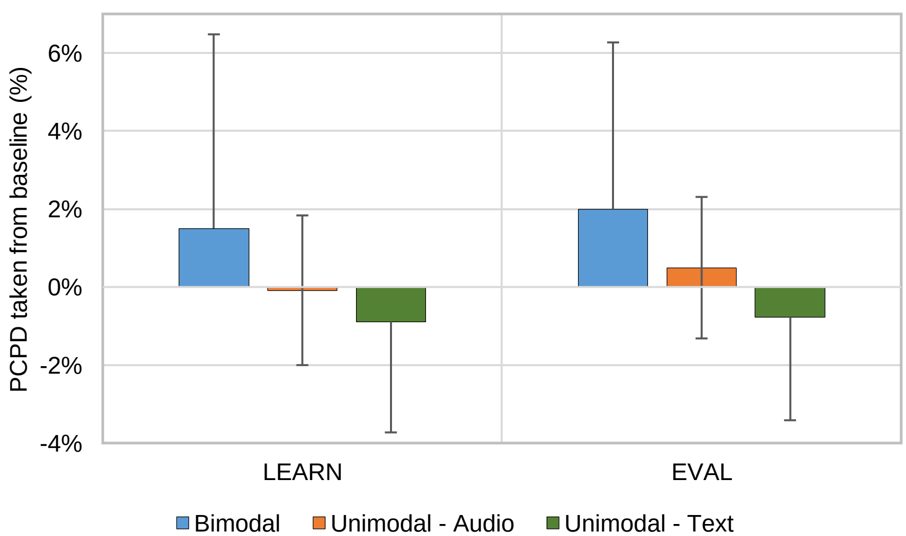


Figure 10. Percentage change in pupil diameter taken from baseline (last ten seconds of each task) for the instructional flight task and evaluation flight task for each modality group. *Error bars represent SD.*

Valence

The emotional valence analysis revealed a significant main effect of the modality on the DV for the instructional flight task (F(2, 25) = [4.89], p = 0.0161). The pairwise comparisons did not reveal statistically significant differences between the bimodal and audio conditions (F(1, 17) = [4.99], p = 0.0784), between the bimodal and the text conditions (F(1, 17) = [6.01], p = 0.0759), nor between the audio and the text conditions (F(1, 16) = [0.15], p = 0.7005) (Figure 11).

A two-way type III ANOVA revealed a main effect of the modality on emotional valence for the evaluation flight task (F(2, 22) = [5.71], p = 0.0101). In this case, the pairwise comparisons indicated that the emotional valence was significantly higher for the text condition when compared with the bimodal condition (F(1, 15) = [7.33], p = 0.0486), but no statistically significant differences were found when the bimodal condition group's mean emotional valence was compared with the unimodal audio condition group (F(1, 15) = [5.04], p = 0.0804) or when the unimodal audio and text conditions' mean emotional valence were compared together (F(1, 14) = [0.66], p = 0.4289).

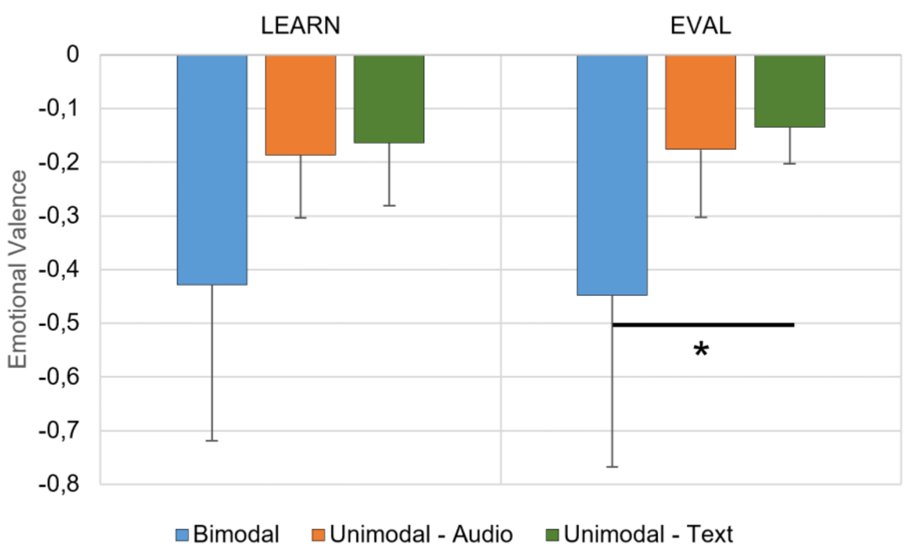


Figure 11. Mean emotional valence per modality group during the instructional flight task (left) and the evaluation flight task (right)*. Error bars represent SD, * indicates main effects p < 0.05, ** indicates main effects p < 0.01 and *** indicates main effects p < 0.001.*

The mean ratings show that the mean valence per condition was more positive for the unimodal text condition during the training task with a virtual instructor, and it was also the case during the solo flight evaluation task.

We investigated the effects of the flight instruction sensory modality on emotional valence results for individual flight segments during the instructional flight task (Figure 12). The two-way analysis performed allowed the detection of a main effect of sensory modality on the mean emotional valence for each of the three flight segments of the instructional flight task: the departure segment ($F(2, 22) = [3.79]$, $p = 0.0386$), the navigation segment ($F(2, 22) = [5.74]$, $p = 0.0099$), and the arrival segment ($F(2, 22) = [6.04]$, $p = 0.0081$). The pairwise comparisons however only revealed a statistically significant difference of mean emotional valence between the bimodal and the unimodal text condition during the navigation flight segment ($F(1, 15) = [7.49]$, $p = 0.0479$) and during the arrival flight segment ($F(1, 15) = [8.65]$, $p = 0.0303$).

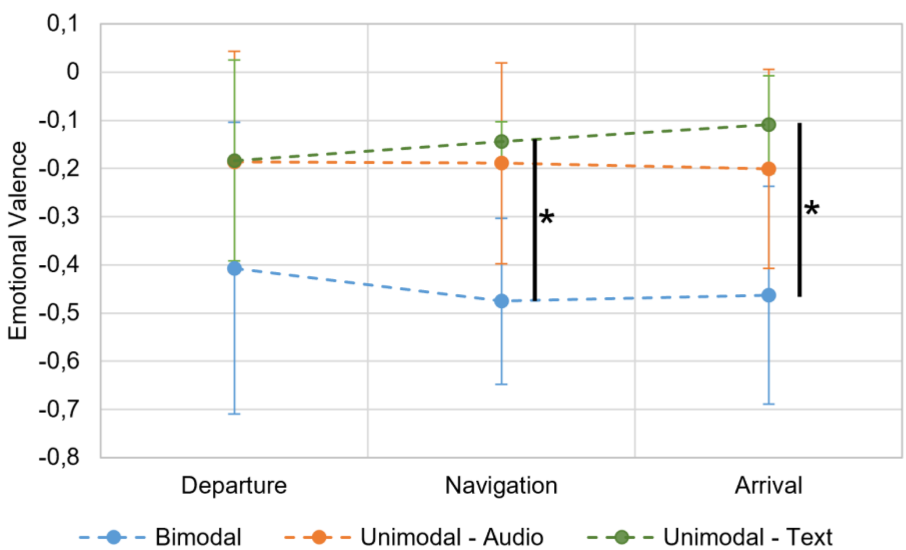
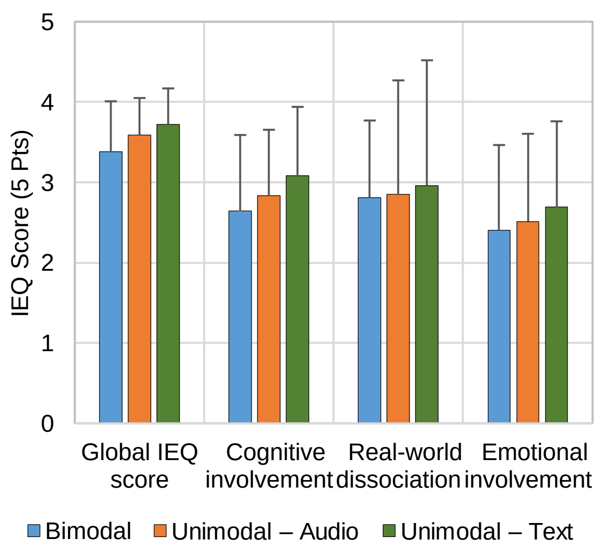


Figure 12. Mean emotional valence for the departure, navigation and arrival flight segments of the instructional flight task for each modality group. *Error bars represent SD, * indicates main effects p < 0.05, ** indicates main effects p < 0.01 and *** indicates main effects p < 0.001.*

Immersion

The analyzes performed did not allow the detection of significant differences between the bimodal, unimodal audio and unimodal text conditions for the results of perceived immersion (F(2, 23) = [0.91], p = 0.4151), including its sub-scale factors cognitive involvement (F(2, 23) = [0.96], p = 0.3981), real-world dissociation (F(2, 23) = [0.24], p = 0.7886) and emotional involvement (F(2, 23) = [0.76], p = 0.4794) (Figure 13). The sub-scale factors of challenge and control were not used, as the Cronbach's alpha could not get higher than 0.1 and 0.6, respectively. The sub-scale factor of real-world dissociation had a Cronbach's alpha of 0.4 with its seven original items; only three items were therefore considered (i.e., *To what extent did you forget about your everyday concerns?*; *To what extent did you feel as though you were separated from your real-world environment?*; *To what extent was your sense of being in the game environment stronger than your sense of being in the real world?*). However not significant, a trend across the scale's sub-factors points in favor of higher perceived immersion for the text unimodal condition.

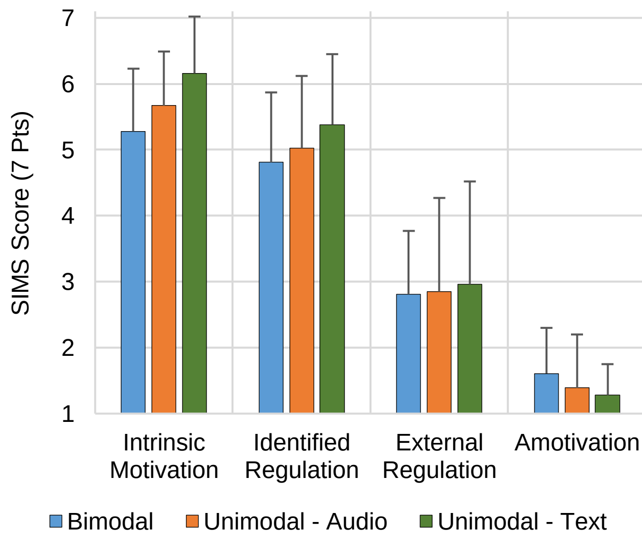


Measure	Cronbach's alpha
Global IEQ score	0.87
Cognitive involvement	0.78
Real-world dissociation	0.74
Emotional involvement	0.72

Figure 13. Mean perceived immersion (IEQ) per modality group (left) and internal consistency results for each scale and sub-scale factor (right). *Error bars represent SD.*

Motivation

The analysis performed did not allow the detection of a main effect of the modality across groups on the perceived motivation of pilots, for the scale's factors of intrinsic motivation ($F(2, 23) = [2.13]$, $p = 0.1418$), identified regulation ($F(2, 23) = [0.6]$, $p = 0.5568$), external regulation ($F(2, 23) = [0.03]$, $p = 0.974$) and amotivation ($F(2, 23) = [1.01]$, $p = 0.381$). The factor of external regulation had a Cronbach's alpha of 0.62 with its four original items; therefore, only three items were considered (i.e., *Because I am supposed to do it.*; *Because it is something that I have to do.*; *Because I don't have any choice.*). A consistent trend points towards higher perceived motivation for the text unimodal condition (Figure 14).



Measure	Cronbach's alpha
Intrinsic motivation	0.88
Identified regulation	0.70
External regulation	0.67
Amotivation	0.89

Figure 14. Mean perceived motivation (SIMS) per modality group (left) and internal consistency results for each scale sub-factor (right). *Error bars represent SD.*

Pilot learning performance

The learning performance mean ratings (and standard deviations) are presented in the **Figure 15** below.

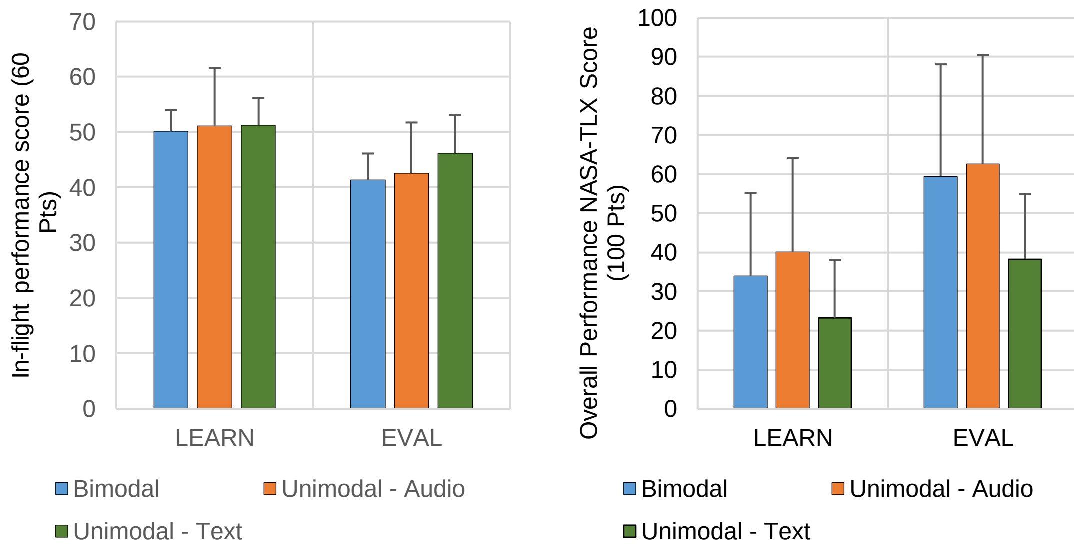


Figure 15. Learning performance mean and standard deviation for the instructional flight task and the evaluation flight task for each modality group. *The observational performance score (left) varies from 0 to 60 (Poor / Good). The extrinsic performance score (right) varies from 0 to 100 (Good / Poor).*

The learning performance assessed by an external pilot evaluator for the instructional flight task (Chi square = 0,512, df = 2, p = 0,774) and the evaluation flight task (Chi square = 1,1455, df = 2, p = 0,564) did not differ significantly between conditions. A main effect of flight instruction modality on pilots' perceived overall performance (retrieved from the NASA-TLX) was not detected for the instructional flight task (F(2, 52) = [2.5], p = 0.0917) and the evaluation flight task (F(2, 23) = [2.27], p = 0.1258).

We wanted to investigate whether the flight instruction modality affected different flight segments differently. Our results did not reveal a significant difference of objective performance between the three modality groups during the "Departure" flight segment (Chi square = 1.1431, df = 2, p = 0.5647), the "Navigation" flight segment (Chi square = 0.2.5315, df = 2, p = 0,282), and the "Arrival" flight segment (Chi square = 0.2499, df =

2, p = 0.8825) of the instructional flight task. Similarly, a Kruskal Wallis Test did not reveal any statistically significant difference between the three conditions for the "Departure" flight segment (Chi square = 2.6266, df = 2, p = 0.2689), the "Navigation" flight segment (Chi square = 3.3632, df = 2, p = 0.1861), and the "Arrival" flight segment (Chi square = 2.9219, df = 2, p = 0.232) of the instructional flight task. The mean ratings (and standard deviation) are shown in the **Figure 16** below.

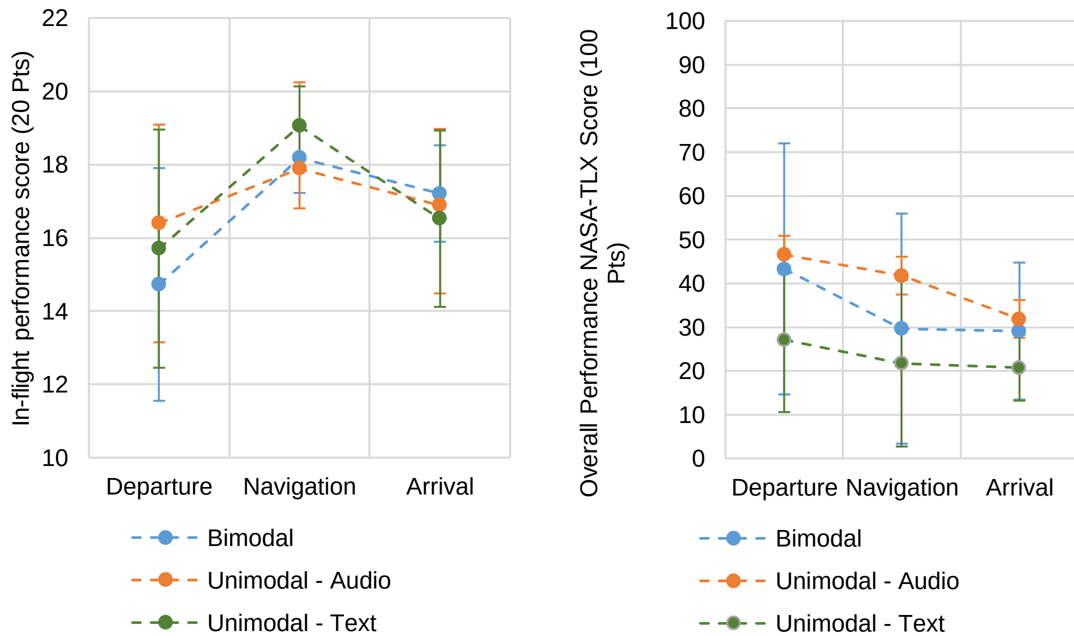


Figure 16. Learning performance mean and standard deviation for the departure, navigation and arrival flight segments of the instructional flight task for each modality group. *The observational performance score varies from 0 to 20 (Poor / Good). The subjective performance score varies from 0 to 100 (Good / Poor).*

Discussion

Previous work has demonstrated that the flight instructor is the greatest determining factor in the success of training of a pilot (Myers III, Starr et Mullins, 2018). The current study was a first effort to evaluate and determine the optimal way to fully integrate flight instruction into the scenario of a FS so that it can be used independently by student pilots in order to enhance their learning performance and cognitive learning state throughout their training. Previous studies aimed at investigating the effects of presenting instructional material in different sensory modalities in an aircraft (John R Helleberg et Wickens, 2003; John Helleberg et Wickens, 2001; Lancaster et Casali, 2008; Latorella, 1998; McGann *et al.*, 1998; Rehmann, 1997), other have used mixed methods to monitor various cognitive states of pilots in an aircraft and/or FS (e.g., (Babu *et al.*, 2019; Blanco *et al.*, 2018; Wanyan, Zhuang et Zhang, 2014) (Bao, Sun et Liu, 2019); Kinney et O'Hare (2020); (Christophe Antony Lounis *et al.*, 2020; Muehlethaler et Knecht, 2016). The purpose of this study was to use a mixed methods research paradigm to assess how different sensory modalities of presentation of instructions (i.e., synthesized instructor speech, textual instructor speech and textual flight objectives) influence the cognitive learning states of pilots in a mainstream FS (i.e., MFS 2020), as described by pilots' cognitive load, attentional strategies, and motivational and affective states during instructional and evaluative simulated flights.

We posited that presenting flight instructions in different sensory modalities would impact the cognitive learning state of the participants during their interaction with the FS. Additionally, this impact would reflect on the pilots' learning performance during an evaluation flight task. The pilots performed two tasks in the FS. A first task constituted the first exposure of the subjects to the VE. During this task, flight instructions were provided to participants in various modalities to teach them how to fly in VFR. During the integration of new theory during training, increased cognitive, attentional, motivational or affective demands do not necessarily mean suboptimal learning, although it can. Hence, we predicted that the optimal way of presenting flight instructions would result in enhanced cognitive learning states during a second evaluation flight task. A general tendency of the results indicates that the use of the single visual channel (i.e., textual material) allowed the improvement of the cognitive learning states of pilots, in terms of visual attention, cognitive load, motivational and affective processes. Our results did not reveal a significant difference of learning performance across experimental groups. We discuss the implication of these results below.

Pilot cognitive learning states

Our first objective was to assess pilots' cognitive learning state according to the cognitive load, attentional processes, motivational and affective processes involved during an instructional flight task and during an evaluation flight task. Our results revealed a significant effect of the sensory modality on visual transition entropy results during the evaluative flight task. In this case, the visual transition entropy of subjects was significantly lower for subjects of the unimodal text condition and the bimodal condition, when compared to subjects from the unimodal audio condition. The analyzes performed additionally allowed the identification of a significant difference between the emotional valence responses of subjects in the bimodal group and the text unimodal group, where the mean emotional valence was more positive for the text unimodal group during the evaluative flight task. Although not significant, the variation in the results across modality groups during the evaluation flight task suggests in a quite compelling way that the unimodal text condition was more appropriate, in terms of objective cognitive load, perceived immersion and perceived motivation, to enhance pilots' cognitive states during training. This strengthens the tendency according to which choosing the textual modality to provide flight instructions to pilots is more suitable for learning.

Sensory modality effect on pilot visual attention

The analyses results of the two attentional metrics, visual transition entropy (GTE/H_{max}) and visual attention dispersion (ambient-focal K coefficients), are perhaps the most interesting of this study. We had expected a decrease in visual transition entropy (i.e., lower GTE/H_{max}) and a rather "ambient" dispersion of subjects' visual attention (i.e., lower K coefficient) for the experimental condition that would yield better training outcomes during the evaluation flight task. Our results indicate that transition entropy results from the bimodal and unimodal text conditions were significantly lower to those from the unimodal audio condition. However, no significant differences of visual attention dispersion were found across groups. Therefore, our hypothesis 1A is only partially supported, as only the visual transition entropy of both the bimodal and unimodal text groups pointed to enhanced attentional states during the evaluation task. Results indicate that ocular behaviors from the bimodal condition and unimodal text condition were comparable in terms of the two visual scan metrics results, meanwhile the subjects from the unimodal audio condition behave differently.

During training with instructional material, the audio condition participants had the lowest visual transition entropy; they had the highest scanning entropy during the evaluation flight task without instructional material. The GTE/H_{max} ratio represents both continuums of "simplicity – complexity" of visual patterns and/or the "organization – disorder" in the visual patterns (Krejtz *et al.*, 2016). Our results suggest that during the instructional flight task, the textual instructional material redirected the participants' gaze (i.e., high salience of visual cues). In this case, it is plausible that higher GTE/H_{max} signifies an increase in the complexity of the participants' visual gaze patterns, rather than an increase in disorder. Indeed, in addition to having to focus their attention on the instructional content presented on-screen, the participants had to keep on monitoring the important signals in the VE to operate the aircraft adequately. For the unimodal audio condition for which there was less information in the visuospatial sketch pad (i.e., low salience of visual cues), participants were able to focus only on the operation of the aircraft while listening to the auditory instructional material. Interestingly, during the return flight where all participants had no access to instructional material to guide their ocular behaviors, the two groups whose GTE/H_{max} levels were the lowest were the two modality groups with visual instructions in the precedent task. These results suggest that the bimodal and unimodal text groups subjects knew better where to look in the VE to complete the task for similar performance results.

During the instructional flight task, the results pertaining to visual attention dispersion indicated that the subjects' mode of attention in the two experimental groups with textual instructional material was more ambient than for the subjects from the unimodal auditory group. It is arguable that the salience of the textual instructional material redirected participants' visual attention to various interface elements that were spatially spread out on-screen, thus increasing the length of saccades and decreasing the duration of fixations. The low salience for these conditions during the evaluation flight task would have caused this increase in the K coefficient (i.e., a more focal visual mode of attention). For the auditory unimodal condition, during the instructional flight task, the absence of textual instructional material left participants looking where they wished to operate the aircraft. Participants were able to gaze focally OTW or at panel instruments, for instance. During the evaluation flight task, participants in the auditory unimodal condition had more ambient dispersion of their attention, at a level roughly equal to that of participants in the other two conditions during the instructional flight task. Considering that auditory condition subjects' level of GTE was higher during this task, suggesting more disorder, when compared to the other two conditions, it is plausible that these participants adopted a more exploratory and therefore ambient dispersion of their visual attention. We then discuss how flight instructions sensory modalities guided participant motivational amd affective states both during and after an evaluative flight task.

The use of complex eye-tracking measures allowed us to better characterize the visual strategies of learners. We highlight the importance of contextualizing complex visual scan metric results to draw valid conclusions. For instance, GTE increases with task demand and visual scene complexity (Krejtz *et al.*, 2015; Schieber et Gilland, 2008; Shiferaw, Downey et Crewther, 2019). Authors have used this measure as an indicator of mental workload (Ephrath *et al.*, 1980; Harris Sr, Glover et Spady Jr, 1986; Tole *et al.*, 1983), whereas a report later suggested using the metric as an indicator of visual scanning efficiency (Shiferaw, Downey et Crewther, 2019). By manipulating the amount of instructional material presented in the visuospatial sketchpad during an instructional flight task, we expected an increase of GTE for the visual conditions that did not indicate suboptimal learning. During the assessment task, however, task demand as well as visual scene complexity were controlled across groups, so it was possible to interpret that a decrease in GTE for the bimodal and unimodal text conditions potentially referred to less disorder, or enhanced visual scanning efficiency, therefore suggesting better training outcomes. Moreover, a study by C. Lounis, Peysakhovich et Causse (2021) described a higher value of GTE as desirable since expert pilots demonstrated greater complexity in their visual scans than novice pilots. In short, ocular measurements are highly complex and vary in accordance with several factors, thus, all task and environment variables must be considered.

Sensory modality effect on pilot subjective and objective cognitive load

We had predicted that the flight instruction sensory modality that would reduce extraneous processing the most would generate lower levels of subjective and objective cognitive load during an evaluation flight task. The objective and subjective cognitive loads of the three experimental groups were not significantly different. We therefore infirm our hypothesis 1B that different flight instruction sensory modalities would yield significantly different cognitive load results across groups.

The mean objective cognitive load was lower both during the instructional flight task and during the evaluation flight task for the unimodal textual condition. Conversely, subjective cognitive load global scores were higher for the unimodal audio and unimodal text conditions, and lower for the bimodal condition. During the instructional flight task, participants in the unimodal text condition gave the highest ratings for each item of the questionnaire except for the perceived performance item (i.e., mental demand, physical demand, temporal demand, frustration and effort). This is in line with previous work that showed that the display of strictly textual navigational instructions increases pilot cognitive load ratings (Lancaster and Casali, 2008). Thus, in the text group, pilots perceived a high cognitive load, whereas their objective cognitive load was in fact the lowest. We then discuss how flight instructions sensory modalities guided participant visual attention strategies.

Sensory modality effect on pilot motivation and affect

As per the CATLM, we posited that the flight instruction sensory modality that would result in significantly increased emotional valence, as well as increased ratings of motivation and immersion, would have enhanced mediative effects on learning. We found that participants from the unimodal text group had significantly more positive emotional valence than the bimodal group participants. We however found no statistically significant differences across groups for the motivation and immersion self-reported questionnaire results. Hence, our hypothesis 1C is only partially supported; our results suggest that the unimodal text condition could enhance cognitive engagement through increased positive emotional valence.

The analysis performed for the emotional valence revealed a significant difference between the bimodal condition and the unimodal text condition, where the experience of the text modality group was more positive overall. When using a technology, a user's emotional experience directly affects their acceptance and satisfaction of the product (Marangunić et Granić, 2015). A user’s affective experience is also a good predictor of subsequent usage behaviors (Wu et Holsapple, 2014). This is a critical aspect that designers must consider, specifically for products that will allow users to improve their skills in proportion to the time spent training with the system. This is the case of pilot training, an area in which hours spent flying are the main indicator of experience/skills levels.

Consequently, users' hedonic and utilitarian perceptions of a system will influence their motivations to use the system (Shih et Liu, 2007). When we examined the results of perceived motivation across groups, we did not find significant differences, but results show that the perceived motivation of pilots was superior in the unimodal text group, which could be a consequence of the emotional valence results. Indeed, considering these results together suggests that the positive affective experience of users in the unimodal text group influenced their ratings of motivation. Again, designers must consider that increased motivation is crucial in a context of self-administered training. Thus, these results seem to indicate that perhaps the unimodal text condition is better suited for delivering flight instructions to student pilots.

The perceived immersion results, although not significantly different between groups, indicated that text unimodal instructional material favored immersion. Again, these results are surprising as research on high/low immersion VR has shown that increasing sensory input through various modalities (e.g., audio, visual, haptic, etc.) increases user immersion, which can in turn foster task engagement and procedural practice (Ke et Carafano, 2016). Furthermore, the presentation of a virtual instructor's speech in a textual and non-vocal way is not representative of reality, which we anticipated would be a threat to user immersion. This is an interesting result as it points to the pertinence of evaluating immersion through experience and not only through the characterization of the properties of a VE. In this experiment, it is possible that the increased efficiency afforded by the text-based flight objectives display overcame the advantages afforded by the inclusion of flight instructions in the same modality as in a real-life setting, resulting in no reduction of perceived immersion for the unimodal text condition participants.

Pilot learning performance

Our second objective was to assess the flight performance by an external evaluator and the perceived flight performance (NASA-TLX) of participants during an instructional flight task and an evaluation flight task. We found no statistically significant difference between the three experimental groups for both tasks, thus we reject our first hypothesis that different flight instruction sensory modality would generate different levels of learning performance. While the observed flight performance results were very similar between groups, the mean perceived performance of subjects in the unimodal text group was superior. We assume that this difference in perceived performance is at least partially due to the permanence on-screen of the flight objectives display, which gave real-time feedback to the textual group subjects about their current state of performance (Rehmann, 1997). Our results are supported by previous work that demonstrated that using performance-only measures was insufficient to assess the multidimensional cognitive learning states of pilots (Gisler *et al.*, 2021).

Sensory modality theories

The CTML states that multimodal instructional material reduces cognitive load, thus facilitating deep learning (Richard Mayer et Mayer, 2005). Our results suggested that the textual instructional material improved learners' cognitive learning states during the evaluation flight task. To explain our results, we must first revisit the validity criteria of the CTML. The modality principle of CTML states that dividing instructional material into two modalities expands the working memory of learners, if the information presented in the auditory and visual conditions is not redundant and if it does not cause split-attention (Ginns, 2005). Redundancy and split-attention are two closely related concepts often governed by the same design principles. That is to say, if concepts are not well understood in isolation, a split-attention is likely to occur and conversely, if they are understood in isolation, a redundancy effect is likely to take place (Tzu-Chien Liu *et al.*, 2012).

The redundancy principle states that information in a text form presented simultaneously with identical verbal information becomes redundant, increasing the load on working memory and interfering with learning (Sweller, 2005). In this study, the characteristics of the information transmitted in each modality were non-identical. Synthesized and written instructor speech provided contextualized instructions and text-based flight objectives provided uncontextualized instructions; therefore, the non-redundancy condition was respected.

The split-attention effect can be defined as the need for learners to divide their attention between several sources of information that must be integrated before they can be understood (ARSLAN, 2012). In a context of training in a complex FS, information coming from various sources are often complementary. Take the example of a virtual instructor explaining a motor task to a student pilot while he controls an aircraft. The pilot must plan and perform a motor task, while receiving instant feedback from multiple sources in the aircraft and potentially their instructor. Some level of split-attention may inevitably occur. Nevertheless, the detrimental effect of split-attention lessens as low-level learners gain expertise (Kalyuga, 2009). In the present study, the subjects were mainly novices and intermediates. It is very plausible that a split-attention effect induced by the sources of information of distinct modalities had a deleterious effect on the cognitive load, attentional strategies, motivational and affective states of the learners of the bimodal sensory group, canceling out the benefits of a bimodality predicted by the CTML. Considering that split-attention may occur when training in a FS, our results underline the importance of designing instructional material suited to the level of expertise of the learner.

Furthermore, according to Richard Mayer et Mayer (2005), the modality principle is expected to apply when the instructional material is complicated rather than simple, when the presentation is system-paced rather than self-paced, when graphs are dynamic rather than static, when learners have a low level of knowledge rather than a high level, when segments are short rather than long, and when the terms used are familiar rather than unfamiliar. We may first mention that this long list of criteria makes it difficult to apply the principle of modality to evaluate systems with high ecological validity and in a context which is not that of a highly controlled laboratory. Nevertheless, the system used and the tasks selected for this study met these criteria.

Explanations that may have contributed to our results pointing to the unimodal text condition as enhancing the cognitive learning state of pilots include the permanence of written instructions, an auditory pre-emption effect, and the fact that spatial tasks are more compatible with visual inputs than auditory inputs.

First, the fact that written instructions support referencing capability could have improved the learning experience of learners (Rehmann, 1997). For this purpose, several participants have anecdotally reported that they appreciated being able to refer in real-time to the flight objectives display to situate themselves in relation to the flight objectives to achieve during the tasks.

Second, our results are compatible with previous work that described an auditory pre-emption effect as having a detrimental consequence on the cognitive learning states of subjects (John R Helleberg et Wickens, 2003; Lancaster et Casali, 2008). In this study, an auditory pre-emption effect could have happened for the two experimental groups where audio instructional material was presented, the bimodal condition and the unimodal audio condition. During the instructional flight task, the visual attention of the subjects in the audio group was substantially more focal than for the subjects in the other groups. It is plausible that for this group, a pre-emption effect took place, to the detriment of the ongoing task of traffic monitoring and flight-path tracking. If a user relocated their attention to a synthesized instructor speech and stopped monitoring signals on the instrument panel, one could expect focal visual attention, as observed in our results.

Third, the flight instructions provided to subjects being mainly spatial/navigational in nature, a plausible explanation of our results is the rationale standing behind the SCR compatibility model. This model describes a compatibility between visual instructions of a spatial nature, in particular when these instructions are responded to motorically (Fitts et Deininger, 1954; Fitts et Seeger, 1953; Wickens *et al.*, 2021). For voice-distributed flight instructions, a translation of the vocal command into a spatial representation must be performed, increasing the information processing load. Operating an aircraft under VFR is perhaps the flight scenario in which the visual and navigational representational requirements are highest, as pilots must continually monitor the airfield, translate and integrate terrestrial knowledge with adapted motor responses. The facilitation of processing enabled by the visuospatial network therefore could have had a noticeable effect on our results. It would be relevant to measure the magnitude of this effect during pilot training in a simulator under other types of regulations where flight instructions are treated differently (e.g. under Instrument Flight Rules). Designers should also consider that this compatibility effect is exacerbated when visual representations are pictorial rather than textual (Healy *et al.*, 2013; Nelson, Reed et Walling, 1976). During the "departure" segment of the instructional flight task, the virtual instructor provided the users with a summary of the flight itinerary. Several participants spontaneously mentioned that they would have appreciated being able to observe a graphical representation of the flight route before take-off.

Limitations

There are limitations to this study. First, we accepted eye-tracker calibrations when their accuracy was greater than 90%. The 10% error margin could have impacted the eye tracking results, especially for adjacent AOIs. Secondly, it is important to note that the eye-tracker measures "overt" attention and not "covert" attention. That is, if a subject's attention is redirected to an object in the peripheral field without the eyes moving, then eye tracking measurements cannot measure this redirection of attention. Third, the experimental setup included a 2D computer monitor on which the simulation was displayed. The participants had to use a joystick on the yoke to be able to perceive the VE panoramically. The use of this joystick was not instinctive for several participants who mentioned in interview that it was difficult to perceive landmarks OTW using this function. Using a VR headset rather than a 2D computer screen would be more instinctive and would potentially increase immersion. We chose to use a computer monitor since we envisioned this to be the most popular experimental setup among MFS 2020 end-users. Fourth, we used an MFS 2020 built-in task scenario, which involved participant navigation between two airports where they had never flown before. In a real-life setting, pilots train at the same airports and a novelty effect is therefore small. Additionally, when flying in VFR, pilots usually prepare their flight plan themselves, which implies that they already have a general idea of the important landmarks and directions to their flight route. During the experimental tasks, the flight plan was displayed to users on the "navigation log" in a way that could have been less familiar or instinctive, thus complexifying the task and negatively impacting the results.

Conclusion

This study used a multidimensional approach to demonstrate how the incorporation of flight instruction into a FS scenario impacts pilots' cognitive learning states as described by their cognitive load, attentional strategies, and motivational and affective states, as well as pilots' in-flight performance during an instructional flight task and during an evaluation flight task. This research underlines the importance of measuring the learning experience of users using various tools and types of measurements. Indeed, through the monitoring of different cognitive states, we were able to deepen our understanding of how different flight instruction sensory modalities mediate learning. However, performance-only measures did not allow the detection of a significant difference between the three experimental conditions.

Our results showed that the visual transition entropy of the bimodal and unimodal text groups was significantly inferior than the audio unimodal group during the evaluation flight task. In addition, results indicated that the affective experience of participants in the text flight instruction group was significantly more positive than that of participants in the bimodal group during the evaluation flight task. Although no significant difference was detected between the three experimental groups, the mean ratings of various measures seemed to point towards the unimodal text condition as optimal for learning (lower implicit cognitive load, higher perceived immersion, higher perceived motivation). Our results show that in a VE with high-element interactivity, bimodal flight instructions can exacerbate the deleterious effects of split-attention on learning, rather than decreasing the cognitive load of subjects as suggested by the CTML. In a VFR flight training context, the high visual demand of operating the aircraft, in addition to providing users with exclusively auditory flight instructions, confers a bimodality to the instructional task, therefore, the same deleterious effect of split-attention was observed for the bimodal condition as well as the unimodal audio condition.

The evaluation of pilots' learning state during the instructional flight task revealed that the measurement levels could vary differently depending on the flight segment. Indeed, over the total duration of a training flight, the tasks/maneuvers/procedures to be performed by an aircraft operator vary. Hence, future work should aim at better specifying how different flight instruction modalities impact the learning experience of pilots according to the low-level tasks in progress. Our results suggest that the positive effects of the modality principle are not observed when pilots perform multiple tasks concurrently. Nevertheless, it could perhaps be appropriate to provide flight instructions in more than one sensory modality when the pilot is on the ground (e.g., departure) and all his attention is redirected towards a synthesized instructor speech. Future work should also assess how different task scenarios influence the learning state of pilots (e.g., sets of regulations, weather conditions, airports, aircraft types, etc.). Moreover, a longitudinal study may focus on determining whether each sensory modality of flight instruction allows the creation of positive training transfer. That is, flight instructions displayed to a user in a simulated VE should facilitate the transition from simulator to reality.

Acknowledgments. We would like to thank the pilots who participated in the experiment. The authors thank the Tech3Lab technical assistants for their valuable help during data collection, CAE Inc. for its precious collaboration and funding as well as the NSERC-PROMPT Industrial Research Chair in UX.

References

1. Agrawal, Sarvesh, Adèle Simon, Søren Bech, Klaus Bæntsen et Søren Forchhammer (2020). « Defining immersion: Literature review and implications for research on audiovisual experiences », *Journal of the Audio Engineering Society*, vol. 68, no 6, p. 404-417.
2. Alexander, Amy L, Tad Brunyé, Jason Sidman et Shawn A Weil (2005). « From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games », *DARWARS Training Impact Group*, vol. 5, p. 1-14.
3. Allsop, Jonathan et Rob Gray (2014). « Flying under pressure: Effects of anxiety on attention and gaze behavior in aviation », *Journal of Applied Research in Memory and Cognition*, vol. 3, no 2, p. 63-71.

4. ARSLAN, Pelin YUKSEL (2012). « A review of multimedia learning principles: Split-attention, modality, and redundancy effects », *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, vol. 8, no 1, p. 114-122.
5. Attard-Johnson, J., O. Ciardha C et M. Bindemann (2019). « Comparing methods for the analysis of pupillary response », *Behav Res Methods*, vol. 51, no 1, p. 83-95.
6. Attard-Johnson, Janice, Caoilte Ó Ciardha et Markus Bindemann (2019). « Comparing methods for the analysis of pupillary response », *Behavior Research Methods*, vol. 51, no 1, p. 83-95.
7. Babu, M. D., D. V. JeevithaShree, G. Prabhakar, K. P. S. Saluja, A. Pashilkar et P. Biswas (2019). « Estimating pilots' cognitive load from ocular parameters through simulation and in-flight studies », *J Eye Mov Res*, vol. 12, no 3.
8. Baddeley, Alan (1992). « Working memory », *Science*, vol. 255, no 5044, p. 556-559.
9. Bailenson, Jeremy N, Nick Yee, Jim Blascovich, Andrew C Beall, Nicole Lundblad et Michael Jin (2008). « The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context », *The Journal of the Learning Sciences*, vol. 17, no 1, p. 102-141.
10. Bao, Jun-Ping, You-Chao Sun et Xin-Chao Liu (2019). « Research on pilot situation awareness in noise environment », communication présentée au *2019 4th International Conference on Electromechanical Control Technology and Transportation (ICECTT)*,
11. Beatty, Jackson (1982). « Task-evoked pupillary responses, processing load, and the structure of processing resources », *Psychological bulletin*, vol. 91, no 2, p. 276.
12. Binda, Paola et Paul D Gamlin (2017). « Renewed attention on the pupil light reflex », *Trends in Neurosciences*, vol. 40, no 8, p. 455-457.
13. Blanco, Justin A., Michael K. Johnson, Kyle J. Jaquess, Hyuk Oh, Li-Chuan Lo, Rodolphe J. Gentili, *et al.* (2018). « Quantifying cognitive workload in simulated flight using passive, dry eeg measurements », *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no 2, p. 373-383.
14. Brooks, Lee R (1967). « The suppression of visualization by reading », *The Quarterly journal of experimental psychology*, vol. 19, no 4, p. 289-299.
15. Butcher, Kirsten R (2014). « The multimedia principle ».

16. Charlton, Samuel G (2002). « Measurement of cognitive states in test and evaluation », *Handbook of human factors testing and evaluation*, p. 115-122.
17. Dalgarno, Barney et Mark JW Lee (2010). « What are the learning affordances of 3-d virtual environments? », *British Journal of Educational Technology*, vol. 41, no 1, p. 10-32.
18. Dede, Chris (2009). « Immersive interfaces for engagement and learning », *science*, vol. 323, no 5910, p. 66-69.
19. Diaz-Piedra, Carolina, Hector Rieiro, Alberto Cherino, Luis J Fuentes, Andres Catena et Leandro L Di Stasi (2019). « The effects of flight complexity on gaze entropy: An experimental study with fighter pilots », *Applied ergonomics*, vol. 77, p. 92-99.
20. Dirican, Ahmet Cengizhan et Mehmet Göktürk (2011). « Psychophysiological measures of human cognitive states applied in human computer interaction », *Procedia Computer Science*, vol. 3, p. 1361-1367.
21. Ekman, Paul et Wallace V Friesen (1978). « Facial action coding system », *Environmental Psychology & Nonverbal Behavior*.
22. Fitts, Paul M et Richard L Deininger (1954). « Sr compatibility: Correspondence among paired elements within stimulus and response codes », *Journal of experimental psychology*, vol. 48, no 6, p. 483.
23. Fitts, Paul M et Charles M Seeger (1953). « Sr compatibility: Spatial characteristics of stimulus and response codes », *Journal of experimental psychology*, vol. 46, no 3, p. 199.
24. Ginns, Paul (2005). « Meta-analysis of the modality effect », *Learning and instruction*, vol. 15, no 4, p. 313-331.
25. Glaholt, Mackenzie G (2014). « Eye tracking in the cockpit: A review of the relationships between eye movements and the aviators cognitive state ».
26. Gottfried, Adele E (1990). « Academic intrinsic motivation in young elementary school children », *Journal of Educational psychology*, vol. 82, no 3, p. 525.
27. Guay, Frédéric, Robert J Vallerand et Céline Blanchard (2000). « On the assessment of situational intrinsic and extrinsic motivation: The situational motivation scale (sims) », *Motivation and emotion*, vol. 24, no 3, p. 175-213.

28. Hamel, Robert F (1974). « Female subjective and pupillary reaction to nude male and female figures », *The Journal of psychology*, vol. 87, no 2, p. 171-175.
29. Harris Sr, Randall L, Bobby J Glover et Amos A Spady Jr (1986). *Analytical techniques of pilot scanning behavior and their application*.
30. Hart, Sandra G et Lowell E Staveland (1988). « Development of nasa-tlx (task load index): Results of empirical and theoretical research », dans *Advances in psychology*, vol 52, Elsevier, p. 139-183.
31. Healy, Alice F, Vivian I Schneider, Blu McCormick, Deanna M Fierman, Carolyn J Buck-Gengler et Immanuel Barshi (2013). « Which modality is best for presenting navigation instructions? », *Journal of Applied Research in Memory and Cognition*, vol. 2, no 3, p. 192.
32. Heitz, Richard P et Randall W Engle (2007). « Focusing the spotlight: Individual differences in visual attention control », *Journal of Experimental Psychology: General*, vol. 136, no 2, p. 217.
33. Helleberg, John R et Christopher D Wickens (2003). « Effects of data-link modality and display redundancy on pilot performance: An attentional perspective », *The International Journal of Aviation Psychology*, vol. 13, no 3, p. 189-210.
34. Jennett, Charlene, Anna L Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, *et al.* (2008). « Measuring and defining the experience of immersion in games », *International journal of human-computer studies*, vol. 66, no 9, p. 641-661.
35. Jeung, Hyun-Ju, Paul Chandler et John Sweller (1997). « The role of visual indicators in dual sensory mode instruction », *Educational Psychology*, vol. 17, no 3, p. 329-345.
36. Kalyuga, Slava (2008). *Managing cognitive load in adaptive multimedia learning*, IGI Global.
37. Kalyuga, Slava (2009). « The expertise reversal effect », dans *Managing cognitive load in adaptive multimedia learning*, IGI Global, p. 58-80.
38. Kalyuga, Slava, Paul Chandler et John Sweller (1999). « Managing split-attention and redundancy in multimedia instruction », *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, vol. 13, no 4, p. 351-371.

39. Ke, Fengfeng et Peter Carafano (2016). « Collaborative science learning in an immersive flight simulation », *Computers & Education*, vol. 103, p. 114-123.
40. Kinney, L. et D. O'Hare (2020). « Responding to an unexpected in-flight event: Physiological arousal, information processing, and performance », *Hum Factors*, vol. 62, no 5, p. 737-750.
41. Komogortsev, Oleg V, Denise V Gobert, Sampath Jayarathna et Sandeep M Gowda (2010). « Standardization of automated analyses of oculomotor fixation and saccadic behaviors », *IEEE Transactions on biomedical engineering*, vol. 57, no 11, p. 2635-2645.
42. Krejtz, Krzysztof, Andrew Duchowski, Izabela Krejtz, Agnieszka Szarkowska et Agata Kopacz (2016). « Discerning ambient/focal attention with coefficient k », *ACM Transactions on Applied Perception (TAP)*, vol. 13, no 3, p. 1-20.
43. Krejtz, Krzysztof, Andrew Duchowski, Tomasz Szmidt, Izabela Krejtz, Fernando González Perilli, Ana Pires, *et al.* (2015). « Gaze transition entropy », *ACM Transactions on Applied Perception (TAP)*, vol. 13, no 1, p. 1-20.
44. Laeng, Bruno et Dag Alnaes (2019). « Pupillometry », dans *Eye movement research*, Springer, p. 449-502.
45. Lancaster, Jeff A et John G Casali (2008). « Investigating pilot performance using mixed-modality simulated data link », *Human factors*, vol. 50, no 2, p. 183-193.
46. Léger, Pierre-Majorique, Francois Courtemanche, Marc Fredette et Sylvain Sénécal (2019). « A cloud-based lab management and analytics software for triangulated human-centered research », dans *Information systems and neuroscience*, Springer, p. 93-99.
47. Lintern, Gavan (1991). « An informational perspective on skill transfer in human-machine systems », *Human Factors*, vol. 33, no 3, p. 251-266.
48. Liu, Dejian, Chris Dede, Ronghuai Huang et John Richards (2017). *Virtual, augmented, and mixed realities in education*, Springer.
49. Liu, Tzu-Chien, Yi-Chun Lin, Meng-Jung Tsai et Fred Paas (2012). « Split-attention and redundancy effects on mobile learning in physical environments », *Computers & Education*, vol. 58, no 1, p. 172-180.

50. Loijens, Leanne et Olga Krips (2018). « Facereader methodology note », *A white paper by Noldus Information Technology*.

51. Lounis, C., V. Peysakhovich et M. Causse (2021). « Visual scanning strategies in the cockpit are modulated by pilots' expertise: A flight simulator study », *PLoS One*, vol. 16, no 2, p. e0247061.

52. Lounis, Christophe Antony, Almoctar Hassoumi, Olivier Lefrancois, Vsevolod Peysakhovich et Mickael Causse (2020). « Detecting ambient/focal visual attention in professional airline pilots with a modified coefficient k: A full flight simulator study », communication présentée au *Symposium on Eye Tracking Research and Applications*,

53. Makransky, Guido, Thomas S Terkildsen et Richard E Mayer (2019). « Adding immersive virtual reality to a science lab simulation causes more presence but less learning », *Learning and instruction*, vol. 60, p. 225-236.

54. Marangunić, Nikola et Andrina Granić (2015). « Technology acceptance model: A literature review from 1986 to 2013 », *Universal access in the information society*, vol. 14, no 1, p. 81-95.

55. Mayer, Richard E (2003). « The promise of multimedia learning: Using the same instructional design methods across different media », *Learning and instruction*, vol. 13, no 2, p. 125-139.

56. Mayer, Richard E (2005). « Cognitive theory of multimedia learning », *The Cambridge handbook of multimedia learning*, vol. 41, p. 31-48.

57. Mayer, Richard E (2014). « Incorporating motivation into multimedia learning », *Learning and instruction*, vol. 29, p. 171-173.

58. Mayer, Richard E et Roxana Moreno (1998). « A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory », *Journal of educational psychology*, vol. 90, no 2, p. 312.

59. Mayer, Richard E et Celeste Pilegard (2005). « Principles for managing essential processing in multimedia learning: Segmenting, pretraining, and modality principles », *The Cambridge handbook of multimedia learning*, p. 169-182.

60. Mayer, Richard et Richard E Mayer (2005). *The cambridge handbook of multimedia learning*, Cambridge university press.

61. McGann, Alison, Dan Morrow, Michelle Rodvold et Margaret-Anne Mackintosh (1998). « Mixed-media communication on the flight deck: A comparison of voice, data link, and mixed atc environments », *The International Journal of Aviation Psychology*, vol. 8, no 2, p. 137-156.
62. McGuinness, CAROL (1990). « Talking about thinking: The role of metacognition in teaching thinking », *Lines of thinking*, vol. 2, p. 310-312.
63. Moreno, Roxana (2005). « Instructional technology: Promise and pitfalls », *Technology-based education: Bringing researchers and practitioners together*, p. 1-19.
64. Moreno, Roxana (2006). « Does the modality principle hold for different media? A test of the method-affects-learning hypothesis », *Journal of Computer Assisted Learning*, vol. 22, no 3, p. 149-158.
65. Moreno, Roxana (2007). « Optimising learning from animations by minimising cognitive load: Cognitive and affective consequences of signalling and segmentation methods », *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, vol. 21, no 6, p. 765-781.
66. Moreno, Roxana (2009). « Learning from animated classroom exemplars: The case for guiding student teachers' observations with metacognitive prompts », *Educational Research and Evaluation*, vol. 15, no 5, p. 487-501.
67. Moreno, Roxana et Richard E Mayer (1999). « Cognitive principles of multimedia learning: The role of modality and contiguity », *Journal of educational psychology*, vol. 91, no 2, p. 358.
68. Moreno, Roxana et Richard E Mayer (2000). « Engaging students in active learning: The case for personalized multimedia messages », *Journal of educational psychology*, vol. 92, no 4, p. 724.
69. Moreno, Roxana, Richard E Mayer, Hiller A Spires et James C Lester (2001). « The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? », *Cognition and instruction*, vol. 19, no 2, p. 177-213.

70. Morin, Robert E et David A Grant (1955). « Learning and performance on a key-pressing task as function of the degree of spatial stimulus-response correspondence », *Journal of experimental psychology*, vol. 49, no 1, p. 39.

71. Morris, Carl C (1990). « Retrieval processes underlying confidence in comprehension judgments », *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 16, no 2, p. 223.

72. Mousavi, Seyed Yaghoub, Renae Low et John Sweller (1995). « Reducing cognitive load by mixing auditory and visual presentation modes », *Journal of educational psychology*, vol. 87, no 2, p. 319.

73. Muehlethaler, Céline M. et Chiara P. Knecht (2016). « Situation awareness training for general aviation pilots using eye tracking », *IFAC-PapersOnLine*, vol. 49, no 19, p. 66-71.

74. Myers III, Paul L, Arnold W Starr et Kadie Mullins (2018). « Flight simulator fidelity, training transfer, and the role of instructors in optimizing learning », *International Journal of Aviation, Aeronautics, and Aerospace*, vol. 5, no 1, p. 6.

75. Nelson, Douglas L, Valerie S Reed et John R Walling (1976). « Pictorial superiority effect », *Journal of experimental psychology: Human learning and memory*, vol. 2, no 5, p. 523.

76. Noetel, Michael, Shantell Griffith, Oscar Delaney, Nicola Rose Harris, Taren Sanders, Philip Parker, *et al.* (2022). « Multimedia design for learning: An overview of reviews with meta-meta-analysis », *Review of Educational Research*, vol. 92, no 3, p. 413-454.

77. Nunnally, Jum C, Paul D Knott, Albert Duchnowski et Ronald Parker (1967). « Pupillary response as a general measure of activation », *Perception & psychophysics*, vol. 2, no 4, p. 149-155.

78. O'Neil, HF, Richard E Mayer, Howard E Herl, Colleen Niemi, Karen Olin et Richard Allen Thurman (2000). « Instructional strategies for virtual aviation training environments », *Aircrew training and assessment*, p. 105-130.

79. Paivio, Allan (1990). *Mental representations: A dual coding approach*, Oxford University Press.

80. Park, Babette, Roxana Moreno, Tina Seufert et Roland Brünken (2011). « Does cognitive load moderate the seductive details effect? A multimedia study », *Computers in Human Behavior*, vol. 27, no 1, p. 5-10.
81. Pirovano, Giacomo, Sheryl Roberts, Susanne Kossatz et Thomas Reiner (2020). « Optical imaging modalities: Principles and applications in preclinical research and clinical settings », *Journal of Nuclear Medicine*, vol. 61, no 10, p. 1419-1427.
82. Pociask, Fredrick D et Gary Morrison (2004). « The effects of split-attention and redundancy on cognitive load when learning cognitive and psychomotor tasks », *Association for Educational Communications and Technology*.
83. Rehmann, Albert J (1997). *Human factors recommendations for airborne controller-pilot data link communications (cpdls) systems: A synthesis of research results and literature*, FEDERAL AVIATION ADMINISTRATION TECHNICAL CENTER ATLANTIC CITY NJ.
84. Reinwein, Joachim (2012). « Does the modality effect exist? And if so, which modality effect? », *Journal of psycholinguistic research*, vol. 41, no 1, p. 1-32.
85. Scott, Thomas R, William H Wells et Dorothy Z Wood (1967). « Pupillary response and sexual interest reexamined », *Journal of Clinical Psychology*.
86. Shiferaw, Brook, Luke Downey et David Crewther (2019). « A review of gaze entropy as a measure of visual scanning efficiency », *Neuroscience & Biobehavioral Reviews*, vol. 96, p. 353-366.
87. Shih, Yi-Hsuen et Min Liu (2007). « The importance of emotional usability », *Journal of Educational Technology Systems*, vol. 36, no 2, p. 203-218.
88. Slater, Mel et Sylvia Wilbur (1997). « A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments », *Presence: Teleoperators & Virtual Environments*, vol. 6, no 6, p. 603-616.
89. Sweller, John (2005). « The redundancy principle in multimedia learning », *The Cambridge handbook of multimedia learning*, vol. 159, p. 167.
90. Tindall-Ford, Sharon, Paul Chandler et John Sweller (1997). « When two sensory modes are better than one », *Journal of experimental psychology: Applied*, vol. 3, no 4, p. 257.

91. Tole, JR, AT Stephens, M Vivaudou, AR Ephrath et LR Young (1983). *Visual scanning behavior and pilot workload*, NASA.
92. Unema, Pieter JA, Sebastian Pannasch, Markus Joos et Boris M Velichkovsky (2005). « Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration », *Visual cognition*, vol. 12, no 3, p. 473-494.
93. Valverde, Horace H (1973). « A review of flight simulator transfer of training studies », *Human Factors*, vol. 15, no 6, p. 510-522.
94. Van De Merwe, Koen, Henk Van Dijk et Rolf Zon (2012). « Eye movements as an indicator of situation awareness in a flight simulator experiment », *The International Journal of Aviation Psychology*, vol. 22, no 1, p. 78-95.
95. van der Wel, Pauline et Henk van Steenbergen (2018). « Pupil dilation as an index of effort in cognitive control tasks: A review », *Psychonomic bulletin & review*, vol. 25, no 6, p. 2005-2015.
96. van Dijk, Henk, Koen van de Merwe et Rolf Zon (2011). « A coherent impression of the pilots' situation awareness: Studying relevant human factors tools », *The International Journal of Aviation Psychology*, vol. 21, no 4, p. 343-356.
97. Wanyan, X., D. Zhuang et H. Zhang (2014). « Improving pilot mental workload evaluation with combined measures », *Biomed Mater Eng*, vol. 24, no 6, p. 2283-2290.
98. Weber, Pavel, Franca Rupprecht, Stefan Wiesen, Bernd Hamann et Achim Ebert (2021). « Assessing cognitive load via pupillometry », dans *Advances in artificial intelligence and applied cognitive computing*, Springer, p. 1087-1096.
99. Wickens, Christopher D, William S Helton, Justin G Hollands et Simon Banbury (2021). *Engineering psychology and human performance*, Routledge.
100. Wickens, Christopher D et Yili Liu (1988). « Codes and modalities in multiple resources: A success and a qualification », *Human factors*, vol. 30, no 5, p. 599-616.
101. Wittrock, Merlin C (1989). « Generative processes of comprehension », *Educational psychologist*, vol. 24, no 4, p. 345-376.

102. Wu, Jiming et Clyde Holsapple (2014). « Imaginal and emotional experiences in pleasure-oriented it usage: A hedonic consumption perspective », *Information & Management*, vol. 51, no 1, p. 80-92.

Chapter 4. Conclusion

Rappel de la question de recherche

En nous basant sur la TCAM, l'objectif de ce mémoire était d'identifier et d'optimiser une caractéristique du design d'interface des simulateurs en RV 3D sur ordinateur pour permettre l'amélioration des états cognitifs et de la performance d'apprentissage d'experts en formation. Nous avons formulé la question de recherche suivante :

Dans quelle(s) modalité(s) sensorielle(s) le matériel de formation d'une simulation en RV 3D sur ordinateur doit-il être présenté pour permettre d'améliorer significativement les états cognitifs d'apprentissage et la performance d'apprentissage de professionnels ?

Principaux résultats

Globalement, les résultats montrent que l'inclusion de matériel de formation sous forme de texte place les sujets dans un meilleur état d'apprentissage que lorsque les instructions sont présentées sous forme d'audio, ou sous la forme d'un agencement de texte et d'audio. Spécifiquement, les participants de la condition texte étaient dans des états cognitifs d'apprentissage plus optimaux pendant la tâche d'évaluation en vol, tel que démontré par une efficacité du balayage visuel[14] significativement supérieure et une valence émotionnelle significativement supérieure. De surcroit, la condition texte induit une charge cognitive implicite inférieure, une immersion perçue supérieure et une motivation perçue supérieure sans différence significative. Ces résultats font objection au principe de modalité de la TCAM qui stipule que des individus apprennent mieux lorsqu'un message multimédia est prononcé plutôt qu'écrit (Richard Mayer et Mayer, 2005). Une explication possible pour nos résultats vient du fait de la division de l'attention des sujets entre les tâches instructionnelles et les tâches d'opération de l'équipement hautement spécialisé. La littérature indique en effet qu'une attention divisée pendant l'apprentissage multimédia

[14] Traduction de « Visual Transition Entropy »

est incompatible avec le principe de modalité (Pociask et Morrison, 2004). Finalement, aucune différence de performance d'apprentissage ne fut détectée entre les trois groupes expérimentaux. Ceci suggère que l'utilisation isolée de mesures de performance issus des paramètres de l'avion soit insuffisante pour permettre de tirer des conclusions sur les caractéristiques d'interface à sélectionner pour améliorer les résultats d'apprentissage (Gisler *et al.*, 2021).

Contributions de l'étude

Contributions théoriques

Lors de l'entrainement avec un simulateur où ont lieu des tâches simultanées (e.g., le traitement de matériel de formation et l'opération d'équipement hautement spécialisé), les multiples états d'apprentissage cognitifs, attentionnels, motivationnels et affectifs doivent guider les choix de design. Dans cette étude, nous avons démontré que lorsque la demande attentionnelle visuelle pour la surveillance des signaux du tableau de bord est élevée, il parait désirable d'éviter de diviser et réorienter l'attention d'un apprenant en utilisant la voie auditive pour transmettre des instructions (Latorella, 1998). Diverses caractéristiques propres à la modalité sensorielle auditive peuvent avoir un effet délétère sur l'expérience d'apprentissage, notamment, un effet de pré-emption sur la modalité visuelle et la nature éphémère du message auditif (Wickens, Dixon et Seppelt, 2005). À l'inverse, la possibilité de visualiser à l'écran le matériel pédagogique offre la possibilité aux apprenants de sélectionner le moment opportun pour rediriger leur attention d'une tâche à l'autre, réduisant les effets néfastes d'une division de l'attention. Bref, ce travail appuie que l'apparition d'une attention divisée pendant l'apprentissage ne soit pas compatible avec le principe de modalité de la CTML (Richard Mayer et Mayer, 2005).

Contributions pratiques

Les résultats des deux études ont plusieurs implications pour les designers de simulations destinées à l'entrainement de personnel hautement spécialisé.

Dans un premier temps, les résultats de la première étude suggèrent que les designers de simulation pour l'entrainement de professionnels ont intérêt à intégrer un didacticiel introductif au système afin d'augmenter ses capacités d'apprentissage. Un taux d'échec de seulement 2,73% dans les tâches de rétention indique que les sujets se rappelaient la quasi-totalité des fonctionnalités d'interface qui leur furent présentées pendant le didacticiel, ce qui appuie l'efficacité de la solution. Afin de guider les concepteurs, nous présentons dans le **Tableau 7** ci-bas, les paramètres du didacticiel introductif qui furent mis en évidence pendant l'évaluation d'utilisabilité ainsi que des recommandations concrètes ayant le potentiel d'augmenter les capacités d'apprentissage du simulateur. Les suggestions effectuées sont le résultat d'une analyse qualitative fine des points de friction de valence émotionnelle, interprétée à la lumière des principes de design de la TCAM. Or, pour chaque recommandation mentionnée, les concepteurs devraient prioriser les tests utilisateurs si leurs ressources les permettent.

Élément de design multimédia observé	**Recommandations pour les concepteurs d'interface**
Texte intégral imposé avec narration L'information auditive est simultanément affichée à l'écran dans des encadrés de texte (redondance, modalité).	· Tester et comparer la présentation de matériel de formation dans une seule modalité et dans deux modalités (Arslan, 2012). · Rendre la narration disponible au format texte pour suivre les bonnes pratiques d'accessibilité (Seale, 2013) et pour tenir compte des préférences et contextes personnels. · Rendre possible l'activation/désactivation de la fonction textuelle ou audio. Si le matériel est présenté dans deux modalités, la présentation visuelle sous forme de sous-titres facilite la compréhension qu'il s'agit de la même information.
Rythme imposé Le rythme du narrateur est imposé. Ce faisant, il impose la durée d'affichage du contenu textuel, sans possibilité de modifier le rythme ou de faire des pauses.	· Donner plus de contrôle à l'utilisateur en lui offrant la possibilité de naviguer d'une leçon courte à l'autre, suggérant et non imposant un chemin déterminé (e.g., retour en arrière, saut de leçons, pauses possibles, contrôle pour l'ordre de présentation des segments, etc.) (Lamontagne *et al.*, 2021). · Entre les segments, rendre actives toutes les fonctions d'interface pour que l'utilisateur se familiarise et explore l'interface.

Voix narrative La narration est effectuée par une voix informatisée qui livre du contenu formel (voix, personnalisation).	· Tester et comparer une narration humaine enregistrée et une voix automatisée (Richard E Mayer, 2005). · Tester et comparer du contenu livré dans un ton formel et informel/conversationnel (Richard E Mayer, 2005).
Éléments sonores provenant de l'VE pendant la narration Un bip sonore originaire d'un objet de l'VE est présent pendant toute la durée du didacticiel introductif (cohérence).	· Pendant le didacticiel, conserver uniquement les bruits ambiants de la simulation lorsque ceux-ci sont liés directement ou indirectement avec la fonctionnalité d'interface en cours de présentation dans le segment et retirer les éléments distrayants (Moreno et Mayer, 2000).
Instructions d'utilisation de fonctions d'interface maitrisées Des étapes du didacticiel servent à expliquer comment utiliser des boutons connus des participants, comme un bouton « pause » ou « retour » (cohérence).	· Revoir l'architecture de l'information présentée dans le didacticiel afin de classifier et catégoriser les segments selon un ordre logique (e.g., séparer les segments qui présentent des objets de l'EV de ceux qui expliquent comment naviguer l'interface). · Permettre le saut de sections.
Fonctions bloquées pendant le didacticiel Diverses fonctionnalités d'interface sont bloquées pendant le didacticiel, alors que d'autres sont mises en avant à l'aide d'éléments contrastants (segmentation, signalement).	· Maintenir actives les fonctionnalités en cours de présentation pour permettre la pratique et rendre inactives les fonctionnalités d'interface n'ayant aucun lien direct ou indirect avec celle(s) en cours de présentation. · Conduire un test utilisateur pour déterminer l'étendue du spectre des fonctionnalités actives (e.g. pour permettre la contextualisation des fonctions en cours de présentation; pour permettre de tester les fonctionnalités présentées aux segments précédents). · Signaler clairement l'indisponibilité des fonctionnalités d'interface désactivées. Conserver le signalement clair des éléments en cours de présentation (Alpizar, Adesope et Wong, 2020). Favoriser l'utilisation de couleurs, de flèches, d'éléments clignotants et d'ombres.

Tableau 7. *Éléments de design multimédia observés et recommandations de design d'interface associées.*

L'Article 2 de ce mémoire démontre que pour la conception d'un simulateur destiné à l'entrainement d'experts, les designers devraient prioriser la modalité sensorielle visuelle sous forme de texte pour présenter du matériel pédagogique, si les conditions suivantes sont remplies : pendant l'entrainement, l'attention des sujets est divisée et la tâche primaire comporte une forte demande d'attention visuelle. Nos résultats soulignent l'importance de considérer divers aspects de l'environnement d'apprentissage, notamment, la présence de tâches multiples sérielles ou parallèles. Si un entrainement en simulateur comporte plusieurs tâches sérielles, le niveau de difficulté, les demandes cognitives et les caractéristiques de chaque tâche varieront, d'où l'importance de concevoir le matériel pédagogique spécifiquement pour chaque segment de l'entrainement. S'il y a présence de tâches parallèles, les concepteurs doivent vérifier s'il y a division de l'attention, ils doivent estimer la charge d'attention visuelle requise pour chaque tâche ou le niveau de disruption du signal auditif. Le niveau d'expertise doit également être considéré.

Contributions méthodologiques

Premièrement, dans un contexte d'évaluation UX ayant comme participants des experts, ce travail suggère l'utilisation d'un questionnaire psychométrique validé d'auto-efficacité perçue pouvant être administré pendant la session expérimentale tel qu'un contrôle. Divers questionnaires de ce type continuent d'être développés et validés dans plusieurs domaines, notamment en éducation et pour l'évaluation des compétences au travail (Dybowski, Kriston et Harendza, 2016; Heyne *et al.*, 1998; Kang *et al.*, 2019; Tsai *et al.*, 2020). Nous avons trouvé que les résultats d'auto-efficacité perçue corrélaient fortement avec ceux de charge cognitive perçue, suggérant un manque de connaissances ou de compétences des sujets pour l'évaluation d'un simulateur spécialisé en anesthésie (Dybowski, Kriston et Harendza, 2016; Heyne *et al.*, 1998). Sans ce contrôle, l'interprétation des résultats aurait pu s'avérer erronée. En outre, dans la mesure du possible, le questionnaire devrait être administré avant chaque tâche expérimentale pour éviter que la performance perçue des sujets ne biaise leurs évaluations.

Deuxièmement, des aspects traditionnellement sous-spécifiés des travaux s'appuyant sur la CTML concernent les aspects motivationnels et hédoniques de l'apprentissage réalisé dans un environnement multimédia. Ce travail utilise une approche multidimensionnelle pour capter des états cognitifs, attentionnels, motivationnels et affectifs pendant l'entrainement des sujets. Cette approche assure la captation des processus d'apprentissage importants pour que les chercheurs effectuent une interprétation des données adéquate ainsi que des recommandations de design fondées et justes.

Les méthodes d'analyses employées pour capter et mesurer les états cognitifs des sujets au cours de la deuxième étude de ce mémoire peuvent être répliquées au cours d'évaluation d'utilisabilité traditionnelles. Nous avons utilisé des mesures quantitatives ayant la capacité d'informer les chercheurs de l'état attentionnel des sujets. Spécifiquement, nous avons mesuré l'efficacité du balayage visuel ainsi que la dispersion de l'attention visuelle des participants. Ces mesures peuvent indiquer aux chercheurs ce que les utilisateurs perçoivent réellement tout en caractérisant leurs stratégies visuelles. De telles mesures oculométriques objectives et non-intrusives génèrent des résultats facilement interprétables. Dès l'implémentation, ces analyses peuvent aisément être modifiées et réutilisées pour divers contextes d'évaluations UX.

Limites et pistes de recherches futures

Une limite de ce travail concerne la méthode employée pour évaluer l'effet de la manipulation d'un paramètre d'une interface utilisateur complexe. Nous avons évalué comment la modalité sensorielle de présentation de matériel de formation dans un simulateur destiné à l'entrainement de professionnels permettait l'amélioration de leur états cognitifs et de leur performance d'apprentissage. Pour la plupart des variables de l'étude, les données collectées étaient statistiquement équivalentes entre les groupes expérimentaux. Cet écart inter-groupe aurait pu être légèrement amplifié de diverses façons, notamment en augmentant le nombre de participants par groupe, en comparant les modalités sensorielles d'éléments d'interface au travers de plus d'un scénario de tâches et en testant les simulateurs d'entrainement pour professionnels issus de différentes industries. En outre, les effets combinés de plusieurs paramètres de l'interface peuvent interagir. Ainsi, une recommandation de design fondée sur un test où une seule caractéristique fut manipulée peut devenir désuète. Pour une évaluation subséquente, nous suggérons la manipulation de plus d'une caractéristique d'interface pour amplifier l'effet mesuré sur l'expérience d'apprentissage des utilisateurs. Les variables indépendantes sélectionnées peuvent être déterminées en réalisant une évaluation utilisateur reprenant les principes de design de la TCAM (Arslan, 2012; Richard E Mayer, 2005; Richard E Mayer et Fiorella, 2014; Richard E Mayer et Pilegard, 2005). Le nouveau design expérimental doit prendre en compte le nombre de conditions expérimentales et le nombre de participants requis par groupe pour conduire une expérimentation inter-sujets qui atteint un seuil minimal de puissance statistique, malgré l'éventuelle rareté des professionnels (Brysbaert, 2019).

Comme pistes de recherche futures, il serait intéressant de mesurer l'effet de l'entrainement dans le temps au cours d'une étude longitudinale réalisée en plusieurs sessions expérimentales. De surcroit, il serait pertinent de vérifier si les effets de l'entrainement avec le simulateur se traduisent par un transfert de l'entrainement positif ou négatif lors du passage à l'entrainement dans un avion (Blaiwes, Puig et Regan, 1973).

Tel que suggéré plus tôt, une étude ultérieure pourrait servir à tester l'application de principes de design issus de la TCAM pour la conception de simulateurs d'entrainement pour professionnels (Richard Mayer et Mayer, 2005). En outre, il serait intéressant de mesurer le réel bénéfice associé à l'augmentation du niveau de fidélité de simulateurs destinés à l'entrainement. Par exemple, à ressources limitées, les concepteurs devraient-ils prioriser l'augmentation du niveau de fidélité physique, cognitif ou fonctionnel des simulateurs (Myers III, Starr et Mullins, 2018)?

Finalement, au cours d'un test utilisateur subséquent, des mesures complémentaires pourraient servir à mieux mesurer les états cognitifs d'apprentissage et de performance de professionnels. Des analyses oculométriques supplémentaires pourraient servir à décrire plus en profondeur les stratégies visuelles et les comportements oculaires des apprenants. Ces mesures incluent la complexité Lempel-Ziv, les séquences N-grams, la matrice de probabilité de transition ou la matrice de densité de transition (Lounis, Peysakhovich et Causse, 2021). Par ailleurs, un avantage d'utiliser des outils psychophysiologiques en remplacement aux outils psychométriques est la réduction du temps de test destiné à la complétion de questionnaires pour l'ajout de temps servant à la collecte de données objectives non-intrusives. En complément aux scores de performance générés par le simulateur pour lesquels les évaluateurs externes ont souvent peu de contrôle, des métriques de performance dérivées directement des paramètres de l’avion pourraient servir à évaluer avec un plus fin niveau de granularité la performance des sujets professionnels (e.g., déviation en cap, en altitude, en vitesse). Ces mesures devraient être collectées conjointement à celles décrivant l’état d’apprentissage multidimensionnel des sujets.

Bibliographie

1. Alpizar, D., Adesope, O. O., & Wong, R. M. (2020). A meta-analysis of signaling principle in multimedia learning environments. *Educational Technology Research and Development*, *68*(5), 2095-2119.
2. Arslan, P. Y. (2012). A Review of Multimedia Learning Principles: Split-Attention, Modality, and Redundancy Effects. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, *8*(1), 114-122.
3. Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556-559.
4. Bandura, A. (1997). Insights. self-efficacy. *Harvard Mental Health Letter*, *13*(9), 4-6.
5. Blaiwes, A. S., Puig, J. A., & Regan, J. J. (1973). Transfer of training and the measurement of training effectiveness. *Human factors*, *15*(6), 523-533.
6. Bostrom, R., Olfman, L., & Sein, M. The importance of learning style in end-user training. MIS Q. 14 (1), 101–119 (1990). In.
7. Broder, A. (2002). A taxonomy of web search. ACM Sigir forum,
8. Brown, E., & Cairns, P. (2004). A grounded investigation of game immersion. CHI'04 extended abstracts on Human factors in computing systems,
9. Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of cognition*.
10. Butcher, K. R. (2014). The multimedia principle.
11. Charles, R. L., & Nixon, J. (2019). Measuring mental workload using physiological measures: A systematic review. *Applied ergonomics*, *74*, 221-232.
12. Chittaro, L., & Buttussi, F. (2015). Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety. *IEEE transactions on visualization and computer graphics*, *21*(4), 529-538.
13. Cockburn, A., Quinn, P., & Gutwin, C. (2015). Examining the peak-end effects of subjective experience. Proceedings of the 33rd annual ACM conference on human factors in computing systems,

14. Council, N. R., Glaser, R., Chudowsky, N., & Pellegrino, J. W. (2001). *Knowing what students know: The science and design of educational assessment*. National Academies Press.
15. Courtemanche, F., Léger, P., Fredette, M., & Sénécal, S. (2022). COBALT-Photobooth: Integrated UX Data System. *Montréa: HEC Montréal*.
16. Dargent, T., Karran, A., Léger, P.-M., Coursaris, C. K., & Sénécal, S. The Influence of Task Types on User Experience after a Web Interface Update.
17. Dawson, M. E., Schell, A. M., & Filion, D. L. (2017). The electrodermal system.
18. de Guinea, A. O., Titah, R., & Léger, P.-M. (2014). Explicit and implicit antecedents of users' behavioral beliefs in information systems: A neuropsychological investigation. *Journal of Management Information Systems*, *30*(4), 179-210.
19. Dybowski, C., Kriston, L., & Harendza, S. (2016). Psychometric properties of the newly developed Physician Teaching Self-efficacy Questionnaire (PTSQ). *BMC medical education*, *16*(1), 1-11.
20. Egle, J. P., Malladi, S. V. S., Gopinath, N., & Mittal, V. K. (2015). Simulation training improves resident performance in hand-sewn vascular and bowel anastomoses. *Journal of Surgical Education*, *72*(2), 291-296.
21. Eich, E., Kihlstrom, J. F., Bower, G. H., Niedenthal, P. M., & Forgas, J. P. (2000). *Cognition and emotion*. Oxford University Press on Demand.
22. Ekman, P., & Friesen, W. V. (1978). Facial action coding system. *Environmental Psychology & Nonverbal Behavior*.
23. Feng, Z., González, V. A., Amor, R., Lovreglio, R., & Cabrera-Guerrero, G. (2018). Immersive virtual reality serious games for evacuation training and research: A systematic literature review. *Computers & Education*, *127*, 252-266.
24. Fowler, J. (2008). Experiential learning and its facilitation. *Nurse Education Today*, *28*(4), 427-433.
25. Ganglbauer, E., Schrammel, J., Deutsch, S., & Tscheligi, M. (2009). Applying psychophysiological methods for measuring user experience: possibilities, challenges and feasibility. Workshop on user experience evaluation methods in product development,

26. Giroux-Huppé, C., Sénécal, S., Fredette, M., Chen, S. L., Demolin, B., & Léger, P.-M. (2019). Identifying psychophysiological pain points in the online user journey: the case of online grocery. International Conference on Human-Computer Interaction,
27. Giroux, F., Léger, P.-M., Brieugne, D., Courtemanche, F., Bouvier, F., Chen, S.-L., Tazi, S., Rucco, E., Fredette, M., & Coursaris, C. (2021). Guidelines for collecting automatic facial expression detection data synchronized with a dynamic stimulus in remote moderated user tests. International Conference on Human-Computer Interaction,
28. Gisler, J., Schneider, J., Handali, J., Holzwarth, V., Hirt, C., Fuhl, W., Vom Brocke, J., & Kunz, A. (2021). Indicators of Training Success in Virtual Reality Using Head and Eye Movements. 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct),
29. Glaholt, M. G. (2014). Eye tracking in the cockpit: a review of the relationships between eye movements and the aviators cognitive state.
30. Goodyear, P., & Retalis, S. (2010). Learning, technology and design. In *Technology-Enhanced learning* (pp. 1-27). Brill.
31. Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). Elsevier.
32. Hays, R. T., Jacobs, J. W., Prince, C., & Salas, E. (1992). Flight simulator training effectiveness: A meta-analysis. *Military psychology*, *4*(2), 63-74.
33. Helleberg, J. R., & Wickens, C. D. (2003). Effects of data-link modality and display redundancy on pilot performance: An attentional perspective. *The International Journal of Aviation Psychology*, *13*(3), 189-210.
34. Heyne, D., King, N., Tonge, B., Rollings, S., Pritchard, M., Young, D., & Myerson, N. (1998). The Self-efficacy Questionnaire for School Situations: Development and Psychometric Evaluation. *Behaviour Change*, *15*(1), 31-40. https://doi.org/10.1017/S081348390000588X
35. Hippe, D. S., Umoren, R. A., McGee, A., Bucher, S. L., & Bresnahan, B. W. (2020). A targeted systematic review of cost analyses for implementation of

simulation-based education in healthcare. *SAGE Open Medicine*, *8*, 2050312120913451.

36. Isaranuwatchai, W., Brydges, R., Carnahan, H., Backstein, D., & Dubrowski, A. (2014). Comparing the cost-effectiveness of simulation modalities: a case study of peripheral intravenous catheterization training. *Advances in Health Sciences Education*, *19*(2), 219-232.
37. ISO. (2018). 9241-11: 2018 Ergonomics of Human-System Interaction—Part 11: Usability: Definitions and Concepts. *International Organization for Standardization. https://www. iso. org/obp/ui/# iso: std: iso*, *9241*(11).
38. ISO, B., & STANDARD, B. (2010). Ergonomics of human-system interaction. *British Standatds Institution*.
39. Izard, S. G., Juanes, J. A., García Peñalvo, F. J., Estella, J. M., Ledesma, M., & Ruisoto, P. (2018). Virtual reality as an educational and training tool for medicine. *Journal of medical systems*, *42*(3), 1-5.
40. Kang, Y.-N., Chang, C.-H., Kao, C.-C., Chen, C.-Y., & Wu, C.-C. (2019). Development of a short and universal learning self-efficacy scale for clinical skills. *PLoS One*, *14*(1), e0209155.
41. Lackmann, S., Léger, P.-M., Charland, P., Aubé, C., & Talbot, J. (2021). The influence of video format on engagement and performance in online learning. *Brain Sciences*, *11*(2), 128.
42. Lamontagne, C., Sénécal, S., Fredette, M., Chen, S. L., Pourchon, R., Gaumont, Y., Grandpré, D. D., & Léger, P.-M. (2019). User test: how many users are needed to find the psychophysiological pain points in a journey map? International Conference on Human Interaction and Emerging Technologies,
43. Lamontagne, C., Sénécal, S., Fredette, M., Labonté-LeMoyne, É., & Léger, P.-M. (2021). The effect of the segmentation of video tutorials on user's training experience and performance. *Computers in Human Behavior Reports*, *3*, 100071.
44. Lee, E. A.-L., & Wong, K. W. (2014). Learning with desktop virtual reality: Low spatial ability learners are more positively affected. *Computers & Education*, *79*, 49-58.

45. Léger, P.-M., Courtemanche, F., Fredette, M., & Sénécal, S. (2019). A cloud-based lab management and analytics software for triangulated human-centered research. In *Information Systems and Neuroscience* (pp. 93-99). Springer.
46. Léger, P.-M., Sénecal, S., Courtemanche, F., de Guinea, A. O., Titah, R., Fredette, M., & Labonte-LeMoyne, E. (2014). Precision is in the eye of the beholder: Application of eye fixation-related potentials to information systems research.
47. Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *Int. J. Hum. Comput. Interact.*, *7(1)*, 57-78. https://doi.org/https://doi.org/10.1080/10447319509526110
48. LibreTexts, M. (2020). *12.1C: Sensory Modalities*. Retrieved August 8th from https://remotedesktop.google.com/access/session/4b6b207e-0adb-4c23-add4-229b44fd0f79
49. Loijens, L., & Krips, O. (2018). *FaceReader Methodology Note. A white paper by Noldus Information Technology*.
50. Lounis, C., Peysakhovich, V., & Causse, M. (2021). Visual scanning strategies in the cockpit are modulated by pilots' expertise: A flight simulator study. *PLoS One*, *16*(2), e0247061. https://doi.org/10.1371/journal.pone.0247061
51. Makled, E., Yassien, A., Elagroudy, P., Magdy, M., Abdennadher, S., & Hamdi, N. (2019). PathoGenius VR: VR medical training. Proceedings of the 8th ACM International Symposium on Pervasive Displays,
52. Makransky, G., Borre-Gude, S., & Mayer, R. E. (2019). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning*, *35*(6), 691-707.
53. Makransky, G., & Petersen, G. B. (2019). Investigating the process of learning with desktop virtual reality: A structural equation modeling approach. *Computers & Education*, *134*, 15-30. https://doi.org/10.1016/j.compedu.2019.02.002
54. Manakhov, P., & Ivanov, V. D. (2016). Defining usability problems. Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems,
55. Mandal, S. (2013). Brief introduction of virtual reality & its challenges. *International Journal of Scientific & Engineering Research*, *4*(4), 304-309.

56. Martin, A. R. (1986). Stress in residency. *Journal of general internal medicine*, *1*(4), 252-257.
57. Mayer, R., & Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. Cambridge university press.
58. Mayer, R. E. (2005). Principles of multimedia learning based on social cues: Personalization, voice, and image principles.
59. Mayer, R. E. (2014). Incorporating motivation into multimedia learning. *Learning and instruction*, *29*, 171-173.
60. Mayer, R. E., & Fiorella, L. (2014). 12 principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In *The Cambridge handbook of multimedia learning* (Vol. 279). Cambridge University Press New York, NY.
61. Mayer, R. E., & Pilegard, C. (2005). Principles for managing essential processing in multimedia learning: Segmenting, pretraining, and modality principles. *The Cambridge handbook of multimedia learning*, 169-182.
62. McComas, J., MacKay, M., & Pivik, J. (2002). Effectiveness of virtual reality for teaching pedestrian safety. *CyberPsychology & Behavior*, *5*(3), 185-190.
63. Moreno, R. (2005). Instructional technology: Promise and pitfalls. *Technology-based education: Bringing researchers and practitioners together*, 1-19.
64. Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning*, *22*(3), 149-158.
65. Moreno, R. (2007). Optimising learning from animations by minimising cognitive load: Cognitive and affective consequences of signalling and segmentation methods. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *21*(6), 765-781.
66. Moreno, R. (2009). Learning from animated classroom exemplars: The case for guiding student teachers' observations with metacognitive prompts. *Educational Research and Evaluation*, *15*(5), 487-501.

67. Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of educational psychology, 92*(1), 117.
68. Myers III, P. L., Starr, A. W., & Mullins, K. (2018). Flight simulator fidelity, training transfer, and the role of instructors in optimizing learning. *International Journal of Aviation, Aeronautics, and Aerospace, 5*(1), 6.
69. Naismith, L. M., Cheung, J. J., Ringsted, C., & Cavalcanti, R. B. (2015). Limitations of subjective cognitive load measures in simulation-based procedural training. *Medical education*, *49*(8), 805-814.
70. O'Brien, H. L., & Lebow, M. (2013). Mixed-methods approach to measuring user experience in online news interactions. *Journal of the American Society for Information Science and Technology*, *64*(8), 1543-1556.
71. Ortiz, G. A. (1994). Effectiveness of PC-based flight simulation. *The International Journal of Aviation Psychology*, *4*(3), 285-291.
72. Paas, F., Tuovinen, J. E., Van Merrienboer, J. J., & Aubteen Darabi, A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research and Development*, *53*(3), 25-34.
73. Pociask, F. D., & Morrison, G. (2004). The Effects of Split-Attention and Redundancy on Cognitive Load When Learning Cognitive and Psychomotor Tasks. *Association for Educational Communications and Technology*.
74. Qin, H., Rau, P.-L. P., & Salvendy, G. (2010). Effects of different scenarios of game difficulty on player immersion. *Interacting with computers*, *22*(3), 230-239.
75. Redifer, J. L., Bae, C. L., & Zhao, Q. (2021). Self-efficacy and performance feedback: Impacts on cognitive load during creative thinking. *Learning and instruction*, *71*, 101395.
76. Riener, R., & Harders, M. (2012). VR for medical training. In *Virtual reality in medicine* (pp. 181-210). Springer.
77. Rochon, L.-J., Karran, A. J., Bouvier, F., Coursaris, C. K., Sénécal, S., Delisle, J.-F., & Léger, P.-M. (2021). Improving Learnability Capabilities in Desktop VR Medical Applications. International Conference on Human-Computer Interaction,

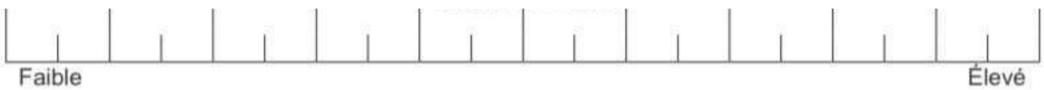
78. Rodwin, B. A., Bilan, V. P., Merchant, N. B., Steffens, C. G., Grimshaw, A. A., Bastian, L. A., & Gunderson, C. G. (2020). Rate of preventable mortality in hospitalized patients: a systematic review and meta-analysis. *Journal of general internal medicine*, *35*(7), 2099-2106.
79. Rose, F. D., Attree, E. A., Brooks, B. M., Parslow, D. M., & Penn, P. R. (2000). Training in virtual environments: transfer to real world tasks and equivalence to real task training. *Ergonomics*, *43*(4), 494-511.
80. Schroeder, R. (2008). Defining virtual worlds and virtual environments. *Journal For Virtual Worlds Research*, *1*(1).
81. Schunk, D. H., & DiBenedetto, M. K. (2016). Self-efficacy theory in education. In *Handbook of motivation at school* (pp. 34-54). Routledge.
82. Seale, J. (2013). *E-learning and disability in higher education: accessibility research and practice*. Routledge.
83. Sewell, J. L., Boscardin, C. K., Young, J. Q., Ten Cate, O., & O'Sullivan, P. S. (2017). Learner, patient, and supervisor features are associated with different types of cognitive load during procedural skills training: implications for teaching and instructional design. *Academic Medicine*, *92*(11), 1622-1631.
84. Smith, S., & Ericson, E. (2009). Using immersive game-based virtual reality to teach fire-safety skills to children. *Virtual reality*, *13*(2), 87-99.
85. Standardization, I. O. f. (2010). *Ergonomics of Human-system Interaction: Part 210: Human-centred Design for Interactive Systems*. ISO.
86. Stephanidis, C., Kurosu, M., Chen, J. Y., Fragomeni, G., Streitz, N., Konomi, S. i., Degen, H., & Ntoa, S. (2021). *HCI International 2021-Late Breaking Papers: Multimodality, eXtended Reality, and Artificial Intelligence: 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings* (Vol. 13095). Springer Nature.
87. Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37-76). Elsevier.
88. Sweller, J., Ayres, P., & Kalyuga, S. (2011). Measuring cognitive load. In *Cognitive load theory* (pp. 71-85). Springer.

89. Tsai, C.-L., Cho, M.-H., Marra, R., & Shen, D. (2020). The self-efficacy questionnaire for online learning (SeQoL). *Distance Education*, *41*(4), 472-489.
90. Van Merriënboer, J. J., & Sweller, J. (2010). Cognitive load theory in health professional education: design principles and strategies. *Medical education*, *44*(1), 85-93.
91. Vasile, C., Marhan, A.-M., Singer, F. M., & Stoicescu, D. (2011). Academic self-efficacy and cognitive load in students. *Procedia-Social and Behavioral Sciences*, *12*, 478-482.
92. Vasseur, A., Léger, P.-M., Courtemanche, F., Labonte-Lemoyne, E., Georges, V., Valiquette, A., Brieugne, D., Rucco, E., Coursaris, C., & Fredette, M. (2021). Distributed remote psychophysiological data collection for UX evaluation: a pilot project. International Conference on Human-Computer Interaction,
93. Verified Market Research. (2022). *Virtual Training And Simulation Market Size And Forecast*. Retrieved July 1st, 2022 from https://www.verifiedmarketresearch.com/product/virtual-training-and-simulation-market/
94. Wang, P., Wu, P., Wang, J., Chi, H.-L., & Wang, X. (2018). A critical review of the use of virtual reality in construction engineering education and training. *International journal of environmental research and public health*, *15*(6), 1204.
95. Wickens, C. D., Dixon, S. R., & Seppelt, B. (2005). Auditory preemption versus multiple resources: Who wins in interruption management? Proceedings of the Human Factors and Ergonomics Society Annual Meeting,
96. Zulkosky, K. (2009). Self-efficacy: a concept analysis. Nursing forum,

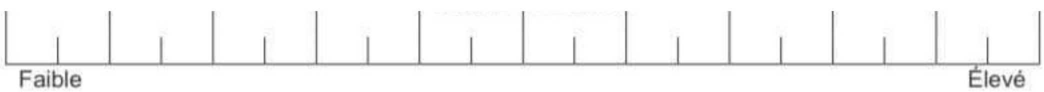
Annexe - Questionnaires

1.1 NASA-TLX

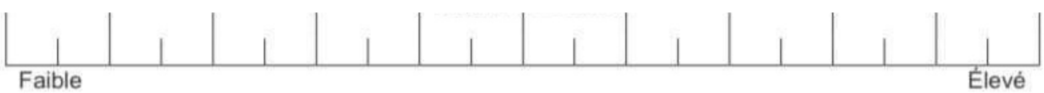
Quelle a été l’importance de l’exigence mentale requise pour les tâches précédentes?
(1- Peu exigeant, 10-Très exigeant)



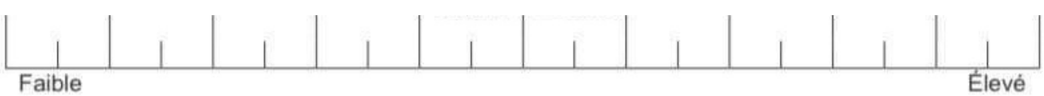
Quelle a été l’importance de l’exigence physique requise pour les tâches précédentes?
(1- Peu exigeant, 10-Très exigeant)



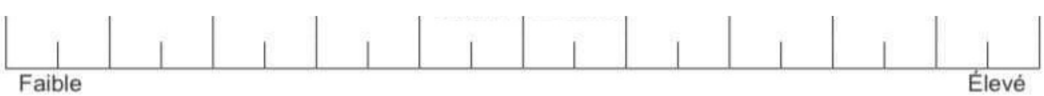
Quelle a été l’importance de la pression ressentie par rapport au temps de réaction nécessaire?



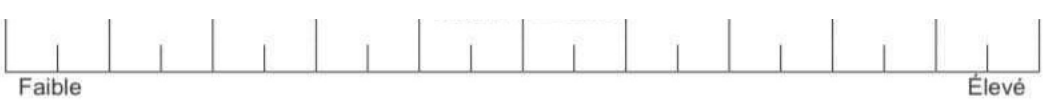
Quel niveau de réussite pensez-vous avoir atteint pour les tâches précédentes?



Quel degré d'effort (physique ou mental) avez-vous dû fournir pour effectuer les tâches précédentes?



Pendant l’exécution des tâches précédentes, à quel point vous êtes-vous senti stressé(e), irrité(e), non-confiant(e)?



1.2 PSSUQ

Lisez chaque énoncé et indiquez dans quelle mesure vous êtes en accord ou en désaccord, en vous référant au tutoriel que vous venez de compléter.

Fortement en désaccord	En désaccord	Plutôt en désaccord	Neutre	Plutôt en accord	En accord	Fortement en accord
1	**2**	**3**	**4**	**5**	**6**	**7**

Durant la simulation							
De manière générale, je suis satisfait.e de la facilité d'utilisation de ce site web	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Il était simple d'utiliser ce site web	**1**	**2**	**3**	**4**	**5**	**6**	**7**
En utilisant ce site web, j'ai complété rapidement les tâches et scénarios demandés	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Je me suis senti.e à l'aise en utilisant ce site web	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Il a été facile d'apprendre à utiliser ce site web	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Je crois que je pourrais rapidement devenir productif.ve en utilisant ce site web	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Les messages d'erreurs générés par le site web permettaient de résoudre les problèmes	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Lorsque je faisais une erreur en utilisant le site web, je pouvais la rattraper facilement et rapidement	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Les informations fournies sur ce site web étaient claires	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Il était facile de trouver les informations dont j'avais besoin	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Les informations étaient efficaces pour m'aider à accomplir les tâches et les scénario	**1**	**2**	**3**	**4**	**5**	**6**	**7**
L'organisation des informations à l'écran était claire	**1**	**2**	**3**	**4**	**5**	**6**	**7**
L'interface de ce site web était agréable	**1**	**2**	**3**	**4**	**5**	**6**	**7**
J'ai aimé utiliser l'interface de ce site web	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Ce site web a toutes les fonctionnalités et les capacités que j'en attends	**1**	**2**	**3**	**4**	**5**	**6**	**7**
De manière générale, je suis satisfait.e de ce site web	**1**	**2**	**3**	**4**	**5**	**6**	**7**

1.3 L-SES

En vous référant aux trois tâches précédentes, lisez chaque énoncé et dites dans quelle mesure vous êtes en accord ou en désaccord.

Fortement en désaccord	En désaccord	Neutre	En accord	Fortement en accord
1	**2**	**3**	**4**	**5**

Après la simulation					
Je peux me rappeler comment préformer l'habileté clinique	**1**	**2**	**3**	**4**	**5**
Je comprends le contenu de l'habileté clinique et je peux le démontrer aux autres	**1**	**2**	**3**	**4**	**5**
Je peux expliquer verbalement l'usage et le principe d'opération de l'habileté clinique	**1**	**2**	**3**	**4**	**5**
Je peux expliquer verbalement la séquence et l'interrelation entre chaque étape	**1**	**2**	**3**	**4**	**5**

1.4 SIMS – Partie 1

Lisez chaque énoncé et indiquez le chiffre qui correspond le mieux à la raison pour laquelle vous seriez engagé(e) à nouveau dans une activité d'entrainement en simulateur de vol comme celle effectuée aujourd'hui.

Pourquoi vous engageriez-vous dans cette activité dans le futur?

Ne correspond pas	Correspond très peu	Correspond peu	Correspond modérément	Correspond assez	Correspond beaucoup	Correspond exactement
1	**2**	**3**	**4**	**5**	**6**	**7**

Durant la simulation							
Parce que je pense que cette activité est intéressante	1	2	3	4	5	6	7
Parce que je le ferais pour mon propre bien	1	2	3	4	5	6	7
Parce que je serais supposé(e) le faire	1	2	3	4	5	6	7
Il doit y avoir de bonnes raisons de faire cette activité, mais personnellement je n'en vois aucun	1	2	3	4	5	6	7
Parce que je pense que cette activité est plaisante	1	2	3	4	5	6	7
Parce que je pense que cette activité est bonne pour moi	1	2	3	4	5	6	7
Parce qu'il s'agirait de quelque chose que j'ai besoin de faire	1	2	3	4	5	6	7
Je ferais cette activité, mais je ne suis pas certain(e) si ça en vaut la peine	1	2	3	4	5	6	7

1.4 SIMS – Partie 2

Lisez chaque énoncé et indiquez le chiffre qui correspond le mieux à la raison pour laquelle vous seriez engagé(e) à nouveau dans une activité d'entrainement en simulateur de vol comme celle effectuée aujourd'hui.

Pourquoi vous engageriez-vous dans cette activité dans le futur?

Ne correspond pas	Correspond très peu	Correspond peu	Correspond modérément	Correspond assez	Correspond beaucoup	Correspond exactement
1	**2**	**3**	**4**	**5**	**6**	**7**

Durant la simulation							
Parce que cette activité est amusante	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Par décision personnelle	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Parce que je n'aurais pas le choix	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Je ne sais pas; Je ne vois pas ce que cette activité m'apporterait	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Parce que je me sens bien lorsque je fais cette activité	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Parce que je pense que cette activité est importante pour moi	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Parce que je sens que je dois le faire	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Je ferais cette activité, mais je ne suis pas sûr(e) que ce soit une bonne chose de continuer	**1**	**2**	**3**	**4**	**5**	**6**	**7**

1.5 IEQ – Partie 1

Veuillez répondre aux questions suivantes en choisissant le numéro correspondant. Plus particulièrement, rappelez-vous que ces questions vous interrogent sur ce que vous avez ressenti à la fin de cette expérience d'entrainement en simulateur de vol.

Très légèrement	Légèrement	Modérément	Fortement	Très fortement
1	**2**	**3**	**4**	**5**

Durant la simulation					
Dans quelle mesure la simulation a-t-elle capté votre attention?	**1**	**2**	**3**	**4**	**5**
Dans quelle mesure vous êtes-vous senti(e) concentré(e) en jouant à la simulation?	**1**	**2**	**3**	**4**	**5**
Combien d'effort avez-vous consacré à la simulation?	**1**	**2**	**3**	**4**	**5**
Dans quelle mesure pensez-vous avoir donné le meilleur de vous-mêmes?	**1**	**2**	**3**	**4**	**5**
Dans quelle mesure avez-vous perdu la notion du temps?	**1**	**2**	**3**	**4**	**5**
Dans quelle mesure vous êtes-vous senti dans le monde réel lors de votre participation à la simulation?	**1**	**2**	**3**	**4**	**5**
Dans quelle mesure avez-vous oublié vos préoccupations de la vie quotidienne?	**1**	**2**	**3**	**4**	**5**
Dans quelle mesure étiez-vous conscient(e) de vous-même dans votre environnement?	**1**	**2**	**3**	**4**	**5**

1.5 IEQ – Partie 2

Veuillez répondre aux questions suivantes en choisissant le numéro correspondant. Plus particulièrement, rappelez-vous que ces questions vous interrogent sur ce que vous avez ressenti à la fin de cette expérience d'entrainement en simulateur de vol.

Très légèrement	Légèrement	Modérément	Fortement	Très fortement
1	**2**	**3**	**4**	**5**

Durant la simulation					
Dans quelle mesure avez-vous remarqué des évènements se déroulant autour de vous?	**1**	**2**	**3**	**4**	**5**
Avez-vous ressenti le besoin à un moment d'arrêter la simulation pour regarder ce qui se passait autour de vous?	**1**	**2**	**3**	**4**	**5**
Dans quelle mesure avez-vous eu le sentiment d'interagir avec l'environnement de la simulation?	**1**	**2**	**3**	**4**	**5**
Dans quelle mesure vous êtes-vous senti(e) séparé(e) du monde réel?	**1**	**2**	**3**	**4**	**5**
Dans quelle mesure avez-vous senti que la simulation était quelque chose que vous viviez, plutôt que d'être simplement quelque chose que vous faisiez?	**1**	**2**	**3**	**4**	**5**
Dans quelle mesure votre sentiment d'être dans l'environnement de la simulation était-il plus fort que votre sentiment d'être dans le monde réel?	**1**	**2**	**3**	**4**	**5**
À un moment donné, vous êtes-vous retrouvé(e) tellement impliqué(e) dans la simulation que vous ne saviez même pas que vous utilisiez des commandes?	**1**	**2**	**3**	**4**	**5**
Dans quelle mesure avez-vous eu l'impression de bouger dans la simulation selon votre propre volonté?	**1**	**2**	**3**	**4**	**5**

1.5 IEQ – Partie 3

Veuillez répondre aux questions suivantes en choisissant le numéro correspondant. Plus particulièrement, rappelez-vous que ces questions vous interrogent sur ce que vous avez ressenti à la fin de cette expérience d'entrainement en simulateur de vol.

Très légèrement	Légèrement	Modérément	Fortement	Très fortement
1	**2**	**3**	**4**	**5**

Durant la simulation					
Dans quelle mesure avez-vous trouvé la simulation difficile?	**1**	**2**	**3**	**4**	**5**
Y a-t-il eu des moments pendant la simulation où vous vouliez simplement abandonner?	**1**	**2**	**3**	**4**	**5**
Dans quelle mesure vous êtes-vous senti(e) motivé(e) pendant la simulation?	**1**	**2**	**3**	**4**	**5**
Dans quelle mesure avez-vous trouvé la simulation facile?	**1**	**2**	**3**	**4**	**5**
Dans quelle mesure avez-vous eu l'impression de progresser vers la fin du la simulation?	**1**	**2**	**3**	**4**	**5**
Dans quelle mesure pensez-vous avoir performé dans la simulation?	**1**	**2**	**3**	**4**	**5**
Dans quelle mesure vous êtes-vous senti(e) émotionnellement attaché(e) à la simulation?	**1**	**2**	**3**	**4**	**5**
Dans quelle mesure étiez-vous intéressé(e) à voir comment les évènements de la simulation évolueraient?	**1**	**2**	**3**	**4**	**5**

1.5 IEQ – Partie 4

Veuillez répondre aux questions suivantes en choisissant le numéro correspondant. Plus particulièrement, rappelez-vous que ces questions vous interrogent sur ce que vous avez ressenti à la fin de cette expérience d'entrainement en simulateur de vol.

Très légèrement	Légèrement	Modérément	Fortement	Très fortement
1	**2**	**3**	**4**	**5**

Durant la simulation					
Dans quelle mesure vouliez-vous “gagner” dans la simulation?	**1**	**2**	**3**	**4**	**5**
Étiez-vous en suspens quant à savoir si vous "gagneriez ou perdriez" dans la simulation?	**1**	**2**	**3**	**4**	**5**
À un moment donné, vous êtes-vous retrouvé(e) tellement impliqué(e) que vous vouliez parler directement à la simulation?	**1**	**2**	**3**	**4**	**5**
Dans quelle mesure avez-vous apprécié le graphisme et les images de la simulation?	**1**	**2**	**3**	**4**	**5**
À quel point diriez-vous que vous avez apprécié la simulation?	**1**	**2**	**3**	**4**	**5**
Lorsque vous avez été interrompu(e), avez-vous été déçu(e) que la simulation soit terminée?	**1**	**2**	**3**	**4**	**5**
Aimeriez-vous réutiliser ce simulateur?	**1**	**2**	**3**	**4**	**5**