

Université de Montréal

**Analyse visuelle et cérébrale de l'état cognitif d'un
apprenant**

Par
Asma Ben Khedher

Département d'informatique et de recherche opérationnelle
Faculté des arts et des sciences

Thèse présentée
en vue de l'obtention du grade de Philosophiae Doctor (Ph.D.)
en Informatique

Février, 2019

© Asma Ben Khedher, 2019

Résumé

Un état cognitif peut se définir comme étant l'ensemble des processus cognitifs inférieurs (par exemple : perception et attention) et supérieurs (par exemple : prise de décision et raisonnement), nécessitant de la part de l'être humain toutes ses capacités mentales en vue d'utiliser des connaissances existantes pour résoudre un problème donné ou bien d'établir de nouvelles connaissances. Dans ce contexte, une attention particulière est portée par les environnements d'apprentissage informatisés sur le suivi et l'analyse des réactions émotionnelles de l'apprenant lors de l'activité d'apprentissage. En effet, les émotions conditionnent l'état mental de l'apprenant qui a un impact direct sur ses capacités cognitives tel que le raisonnement, la prise de décision, la mémorisation, etc. Dans ce contexte, l'objectif est d'améliorer les capacités cognitives de l'apprenant en identifiant et corrigeant les états mentaux défavorables à l'apprentissage en vue d'optimiser les performances des apprenants.

Dans cette thèse, nous visons en particulier à examiner le raisonnement en tant que processus cognitif complexe de haut niveau. Notre objectif est double : en premier lieu, nous cherchons à évaluer le processus de raisonnement des étudiants novices en médecine à travers leur comportement visuel et en deuxième lieu, nous cherchons à analyser leur état mental quand ils raisonnent afin de détecter des indicateurs visuels et cérébraux permettant d'améliorer l'expérience d'apprentissage. Plus précisément, notre premier objectif a été d'utiliser les mouvements des yeux de l'apprenant pour évaluer son processus de raisonnement lors d'interactions avec des jeux sérieux éducatifs. Pour ce faire, nous avons analysé deux types de mesures oculaires à savoir : des mesures statiques et des mesures dynamiques. Dans un premier temps, nous avons étudié la possibilité d'identifier automatiquement deux classes d'apprenants à partir des différentes mesures statiques, à travers l'entraînement d'algorithmes d'apprentissage machine. Ensuite, en utilisant les mesures dynamiques avec un algorithme d'alignement de séquences issu de la bio-informatique, nous avons évalué la séquence logique visuelle suivie par l'apprenant en cours de raisonnement pour vérifier s'il est en train de suivre le bon processus de raisonnement ou non.

Notre deuxième objectif a été de suivre l'évolution de l'état mental d'engagement d'un apprenant à partir de son activité cérébrale et aussi d'évaluer la relation entre l'engagement et

les performances d'apprentissage. Pour cela, une étude a été réalisée où nous avons analysé la distribution de l'indice d'engagement de l'apprenant à travers tout d'abord les différentes phases de résolution du problème donné et deuxièmement, à travers les différentes régions qui composent l'interface de l'environnement. L'activité cérébrale de chaque participant a été mesurée tout au long de l'interaction avec l'environnement. Ensuite, à partir des signaux obtenus, un indice d'engagement a été calculé en se basant sur les trois bandes de fréquences α , β et θ .

Enfin, notre troisième objectif a été de proposer une approche multimodale à base de deux senseurs physiologiques pour permettre une analyse conjointe du comportement visuel et cérébral de l'apprenant. Nous avons à cette fin enregistré les mouvements des yeux et l'activité cérébrale de l'apprenant afin d'évaluer son processus de raisonnement durant la résolution de différents exercices cognitifs. Plus précisément, nous visons à déterminer quels sont les indicateurs clés de performances à travers un raisonnement clinique en vue de les utiliser pour améliorer en particulier, les capacités cognitives des apprenants novices et en général, l'expérience d'apprentissage.

Mots-clés : Mouvements des yeux, activité cérébrale, comportement visuel, émotions, processus de raisonnement, analyse de l'état mental d'un apprenant, jeux sérieux, évaluation des performances d'apprentissage, alignement de séquences.

Abstract

A cognitive state can be defined as a set of inferior (e.g. perception and attention) and superior (e.g. perception and attention) cognitive processes, requiring the human being to have all of his¹ mental abilities in an effort to use existing knowledge to solve a given problem or to establish new knowledge. In this context, a particular attention is paid by computer-based learning environments to monitor and assess learner's emotional reactions during a learning activity. In fact, emotions govern the learner's mental state that has in turn a direct impact on his cognitive abilities such as reasoning, decision-making, memory, etc. In this context, the objective is to improve the cognitive abilities of the learner by identifying and redressing the mental states that are unfavorable to learning in order to optimize the learners' performances.

In this thesis, we aim in particular to examine the reasoning as a high-level cognitive process. Our goal is two-fold: first, we seek to evaluate the reasoning process of novice medical students through their visual behavior and second, we seek to analyze learners' mental states when reasoning to detect visual and cerebral indicators that can improve learning outcomes. More specifically, our first objective was to use the learner's eye movements to assess his reasoning process while interacting with educational serious games. For this purpose, we have analyzed two types of ocular metrics namely, static metrics and dynamic metrics. First of all, we have studied the feasibility of using static metrics to automatically identify two groups of learners through the training of machine learning algorithms. Then, we have assessed the logical visual sequence followed by the learner when reasoning using dynamic metrics and a sequence alignment method from bio-informatics to see if he/she performed the correct reasoning process or not.

Our second objective was to analyze the evolution of the learner's engagement mental state from his brain activity and to assess the relationship between engagement and learning performance. An experimental study was conducted where we analyzed the distribution of the learner engagement index through first, the different phases of the problem-solving task and

¹ The use of the masculine gender (his or he) in this manuscript is employed solely to facilitate reading, it refers also to feminine gender (her or she).

second, through the different regions of the environment interface. The cerebral activity of each participant was recorded during the whole game interaction. Then, from the obtained signals, an engagement index was computed based on the three frequency bands α , β et θ .

Finally, our third objective was to propose a multimodal approach based on two physiological sensors to provide a joint analysis of the learner's visual and cerebral behaviors. To this end, we recorded eye movements and brain activity of the learner to assess his reasoning process during the resolution of different cognitive tasks. More precisely, we aimed to identify key indicators of reasoning performance in order to use them to improve the cognitive abilities of novice learners in particular, and the learning experience in general.

Keywords : Eye movements, cerebral activity, visual behavior, emotions, reasoning process, learner's mental state analysis, serious games, learning performance assessment, sequence alignment.

Table des matières

Résumé.....	1
Abstract.....	3
Liste des tableaux.....	10
Liste des figures	11
Liste des sigles	13
Remerciements.....	16
Chapitre 1.....	17
Introduction.....	17
1.1 Contexte général	17
1.2 Motivations	19
1.3 Objectifs de recherche.....	22
1.4 Organisation du document	25
Chapitre 2.....	28
Les environnements d'apprentissage	28
2.1 Introduction.....	28
2.2 Historique et définition	29
2.2.1 Enseignement Assisté par Ordinateur (EAO).....	29
2.2.2 Systèmes Tutoriels Intelligents	30
2.3 Intégration des émotions dans les STI	31
2.3.1 Émotions : définition.....	31
2.3.2 Systèmes Tutoriels Émotionnellement Intelligents	32
2.4 Jeux sérieux.....	33
Chapitre 3.....	38
Étude de l'état mental d'un apprenant et analyse du processus de raisonnement	38

3.1 Introduction.....	38
3.2 Étude de l'état mental	39
3.2.1 Engagement mental.....	39
3.2.2 Charge mentale de travail	40
3.3 Approches de représentation.....	41
3.3.1 Les approches quantitatives	41
3.3.2 Les approches qualitatives	42
3.3.3 Les senseurs physiologiques	44
3.4 Analyse du raisonnement.....	52
3.4.1 Raisonnement : définition.....	52
3.4.2 Types de raisonnement	53
3.4.3 Modélisation du raisonnement.....	56
3.4.4 Étude de l'existant.....	58
3.5 Conclusion	60
Chapitre 4.....	64
Analyse des mouvements des yeux d'un apprenant lors d'une activité de résolution de problème	64
Predicting User Learning Performances from Eye Movements during Interaction with a Serious Game.....	68
Abstract.....	68
4.1 Introduction.....	68
4.2 Previous Work	69
4.3 Crystal Island	70
4.4 Experimental Methodology	71
4.4.1 Apparatus	71
4.4.2 Participants.....	72
4.4.3 Protocol	72
4.4.4 Eye Movement Analysis	73
4.5 Results and Discussion	75
4.5.1 General Statistics	75

4.5.2 Classification.....	77
4.6 Conclusion	78
Acknowledgements.....	79
Chapitre 5.....	81
Étude exploratoire du comportement visuel d'un apprenant lors d'une activité de raisonnement dans un environnement d'apprentissage	81
Static and dynamic eye movement metrics for students' performance assessment.....	86
Abstract.....	86
5.1 Introduction.....	87
5.2 Related work	88
5.3 Experimental design.....	90
5.3.1 Protocol	90
5.3.2 Participants and apparatus.....	91
5.4 Results and discussion	92
5.4.1 Fixation-based metrics and students' performance assessment across the AOIs	93
5.4.2 Performance assessment using scan path metric	96
5.5 Conclusion	100
Acknowledgements.....	100
Chapitre 6.....	102
Une approche d'évaluation du raisonnement de l'apprenant basée sur l'alignement de séquences	102
Local Sequence Alignment for Scan Path Similarity Assessment.....	106
Abstract.....	106
6.1 Introduction.....	107
6.2 State of the art	109
6.3 Sequence alignment algorithm.....	112
6.3.1 Local alignment	113
6.3.2 Smith-Waterman algorithm	114
6.4 Experimental protocol.....	116

6.4.1 Amnesia	116
6.4.2 Medical cases	116
6.4.3 Participants and apparatus.....	117
6.5 Data analysis	117
6.5.1 Students' reasoning process.....	118
6.5.2 Sequence comparison.....	119
6.6 Results and discussion	122
6.6.1 Association of alignment metrics and the number of attempt features.....	122
6.6.2 Performance assessment through the attempts	123
6.7 Conclusion	124
Acknowledgement	125
Chapitre 7.....	127
Analyse de l'état mental de l'apprenant lors d'une activité de raisonnement	127
Tracking Students' Mental Engagement using EEG Signals during an Interaction with a Virtual Learning Environment	131
Abstract.....	131
7.1 Introduction.....	132
7.2 Related work	133
7.3 Experimental design.....	135
7.3.1 Amnesia: an educational video game	135
7.3.2 EEG recordings	136
7.4 Results and discussion	137
7.4.1 Engagement index evolution through the resolution process	137
7.4.2 Engagement index distribution across the AOIs.....	139
7.4.3 Engagement index and performance.....	142
7.5 Conclusion	144
Acknowledgements.....	144
Chapitre 8.....	146

Une approche multimodale à base de senseurs physiologiques pour une analyse visuelle et cérébrale de l'état d'un apprenant durant un processus de raisonnement.....	146
Assessing Students' Clinical Reasoning using Gaze and EEG features.....	150
Abstract.....	150
8.1 Introduction.....	150
8.2 Previous work	151
8.3 Experimental design and methodology.....	152
8.3.1 Visual behavior	153
8.3.2 Brain activity.....	155
8.4 Results and discussion	157
8.4.1 Visual behavior	157
8.4.2 Brain activity.....	158
8.4.3 Relationship between Gaze and EEG Data.....	159
8.5 Conclusion	160
Acknowledgments.....	160
Chapitre 9.....	162
Conclusion	162
9.1 Contributions.....	162
9.2 Travaux futurs.....	165
Bibliographie.....	169
Publications.....	ii
Articles de revue	ii
Articles de conférence.....	ii
Articles de workshops.....	iii

Liste des tableaux

Table 4.1 Participants' statistics.	72
Table 4.2 Description of eye movement measures.	75
Table 4.3 Correlation results.	76
Table 5.1 Participants' statistics.	91
Table 5.2 Descriptive statistics for the fixation duration metric across all the AOIs, Means (Standard Deviations). ANOVAs are reported in italic ($p < 0.001$).	94
Table 5.3 MANOVAs' results summary between the symptoms' fixation duration and the diagnosis success/failure.	95
Table 5.4 Descriptive statistics for the response attempts in each medical case.	98
Table 6.1 Bivariate correlational results.	123
Table 7.1 Descriptive statistics for the medical case resolution.	139
Table 7.2 Descriptive statistics for the engagement index per medical case, Means (Standard Deviation).	140
Table 7.3 ANOVAs' results summary.	143

Liste des figures

Figure 2.1 Architecture d'un STI (Burns and Capps 1988).....	30
Figure 2.2 La différence entre émotion, affect et humeur (Hume 2012).....	32
Figure 2.3 Capture d'écran du jeu Crystal Island	34
Figure 2.4 Capture d'écran du jeu Virtual U	35
Figure 2.5 Exemple d'exercice logique	36
Figure 2.6 Interface de résolution des cas médicaux dans le jeu Amnesia.....	36
Figure 3.1 Exemples de senseurs physiologiques [a] capteur de pression sanguine [b] bracelet Q-sensor	45
Figure 3.2 Exemple de fréquences du signal EEG	46
Figure 3.3 Dispositif matériel du pack Emotiv Epoc: [a] le casque, [b] le récepteur USB et [c] le kit d'hydratation des senseurs	47
Figure 3.4 Positionnement des électrodes.....	48
Figure 3.5 Exemples de traceurs de regard: [a] SMI glasses [b] Dikablis glasses [c] EyeLink-1000[d] Tobii Tx-300. Les deux premiers sont mobiles et les deux derniers sont fixes.	50
Figure 3.6 Exemples de fixations et de saccades.....	51
Figure 3.7 Description du processus de raisonnement hypothético-déductif (Nendaz et al. 2005)	55
Figure 3.8 Un exemple de protocole, de modèle de résolution et d'une correspondance entre les deux. Les chiffres dans des carrées représentent l'ordre dans lequel l'étudiant a regardé les éléments de l'équation.	58
Figure 4.1 Crystal Island serious game.....	71
Figure 4.2 Example of a book with 3 areas of interest: each one represents respectively a question in the quiz.	74
Figure 5.1 The six areas of interest identified in all the medical cases.	92
Figure 5.2 A screenshot of one medical case. The red colour represents the most fixated areas of interest within the screen, followed by yellow and then green for the areas with less fixation intensity.....	93
Figure 5.3 Comparison of alignment metrics for the three different attempts. (a) Score (b) Gap (c) Mismatch	99

Figure 6.1 Editing operations for aligning two sequences.....	113
Figure 6.2 An alignment of two sequences with the elementary scores for each two characters.	113
Figure 6.3 Similarity matrix using the local alignment (the red arrows represent the optimal alignment path).	115
Figure 6.4 An example of scan path over the six AOIs.	118
Figure 6.5 The hypothetico-deductive reasoning process.....	120
Figure 6.6 Two examples of sequence alignment with the corresponding similarity matrices. (a) A perfect match between the sequences. (b) An alignment with two errors “ ” means a match and “—” means a gap.	121
Figure 6.7 Comparison of alignment metrics for the three different attempts. (a) Score (b) Gap (c) Mismatch.	124
Figure 7.1 EEG engagement evolution in each medical case.	138
Figure 7.2 Distribution of students’ engagement index across the different phases of the resolution process.....	139
Figure 8.1 Screenshot of Amnesia with the AOI. The red color represents the most fixated areas, followed by yellow and then green with lower intensity.	155
Figure 8.2 Screenshot from the acquisition module used for the EEG data	156

Liste des sigles

ADN : Acide Désoxyribonucléique
AET : Average Exposure Time
ANOVA : Analysis Of Variance
AOI : Area Of Interest
BVP : Blood Volume Pressure
CMS: Common Mode Sense
DRL: Drive Right Leg
EAO : Enseignement Assisté par Ordinateur
EIAO : Enseignement Intelligemment Assisté par Ordinateur
EEG : ElectroEncephaloGram
EIAH : Environnements Informatique pour l'Apprentissage Humain
EIAO : Enseignement Intelligemment Assisté par Ordinateur
ET : Eye Tracker
FFT : Fast Fourier Transform
GSR : Galvanic Skin Response
HERON : Higher Educational Research ON tutoring systems
IHM : Interaction Homme-Machine
JS : Jeux Sérieux
M : Mean
MANOVA : Multivariate Analysis Of VAriance
MFD : Mean Fixation Duration
NASA-TLX : Nasa Task Load Index
NbAtmpDiag : Number of Attempts in the Diagnostic
NbAtmpTreat : Number of Attempts in the Treatment
NbGap: Number of Gaps
NbMa : Number of Matches
NbMis : Number of Mismatches
NF : Number of Fixations
NRe : Number of Revisits

RF : Random Forest

SD : Standard Deviation

STEI : Systèmes Tutoriels Émotionnellement Intelligents

STI : Systèmes Tutoriels Intelligents

TTFF : Time To First Fixation

À mes très chers parents

Remerciements

Je voudrais exprimer mes remerciements à mon directeur de recherche, le professeur Claude Frasson pour les conseils avisés et précieux qu'il m'a fournis durant mes travaux, pour sa supervision et son expertise qui m'ont beaucoup aidée lors de mes recherches.

Je remercie le SSHRC (*Social Science and Human Research Council*) à travers le projet LEADS (*Learning Environment Across Disciplines*) et le NSERC (*National Science and Engineering Research Council*) pour avoir supporté financièrement mon projet.

J'ai eu la chance de rencontrer et de collaborer avec de nombreuses personnes tout au long de cette thèse qui m'ont aidée de près ou de loin dans l'ensemble de mes travaux notamment mes deux collègues de recherche Hamdi et Maher ainsi que le Docteur Issam Tanoubi, anesthésiste à l'hôpital Maisonneuve-Rosemont et directeur de la recherche en éducation médicale basée sur la simulation au centre de simulation médicale de l'Université de Montréal.

Je remercie tout particulièrement Imène Jraïdi, chercheuse associée au sein du laboratoire HERON à l'Université de Montréal, qui a assuré une assistance continue et utile pour la réalisation des objectifs de cette thèse. Grâce aux longues discussions et réunions de travail que nous avons fréquemment effectuées, ses suggestions et conseils m'ont beaucoup aidée dans la rédaction des articles scientifiques ainsi que dans l'amélioration de la qualité de ma thèse.

Finalement, j'exprime ma gratitude à mes parents qui ont toujours cru en moi et m'ont poussée à montrer le meilleur de moi-même. Je remercie également mon cher mari pour sa patience, son amour et ses encouragements tout au long de ce parcours académique. Une petite pensée aussi à mon fils adoré qui j'espère sera fier de sa maman.

Chapitre 1.

Introduction

1.1 Contexte général

Cette thèse se situe dans le cadre des travaux de recherche menés par le laboratoire HERON (Higher Educational Research ON tutoring systems). Ce groupe de recherche est multidisciplinaire, incluant l'intelligence artificielle, la psychologie cognitive, la neuroscience et leurs applications dans la conception et le développement d'environnements d'apprentissage tels que les Systèmes Tutoriels Intelligents (STI) et les jeux sérieux (JS) intelligents. Ces environnements basés sur l'Interaction Homme-Machine (IHM) ont connu un grand essor cette dernière décennie dans le milieu académique traditionnel. En effet, plusieurs écoles et universités de nos jours offrent un apprentissage hybride qui combine les deux types d'enseignement à savoir l'enseignement à base de cours magistraux en classe et l'enseignement assisté par ordinateurs.

Les jeux sérieux se distinguent des autres environnements d'apprentissage par leur aspect ludique. En effet, les JS ont cette capacité à offrir à l'utilisateur d'un côté un aspect éducatif dont le but principal est l'apprentissage et d'un autre côté un aspect attractif et plus agréable qui combine différents modes de présentations (audio, image, texte, etc.). Cette dernière caractéristique permet de garder l'apprenant concentré en l'incitant tout naturellement à être attentif et intéressé par la tâche qu'il entreprend. De plus, dans la plupart des jeux vidéo, il y a le concept de primes où le joueur est récompensé pour son efficacité à travers des points en bonus. Cet aspect favorise la motivation et l'estime de soi chez l'apprenant le poussant ainsi à se dépasser davantage et montrer le meilleur de lui-même.

Des études récentes montrent que le processus d'apprentissage est fortement lié aux émotions ressenties par l'apprenant (Clore and Huntsinger 2007; Loewenstein and Lerner 2003; Mega et al. 2014). Ces émotions conditionnent en fait l'état mental de l'apprenant au moment d'exécuter un raisonnement cognitif. L'état mental est ainsi la capacité d'un apprenant à acquérir ou non de nouvelles connaissances. Il a un impact direct sur tous les processus cognitifs tel que l'attention, le raisonnement et la prise de décision. Certains travaux de recherche soulignent qu'un état mental négatif est défavorable à l'apprentissage (Acee et al. 2010; Olafson and Ferraro 2001; Ranellucci et al. 2015). Par exemple, si l'apprenant est dans un état de confusion et d'anxiété parce qu'il n'arrive pas à assimiler le contenu pédagogique présenté malgré les efforts entretenus, ceci peut le mener à un blocage pouvant affecter considérablement ses résultats. À l'opposé, si le matériel d'apprentissage est facile, l'étudiant peut perdre sa motivation d'apprendre et ainsi se désintéresser du cours et s'intéresser à d'autres activités non reliées à l'apprentissage. D'autres travaux au contraire, ont prouvé qu'un état mental négatif a tendance à avoir des effets positifs variables sur l'apprentissage (Pekrun et al. 2012; Zeidner 2007). Par exemple Pekrun et al. (2012) l'ont associé à une augmentation de la motivation extrinsèque chez l'apprenant qui le pousse à investir plus d'effort pour éviter les échecs. Lane et ses collègues (2005) ont démontré qu'un état mental négatif peut améliorer les performances des apprenants, la défaite les pousse ainsi à fournir davantage d'efforts pour réussir (Lane et al. 2005).

Dans cette thèse, nous n'avons pas mesuré directement les émotions mais nous sommes intéressés aux états mentaux générés par ces dernières, en particulier deux états mentaux à savoir l'engagement et la charge mentale de travail que nous définirons en détail dans le troisième chapitre. Dans ce contexte que les chercheurs ces dernières décennies ont voué un intérêt grandissant à l'étude de l'état mental vu cette interdépendance qui existe entre état mental et processus cognitifs et leur impact considérable sur les performances d'apprentissage. En effet, un état mental positif stimule positivement les processus cognitifs de l'apprenant en l'incitant à fournir un effort pour chercher dans sa mémoire les connaissances déjà acquises en vue de construire de nouvelles connaissances. Par exemple, si un apprenant est engagé ceci favoriserait son raisonnement et les décisions positives à prendre afin de résoudre correctement un problème donné. À l'opposé, si un apprenant est désengagé ou surchargé mentalement il ne sera plus

impliqué dans le processus de résolution de problème ni de suivre le bon processus de raisonnement. Par conséquent, nous pouvons constater que l'état mental a un impact sur les processus cognitifs de l'individu qui à leur tour ont tendance à favoriser/défavoriser l'apprentissage. C'est pour cette raison qu'il est nécessaire d'une part, de suivre l'état mental de l'apprenant afin de détecter les états mentaux négatifs susceptibles d'affecter les processus cognitifs de l'apprenant et d'entraver le bon déroulement de son apprentissage. D'une autre part, il est nécessaire de suivre le processus de construction de la connaissance en étudiant les processus cognitifs qui ont contribué à cela. L'objectif est de pouvoir optimiser les capacités cognitives de l'apprenant en corrigeant ses états mentaux en vue d'améliorer ses performances et enrichir son expérience d'apprentissage.

1.2 Motivations

Le sujet de cette thèse s'articule autour de deux thématiques : (1) l'analyse du processus cognitif de raisonnement en se basant sur le comportement visuel de l'apprenant et (2) le suivi de l'état mental de l'apprenant.

C'est à travers l'approche cognitiviste que les chercheurs se sont concentrés sur comment l'information est traitée dans le système cognitif humain (Weil-Barais and Dubois 1994). L'apport du cognitivisme par rapport aux autres courants qui l'ont précédé est qu'il s'intéresse au fonctionnement de l'esprit humain et aux mécanismes mentaux et processus cognitifs mis en évidence lors de l'exécution d'une tâche cognitive particulière. Il existe deux types de processus cognitifs à savoir, les processus inférieurs tels que la perception, l'attention et la mémoire et les processus supérieurs tels que la prise de décision, la résolution de problèmes et le raisonnement (Delécraz 2017).

Dans ce travail de recherche, nous nous focaliserons tout particulièrement sur le raisonnement comme un processus cognitif complexe présent dans l'apprentissage qui permet à partir de données initiales d'inférer une nouvelle conclusion ou de vérifier la validité de faits existants. Le raisonnement peut être vu comme un système de boîte noire dont nous ignorons le fonctionnement. En effet, il n'existe pas de schéma formel préétabli pour vérifier si un raisonnement est logique ou non vu que c'est un processus relié aux pensées propres de l'individu. Rares sont les travaux qui ont étudié le raisonnement comme processus cognitif à

part entière. Toutefois, les travaux qui l'ont fait, l'ont analysé uniquement en surface pas en profondeur. En effet, ces travaux se sont intéressés à l'étude des mesures oculaires qui sont en relation indirecte avec le raisonnement sans réellement se focaliser sur l'étude de la logique elle-même du processus de raisonnement ni comment une personne raisonne (Renkewitz and Jahn 2012; Scholz et al. 2015; Vachon and Tremblay 2014). Par exemple, Muldner et al. (2009) ont mesuré la taille de la pupille des participants durant l'activité de raisonnement (Muldner et al. 2009). D'autres ont étudié le temps de fixation suggérant que le temps passé à fixer l'information recherchée peut refléter un effort de traitement élevé (Ball et al. 2006). D'où la nécessité d'une analyse interne qui consiste à suivre d'une manière détaillée et approfondie la logique derrière le raisonnement d'un individu; c'est-à-dire le cheminement successif de l'ensemble des étapes qui permettent de transformer des faits en conclusion valide.

Diverses sources d'informations ont été utilisées dans la littérature pour tenter de modéliser et analyser le processus de raisonnement. Des mesures indirectes comme les questionnaires d'auto-évaluation, les expressions faciales, l'électroencéphalographie, la taille de la pupille, etc. (Ball et al. 2006; Muldner et al. 2009). Cependant, ces mesures nous renseignent sur l'état mental de l'apprenant lors du raisonnement mais pas comment il a raisonné. Pareillement l'évaluation des performances d'apprentissage peut être un bon indicateur mais qui nous renseigne juste sur la réussite ou l'échec d'un processus de raisonnement. Nonobstant l'efficacité de ces méthodes, les informations qu'elles fournissent ne sont pas toujours suffisantes pour identifier les situations où l'apprenant a besoin d'aide durant son raisonnement. Autrement dit, est ce que ces méthodes sont capables de détecter les situations où l'apprenant est égaré ou en train de suivre un raisonnement erroné ?

Dans cette même thématique, plusieurs chercheurs en sciences cognitives et en IHM ont souligné l'importance de l'utilisation des mouvements des yeux comme source d'information pour évaluer les performances cognitives des individus. En effet, une relation étroite existe entre les mouvements des yeux et la cognition humaine; les mesures de suivi oculaire à base de fixations et de saccades sont des indicateurs de l'effort mental fourni par l'individu durant le processus visuel (S.-C. Chen et al. 2014; Kiili et al. 2014; Tsai et al. 2012). Par exemple, la durée de fixation est supposée refléter le temps de traitement et la difficulté de compréhension. Ce qui nous a poussés dans notre travail de recherche à utiliser les mouvements des yeux afin

de suivre et d'évaluer le processus de raisonnement des apprenants lors d'une tâche de résolution de problème.

La deuxième thématique de notre recherche porte sur l'analyse de l'état mental de l'apprenant. En effet, l'un des principaux objectifs des environnements d'apprentissage est de pouvoir suivre l'évolution de l'état mental de l'apprenant en vue d'essayer de le maintenir le plus longtemps possible motivé et attentif. En revanche différentes situations peuvent survenir durant l'activité d'apprentissage qui peuvent nuire d'une façon directe aux performances de l'apprenant telles que le désintérêt, le désengagement, l'ennui, la frustration, etc. Ainsi il est primordial de suivre constamment l'état mental de l'apprenant afin de pouvoir détecter le plus tôt possible une éventuelle tendance négative ce qui permet d'intervenir et offrir l'aide nécessaire au besoin. De ce fait, les chercheurs ont eu recours à une variété d'outils de détection. Le moyen le plus facile et le plus répandu était le fait de demander directement à l'apprenant de décrire son état mental en utilisant des questionnaires d'auto-évaluation. Toutefois cette méthode de détection ne fournit pas vraiment des réponses assez fiables, d'où le recours à d'autres outils de détection plus objectifs qui se basent sur le comportement externe de l'apprenant (posture, expressions faciales, etc.). Par exemple, un froncement des sourcils est synonyme de colère ou encore une intensité forte exercée sur la souris peut représenter un état de frustration. Ce type de senseurs ne permet cependant pas d'obtenir des informations précises sur l'état mental de l'apprenant dans tous les cas. En effet, l'apparence externe d'un individu ne reflète pas toujours son état réel ce qui influe sur la fiabilité des informations recueillies.

Dans cette même thématique, nous avons constaté qu'il est nécessaire d'utiliser des moyens de détection plus efficaces qui sont capables de détecter l'état mental d'un apprenant à partir de comportements non observables liés à ses comportements physiologiques. En effet, les senseurs physiologiques et tout particulièrement les capteurs cérébraux ont pu remédier aux limites rencontrées avec les autres senseurs et ainsi révolutionner les environnements d'apprentissage. Par ailleurs, d'autres travaux de recherche ont accordé une attention particulière aux approches multimodales combinant plusieurs outils de détection afin de fournir un modèle d'évaluation de l'état mental plus robuste (Harley et al. 2015; Rodrigue et al. 2015; Slanzi et al. 2017). En effet, l'utilisation de différents capteurs apporte une valeur ajoutée et peut aider à pallier les problèmes issus de l'utilisation d'une seule modalité de détection.

Ceci nous amène à penser qu'une analyse combinée du comportement visuel et de l'état mental de l'apprenant serait souhaitable afin de faire face aux limites de chacun et offrir ainsi une évaluation plus efficace du processus de raisonnement de l'apprenant.

1.3 Objectifs de recherche

Notre objectif dans cette thèse est de surveiller et d'analyser le raisonnement d'un apprenant lors de la résolution de problèmes dans deux types d'environnements d'apprentissage (jeux sérieux éducatifs) en étudiant son état mental et son comportement visuel. Nos travaux de recherche visent essentiellement à déterminer quels sont les indicateurs cérébraux et visuels qui permettent d'évaluer le processus de raisonnement de l'apprenant afin de les utiliser pour améliorer l'expérience d'apprentissage. Plus précisément, nous proposons de réaliser les trois objectifs suivants :

1. **Proposer une méthode d'évaluation du processus de raisonnement de l'apprenant en utilisant les mouvements des yeux:** notre premier objectif consiste à proposer une nouvelle méthode d'évaluation des performances à travers le comportement visuel de l'apprenant lors de son interaction avec un jeu sérieux incluant différents exercices cognitifs. Cette méthode consiste à utiliser des mesures oculaires statiques et dynamiques pour analyser et évaluer le processus de raisonnement.
2. **Suivre l'évolution de l'état mental d'un apprenant :** notre deuxième objectif consiste à mesurer l'état d'engagement des apprenants à partir de leur activité cérébrale. Plus précisément notre objectif consiste à suivre l'expérience d'interaction de l'apprenant afin de détecter les phases qui sont susceptibles d'affecter son état mental et par conséquent son apprentissage.
3. **Proposer une approche multimodale à base de senseurs physiologiques pour une analyse conjointe du comportement visuel et de l'état mental de l'apprenant :** notre troisième objectif est de proposer une approche multimodale qui permet une analyse visuelle et cérébrale de l'état cognitif d'un apprenant en utilisant les mouvements des yeux et l'activité cérébrale du cerveau. Cette

approche devrait permettre d'établir un modèle unifié pour l'analyse du processus de raisonnement d'un apprenant visant essentiellement à améliorer les capacités cognitives des apprenants novices et à favoriser leurs résultats d'apprentissage.

Deux expérimentations ont été réalisées pour répondre à nos trois objectifs de recherche et à desquelles trois ensembles de données ont été collectés qui ont fait l'objet de quatre études. La première expérimentation (I) a été faite en utilisant un jeu sérieux connu sous le nom de Crystal Island (J. Rowe et al. 2009). Le jeu est un environnement d'apprentissage de la microbiologie qui a été développé à des fins académiques pour les élèves des écoles de l'état de Caroline du Nord aux États-Unis. L'environnement vise à mesurer les connaissances scientifiques des apprenants, en génétique et en microbiologie en leur demandant de résoudre une énigme. Vingt participants du département d'informatique de l'Université de Montréal ont été recrutés pour l'étude durant laquelle leurs mouvements des yeux ont été enregistrés en utilisant le traceur de regard Tobii Tx300. Les données collectées dans cette expérimentation ont fait l'objet d'une seule étude. La deuxième expérimentation (II) a été faite en utilisant un jeu sérieux nommé Amnesia que nous avons développé aussi à des fins académiques (Ben Khedher and Frasson 2016a). Le jeu est un environnement médical destiné aux étudiants en médecine de premier niveau pour évaluer leurs capacités cognitives en résolvant différents cas cliniques. Une quinzaine de participants du département de médecine de l'Université de Montréal ont été recrutés pour l'étude durant laquelle leurs mouvements des yeux et leurs activités cérébrales ont été enregistrés en utilisant respectivement le traceur de regard Tobii Tx300 et le casque sans fil Emotiv EPOC. Les données collectées dans cette expérimentation ont fait l'objet de trois études : la première utilise uniquement les données oculaires, la deuxième uniquement les données du cerveau et la dernière une fusion entre les deux.

Pour répondre à notre premier objectif, nous avons utilisé l'expérimentation (I) pour examiner si les données oculaires sont capables de prédire les performances d'apprentissage. Dans le jeu Crystal Island et pour résoudre l'énigme les apprenants doivent lire des livres scientifiques, faire des analyses en laboratoires ou encore interagir avec d'autres personnages du jeu. Toutefois, dans cette première étude nous nous sommes juste basés sur la partie du jeu où les apprenants devaient lire des livres et répondre à des questionnaires à choix multiples.

Nous avons collecté les mouvements des yeux de chaque apprenant et avons étudié la possibilité d'identifier deux catégories d'apprenants en utilisant seulement des variables telles que la durée de fixation ou encore le nombre de fixations.

Une deuxième étude a été élaborée en utilisant les données de l'expérimentation (II) pour examiner la faisabilité d'utiliser les mouvements des yeux afin d'évaluer le processus de raisonnement d'un apprenant. Nous nous sommes intéressés à l'étude de deux types de mesures : mesures statiques (par exemple : nombre de fixation et durée de fixation) et mesures dynamiques (par exemple : scan path). En effet, nous avons enregistré le comportement visuel de chaque apprenant et avons identifié par la suite la séquence visuelle qui correspond au cheminement logique suivi au cours du raisonnement. Puis, nous avons utilisé un algorithme d'alignement de séquences utilisé en microbiologie pour évaluer si l'apprenant est en train de suivre le bon ou le mauvais chemin de raisonnement. Cet algorithme permet de comparer deux séquences afin de détecter s'il existe des similarités entre elles.

Pour réaliser notre deuxième objectif, une troisième étude dédiée au suivi de l'état mental de l'apprenant a été élaborée en utilisant les données issues de l'expérimentation (II). Plus précisément, nous avons proposé de mesurer l'état d'engagement de l'apprenant durant son interaction avec Amnesia. L'engagement est le degré d'effort mental et de persévérance accompli par un individu pour arriver à son objectif en ayant recours à des stratégies cognitives et métacognitives. Un indice d'engagement mental a été mesuré en utilisant la méthodologie de Pope et al. (1995). Dans un premier temps, l'activité cérébrale des apprenants a été enregistrée en utilisant des capteurs cérébraux appelés électroencéphalogramme (EEG). Ensuite, à partir des données brutes obtenues, l'index d'engagement a été établi en se basant sur les trois bandes de fréquences α , β et θ (Pope et al. 1995).

Enfin pour répondre à notre troisième objectif, nous avons proposé une approche multimodale en utilisant l'activité cérébrale et les mouvements des yeux pour évaluer le processus de raisonnement chez l'apprenant durant l'exécution d'une activité purement cognitive. Plus précisément, nous avons analysé les indicateurs d'état mental et du comportement visuel afin de déterminer quels sont les indicateurs clés du raisonnement en vue de les utiliser pour améliorer l'expérience d'apprentissage des participants. Pour évaluer notre approche, une quatrième étude a été réalisée dans laquelle nous avons synchronisé les données

obtenues de l'expérimentation (II) faite avec le JS éducatif Amnesia. De plus, un index de charge mentale de travail, mesuré en utilisant la méthodologie de Chaouachi et al. (2011), a été calculé à l'aide d'un modèle prédictif préformé et validé (Chaouachi et al. 2011). La charge mentale de travail correspond à l'ensemble des fonctions mentales et cognitives telles que la compréhension, l'effort de concentration, l'attention ou encore le stress, mobilisés par l'individu pour l'accomplissement d'une tâche donnée.

Deux modalités de variables ont été utilisées pour les différentes analyses statistiques à savoir : des variables physiologiques reliées aux mesures de l'état mental et des variables oculaires reliées à l'activité visuelle.

1.4 Organisation du document

Les chapitres 2 et 3 suivants sont consacrés à l'analyse de la littérature. Dans le chapitre 2, nous présentons le cadre de cette thèse à savoir, les environnements d'apprentissage, en décrivant un bref historique de leur évolution jusqu'au développement des systèmes tutoriels intelligents. Par la suite, nous abordons l'intégration des émotions dans les STI donnant ainsi lieu à un nouveau type d'environnement d'apprentissage qui sont les Systèmes Tutoriels Émotionnellement Intelligents (STEI). Vers la fin, nous nous attardons sur un autre type d'environnement qui est les jeux sérieux en présentant particulièrement les JS éducatifs.

Dans le chapitre 3, nous présentons le concept d'état mental en décrivant les différentes approches de modélisation qui existent dans la littérature en mettant l'accent sur l'EEG et les mouvements des yeux. Ensuite, nous présentons le processus de raisonnement, ses types et ses approches de modélisation en faisant une synthèse des travaux existants pour pouvoir positionner nos contributions par rapport à la littérature.

Les chapitres 4, 5, 6, 7 et 8 sont dédiés à nos contributions. Ces chapitres sont présentés sous la forme de cinq articles de recherche : le premier article a été accepté dans la conférence *EdMedia* (World Conference On Educational Media and Technology). Le deuxième article a été publié dans le journal *Smart Learning Environment*. Le troisième article a été publié dans le journal *International Journal of Information and Education Technology*. Une version de cet article a été acceptée dans la conférence *International Conference on Intelligent Computing*. Le

quatrième article a été publié dans le journal *of Intelligent Learning Systems and Applications*. Enfin, le dernier article a été accepté à la conférence *Intelligent Tutoring Systems*.

Dans le chapitre 4, nous présentons une étude exploratoire réalisée dans un environnement d'apprentissage pour examiner si les mouvements des yeux sont capables de prédire deux groupes d'apprenants en se basant sur leurs performances lors d'une tâche de résolution de problèmes. Dans le chapitre 5, nous introduisons deux types de mesures statiques et dynamiques calculées à partir des mouvements des yeux des apprenants pour examiner laquelle de ces métriques a un impact sur les performances d'apprentissage. Le chapitre 6 est dédié à la présentation d'une méthode d'alignement de séquences basée sur le codage de caractères que nous avons utilisé pour évaluer le processus de raisonnement des apprenants. Le chapitre 7 sera dédié à la surveillance de l'état mental d'un apprenant en présentant comment le niveau d'engagement des apprenants évolue durant toutes les phases de résolution de l'exercice cognitif. Le chapitre 8 décrit notre approche multimodale à base d'EEG et traceur de regard capable de fournir des indicateurs cérébraux et visuels pour l'évaluation du processus de raisonnement de l'apprenant.

En conclusion, le chapitre 9 énumère les contributions de cette thèse et présente les limites de nos travaux de recherche ainsi que les travaux futurs à envisager.

Chapitre 2.

Les environnements d'apprentissage

2.1 Introduction

Ce chapitre décrit un bref historique de l'évolution des environnements informatiques d'apprentissage. Nous commencerons par présenter les systèmes d'enseignement assisté par ordinateur, qui représentent la première génération de systèmes éducatifs. Ces systèmes ont rencontré plusieurs problèmes dus à la rigidité de la pédagogie suivie qui privilégiait le contenu éducatif au détriment de l'apprenant. En effet, ils avaient du mal à détecter les erreurs faites par l'apprenant durant son apprentissage pour pouvoir lui fournir une aide individualisée. Avec tous les progrès réalisés ensuite dans le domaine de l'intelligence artificielle, l'apprentissage est désormais axé sur l'apprenant. Ce qui a donné naissance à un nouveau moyen d'apprentissage connu sous le nom de systèmes tutoriels intelligents qui s'adaptent aux besoins des apprenants et leur offrent un enseignement personnalisé selon les besoins de chacun.

Dans la troisième section du chapitre, nous décrirons comment les émotions ont pris place au sein des systèmes tutoriels intelligents en présentant l'intelligence émotionnelle. Nous nous attarderons ensuite dans la dernière section sur la description du concept des jeux sérieux qui constituent un autre moyen d'enseignement offrant une expérience physique et physiologique plaisante pour le sujet. Nous présenterons trois exemples de jeux sérieux, entre autres le jeu *Amnesia* que nous avons développé pour des fins expérimentales et que nous avons utilisé comme environnement de développement principal dans cette thèse.

2.2 Historique et définition

Les Environnements Informatique pour l'Apprentissage Humain (EIAH) sont conçus pour favoriser la construction de connaissances chez les apprenants en les aidant dans la coordination, l'accompagnement et l'évaluation dans des situations d'apprentissage en mobilisant des agents humains (apprenants, enseignants) et artificiels ainsi que des ressources multimédia. Ils peuvent offrir trois types d'interaction avec l'apprenant (Tchounikine 2002). Les EIAH peuvent être utilisés comme outil de présentation de l'information, outil de traitement de l'information qui permet un apprentissage à distance personnalisé ou encore un outil de communication qui permet une interaction homme-machine. Les EIAH sont utilisés dans plusieurs domaines de recherche tels que l'intelligence artificielle pour modéliser les connaissances et le raisonnement de l'apprenant ou encore la psychologie cognitive afin d'étudier le comportement affectif et cognitif de l'apprenant.

2.2.1 Enseignement Assisté par Ordinateur (EAO)

L'utilisation des ordinateurs pour la formation a commencé vers la fin des années 60 et le début des années 70. La première application était un cours de poésie donné sur des terminaux avec des hyperliens. Cependant, ce type d'enseignement a plusieurs limites liées à la rigidité des programmes. En effet, il y a une absence de théories éducatives ainsi que de matériels didactiques. Le contenu est préétabli à l'avance par le concepteur du système et exposé aux apprenants d'une façon statique. Ensuite, vers le début des années 80, est apparu l'EIAO (Enseignement Intelligemment Assisté par Ordinateur) qui a vu le jour grâce aux progrès faits en intelligence artificielle, éducation et psychologie cognitive. L'objectif était de s'inspirer des avantages de ces trois disciplines afin de simuler l'enseignant humain en tant qu'expert pédagogue et expert du domaine.

L'EIAO est centré sur l'apprenant où ce dernier n'est plus un récepteur passif mais un acteur actif dans le processus d'apprentissage. C'est ce qui a donné naissance aux Systèmes Tutoriels Intelligent.

2.2.2 Systèmes Tutoriels Intelligents

Les STI sont des systèmes informatiques qui se basent sur les meilleures méthodes de l'enseignement traditionnel afin de simuler le comportement du tuteur pour offrir un enseignement adapté et une rétroaction personnalisée aux apprenants en tenant compte de leur niveau d'intelligence. Ils visent à assister les apprenants dans leurs processus d'apprentissage et intervenir en cas de besoin; par exemple ils peuvent détecter quand l'étudiant fait une erreur ou tout simplement n'est plus motivé et déclencher ainsi des stratégies d'aide. Un STI est composé de quatre modules principaux comme le montre la figure 2.1. Chaque module est responsable de certaines fonctionnalités précises du STI en étant les 4 interdépendants (Wenger 1987).

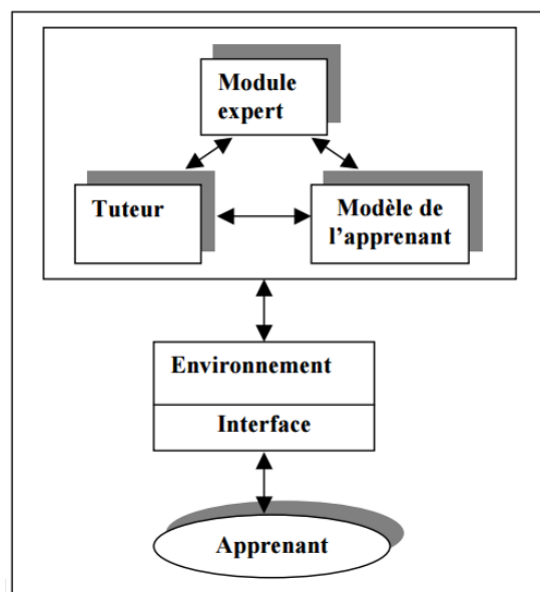


Figure 2.1 Architecture d'un STI (Burns and Capps 1988)

- **Le module expert** : appelé aussi module du domaine ou encore curriculum. Il englobe les connaissances du domaine à enseigner ainsi que les compétences à acquérir relatives à la matière présentée tel que les questions, les explications, etc. Ce module intègre aussi des modèles de représentations relatifs au niveau de connaissances et performances de l'apprenant afin de générer les stratégies d'adaptation nécessaires.
- **Le module apprenant** : appelé aussi module de diagnostic ou modèle de l'apprenant, il contient toutes les informations nécessaires relatives à l'apprenant tel que ses connaissances ainsi que son niveau de progression (que peut faire l'apprenant ? Qu'est

ce qu'il connaît ? Qu'est ce qu'il a déjà appris ?). Ce module est chargé de modéliser et évaluer l'état de connaissances de l'apprenant sur lequel se base l'adaptation du matériel pédagogique.

- **Le module tuteur** : appelé aussi modèle pédagogique ou modèle d'interaction puisqu'il est fortement lié aux modules de l'apprenant et de l'expert. En effet, en se basant sur les informations des deux autres modules, il conçoit son plan d'enseignement. D'un côté, Il interagit avec le module apprenant pour identifier les besoins et les connaissances de l'apprenant et de l'autre côté, avec le module expert pour obtenir le contenu didactique. Ces informations seront utilisées pour appliquer la meilleure stratégie pédagogique d'intervention (quand et comment intervenir) en vue d'améliorer l'apprentissage.
- **Le module interface** : il intègre les moyens nécessaires (texte, clavier, menus, ...) pour faciliter la communication entre l'apprenant et le STI.

2.3 Intégration des émotions dans les STI

2.3.1 Émotions : définition

«Everyone knows what an emotion is until he is asked to give a definition. At that moment, it seems that no one knows» (Fehr and Russell 1984). Cette citation prouve qu'il est difficile de trouver une définition claire et unanime pour le terme *émotion*. De plus, il y a toujours une confusion entre les termes émotion, affect ou humeur. La figure 2.2 présente ces trois terminologies.

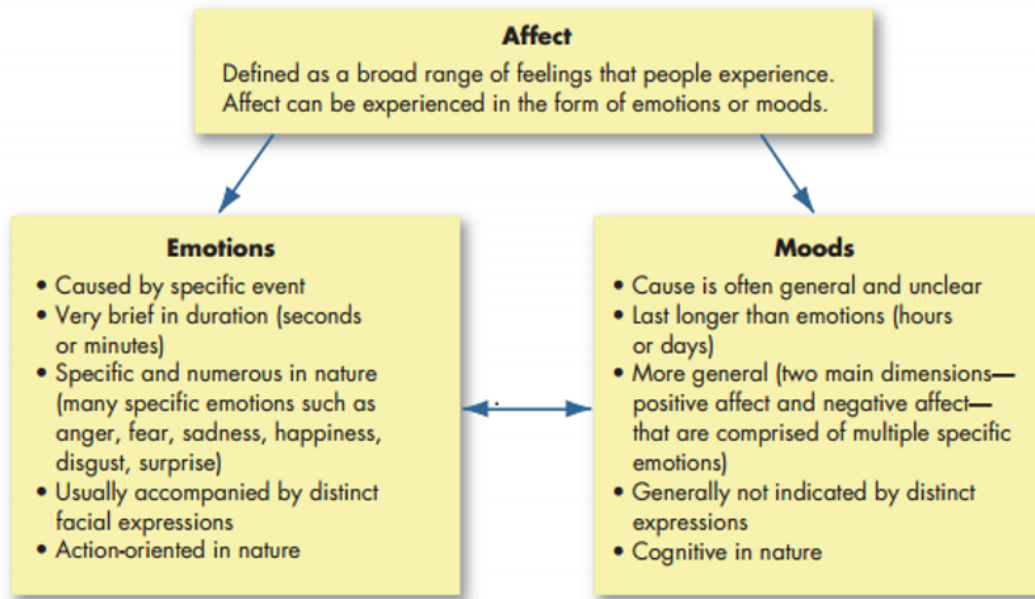


Figure 2.2 La différence entre émotion, affect et humeur (Hume 2012)

2.3.2 Systèmes Tutoriels Émotionnellement Intelligents

Les émotions sont omniprésentes dans la vie quotidienne de l'être humain. Elles sont fortement reliées à ses fonctions cognitives tels que la prise de décision, la mémorisation et la perception (Clare and Huntsinger 2007; Loewenstein and Lerner 2003). Cette association a permis aux émotions d'être intégrées dans les STI surtout après l'émergence des deux courants de recherche; l'intelligence émotionnelle et l'informatique affective. Avant les STI ne prenaient pas en considération l'état émotionnel des individus lors de l'apprentissage humain. En effet, ils n'avaient pas cette aptitude que les tuteurs humains ont pour détecter les émotions d'un apprenant. Par exemple, lors d'une séance de cours classique (face à face avec le tuteur) un tuteur est capable de déchiffrer si un apprenant est désorienté (c'est-à-dire il a du mal à comprendre le contenu du cours) ou encore s'il s'ennuie et il n'est plus réceptif alors il va essayer de l'aider. De ce fait, les STI ont compris l'importance des émotions et leur influence négative ou positive qu'elle soit sur le comportement de l'apprenant. Cette vision a été à l'origine de l'apparition d'une nouvelle génération d'environnements d'apprentissage à savoir les STEI (D'mello et al. 2005; Ochs and Frasson 2004).

Les STEI se distinguent des STI par leur capacité à reconnaître et analyser les émotions de l'apprenant dans un objectif d'adaptation. L'adaptation consiste à appliquer des stratégies de régulation d'émotions dont le but est d'amener l'apprenant à une attitude positive qui sera favorable pour son apprentissage. Si le système détecte que l'apprenant n'est plus intéressé par le contenu proposé, le tuteur peut par exemple lui envoyer des messages de motivation ou encore lui présenter un contenu plus attrayant.

2.4 Jeux sérieux

Les jeux vidéo classiques étaient conçus purement pour un but ludique en utilisant des interfaces et des éléments graphiques qui attirent l'intérêt et l'attention du joueur. Cependant cette attention peut être une arme à double tranchants si le joueur bascule dans l'addiction. En effet, cet aspect virtuel suscite dans beaucoup de joueurs une forte estime de soi et un excès de motivation qui peuvent les conduire petit à petit à se déconnecter du monde réel et de l'environnement qui les entoure. Une autre limite des jeux vidéo est qu'il y en a ceux qui ont un contenu violent qui peut avoir un impact néfaste sur le comportement du joueur. Anderson et Bushman (2001) ont démontré que les joueurs qui jouent à ce genre de jeux ont tendance à avoir des pensées agressives et un comportement antisocial (Anderson and Bushman 2001). Ceci a poussé les industries du jeu à développer un nouveau genre de jeux vidéo appelé les jeux vidéo sérieux « *serious games* » qui offrent une valeur ajoutée aux joueurs grâce à une perception plus optimiste et réfléchie du contenu proposé. Cette nouvelle aire de jeu n'a plus comme objectif principal, uniquement le divertissement mais aussi l'apprentissage, la communication, l'enseignement, l'information, etc.

Une variété de définitions existe pour définir les JS, nous proposons celle de Zyda (2005) qui a défini un jeu sérieux comme «un défi cérébral, joué avec un ordinateur selon les règles spécifiques, qui utilise le divertissement en tant que valeur ajoutée pour la formation et l'entraînement dans les milieux institutionnels ou privés, dans les domaines de l'éducation, de la santé, de la sécurité civile, ainsi qu'à des fins de stratégie de communication» (Zyda 2005). Les jeux sérieux ont été utilisés dans une multitude de domaines notamment l'éducation ce qui a permis aux jeux éducatifs de voir le jour. En effet, dans le temps où il est devenu très difficile de capter l'attention des étudiants et de les garder motivés tout au long du processus

d'apprentissage, les jeux sérieux éducatifs sont un moyen efficace pour y parvenir. Par exemple, au lieu d'enseigner les mathématiques ou l'histoire d'une manière traditionnelle, les étudiants peuvent le faire à travers un jeu vidéo qui combine plaisir et éducation.

Les jeux éducatifs offrent un environnement attractif (son, images et animation) et interactif pour les apprenants. Ils peuvent interagir avec les personnages du jeu et recevoir des feedbacks qui vont les aider à mieux progresser. Ils peuvent aussi être un moyen pour améliorer les aptitudes cognitives des apprenants. En effet, dans le jeu l'apprenant n'est plus un récepteur mais l'acteur principal de son propre apprentissage. De ce fait, il doit utiliser ses capacités mentales et cérébrales afin de prendre les bonnes décisions et résoudre facilement les problèmes rencontrés. Plusieurs jeux éducatifs commerciaux ou académiques existent dans la littérature, voici quelques exemples :

- **Crystal Island** : est un jeu développé dans un cadre académique dont l'idée est dérivée d'un cours standard de microbiologie enseigné dans les écoles de l'État de Caroline du Nord (J. P. Rowe et al. 2009). Il est utilisé par les élèves comme un complément du cours pour renforcer leurs connaissances en sciences. Le jeu présente un mystère scientifique dans le domaine de la génétique et de la microbiologie. Les événements se déroulent dans une île volcanique où les membres d'une équipe de recherche ont été contaminés par une maladie non identifiée. Le joueur doit découvrir la cause de cette maladie en collectant un maximum de preuves et d'indices. Il peut interagir avec d'autres personnages dans le jeu ou encore mener des expériences en laboratoire



Figure 2.3 Capture d'écran du jeu Crystal Island

- **Virtual U** : est un jeu de simulation qui est téléchargeable gratuitement dont le but est d'apprendre la gestion d'une université. Le joueur endosse le rôle d'un chef d'établissement qui a comme tâches d'utiliser les ressources matérielles et personnelles (corps enseignant, étudiants, etc.) mises à sa disposition afin de bien gérer le campus.



Figure 2.4 Capture d'écran du jeu Virtual U

- **Amnesia** : est un jeu sérieux que nous avons développé en utilisant l'environnement de développement Unity. Il est utilisé par des étudiants novices en première année de médecine pour des fins expérimentales. L'objectif du jeu est d'évaluer les capacités cognitives de ces étudiants tels que la prise de décision et la résolution de problème par le biais de cas cliniques qui ont été conçus en collaboration avec un professionnel de la santé et approuvé par un médecin. L'environnement du jeu est un hôpital virtuel où le joueur endosse le rôle d'un médecin qui a été diagnostiqué par erreur comme amnésique. Il doit prouver qu'il est sain et sauf pour pouvoir sortir du piège. Dans un premier temps, l'apprenant doit résoudre certains exercices cognitifs (*p. ex.* mémorisation et tests logiques) comme le montre la figure 2.5. En deuxième lieu, il doit résoudre correctement six cas médicaux où dans chaque cas, il doit trouver le bon diagnostic ainsi que le traitement approprié (voir figure 2.6). Pour ce faire, il doit analyser une série

d'observations y compris les informations démographiques du patient (nom, prénom, âge, etc.), ses antécédents et ses symptômes (Ben Khedher and Frasson 2016a).

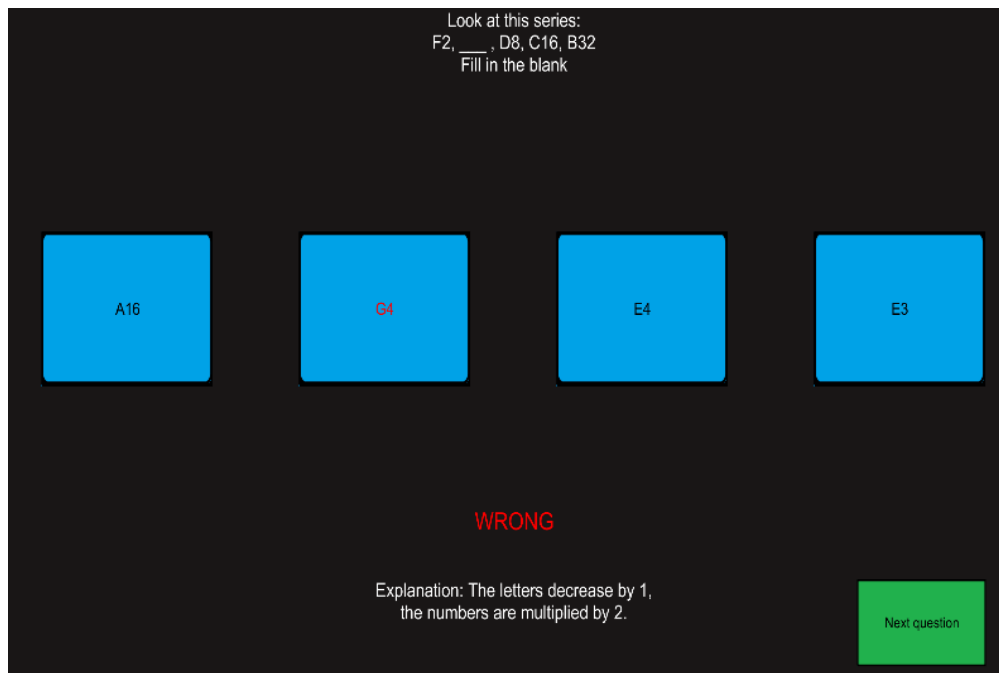


Figure 2.5 Exemple d'exercice logique



Figure 2.6 Interface de résolution des cas médicaux dans le jeu Amnesia

Chapitre 3.

Étude de l'état mental d'un apprenant et analyse du processus de raisonnement

3.1 Introduction

L'objectif des EIAH est de fournir un apprentissage qui prend en compte les différents états mentaux des apprenants durant le processus d'enseignement sachant qu'il existe une relation étroite entre états mentaux et processus cognitifs qui jouent un rôle déterminant dans la réussite ou l'échec de l'apprentissage. Les états mentaux régissent les processus cognitifs de l'individu tel que la prise de décision, le raisonnement ou encore la perception. Par conséquent, il est très important de doter les systèmes d'apprentissage par la capacité de suivre l'état mental de l'apprenant et de détecter les états négatifs afin de permettre aux éducateurs et aux tuteurs de fournir l'assistance adéquate.

Ce chapitre est consacré à l'état de la littérature reliée à nos travaux de recherche. La section 2 sera dédiée à l'analyse de l'état mental de l'apprenant tout en mettant l'accent sur les travaux faits sur l'engagement mental et la charge mentale de travail ainsi que les approches de représentation qui existent dans la littérature. Nous nous attarderons sur l'électroencéphalographie qui est l'outil de mesure que nous avons utilisé dans nos expérimentations pour le suivi de l'état mental. Dans la section 3, nous nous intéresserons à l'analyse du processus de raisonnement en présentant tout d'abord ses différents types et approches de modélisation. Par la suite une petite revue des travaux existants dans la littérature en dégagant les limites de chacun pour pouvoir positionner nos travaux par rapport à l'existant. Finalement, dans la section 4 nous clôturons ce chapitre par un rappel des contributions que nous tenterons de réaliser dans le cadre de cette thèse.

3.2 Étude de l'état mental

Comme nous l'avons mentionné précédemment, l'état mental a une influence directe et considérable sur les processus cognitifs de l'apprenant durant l'activité d'apprentissage. C'est pour cette raison qu'il faut le suivre et l'étudier convenablement pour s'adapter en conséquent. L'engagement, la frustration, l'excitation, la charge mentale et la confusion sont parmi les états mentaux les plus couramment observés et analysés par les chercheurs en particulier dans l'apprentissage. En effet, obtenir une vision précise de l'état mental des apprenants peut permettre aux systèmes d'apprentissage interactifs d'ajuster leur contenu pédagogique ou encore de formuler des stratégies d'aide appropriées afin d'améliorer les résultats des apprenants. Dans cette thèse, nous nous focaliserons sur l'étude de deux états mentaux en particulier à savoir l'engagement et la charge mentale de travail puisque tous deux reflètent le degré d'investissement mental et physique de l'apprenant au cours d'un processus d'apprentissage.

3.2.1 Engagement mental

L'engagement est un processus complexe qui nécessite de la part de l'individu une mobilisation de plusieurs ressources personnelles (physiques, affectives, etc.). Il peut être défini sous trois dimensions : affective, comportementale et cognitive (Connell 1990). *L'engagement affectif* est un engagement émotionnel qui désigne un attachement personnel envers une personne, une organisation, etc. Par exemple, un employé d'une entreprise est engagé affectivement s'il éprouve une satisfaction de son milieu de travail et est heureux d'évoluer en tant que membre de son entreprise (Duchesne 2005). *L'engagement comportemental*, qui reflète la dimension socio-relationnelle. En d'autres termes c'est le degré d'interactivité d'une personne avec son environnement. Plus une personne est active et veut entretenir des relations avec ses pairs, plus elle a un engagement comportemental de qualité. *L'engagement cognitif* est le degré d'effort mental et de persévérance accompli par un individu pour arriver à son objectif en ayant recours à des stratégies cognitives et métacognitives.

Surveiller l'état mental des apprenants revêt un intérêt particulier dans les environnements d'apprentissage informatisés vu qu'il a un impact sur ses performances d'apprentissage. De ce fait, il est primordial d'avoir une surveillance continue du niveau

d'engagement chez l'apprenant afin de permettre aux éducateurs de les aider en cas de besoin pour une acquisition réussie de la connaissance. Dans ce contexte-là, de nombreux chercheurs se sont intéressés à l'étude de l'état d'engagement chez l'apprenant (Benlamine et al. 2015; Chang et al. 2018; Chaouachi et al. 2010; Henrie et al. 2015; Miller 2015) en utilisant plusieurs types d'approches et outils de mesure.

3.2.2 Charge mentale de travail

La charge mentale de travail peut être définie comme étant la *quantité de ressources* cognitives et mentales (concentration, mémorisation à court terme, etc.) nécessaires à l'accomplissement d'une tâche donnée (Sperandio 1980; Tricot 1998). En d'autres termes, la charge mentale représente l'effort mental qu'un apprenant doit déployer pour acquérir une nouvelle connaissance.

La notion de charge mentale de travail est étroitement liée au terme *charge cognitive* qui à son tour s'appuie sur la théorie de la mémoire de travail. Sachant que la capacité de stockage d'information chez l'être humain est limitée, un apprenant peut être confronté à deux états mentaux extrêmes à savoir la surcharge cognitive (« *mental overload* ») et la sous-charge cognitive (« *mental underload* »).

La *surcharge cognitive* est un phénomène où le cerveau humain devient encombré par une quantité trop élevée d'informations qui demandent à être analysées simultanément. Elle peut avoir des effets néfastes et de lourdes conséquences physiques ou matérielles. Par exemple, dans le contexte d'apprentissage la surcharge provoque la *fatigue intellectuelle*. Ainsi lorsqu'un apprenant devient surchargé, ses capacités cognitives ne fonctionnent plus correctement. Il a du mal à raisonner et à prendre les bonnes décisions ce qui a tendance à affecter son niveau et ses résultats. De même, faire face à trop d'informations complexes peut provoquer un désintérêt de la part de l'apprenant. Ce phénomène est appelé *sous-charge cognitive* qui correspond à une baisse d'effort mental fourni lors du traitement d'une tâche particulière. Cet état de sous-charge contribue tout autant à la dégradation des performances des apprenants ainsi qu'à leur taux de réussite scolaire.

Il n'existe pas cependant dans la littérature un consensus ou un outil de mesure prédéfini permettant d'extraire un indicateur précis de niveau de charge mentale de travail. Néanmoins, différentes approches de modélisation et outils de mesure ont été utilisés.

3.3 Approches de représentation

L'état mental de l'apprenant joue un rôle très important quant à sa façon d'interagir avec l'environnement qui l'entoure. En effet, le comportement et les actions d'un apprenant sont régis par son état mental actuel qui peut se manifester de diverses façons : expression faciale, tonalité de la voix, gestuelle, fréquence cardiaque, mouvements des yeux, etc. On distingue deux approches pour de représentation des états mentaux : les approches quantitatives et les approches qualitatives. Les approches quantitatives sont les mesures les plus couramment utilisées où l'apprenant indique lui-même son propre état durant son interaction avec le système. Ce type de mesures est utilisé pour fournir des valeurs de références à utiliser par d'autres outils de mesure plus sophistiquée. Les approches qualitatives quant à elles ne font pas appel à l'opinion de l'apprenant mais identifient son état d'une manière indirecte en se basant sur différents indicateurs comportementaux ou physiologiques.

3.3.1 Les approches quantitatives

Ce sont des mesures subjectives qui sont basées sur des outils d'auto-évaluation tels que les questionnaires et les échelles de mesure. Elles permettent à l'apprenant d'évaluer son expérience d'apprentissage. Les questionnaires peuvent aller de simples questions à choix multiples aux questionnaires psychologiques complexes. NASA-TLX² (« *Nasa Task Load Index* ») est parmi les questionnaires d'auto-évaluation multidimensionnel les plus utilisés permettant de mesurer un score global de charge mentale de travail en utilisant six sous-échelles comme les exigences physiques, les exigences temporelles ou encore le niveau de frustration (Hart and Staveland 1988). Il y a aussi le «Subjective Workload Assessment Technique» qui est une technique d'évaluation multidimensionnelle se basant sur trois niveaux à savoir l'effort mental, le stress psychologique et le temps (Reid and Nygren 1988).

² <http://humansystems.arc.nasa.gov/groups/tlx/index.html> (page consultée le 04 Décembre 2018).

Mega et ses collègues (2014) ont proposé dans leur travail un modèle théorique permettant d'identifier les éléments qui contribuent à la réussite d'un étudiant. Pour ce faire, ils se sont basés sur trois volets importants pour l'apprentissage : les émotions, la cognition et la motivation en utilisant trois différents questionnaires d'auto-évaluation. Parmi les résultats trouvés, ils ont démontré que les émotions positives sont corrélées avec la motivation ainsi qu'avec certains aspects de l'apprentissage auto-régulé. De plus, la motivation aide les apprenants à améliorer leurs résultats scolaires. (Mega et al. 2014)

Dans la littérature, les questionnaires sont aussi utilisés pour mesurer l'état d'engagement des apprenants (P.-S. D. Chen et al. 2010; Fredricks and McColskey 2012). Pintrich and de Groot (1990) ont utilisé un questionnaire d'auto-évaluation afin de mesurer les éléments d'apprentissage auto-régulé des étudiants et comment ils sont reliés aux performances ainsi qu'à l'engagement cognitif. Les résultats ont montré une corrélation positive entre l'auto-efficacité et l'engagement cognitif. En effet, les étudiants qui s'auto-contrôlent sont les plus impliqués sur le plan cognitif et obtiennent de meilleurs résultats que les autres (Pintrich and de Groot 1990).

Ces méthodes sont pratiques, peu coûteuses et faciles à utiliser pour les grands échantillons et l'apprentissage à distance. Cependant, les données obtenues ne sont pas entièrement fiables car les réponses peuvent être biaisées. Puisque les apprenants remplissent eux-mêmes les questionnaires, les données ne seront pas assez objectives pour permettre de tirer des conclusions efficaces sur l'état mental de l'apprenant. Une autre raison est que l'utilisation de ces outils peut entraver le bon déroulement du processus d'apprentissage. En effet, l'abus d'utilisation peut engendrer la perte de concentration chez l'apprenant et le manque d'utilisation peut fausser les résultats par la suite vu qu'il n'y a pas beaucoup de données collectées.

3.3.2 Les approches qualitatives

Différentes mesures qualitatives ont été utilisées par les chercheurs afin de surveiller l'état mental d'un apprenant. Ces mesures sont utiles tout particulièrement dans des conditions où il est difficile d'inclure les apprenants dans l'évaluation ou encore dans des études exploratoires qui nécessitent d'autres indicateurs. Parmi les indicateurs utilisés il y a les observations par des juges externes, les entrevues, les enregistrements vidéo, etc.

Durant le processus d'apprentissage, des observateurs humains (psychologues, tuteurs, etc.) experts évaluent l'état de l'apprenant durant ou après l'interaction avec l'environnement (Alford et al. 2016; S. D'Mello et al. 2008; Wigfield et al. 2008). Une étude a été faite où des juges rapportent chaque 20 secondes les états mentaux d'une quinzaine d'étudiants durant leur interaction avec un jeu vidéo éducatif. Différents états ont été identifiés par chaque juge mais celles qui ont suscité les mêmes observations étaient l'engagement, la confusion et la confiance (Gutica and Conati 2013). Dans le travail de Helme and Clarke (2001), les auteurs ont utilisé des entrevues faites avec des étudiants ainsi que des vidéo enregistrées lors de plusieurs leçons afin d'identifier de possibles indicateurs d'engagement durant un cours de mathématique (Helme and Clarke 2001). Dans le même contexte éducatif, une autre technique d'observation a été utilisée par Turner et al. (2014) durant des séances de cours magistraux. Les auteurs ont eu recours aux groupes de discussion, aux projets, à des travaux pratiques pour conceptualiser l'engagement des étudiants et identifier des indicateurs de désengagement scolaire (Turner et al. 2014).

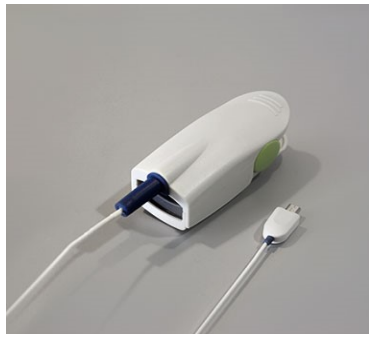
Une autre approche qualitative, souvent utilisée dans la littérature, est l'analyse d'apprentissage connue sous le nom de « learning analytics » (Atherton et al. 2017; B. Chen et al. 2018; Ma et al. 2015; Rienties et al. 2018). Cette approche permet de collecter et d'analyser les traces d'interactions de l'apprenant avec l'environnement d'apprentissage pour fournir des modèles d'apprentissage personnalisé dont le but est d'améliorer l'expérience d'apprentissage de l'étudiant. Parmi les données et traces collectées il y a par exemple le score obtenu, la durée de consultation, nombre de tentatives dans un quizz ou encore les fichiers logs. Par exemple, Lu et al. (2017) ont eu recours à l'analyse d'apprentissage afin d'améliorer le niveau d'engagement des étudiants ainsi que leur résultats d'apprentissage lors d'interactions avec la plateforme MOOCS. La collecte d'information se fait à partir de rapports mensuels envoyés aux tuteurs indiquant les étudiants à risque qui nécessite une intervention rapide. Deux groupes d'apprenants ont été créés pour l'étude; dans le premier les tuteurs se basent sur ces rapports pour fournir l'aide nécessaire aux apprenants et dans le deuxième, les tuteurs se basent sur leurs propres observations. Les résultats ont confirmé que le premier groupe a démontré une amélioration de ces résultats ainsi qu'un niveau d'engagement plus élevé comparé au deuxième groupe (Lu et al. 2017).

Certes ces différentes approches qualitatives permettent une évaluation générale de l'état d'un apprenant, toutefois elles ne peuvent pas être utilisées comme seules sources de données. En effet, nous ne pouvons pas se baser uniquement sur le nombre de cliques effectués ou encore sur les observations d'une tierce personne pour fournir des conclusions précises sur le comportement mental des étudiants. Une autre limite potentielle est que cette approche nécessite la présence d'humains c'est-à-dire du temps et de l'effort ce qui n'est pas toujours facile à obtenir.

3.3.3 Les senseurs physiologiques

Les senseurs physiologiques constituent une autre alternative pour mesurer l'état mental d'un apprenant. Le principal avantage de ces technologies est qu'elles sont capables de détecter les réponses physiologiques d'un individu lors de son apprentissage. En effet, elles fournissent des données quantitatives fiables sur le comportement d'un apprenant ne pouvant pas être directement observées à l'aide des autres outils de mesures citées ci-dessus. Ces données incluent la conductance cutanée, la fréquence cardiaque (Cao et al. 2009; Engström et al. 2005), l'activité cérébrale, la pression sanguine, les mouvements des yeux, etc.

Parmi les capteurs physiologiques fréquemment utilisés (figure 3.1), nous pouvons citer le « *Blood Volume Pressure* » (BVP) servant à mesurer entre autres la fréquence cardiaque et la pression sanguine (Gunes and Pantic 2010; Shen et al. 2009) ou encore le bracelet *Q-Sensor* de *Affectiva* pour mesurer l'activité électrodermale (H.-C. K. Lin et al. 2016; Noordzij et al. 2017; Nourbakhsh et al. 2012). La mesure de la réponse galvanique de la peau « *Galvanic Skin Response* » (GSR) a été utilisée dans beaucoup de travaux de recherche (Nourbakhsh et al. 2012, 2017) notamment dans le domaine de reconnaissances des émotions pour détecter l'excitation (Drachen et al. 2010), le stress (Setz et al. 2010), la frustration (Nasoz et al. 2004), etc.



[a]



[b]

Figure 3.1 Exemples de senseurs physiologiques [a] capteur de pression sanguine [b] bracelet Q-sensor

D'Mello et ses collègues (2007) ont pu identifier deux groupes d'apprenants sur la base de leurs postures en utilisant un système de mesure de la pression corporelle intégré dans un siège. Les résultats ont montré que quand les apprenants s'étendent sur la chaise ou encore manifestent une forme d'agitation (c.-à-d. changements rapides de pression sur le siège); ceci décrit une forme d'ennui qui conduit à un faible niveau d'engagement. Au contraire un haut niveau d'engagement est associé à une augmentation de la pression ainsi que peu de mouvements (S. D'Mello et al. 2007).

Dans le travail de Leyman et al. (2004), les auteurs ont réalisé une expérimentation dans laquelle ils ont mesuré l'activité musculaire (grâce à l'électromyographie) des participants au cours de l'exécution simultanée de deux tâches administratives: une tâche cognitive et une tâche de dactylo. Les résultats ont révélé qu'il existe une relation entre l'activité musculaire et la charge mentale. En effet, plus la tension au niveau des muscles du dos et du cou augmente plus la charge mentale augmente (Leyman et al. 2004).

Outre le rythme cardiaque et les réactions galvaniques de la peau, il y a deux autres indicateurs physiologiques de plus en plus utilisés dans la littérature à savoir *l'activité cérébrale* et *les mouvements des yeux*. En effet, ils fournissent des informations objectives et non-biaisées sur le comportement d'un individu spécifiquement reliées à ses fonctions cognitives comme l'attention ou la charge mentale. Ces deux indicateurs sont mesurés à l'aide de deux technologies qui sont respectivement l'électroencéphalographie (EEG) et les traceurs de regard connus sous

le nom de « Eye Tracker » (ET). Peu couteux et non invasifs, ces deux outils sont utilisés dans divers domaines de recherches telles que la médecine, l'aviation et l'apprentissage. Nous nous attarderons dans ce qui suit sur une description détaillée de chacune de ces techniques puisqu'elles ont été utilisées comme outils de mesure dans nos travaux.

3.2.3.1 L'électroencéphalographie

Le cerveau humain est composé de milliards de neurones reliés entre eux par un nombre gigantesque de connexions synaptiques. La simulation successive (pulsations) de ces neurones déclenche une charge électrique appelée *influx nerveux* qui se traduit sous forme d'ondes cérébrales. La fréquence de ces ondes reflète l'intensité de l'activité cérébrale.

L'électroencéphalographie est la méthode utilisée pour mesurer l'activité cérébrale du cerveau. La première utilisation appliquée aux humains date de 1920 par le physiologiste Hans Berger. Des électrodes sont placées sur la surface de la tête, au niveau du cuir chevelu, permettant d'enregistrer l'activité cérébrale en utilisant un dispositif particulier. La variation dans le temps des potentiels électriques enregistrés à partir de la boîte crânienne est retranscrite par un appareil sous la forme d'un tracé graphique appelé électroencéphalogramme. Le signal EEG obtenu est calculé en hertz (Hz) et peut être décomposé en plusieurs plages de fréquences; les plus principales sont : delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz) et beta (13-22 Hz). La figure 3.2 montre des exemples de signaux EEG sur un intervalle de temps d'une seconde.

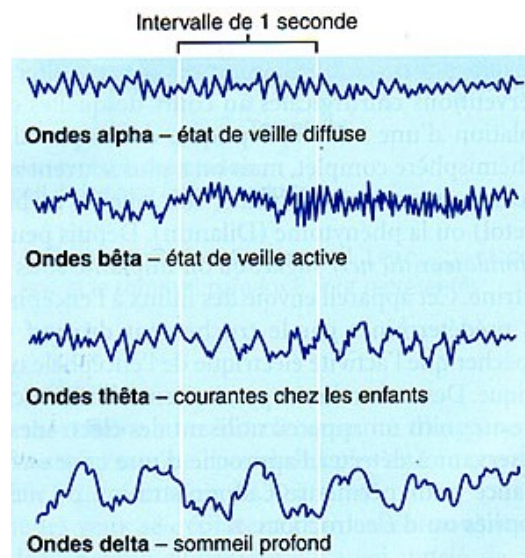


Figure 3.2 Exemple de fréquences du signal EEG

La figure 3.3 [a] montre le dispositif matériel que nous avons utilisé dans toutes nos expérimentations. Le casque Emotiv EPOC est un casque sans fil qui se connecte à toutes les machines grâce à un récepteur USB wifi (figure 3.3 [b]). Il contient 16 électrodes qu'il faut bien mouiller avant chaque utilisation avec une solution saline (figure 3.3 [c]) pour atteindre le bon signal requis. Le casque a un taux d'échantillonnage de 128 Hz.



Figure 3.3 Dispositif matériel du pack Emotiv EPOC: [a] le casque, [b] le récepteur USB et [c] le kit d'hydratation des senseurs

Les électrodes qui composent le casque Emotiv sont positionnées sur le cuir chevelu selon la norme internationale 10-20 (JASPER 1958) comme le montre la figure 3.4. Les deux électrodes CMS « Common Mode Sense » et DRL « Driven Right Leg » sont utilisées comme des points de références. Les 14 autres restantes représentent 14 régions du cerveau humain, par exemple *F7* correspond au lobe frontal gauche ou encore *A2* correspond au lobe de l'oreille droite. Une fois le casque positionné sur la tête, nous devons tout d'abord s'assurer de la qualité du signal. Il faut pour cela vérifier dans le système que toutes les électrodes soient allumées en vert, sinon le signal n'est pas bon. Quand l'électrode est de couleur noire ceci explique que la région du cerveau correspondante est inactive et les couleurs orange, rouge et jaune représentent des régions actives mais avec des impédances incorrectes.

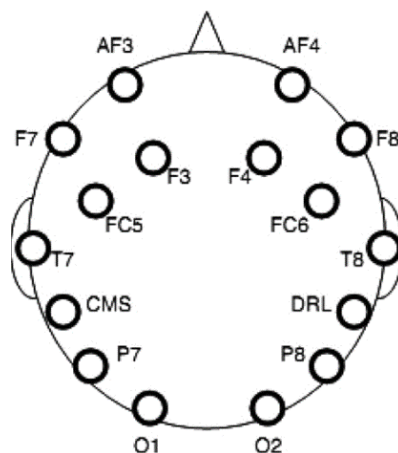


Figure 3.4 Positionnement des électrodes

Les EEG ont prouvé leur efficacité dans un large éventail d'applications notamment : le diagnostic médical (par exemple : la détection de l'épilepsie, la surveillance de l'état de coma des patients) (Ben Hamida et al. 2015; Loo et al. 2016; Roach and Mathalon 2008), la reconnaissance des émotions (par exemple : la détection de l'ennui, la frustration ou encore la confusion chez les individus) (Kim et al. 2018; Soleymani et al. 2016; Zhuang et al. 2017), l'évaluation des performances (par exemple : la prédiction des résultats d'apprentissage, la classification des apprenants) (Jraidi and Frasson 2010; Lujan-Moreno et al. 2016; Van der Hiele et al. 2007) et la modélisation de l'affect (par exemple : la classification des émotions, la modélisation de l'incertitude) (Heraz et al. 2007; Jraidi et al. 2014; X.-W. Wang et al. 2014).

Toutefois, leur utilisation ne se limite pas à ces applications mais il s'est révélé que le EEG est un facteur prédictif de l'état **mental** d'un apprenant (Azevedo 2015; Chaouachi et al. 2015; F.-R. Lin and Kao 2018). Par exemple, Chen et al. (2017) ont utilisé EEG pour évaluer le niveau d'engagement des apprenants dans un environnement d'apprentissage en ligne où ils suivaient des cours par visio-conférence. Le système développé a permis d'identifier les périodes de la conférence qui ont généré un faible niveau d'attention. De plus, les résultats statistiques ont montré que le manque d'attention et d'engagement a un effet négatif sur les performances des apprenants (C.-M. Chen et al. 2017). Dans le même contexte d'apprentissage, une étude a été faite où l'activité cérébrale de 14 apprenants a été mesurée lors de leur interaction avec un système tutoriel intelligent en se basant sur l'électroencéphalographie. L'objectif de ce travail est d'adapter automatiquement les ressources pédagogiques de l'activité d'apprentissage

en fonction de l'état mental de l'apprenant. Lorsqu'un étudiant est désengagé, le système d'adaptation lui présente comme solution d'aide un exemple typique détaillé sur la manière de résoudre un problème précis afin de le maintenir engagé (Chaouachi et al. 2015).

Nonobstant l'efficacité de ces derniers outils de mesure, l'activité cérébrale et surtout l'EEG ont montré des résultats prometteurs quant à la mesure de la charge mentale de travail. Berka et al. (2007) ont utilisé les signaux EEG pour mesurer la charge mentale de travail durant l'exécution d'exercices cognitifs. Les résultats ont prouvé que le niveau de charge de travail augmente au fur et à mesure que le niveau de difficulté des exercices proposés augmente. De plus, les tâches qui sont complexes telles que la résolution de problèmes, le raisonnement analytique ou la mémorisation ont suscité plus de charge cognitive (Berka et al. 2007). Pareillement dans le travail de Hwang et al. (2014), les auteurs ont prouvé que les exercices cognitives complexes engendrent un haut niveau de charge mentale qui résulte à une augmentation de la bande thêta et une diminution de la bande alpha (Hwang et al. 2014). Dans une autre étude faite par Lin et Kao (2018), les auteurs ont eu recours à l'EEG pour détecter le niveau d'effort mental d'un apprenant dans un contexte d'apprentissage en ligne. Une technique d'apprentissage supervisé a été utilisée pour identifier les segments vidéo où les apprenants produisent un effort mental élevé. Par la suite, ces segments peuvent être exploités par les instructeurs afin d'adapter leurs activités d'apprentissage en conséquence. De plus, les étudiants eux-mêmes peuvent avoir recours à ces résultats pour pouvoir auto-réguler leur état mental dans un but commun qui est l'amélioration des résultats d'apprentissage (F.-R. Lin and Kao 2018).

3.2.3.2 Le traceur de regard

Le traceur de regard permet le suivi des mouvements oculaires d'un individu en se basant sur la technique du reflet cornéen. Cette technique consiste à envoyer une lumière infrarouge sur la pupille et de détecter par des capteurs optiques le reflet de cette lumière renvoyé par la cornée de l'œil. Elle sert à mesurer la variation d'intensité du reflet et identifier le centre de la pupille pour nous renseigner sur la position et les déplacements de l'œil.

Il existe différents types de ET sur le marché comme le montre la figure 3.5. Pour nos expérimentations, nous avons utilisé le traceur de regard Tobii Tx-300 qui offre un taux d'échantillonnage de 300 Hz de très haute précision. Les capteurs infrarouges ainsi que la

caméra qui le composent sont branchés sur un écran d'ordinateur de 23 pouces avec une résolution de 1920 x 1080 (voir figure 3.5 [d]). Afin d'obtenir des données oculaires fiables, il faut tout d'abord régler la position de la tête du sujet par rapport au traceur de regard. La distance optimale entre les yeux et le ET doit être approximativement de 65 cm. La deuxième étape préliminaire est la calibration. C'est un processus automatisé qui permet d'enregistrer les caractéristiques des yeux de chaque sujet pendant qu'il suit du regard un point mobile dans l'écran afin de valider si le point de regard mesuré se rapproche du point de regard référence.

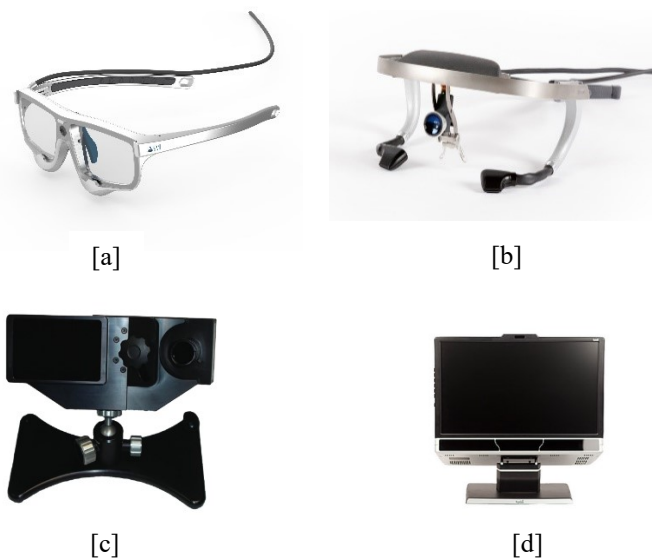


Figure 3.5 Exemples de traceurs de regard: [a] SMI glasses [b] Dikablis glasses [c] EyeLink-1000[d] Tobii Tx-300. Les deux premiers sont mobiles et les deux derniers sont fixes.

Les ET fournissent plusieurs types de mesures qui se basent principalement sur deux notions fondamentales à savoir les fixations et les saccades; décrites dans la figure 3.6. A partir de ces données, nous pouvons mesurer de nombreuses variables telles que le nombre de fixation, la dilatation de la pupille, l'angle de la saccade, le diamètre de la pupille, la durée de fixation, etc.

- **Fixation** : est le fait de maintenir l'œil immobile sur un point particulier de l'écran pour une certaine période de temps. Une fixation est représentée par un cercle. Plus le

diamètre du cercle est grand plus la durée de fixation est longue et plus l'attention de l'utilisateur sur cette zone est grande.

- **Saccade** : correspond à un mouvement rapide de l'œil d'un point de fixation vers un autre. Il est représenté par une ligne qui correspond aux déplacements du regard.

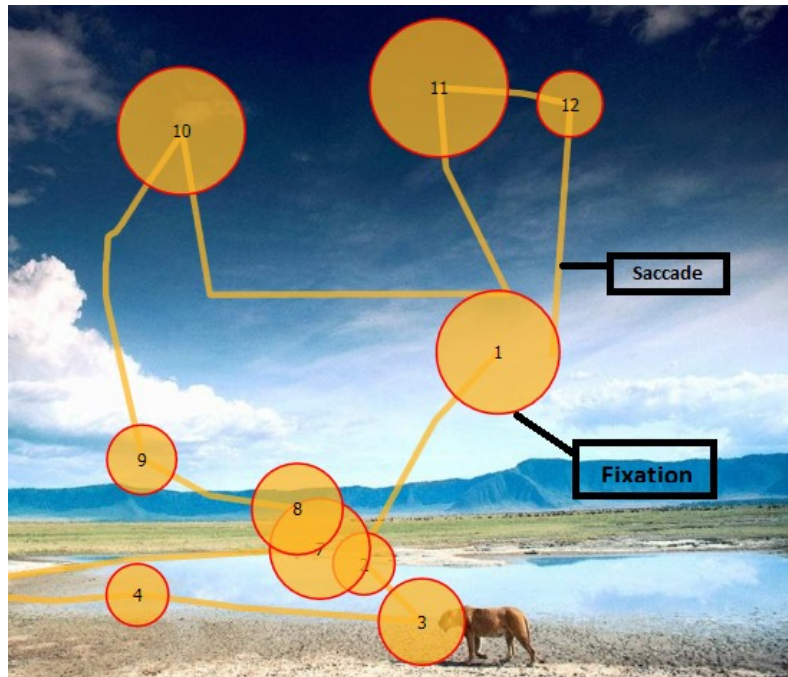


Figure 3.6 Exemples de fixations et de saccades.

Les traceurs de regard ont acquis une grande popularité cette dernière décennie en raison de leur facilité d'utilisation, de leur grande sensibilité et surtout leur caractère non intrusif. Ils sont de plus en plus utilisés dans de nombreux domaines de recherche y compris la visualisation (Lallé et al. 2017; Toker and Conati 2014), la reconnaissance de tâches (Braunagel et al. 2017; Courtemanche et al. 2011; Ishimaru et al. 2014), la détection des émotions (Alhargan et al. 2017; Jaques et al. 2014) et spécialement l'apprentissage (Jung et al. 2018; Luo et al. 2017; Raptis et al. 2017; Trevors et al. 2016; F. Wang et al. 2018).

Les traceurs de regard ont aussi prouvé leur fiabilité dans l'étude de l'état mental de l'apprenant (Lum et al. 2016; Miller 2015; Su and Grauman 2016). Dans le travail de Boucheix et al. (2013), les auteurs ont utilisé les traceurs de regard pour étudier l'incidence de différentes représentations graphiques sur l'engagement et les résultats d'apprentissage des étudiants

(Boucheix et al. 2013). (Siegle et al. 2008) et (S. Chen et al. 2011) ont prouvé que les réponses pupillaires sont des indicateurs fiables de charge mentale de travail.

3.4 Analyse du raisonnement

La psychologie cognitive est une sous discipline de la philosophie qui a émergé en 1960 à la suite du courant béhavioriste. Elle s'intéresse aux processus de traitement de l'information chez l'individu; tout ce qui se rapporte à sa pensée et son intelligence. Les chercheurs de la psychologie cognitive visent à analyser et étudier les mécanismes et stratégies mentales mis en œuvre par l'être humain pour comprendre, assimiler, et acquérir la connaissance. On parle alors de cognition humaine; qui a été définie selon Lemaire (1999) comme une «... *faculté mobilisée dans de nombreuses activités, comme la perception (des objets, des formes, des couleurs ...), les sensations (gustatives, olfactives..), les actions, la mémorisation et le rappel d'informations, la résolution de problèmes, le raisonnement (Inductif et déductif), la prise de décision...*» (Lemaire 1999).

Dans nos travaux de recherche, nous nous sommes basés tout particulièrement sur l'étude du processus cognitif de raisonnement en évaluant les performances des apprenants dans une activité d'apprentissage. Nous présentons dans la partie qui suit le raisonnement ainsi que ces approches de modélisation.

3.4.1 Raisonnement : définition

Le raisonnement est un processus cognitif de haut niveau qui distingue les animaux des êtres humains permettant à partir de connaissances déjà acquises la production de nouvelles connaissances ou l'amélioration de connaissances existantes. L'étude du processus de raisonnement chez l'être humain a depuis longtemps suscité l'intérêt des philosophes ainsi que des scientifiques.

Le philosophe Georges Politzer (1993) mentionne deux grands modes de raisonnement : le raisonnement long qui nécessite des stratégies, une décomposition et une analyse approfondie du problème. Il est utilisé dans des situations inhabituelles ou complexes comme l'argumentation, la résolution de problème, etc. Et le raisonnement court qui identifie directement la solution sans réflexion particulière. Il est utilisé dans les activités quotidiennes

comme produire la conclusion d'un syllogisme ou encore transformer une proposition en une autre équivalente (Politzer 1993).

3.4.2 Types de raisonnement

Festinger (1962) a distingué quatre types de raisonnement cognitif à savoir le raisonnement analogique, le raisonnement abductif, le raisonnement déductif et le raisonnement inductif (Festinger 1962). Nous allons décrire chacun des types en mettant l'accent sur les deux derniers. Nous présenterons aussi l'exemple du raisonnement clinique qui nous a inspiré dans la conception de notre jeu sérieux.

- **Le raisonnement analogique** : consiste à adapter une solution déjà utilisée (la source) face à un nouveau problème (la cible) en s'inspirant du schéma relationnel qui existe entre les éléments de la source. Il est utilisé dans différents domaines tels que les mathématiques, la résolution de problème, la compréhension du langage naturel, etc.
 - ❖ Le raisonnement à base de cas : est un cas particulier du raisonnement analogique qui est une technique de raisonnement par l'exemple où il faut se remémorer un problème déjà résolu dans le passé pour résoudre un nouveau problème (Cordier and Fuchs 2005).
- **Le raisonnement abductif** : l'abduction est un processus permettant d'expliquer un phénomène observé à partir de certains faits sans qu'il y ait des connaissances existantes pour pouvoir l'expliquer. Cette incertitude amène le raisonnement abductif à se baser sur des hypothèses plausibles nécessitant un travail supplémentaire pour vérifier leur exactitude (Catellin 2004).
- **Le raisonnement déductif** : à partir d'un ensemble de prémisses et de faits connus, nous déduisons une conclusion générale. La logique garantit que si toutes les prémisses sont vraies alors la conclusion est vraie (du général au particulier) (Evans 2002). Le raisonnement déductif est devenu un domaine à part entière vers les années 60 dès que les psychologues cognitifs ont commencé à s'intéresser à l'étude de la logique chez les êtres humains. Le philosophe grec Aristote est parmi les premiers théoriciens ayant modélisé le raisonnement avec la notion de syllogisme.

- ❖ Le syllogisme est un modèle de raisonnement qui comporte deux prémisses et une conclusion où la conclusion est déduite des deux premières comme le montre l'exemple suivant :

Prémisse 1 : Tous les hommes sont mortels,

Prémisse 2 : Socrate est un homme,

Conclusion : Donc Socrate est mortel.

- **Le raisonnement inductif** : c'est le cheminement contraire du raisonnement déductif : à partir de faits particuliers observés ou d'informations partielles données, nous dégageons un principe général (du particulier au général) (Varin 2007). L'un des exemples les plus célèbres d'induction est l'exemple de la pomme de Newton. À partir de l'observation de plusieurs expérimentations, le physicien a constaté que tous les objets tombent et a pu généraliser que tous les objets sont attirés vers le centre de la Terre. Ce type de raisonnement est beaucoup utilisé dans les tests d'intelligence qui servent à évaluer les capacités cognitives d'un sujet en utilisant par exemple les séries de chiffres ou encore le test des matrices de Raven.

3.4.2.1 Le raisonnement clinique

C'est un processus cognitif qui consiste à établir un diagnostic et décider d'un traitement à partir de données collectées dans une situation clinique ainsi que de connaissances et expériences préalables (Newble et al. 1995). Il existe deux grandes familles de processus de raisonnement clinique : les processus analytiques et les processus non analytiques dans lesquels nous retrouvons les concepts du raisonnement déductif et inductif.

Le *processus non analytique* se base sur l'intuition du praticien où ce dernier utilise sa mémoire à long terme pour émettre son diagnostic. En effet, il s'agit de repérer au niveau de la mémoire des patrons déjà existants ou encore des cas cliniques similaires déjà traités. Le processus non analytique est un processus de raisonnement essentiellement inductif qui est utilisé par des cliniciens experts.

Le *processus analytique* est un processus itératif qui consiste en premier lieu à générer des hypothèses en se basant sur les informations recueillies en début de l'examen médical.

Ensuite, l'objectif est d'évaluer ces hypothèses (valider ou réfuter) jusqu'à aboutir à un diagnostic final. Le processus analytique est un processus de raisonnement déductif qui est approprié pour les cliniciens novices qui n'ont pas encore acquis l'expérience nécessaire. En effet, en tant qu'étudiants débutants, ils utilisent principalement le raisonnement analytique de type hypothéticodéductif.

- **Le processus hypothético- déductif :** comme le montre la figure 3.7, le raisonnement est représenté par une succession d'étapes détaillées et hiérarchiques. Le mode de raisonnement suivi est tel que l'étudiant commence par émettre des hypothèses précoces dès les premières minutes de la rencontre avec le patient à partir des données démographiques et des symptômes détectés au premier abord. Ensuite un processus itératif commence dont l'objectif est de vérifier les hypothèses émises jusqu'à obtention du diagnostic final. Ce processus consiste à recueillir des données cliniques complémentaires (antécédents, examen de laboratoire, etc.) qui vont servir soit à confirmer ou infirmer les hypothèses du départ. Si les données supplémentaires sont compatibles avec les hypothèses alors ces dernières sont valides et le processus s'arrête sinon le processus recommence avec la génération de nouvelles hypothèses.

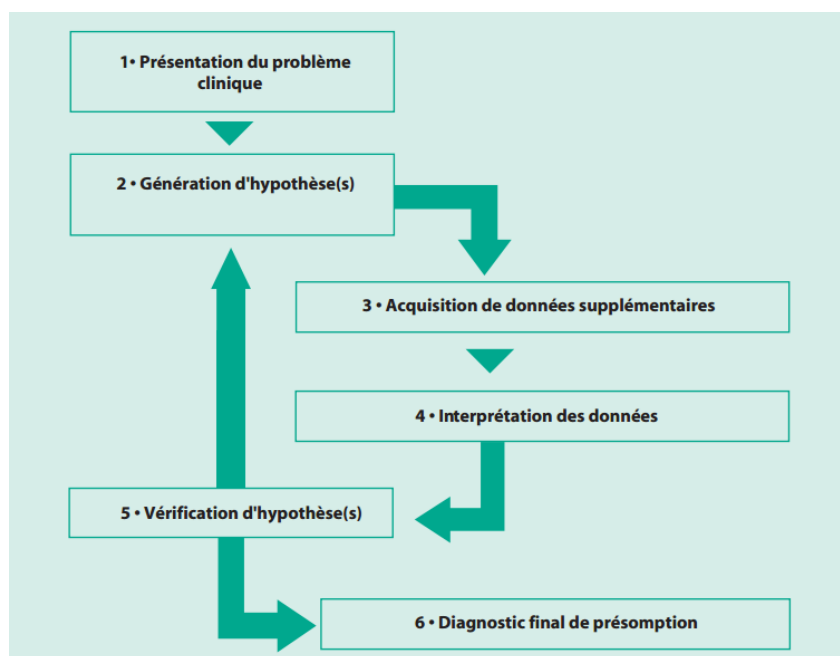


Figure 3.7 Description du processus de raisonnement hypothético-déductif (Nendaz et al. 2005)

3.4.3 Modélisation du raisonnement

Le raisonnement est un processus cognitif très complexe qui attire une attention particulière vu sa nature intuitive et réflexe surtout que chaque être humain est différent. D'où la nécessité d'avoir recours à une approche de modélisation qui est capable d'identifier et d'analyser les démarches suivies lors d'un processus de raisonnement. Toutefois, il n'existe pas jusqu'à nos jours un *modèle unique* pour le représenter.

Plusieurs chercheurs ont tenté d'élaborer des schémas de représentation; une des approches les plus connues et utilisées dans la littérature est **l'analyse de protocole** « *protocol analysis* ». Les protocoles sont des « *séquences d'actions enregistrées au cours de l'exécution de certaines tâches. Ils aident les chercheurs à déterminer les stratégies cognitives impliquées dans la réalisation d'une tâche* » (Salvucci and Anderson 2001). Il existe différentes méthodes d'analyse de protocole que nous présenterons dans ce qui suit.

3.4.3.1 Protocole d'action « *action protocol analysis* »

Cette méthode se focalise sur les actions et le comportement de l'individu lorsqu'il est en train d'interagir avec le système. Les actions qui sont prises en compte sont spécifiquement les clics de la souris et les boutons du clavier. Pendant une période de temps, une séquence de comportements (protocole) peut être identifiée et enregistrée dans un fichier journal «log file» pour une future utilisation (Fu 2001). Prenons l'exemple d'un jeu de puzzle sur un ordinateur où la souris est l'outil le plus utilisé. Cliquer sur une pièce A, la déplacer vers la case voisine vide qui se trouve à gauche, cliquer sur une autre pièce qui se trouvait à droite de A, la déplacer vers la case nouvellement vide, etc. Toutes ces manipulations constituent un protocole d'action.

3.4.3.2 Protocole verbal « *verbal protocol analysis* »

Cette méthode consiste à demander aux individus de décrire verbalement à quoi ils pensent et quelles actions sont faites durant leur interaction avec le système. Il existe deux moyens pour le faire : le « *concurrent protocol* » et le « *retrospective protocol* ». Dans le premier, les personnes doivent penser à voix haute (« *think aloud* ») lorsqu'ils sont en train de raisonner ou d'exécuter une tâche précise. Dans le deuxième, on leur demande de rester silencieux et une vidéo est enregistrée durant le processus de raisonnement. Ensuite, on leur montre la vidéo et ils doivent décrire ce qu'ils ont fait auparavant en parlant à voix haute. Les données collectées peuvent être

utiles pour décrire comment l'utilisateur interagit avec le système ou encore pour identifier les erreurs de logique ou d'incompréhension (Benbunan-Fich 2001).

Forsberg et al. (2014) ont utilisé la méthode « think aloud » afin d'évaluer le raisonnement des infirmières dans la prise de décision clinique. L'étude consiste à utiliser des simulations de patients virtuels et demander aux infirmières de résoudre des cas cliniques complexes. Le protocole obtenu a permis aux auteurs de dégager le cheminement suivi (tester les hypothèses, choisir la bonne stratégie, etc.) par les infirmières durant la prise de décision (Forsberg et al. 2014). Une autre étude a été faite dans le même contexte en utilisant aussi le « think aloud » où des infirmières doivent décrire le raisonnement qu'elles ont suivi pour l'élaboration du planning des soins infirmiers à donner aux personnes âgées pour prévenir l'escarre.

3.4.3.3 Protocole des mouvements oculaires « *eye-movement protocol analysis* »

Nonobstant l'efficacité des protocoles verbaux et d'actions, ils présentent quelques limites. Par exemple, dans le cas du « *retrospective verbal protocol* », si le processus de verbalisation n'est pas fait directement après l'exécution de la tâche, il y a un risque que le sujet oublie quelques passages et cela peut affecter la fiabilité du protocole généré. De plus les mouvements de la souris et les clics ne peuvent pas à eux seuls fournir un protocole précis et efficace. En effet, lorsqu'un participant n'est pas engagé dans son travail, il peut s'amuser à jouer avec la souris ou bien cliquer abondamment. Ce type de comportements peut générer des données erronées. C'est pour cette raison qu'il faut utiliser un moyen efficace, fiable et surtout capable de fournir des données précises et fines. D'où l'utilisation des données oculaires qui ont prouvé leur fiabilité dans plusieurs domaines tels que le marketing, l'imagerie, le web, etc.

Les données oculaires sont utilisées comme une source d'information fournissant une multitude d'indicateurs sur comment et quand les personnes acquièrent l'information et l'utilisent. Néanmoins durant l'exécution d'une tâche particulière, nous pouvons avoir comme sortie une quantité énorme de données (des centaines de protocoles) vu qu'une personne normale peut faire environs trois à quatre mouvements avec les yeux par seconde, ce qui par conséquent demande du temps et de l'effort pour le traitement. C'est pour cette raison que les chercheurs ont essayé d'explorer une nouvelle approche qui consiste à détecter l'ordre logique

des éléments qui ont attiré l'attention de la personne quand elle raisonne. Cette approche est connue sous le nom de « tracing ». Cette technique consiste à faire correspondre le tracé observé des mouvements oculaires d'un sujet avec un modèle prédit d'un processus cognitif (raisonnement, prise de décision, résolution de problème, etc.).

Pour mieux assimiler le principe de cette technique, nous reprenons l'exemple fait par (Salvucci and Anderson 2001). Dans leur expérimentation, les auteurs ont invité les étudiants à résoudre l'équation suivante : $ax/B = A/b$. Le résultat étant le suivant : $x = (A/a) (B/b)$, le modèle de résolution prédit pour cette équation est le suivant : encoder a , encoder A , calculer A/a , encoder B , encoder b , calculer B/b , calculer x . Ensuite, après avoir enregistré les mouvements des yeux de l'étudiant, le « tracing » fait correspondre le modèle de résolution avec le protocole obtenu (séquence visuelle des actions exécutées par un étudiant) comme le montre la figure 3.8.

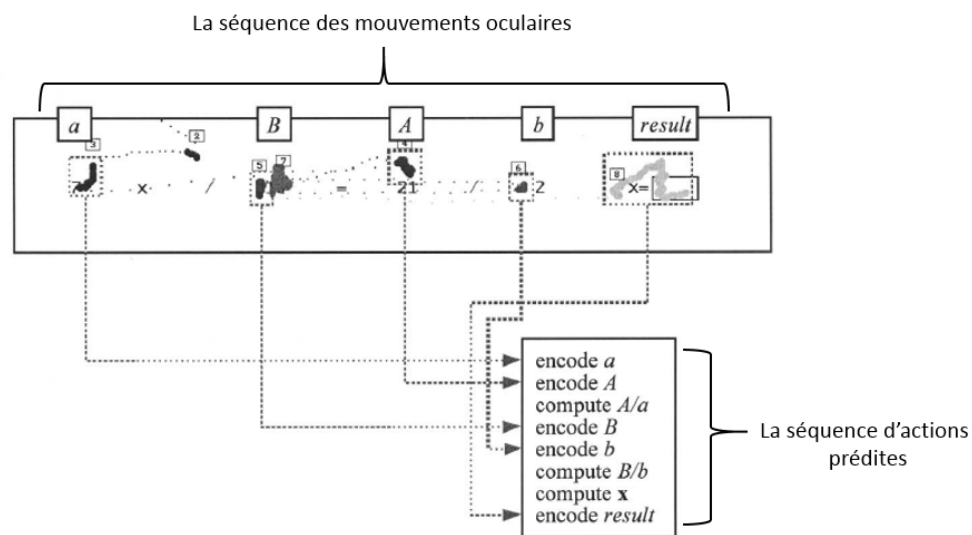


Figure 3.8 Un exemple de protocole, de modèle de résolution et d'une correspondance entre les deux. Les chiffres dans des carrés représentent l'ordre dans lequel l'étudiant a regardé les éléments de l'équation.

3.4.4 Étude de l'existant

Compte tenu des contraintes techniques et méthodologiques dues à la complexité du processus de raisonnement, il est très laborieux de vérifier dans une activité d'apprentissage si un apprenant raisonne bien ou non. Divers outils de mesure ont été utilisés à cette fin, il y a ceux

qui se sont basés sur l'EEG pour mesurer l'énergie fournie par un apprenant lors de son raisonnement, ceux qui ont eu recours à la mesure des performances ou encore les travaux qui ont utilisé les données oculaires comme la taille de la pupille et le temps de fixation (Mulder et al. 2009; Renkewitz and Jahn 2012; Vachon and Tremblay 2014).

Certes ces méthodes ont prouvé leur efficacité, mais elles ne sont pas toujours très fiables car elles fonctionnent en aveugle, c'est-à-dire elles analysent le processus de raisonnement en externe. Or, le raisonnement est un processus complexe qui est intrinsèque à d'autres processus cognitifs comme la prise de décision et la résolution de problèmes. Il peut être vu comme un système de boîte noire dont nous ignorons le fonctionnement. D'où la nécessité d'une analyse interne qui consiste à suivre d'une manière détaillée et approfondie le cheminement logique du raisonnement d'un apprenant. Pour faire face à ces limites, les chercheurs ont eu recours ces dernières années aux données oculaires comme sources d'information afin de tirer des conclusions sur les processus cognitifs et attentionnels de l'individu. En effet, les mouvements des yeux sont utilisés pour étudier les propriétés séquentielles du regard d'un sujet fournissant ainsi des informations plus riches sur la dynamique de son comportement visuel.

Susac et ses collègues (2014) ont utilisé le « scan path » pour analyser les mouvements des yeux de 40 apprenants lorsqu'ils résolvaient des équations mathématiques. Ils ont démontré que les apprenants experts ont une stratégie visuelle bien définie et organisée par rapport aux non-experts qui passent plus de temps à fixer les solutions présentées (Susac et al. 2014). Dans le même contexte, Byun et al. (2014) ont pu différencier entre des apprenants novices et experts lors d'une interaction avec un jeu sérieux. Les résultats ont montré que les experts ont aussi un chemin visuel plus organisé et ont tendance à se focaliser davantage sur les zones pertinentes du jeu alors que les novices ont une trajectoire visuelle aléatoire (Byun et al. 2014). Le « scan path » a aussi été utilisé pour examiner s'il existe une différence entre les enfants et les adultes en termes de leurs trajectoires visuelles (Glady et al. 2013). Une autre étude a été faite où le comportement visuel des apprenants a été analysé lors de la lecture d'un rapport scientifique sur le web ; contenant des données graphiques et textuelles. L'objectif de cette recherche est de déterminer si le niveau de connaissances antérieures des apprenants (élevé et faible) a un impact sur leur processus de lecture. Les résultats ont montré que les deux types d'apprenants ont un profil similaire (c.-à-d. ils privilégient le texte au graphique). Quant au temps de fixation, les

apprenants qui avaient un niveau de connaissances élevé passaient plus de temps à fixer les informations graphiques que les textes explicatifs (Ho et al. 2014).

3.5 Conclusion

Dans ce chapitre, nous avons abordé les deux grands axes sur lesquels se repose cette thèse à savoir l'analyse du processus cognitif de raisonnement et le suivi de l'état mental d'un apprenant. Nous avons tout d'abord, présenté l'état mental en particulier deux états mentaux à savoir l'engagement et la charge mentale de travail. Ensuite, nous avons fait une synthèse des outils de mesure les plus couramment utilisés pour la surveillance de ces deux états. Enfin, nous avons essayé de définir le concept de raisonnement, ses différents types ainsi que ses approches de modélisation. De plus, une revue de la littérature a été présentée pour présenter les travaux existants sur l'analyse du raisonnement ainsi que leur limite.

Dans cette thèse, nous allons tenter en premier lieu d'analyser le processus de raisonnement d'un apprenant lors d'interactions avec deux environnements d'apprentissage. Une première étude exploratoire a été faite pour évaluer les performances d'apprentissage dans un contexte de résolution de problème où des étudiants devaient répondre à des questionnaires à choix multiples après une activité de lecture. Dans ce travail, nous avons opté pour l'utilisation des mesures oculaires statiques basées sur les fixations (*p. ex.* la durée de fixation). Une autre étude plus poussée a été faite afin d'évaluer le processus de raisonnement d'un apprenant lorsqu'il était en train de résoudre un cas clinique. Pour répondre efficacement à notre objectif, nous avons choisi cette fois-ci d'utiliser une mesure dynamique appelée « scan path » qui nous renseigne sur le cheminement séquentiel des étapes qu'un apprenant effectue durant son processus de raisonnement.

En deuxième lieu, nous allons tenter de surveiller l'évolution de l'état mental d'un apprenant lors de son interaction avec le système, ce qui nous permet d'avoir une vision globale et précise de son expérience d'apprentissage. En effet, en utilisant seulement les mouvements des yeux, il y a des situations difficiles à détecter ou des comportements non observables qui peuvent nuire à l'apprentissage tels que la surcharge mentale ou le décrochage. Nous avons utilisé pour cela l'EEG permettant de mesurer en particulier le niveau d'engagement et de charge mentale de travail des apprenants.

Nous proposons dans cette thèse, trois contributions majeures qui seront présentées dans les chapitres suivants sous la forme de cinq articles scientifiques. Dans le premier article (chapitre 4), nous examinerons si les mouvements des yeux sont capables de prédire les performances d'apprentissage des étudiants lors de leur interaction avec le jeu sérieux Crystal Island. Nous proposerons d'étudier quatre variables basées sur la fixation afin d'identifier deux groupes d'apprenants : ceux qui ont réussi à résoudre les questionnaires et ceux qui ont échoué.

Dans le deuxième article (chapitre 5), nous analyserons deux types de mesures oculaires à savoir des mesures statiques et des mesures dynamiques. L'objectif de ce travail sera premièrement d'analyser comment les apprenants explorent visuellement l'environnement d'apprentissage avec lequel ils interagissent en utilisant ces deux types de mesures. En deuxième lieu, nous essayerons d'examiner si ces mesures oculaires ont un impact sur la performance de raisonnement des apprenants lors de la résolution de cas médicaux.

Dans le troisième article (chapitre 6), nous introduirons le concept d'alignement de séquences : une méthode très répandue dans le domaine de la bio-informatique et surtout la génétique afin d'identifier des régions de similarité entre deux séquences d'Acide DésoxyriboNucléique (ADN ou en anglais DNA). Nous nous baserons sur des travaux faits dans le domaine de comparaison de séquences en utilisant une méthode basée sur le codage de caractères dans le but d'identifier si un apprenant est en train de suivre le bon processus de raisonnement. Nous présenterons cette méthode en détail ainsi que l'approche d'alignement local utilisée à travers l'algorithme de Smith-Waterman.

Dans le quatrième article (chapitre 7), nous proposerons d'étudier l'engagement mental des apprenants en utilisant les signaux EEG. L'objectif de ce travail sera d'abord de suivre l'évolution de leur niveau d'engagement afin de déterminer s'il existe des zones particulières qui suscitent un haut niveau d'attention. Si c'est le cas, nous proposerons d'examiner l'impact de ces zones sur les performances d'apprentissage.

Finalement dans le cinquième article (chapitre 8), nous proposerons une approche multimodale où nous combinerons deux senseurs physiologiques à savoir, le ET et l'EEG pour une analyse conjointe des indicateurs d'état mental et du comportement visuel durant un

processus de raisonnement. Notre objectif est de déterminer quels sont les indicateurs clés d'un raisonnement correct en vue de les utiliser pour améliorer les performances d'apprentissage.

Chapitre 4.

Analyse des mouvements des yeux d'un apprenant lors d'une activité de résolution de problème

Dans ce chapitre, nous présentons une première étude exploratoire pour évaluer les performances d'apprentissage de l'apprenant lors de son interaction avec le jeu sérieux éducatif Crystal Island. Notre objectif étant d'examiner si les mouvements des yeux sont un indicateur efficace pour prédire les performances d'apprentissage. En effet, l'un des principaux intérêts des environnements d'apprentissage informatisés est d'essayer constamment de surveiller la progression des apprenants afin de leur offrir l'aide nécessaire au besoin dans le but d'améliorer leur expérience d'apprentissage. Plusieurs approches qualitatives et quantitatives existent dans la littérature dont le but est de fournir des informations fiables sur les résultats et les progrès des apprenants, toutefois, elles présentent chacune certaines limites.

Dans cette première étude, nous avons opté pour les mouvements des yeux comme une mesure physiologique connue pour être étroitement liée avec la cognition humaine et la charge mentale (Kiili et al. 2014; Nourbakhsh et al. 2013). Plus précisément, nous avons utilisé quatre variables de fixation qui sont considérées comme des indicateurs de compréhension chez les humains (Rayner et al. 2006), à savoir : la durée moyenne de fixation, le nombre de fixations, le nombre de revisites et le temps de la première fixation. La première représente la durée moyenne de tous les points de fixation dans une même zone d'intérêt. La deuxième correspond au nombre de fixations enregistrées dans une zone d'intérêt. La troisième identifie le nombre de fois où l'apprenant a refixé une zone d'intérêt en particulier et la dernière représente le temps où la première fixation s'est posée sur la zone d'intérêt.

Notre travail consiste à explorer si en utilisant les mouvements des yeux nous sommes capables de distinguer deux groupes d'apprenants en termes de leur réussite ou échec dans la résolution de plusieurs exercices cognitifs. L'environnement avec lequel les apprenants

interagissent est un jeu sérieux éducatif pour l'apprentissage de la microbiologie connu sous le nom de Crystal Island (J. Rowe et al. 2009). Le but est de faire apprendre aux participants les mystères de la science et de la génétique en leur demandant de déterminer la cause d'une maladie qui a contaminé tous les membres d'un laboratoire de recherche. Pour résoudre cette énigme, ils peuvent lire des livres scientifiques, faire des analyses en laboratoire ou encore parler aux autres personnages du jeu. Pour nos expérimentations, nous nous sommes intéressés particulièrement à la partie lecture où après chaque livre lu, l'apprenant doit répondre à un questionnaire à choix multiples pour évaluer ses connaissances acquises durant la session du jeu.

Une étude expérimentale a été réalisée sur 20 participants volontaires de l'Université de Montréal durant laquelle leurs mouvements des yeux ont été enregistrés en utilisant le traceur de regard Tobii Tx300. L'expérimentation a duré en moyenne 50 minutes par participant pour un total d'un mois. Au cours de l'expérimentation, j'ai procédé à l'installation du traceur de regard, à la phase de calibrage et au lancement du jeu. Deux phases de pré-test et post-test ont aussi été élaborées.

Une fois les données oculaires collectées, la première étape de cette étude consiste à extraire pour chaque zone d'intérêt « Area Of Interest » (AOI) définie les quatre variables de fixation précédemment présentées. Dans un deuxième temps, nous avons fait des analyses statistiques pour observer la relation entre les mouvements des yeux et les performances d'apprentissage en supposant que les apprenants qui ont réussi leur questionnaire ont passé plus de temps dans la lecture des livres. Les résultats de l'analyse de variance « ANalysis Of VAriance » (ANOVA) ont montré, entre autres, qu'il n'y a pas de différences statistiquement significatives entre les deux groupes d'apprenants en termes de durée de fixation suggérant que le temps de fixation n'a pas un impact sur les performances des apprenants. Ceci est en accord avec les travaux de (Clifton et al. 2007; Rayner et al. 2006) qui présument que les longues fixations indiquent des difficultés de compréhension et des perturbations dans la lecture. Ensuite, nous avons étudié la possibilité que les mouvements des yeux soient capables de prédire les performances des apprenants en fonction de leur comportement visuel durant la phase de lecture. Différents classifieurs ont été testés à cet effet en utilisant la technique de la validation

croisée. Les résultats ont montré que l'algorithme de classification les arbres de forêts (*Random Forest*) est celui qui a atteint la meilleure précision.

Le reste de ce chapitre est constitué de l'article de recherche intitulé « *Predicting User Learning Performance From Eye Movements During Interaction With a Serious Game* » accepté à la conférence EdMedia en 2016. Ma contribution essentielle dans cet article consiste à faire une étude exploratoire pour montrer la faisabilité d'utiliser les mouvements des yeux comme source d'information pour évaluer les performances des apprenants lors de leur interaction avec un jeu sérieux éducatif. L'article a été révisé et corrigé par mon directeur de recherche Claude Frasson.

Predicting User Learning Performances from Eye Movements during Interaction with a Serious Game

Asma Ben Khedher, Claude Frasson

Published in : *Proceedings of EdMedia – World Conference on Educational Media and Technology*, 1504-1511 (2016), June 28-30, Vancouver, British Columbia.

Abstract

This paper explored the relationship between eye movements' measures and learners' performance during interaction with Crystal Island, a narrative-centered learning game environment. We gathered gaze data from 20 participants using Tobii Tx300 eye tracker while they were reading books and answering multiple-choices quizzes. Statistical analysis as well as classifications were performed. Random forest classifier reached 70% accuracy and was able to discriminate between the learners who successfully completed the quizzes and the learners who do not, providing thus insight for using eye tracking technique to assess learner's outcomes.

4.1 Introduction

Learning performances represents learners' achievements and progressions, what the students are able to accomplish as well as their skills and level of knowledge (Kobrin et al. 2015). During learning session, it is important to assess the cognitive performance and progress of the learners in order to be able to offer them the appropriate help to overcome their gaps. In fact, enhancing

students learning outcomes is becoming a crucial research domain specially in learning environments. One of the major interests of learning environments is to identify whether the students have well acquired and assimilated the new knowledge. Several methods have emerged in the recent years that provide insights to learning performances and outcomes (Jo et al. 2015; Yu and Jo 2014; Zacharis 2015) as well as physiological measures (Ninaus et al. 2014; Pham and Wang 2015). In this study, we experimented eye tracking and aim to examine whether it is a valuable tool in assessing student learning. Eye tracking is used in different research domains such as activity recognition (Bulling et al. 2011; Courtemanche et al. 2011), affect detection (Jaques et al. 2014), learning (Kardan and Conati 2012; Knoepfle et al. 2009; J. J.-H. Lin and Lin 2014), etc. Lin (2014) investigated the effect of geometry problem's difficulty on gaze data indicating that more the problem was difficult more the fixation on the geometric figures was long. Despite the advances made in this body of research, it is still challenging to provide insights to learners' progress and performance.

This paper contributes to address this issue by exploring the relationship between eye movements and learning performances during interactions with a learning environment. We aim specifically to investigate whether learners' visual behavior can predict their learning performance. In this study, we record and analyze learners' gaze data during their interaction with a game-based learning environment as they are reading scientific books and solving multiple-choices quizzes.

The rest of the paper is organized as follows: in the second section, we discuss some previous works. Then, we describe our learning environment as well as the experimental methodology and the eye movement measures we used. Next, the fifth section is dedicated to the results' presentation and finally we conclude.

4.2 Previous Work

Eye movements are closely related to human cognition and brain activity (S.-C. Chen et al. 2014; Kiili et al. 2014; Tsai et al. 2012). Indeed, eye tracking metrics e.g. fixations (i.e. keep the eye fixated on a particular point of the screen) and saccades (i.e. a rapid eye movement from a point of fixation to another) are indicators of the amount of mental effort being provided during the visual process. For instance, fixation duration is assumed to reflect processing time and

difficulty of comprehension (i.e. the longer a learner looks to an information, the harder it is difficult to understand it) (Just and Carpenter 1976).

Since eye movements are used to understand cognitive processes, we provide an overview of some studies related to reading and problem solving fields (S.-C. Chen et al. 2014; Dimigen et al. 2011; Hyönä 2010; Loboda et al. 2011). Tai et al. (2006) investigated eye movements in a scientific problem-solving environment in order to assess whether there are differences in student's behaviors. The main finding is to be able to discriminate between students with different degree of expertise in science disciplines such as chemistry (Tai et al. 2006). Eivazi and Bednarik (2011) in their work, successfully discriminated between good and poor performers using a real-time eye tracking system. They also performed a classification to predict cognitive states as well as users' performance level in a problem solving task (Eivazi and Bednarik 2011). Researchers in cognitive science and human computer interaction highlight the importance of using eye movements as source of information, however there is still a lack in exploring eye tracking in learning serious games. We aim through this paper to answer these research questions: Are eye movements' measures (e.g. mean fixation duration and number of fixations) able to provide insight into learning performance? Is it possible to predict users' performance results during interaction with a serious game?

4.3 Crystal Island

Crystal Island is a narrative-centered learning environment (see figure 4.1). It is a task-oriented learning environment featuring a science mystery in the domain of genetics and microbiology. The events take place in a volcanic island where research team members were contaminated by an unidentified illness. The player is attempting to discover the cause of this infectious disease. The user is free to perform many actions in order to collect the maximum number of clues and evidence. He can interact with other characters (speak with patients to know their symptoms), run experiments in the laboratory, read books to obtain background information on diseases, collect data about the food recently eaten by the team. After gathering all the relevant information, the final step is that the learner should choose among candidate diagnoses (botulism, cholera, and salmonellosis) the one that matches the patients' symptoms and the

analysis results and also identify the appropriate treatment for the disease in order to solve the mystery (J. Rowe et al. 2009).



Figure 4.1 Crystal Island serious game

4.4 Experimental Methodology

Participants' eye movements were recorded during the whole interaction with Crystal Island. Nevertheless, for this study we aim to examine only the learners' visual behavior when reading books, especially their attention on specific parts of the books called areas of interest AOI. We aim also to study the impact of these specific areas on the learning performances (multiple-choices quizzes' resolution).

4.4.1 Apparatus

We used a commercial eye tracker (Tobii Tx300) with a sampling rate of 300 Hz to record the eye movement. A 23-inch computer monitor (1920 x 1080) resolution is also used to display the experimental stimuli. It integrates the infrared sensors and the camera. Participants were seated in front of the monitor which is placed at a distance of 65 cm approximately from their eyes. Free head movements were allowed during the experiment.

For eye movements' visualization and data synchronization, we used iMotions 5.2 software. The biometric Attention Tool platform of iMotions is an interesting tool that allows to integrate eye tracking data with other data sources such as facial expression, GSR, EEG, etc.

4.4.2 Participants

We collected data from twenty graduate students recruited at the computer science department. Participants were between 17 years to 33 years old ($M = 25.9$, $SD = 4.34$). After the experiment only the eye movements of nineteen participants were analyzed since the remaining one didn't read any book while he was playing. Four of the learners were English speakers and the sixteen remaining were French with abilities to understand written English, since the experiment was in English.

Table 4.1 shows the age distribution of the participants as well as the gaze data quality based on the calibration results.

Table 4.1 Participants' statistics.

	<i>Male</i>	<i>Female</i>
<i>Respondents</i>	18	2
<i>Average age</i>	24.8	27
<i>Std. Dev</i>	4.6	1.4
<i>Data quality</i>	97%	99%

4.4.3 Protocol

As soon as he arrived to the experiment room, the participant was asked to sign a consent form and a personality test namely the Big Five Personality test which defines the taxonomy of the learner's personality into five dimensions: neuroticism, extraversion, openness, agreeableness and conscientiousness (Jr and McCrae 2013). This personality test will be used in future work

to investigate the relationship between the reading tendency of the learner and his personality. Once completing the questionnaire, a back-story of Crystal Island was shown to the participant in order to remind him the game's objectives. The next step consisted in placing him in front of the eye tracker and start the calibration phase. Following this later setup process, the recording session begins by displaying the serious game to the participant during 50mn. After this period, he either resolved the mystery or the game is over. Finally, the last step of the protocol was to fill a post-test questionnaire.

4.4.4.1 Calibration

The Tobii eye tracker uses in most cases a 9-point calibration grid. It consists in evaluating the measured gaze point quality. The calibration results could be poor, good or excellent. Participants had to obtain at least good in order to be part of the study. For this experiment all the participants passed the calibration phase successfully.

4.4.4.2 Post-test

The test contains twelve multiple-choices questions related to the scientific concepts covered in the game. The objective is to validate the participants' level of acquired knowledge after their interaction with the learning game. For each question, there is one correct alternative: 1 point for a correct choice and 0 point for a wrong one.

4.4.4 Eye Movement Analysis

During the game the learner can discover several scientific books, only thirteen of them contain multiple-choices quizzes. Each quiz contains on average 3.46 questions ($SD = 0.97$) and the participants have three answers' attempts per quiz. Participants can explore the island differently, they are free to read all the books or some of them. Thus, the number of read books is variable from one learner to another (on average, participants read 5.68 books ($SD = 1.733$)).

In this experiment, we focused on analyzing the participants' gaze data during this reading phase. And since we focused on specific areas of interest, we defined them before data analysis as shown in figure 4.2 (labeled with rectangles). The image shows an example of a scientific book titled "*How do diseases spread?*" that the learners should read to answer the

quiz. There are three questions in the associated quiz and in order to solve them successfully, the learner should pay attention to some key sentences within the text (AOI). For this book we identified three AOIs: AOI1, AOI2 and AOI3 correspond respectively to question 1, 2 and 3 in the quiz, e.g. the question 2 is “*How infectious disease is transmitted?*” and the answer is “*Through a vector*”. Thus, we can notice that the second area of interest contains clues about the response. Therefore, a total of 45 areas of interest and 97 AOI-based features were defined.

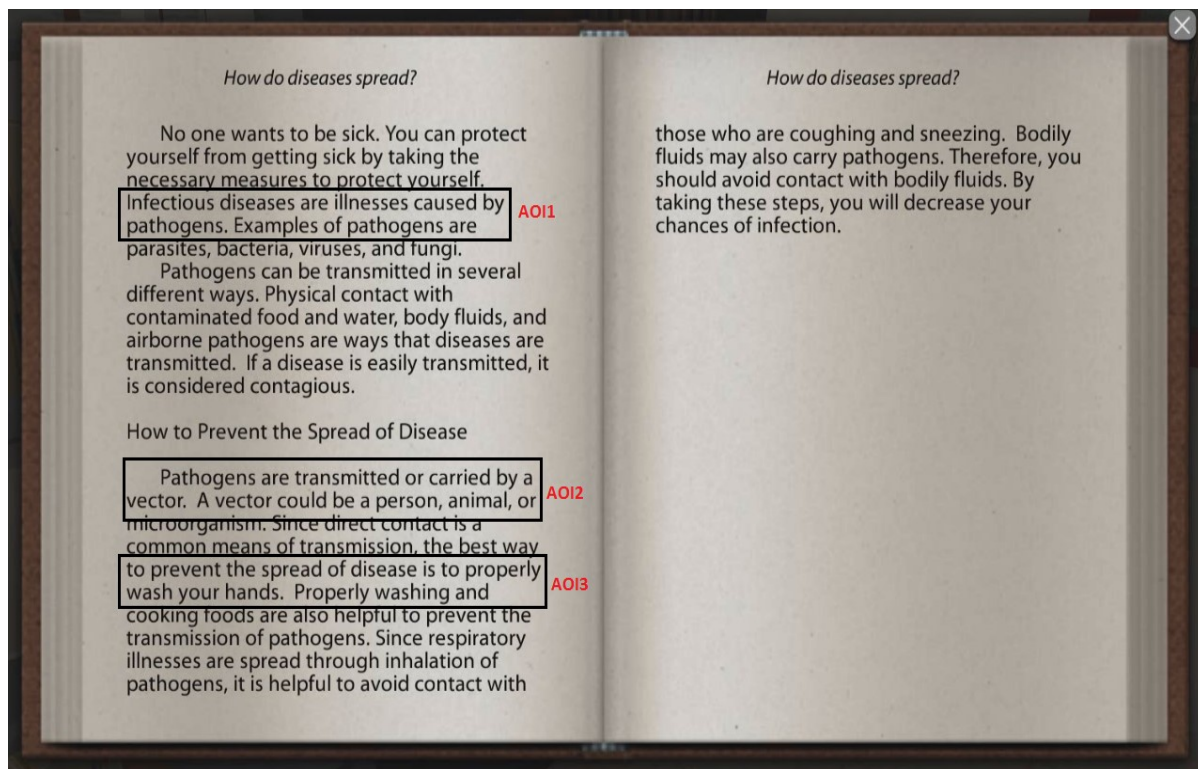


Figure 4.2 Example of a book with 3 areas of interest: each one represents respectively a question in the quiz.

From the obtained data we extracted eye movements’ measures in terms of fixations because they are assumed to be indicators of human comprehension (Rayner, Chace et al., 2006). Fixations also offer more accurate metrics since it is an average over a number of gaze points. The different eye movement measures computed within each AOI are: *mean fixation duration (MFD)*, *number of fixations (NF)*, *number of revisits (NRe)* and *time to first fixation (TTFF)*. They are described in table 4.2.

Table 4.2 Description of eye movement measures.

Measure	Description
Mean fixation duration (ms)	The mean of all fixation points' duration in the AOI.
Number of fixations	The number of fixations recorded when reading inside an AOI.
Number of revisits	How many times the learner returned to re-fixate the AOI.
Time to first fixation (ms)	The time of the first fixation inside the AOI.

4.5 Results and Discussion

4.5.1 General Statistics

We performed a correlation analysis to observe the relationships between the eye movement measures we computed. The results presented in table 4.3 showed statistically significant correlations among the different measures.

For the current study, we aim to predict students' success in learning after reading scientific books during their interaction with Crystal Island. This success can be measured by the learners' results in the different quizzes remembering that each quiz addresses a different microbiology concept. In fact, we aim to investigate whether the fact of looking to the AOIs is an indicator of success. Indeed, the accuracy of the learners' responses demonstrated that 51.55% of them answered successfully to the quizzes and 48.45% failed. Also, with regard to mean fixation duration (MFD) and number of revisits (NRe) as shown in table 4.4, the results revealed that the good performers (Success = Yes) spent on average more time looking to the AOIs than bad performers. These findings confirm that, since the areas of interest provide clues about the right response, paying more attention on AOI (longer MFD) leads to success. However, concerning NRe, the difference between the two groups is no significant, this can suggest two interpretations, either the successful learners revisit the AOI to validate their response, or the unsuccessful ones revisit because they didn't understand the concepts.

Table 4.3 Correlation results³.

<i>Bivariate Correlations between Eye Movement Measures (N = 97)</i>						
	<i>(MFD, NF)</i>	<i>(MFD, NRe)</i>	<i>(MFD, TTFE)</i>	<i>(NF, NRe)</i>	<i>(NF, TTFE)</i>	<i>(NRe, TTFE)</i>
<i>r</i>	.641**	.641**	.739**	.896**	.325**	.326**
<i>p</i>	.000	.000	.000	.000	.001	.001

We also conducted an analysis of variance (ANOVA) in order to examine whether there is significant difference in term of fixation duration (i.e. the time a learner spent fixating an AOI) between the successful and unsuccessful learners. We suggest that the successful learners spent more time within an AOI. The results showed no statistically significant differences between the two groups of learners $F(1, 95) = .205$ $p > .05$. We can interpret from these findings that longer fixation duration is not always a high-level performance indicator (Sharma et al. 2014). Long fixations can be the result of comprehension difficulties and reading disruption (Byrne et al. 1992; Clifton et al. 2007; Rayner et al. 2006). Byrne et al. (1992) also assumed that successful learners tend to read fast.

³ *Correlation is significant at the 0.05 level.

**Correlation is significant at the 0.01 level.

Table 4.4 Statistic results of eye movement measures comparing two groups of learners.

		<i>MFD (ms)</i>			<i>NRe</i>		
		<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>
<i>Success</i>	<i>Yes</i>	50	198.718	120.962	50	3.016	2.472
	<i>No</i>	47	196.333	126.858	47	2.561	3.001

4.5.2 Classification

In order to ascertain whether eye movements are able to predict learning performances, we used Weka data mining toolkit to train and test our dataset. There exist a variety of classifiers in Weka, we tried 4 of them (Naïve Bayes, Random Forest, Multilayer Perceptron and Decision Tree). The role of these classifiers is to discriminate between two classes of learners defined based on the different quizzes' results: success and failure. The k-fold ($k = 15$) cross validation technique was used where in each run the classifier is trained on k-1 subsets and tested on the remaining one.

Table 4.5 Classification accuracy.

<i>Classifier</i>	<i>Accuracy (%)</i>	<i>Kappa</i>
<i>Naïve Bayes</i>	56.70	0.10
<i>Random Forest</i>	70.10	0.38
<i>Multilayer Perceptron</i>	56.70	0.10
<i>Decision Tree</i>	57.73	0.13

The performance of the different classifiers was evaluated in terms of their accuracy (percentage of correctly classified instances) and the Kappa score. Table 4.5 presents the obtained results. Random Forest (RF) is the classifier with the highest accuracy (70.10%), in

addition it achieves higher accuracy than a baseline classifier (ZeroR) that predicts always the most likely class (Success, $Accuracy = 54.63\%$, $Kappa = 0$) as shown in figure 4.3.

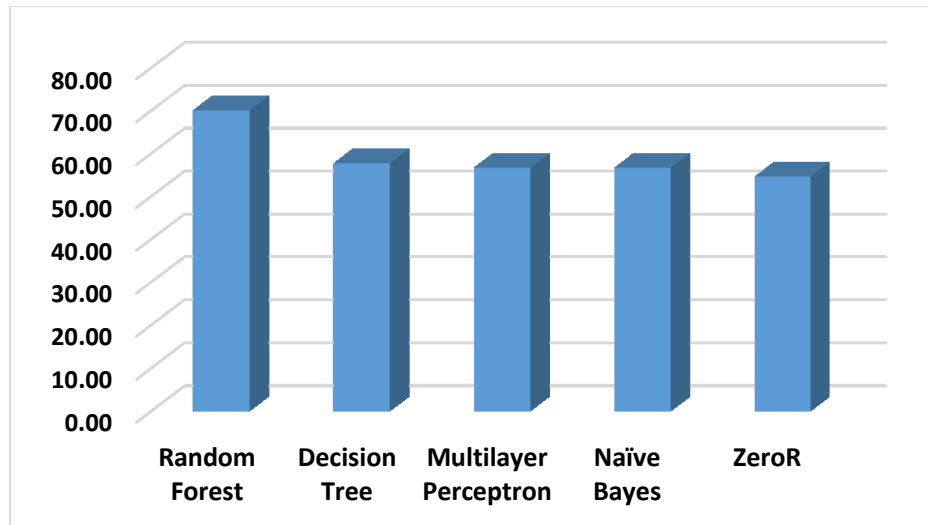


Figure 4.3 Overall accuracy of the 5 classifiers.

4.6 Conclusion

Assessing learners' performances and outcomes are for primary interest in learning environments in order to be able to provide them with the adequate guidance when needed. In this study, we used eye tracking technique to explore the student's visual behavior during a reading task and a multiple-choices quiz solving. The goal of this paper was to investigate the impact of eye movements on the learners' results. Furthermore, the paper objective is to investigate whether gaze data are able to predict learner's level of performance during the game interaction. The results obtained demonstrated that, using eye movements measures e.g. mean fixation duration and number of revisits, a classifier achieves a good accuracy in discriminating between two groups of learners in terms of their results in a multiple-choices quizzes.

Our long-term goal is to apply this study in a real-time interaction with another serious game and also to provide human interaction learning environments with feedbacks concerning the students who learn well and who do not. In fact, these feedbacks will contribute in planning adaptive interventions and help strategies for the learners.

Acknowledgements

We acknowledge SSHRC (Social Science and Human Research Council) through the LEADS project and NSERC (National Science and Engineering Research Council) for funding this research. A special thanks to James Lester and Roger Azevedo from University of North Carolina for their collaboration.

Chapitre 5.

Étude exploratoire du comportement visuel d'un apprenant lors d'une activité de raisonnement dans un environnement d'apprentissage

Dans ce chapitre nous proposons d'explorer le comportement visuel de l'apprenant durant son interaction avec un environnement d'apprentissage qui met en évidence le raisonnement analytique en utilisant deux types de mesures oculaires à savoir des mesures statiques et des mesures dynamiques. L'objectif de cette étude est double : en premier lieu, nous analysons comment les participants explorent visuellement l'environnement d'apprentissage pendant qu'ils examinent une série d'observations pour résoudre différents cas cliniques. Deuxièmement, nous examinons si ces mesures statiques et dynamiques peuvent avoir une influence sur l'expérience d'apprentissage. En d'autres termes existe-t-il une relation entre le comportement visuel de l'apprenant et ses performances?

Plus précisément, nous cherchons à analyser l'activité visuelle des apprenants novices durant leur processus de raisonnement clinique à travers différentes zones d'intérêt définies au préalable. Nous proposons d'explorer deux questions de recherche : (1) Que peuvent révéler les mesures statiques sur l'expérience de l'apprenant durant son processus de raisonnement ? (2) Est-ce que la dynamique visuelle de l'apprenant a un impact sur ses performances ? Les mesures statiques telles que la durée de fixation, le nombre de fixation ou encore la longueur de saccade sont mesurées à partir des données observées selon une dimension temporelle. Elles sont utilisées pour évaluer l'attention visuelle de l'apprenant (Toker et al. 2013). Les mesures dynamiques telles que le « scan path » sont mesurées selon une dimension spatio-temporelle. Le « scan path » (chemin de balayage) est une série de fixations et de saccades représentant la trajectoire visuelle des mouvements des yeux d'un apprenant (Duchowski et al. 2010). Les

mesures dynamiques sont utilisées pour analyser les changements visuels des participants au fil du temps et à travers différents points d'attention.

Une étude expérimentale a été réalisée pour répondre aux deux questions de recherche posées ci-haut. 15 participants ont été recrutés pour cette étude sur une période de deux mois. Chaque session d'un participant dure en moyenne 30 minutes durant lesquels il interagit avec Amnesia, l'environnement d'apprentissage que nous avons développé. Amnesia est un jeu sérieux médical conçu pour évaluer les capacités cognitives des étudiants novices en médecine. J'ai conçu moi-même l'environnement en développant en premier lieu un petit jeu 2D de mémorisation ainsi que des exercices de logique tel que les séries, les analogies et les épreuves d'attention. Ensuite, pour la partie cas médicaux j'ai conçu une interface en essayant de mettre en évidence le processus de raisonnement clinique et surtout le raisonnement hypothético-déductif. J'ai choisi six maladies différentes en préparant pour chacune la liste des symptômes ainsi que le ou les traitements adéquats. En dernier lieu, j'ai vérifié et validé ces cas cliniques par un expert du domaine qui est un docteur et professeur à l'Université de Montréal.

Durant le jeu, l'apprenant joue le rôle d'un médecin qui est pris au piège dans un hôpital où il a été diagnostiqué amnésique par erreur. Sa mission consiste à prouver qu'il ne souffre pas d'amnésie afin de retrouver sa liberté. Pour cela, il doit prouver aux médecins qu'il a encore ses capacités cognitives. La première étape de son parcours consiste à résoudre deux énigmes pour pouvoir sortir de la chambre où il est enfermé. Ensuite, il doit résoudre trois exercices cognitifs dans lequel il doit obtenir un score supérieur à 70%. Enfin, la dernière étape consiste à résoudre six cas médicaux où chaque cas représente une maladie pour laquelle il doit établir le bon diagnostic et le traitement approprié.

Le protocole expérimental de cette étude se compose de trois étapes. Dans la première étape, le participant doit tout d'abord signer un consentement dans lequel le déroulement de l'expérience est expliqué ainsi que le matériel à utiliser. Ensuite, une fois son approbation obtenue, je l'installe devant le traceur de regard pour procéder à la phase de calibrage. La deuxième étape du protocole sera dédiée à l'interaction avec Amnesia. Enfin, dans la troisième étape je demande aux participants de répondre à un questionnaire de faisabilité dans lequel ils doivent évaluer l'aspect ergonomique du jeu et donner éventuellement des commentaires.

Le traceur de regard Tobii Tx-300 a été utilisé pour enregistrer les mouvements des yeux des participants. À partir des données oculaires collectées, nous avons généré deux mesures statiques à savoir, la durée de fixation (en seconde) et le temps de la première fixation (en seconde) et une mesure dynamique à savoir, le « scan path ». Des variables d'interaction ont aussi été collectées durant les expérimentations (p. ex. le nombre de clics sur la souris, les déplacements des participants dans le jeu, etc.), ainsi que des mesures de performances (score dans chaque exercice cognitif, réponses fournies, nombre de cas médicaux résolus, etc.). Par ailleurs, six zones d'intérêt identiques ont été créées dans chaque cas médical afin d'obtenir une analyse détaillée du comportement visuel de l'apprenant. Ces zones représentent six régions distinctes de l'interface de résolution, définies comme suit : la zone « *I* », nommée information, inclut les informations démographiques du patient. La zone « *A* » représente les antécédents du patient, c'est-à-dire les éventuelles allergies et maladies antérieures qu'il a pu avoir. La zone « *S* » inclut tous les symptômes liés à la maladie du cas en question. La zone « *N* », nommée analyse, comprend les données sur la température, la fréquence cardiaque et la pression artérielle du patient. La zone « *D* », nommée diagnostic, inclut les différentes suggestions de réponses parmi lesquelles l'apprenant doit choisir la bonne maladie. Enfin, la zone « *T* », nommée traitement, inclut les différentes suggestions de réponses pour les traitements potentiels de la maladie sélectionnée.

Pour répondre à la première question de recherche à propos des mesures statiques, les résultats ont montré qu'en se basant sur les variables de fixations nous avons pu détecter une tendance visuelle générale pour tous les participants lors de leur résolution des cas médicaux. Nous avons observé que la zone d'intérêt symptômes était la zone la plus fixée par tous les apprenants quant à son importance dans le processus de raisonnement. En revanche, les analyses statistiques ont montré que le fait de fixer cette zone pendant une longue période ne mène pas forcément à de bons résultats. Ces résultats se rapportent à ce qui a été prouvé dans les travaux de (Djamasbi 2014; Poole and Ball 2005) indiquant que les mesures statiques basées sur les fixations ne sont pas toujours positivement corrélées avec les performances. Toutefois et pour répondre à notre deuxième question de recherche, les résultats ont été plus prometteurs. En effet, en utilisant le « scan path » nous avons été capable de discriminer entre les apprenants qui ont exécuté un bon processus de raisonnement et ceux qui ne l'ont pas fait.

La suite de ce chapitre est constituée de l'article de recherche intitulé « *Static and Dynamic Eye Movement Metrics for Students' Performance Assessment* », publié dans le journal *Smart Learning Environments* en 2018. Ma première contribution dans cet article a été de développer le jeu sérieux Amnesia en collaboration avec un étudiant stagiaire de notre laboratoire de recherche en utilisant l'environnement de développement de jeu Unity. Ma deuxième contribution a été d'examiner à travers une étude exploratoire l'impact des mesures statiques et dynamiques sur les performances de l'apprenant via son processus de raisonnement. Imène Jraidi, quant à elle, a contribué avec ses conseils et ses corrections dans la rédaction de l'article, ainsi que dans la phase d'analyses statistiques.

Static and dynamic eye movement metrics for students' performance assessment

Asma Ben Khedher, Imène Jraidi, Claude Frasson

Published in : *Smart Learning Environment*, 5(1), 14 (2018).

<https://doi.org/10.1186/s40561-018-0065-y>

Abstract

This paper reports on the feasibility of using eye tracking as a tool for students' performance assessment in a medical serious game. We are particularly interested in analyzing the relationship between learners' visual behaviour and their performance while solving medical cases. The objective of this study is twofold. First, we analyze how the students visually explore the learning environment across different areas of interest. Second, we examine whether static and dynamic eye tracking metrics can have an impact on students' reasoning performance. Results revealed statistically significant associations between eye movement metrics and students' outcomes. Particularly dynamic metrics better reflected students' analytical reasoning abilities. Our findings have implications for the educational technology community seeking to gain a deeper understanding of the students' learning experience

Keywords: Eye tracking, visual behavior, scan path, students' performance, serious game, analytical reasoning.

5.1 Introduction

In the recent years, there has been an increased interest in tracking users' eye movements and particularly in assessing what can gaze data reveal about the users' experience. In fact, eye-tracking technologies attempt to make the invisible visible. Many researchers in various domains have employed eye tracking as a technique to capture where ocular attention is focused on a visual display at any specific time, as well as the sequence of fixations followed from one area of interest to another.

Eye tracking is used in human-computer interaction to study usability issues (Eraslan et al. 2015; Poole and Ball 2005; Weill-Tessier et al. 2016) and in visualization tasks to provide effective user-adaptive systems (Jaques et al. 2014; Lallé et al. 2017; Toker and Conati 2014). The development of computer-based adaptive systems in learning environments using eye tracking relies principally on the assessment of users' visual behaviour, users' visual pattern and users' learning. Indeed, enhancing students' learning outcomes and facilitating knowledge construction and acquisition is of primary interest. Eye tracking technique has proven its efficiency in educational environments to understand how students evolve through the learning process using different eye gaze metrics (Ben Khedher et al. 2017a; Lallé et al. 2017; Luo et al. 2017; Raptis et al. 2017; Trevors et al. 2016; C.-. Y. Wang et al. 2016).

The majority of eye tracking studies in the literature use static metrics such as fixation count, fixation duration or saccade length (Ho et al. 2014; Kardan and Conati 2012). Yet, some researchers are concerned with the sequential properties of eye movements such as the visual scan path (Lee and Wu 2018; Susac et al. 2014). These dynamic metrics provide reliable information about moment-to-moment changes in the participants' visual behaviour dynamics. This paper proposes to use both types of eye gaze metrics, in order to assess learners' performance as they interact with a medical serious game. We explore the following two research questions. RQ1: What can fixation-based static metrics reveal about students' experience within the learning environment? RQ2: Do students' visual patterns across areas of interest (AOIs) have an impact on their performance?

The rest of the paper is structured as follows. Section 2 presents related work. Section 3 describes our learning environment and the experimental protocol. Section 4 discusses the obtained results. Section 5 concludes and presents directions for future work.

5.2 Related work

The aim of the educational technology communities is to constantly seek to offer a pleasant and effective learning environment in order to keep students immersed and foster learning performance (Jraidi et al. 2014; Jraidi and Frasson 2010; Sawyer et al. 2017). Eye tracking can provide valuable insight into the learners' gaze behaviour during cognitive task resolution using static and dynamic metrics (Takeuchi and Habuchi 2007). Static metrics such as fixation duration, fixation count or saccade length are measured from the observed data according to a temporal dimension. They are used to assess where visual attention is focused view of our long-term goal of designing information visualisation systems that can dynamically adapt to individual user (Toker et al. 2013). Fixation is when the eyes momentarily keep fixating a particular target for a certain duration. Saccade corresponds to a rapid eye movement from one fixation to another (Holland and Komogortsev 2011; Yoshimura et al. 2015). Dynamic metrics such as scan path are measured according to spatial-temporal dimensions. Scan path is a series of fixations and saccades representing the visual trajectory of a student's eye movements across the AOIs (Blascheck et al. 2014; Duchowski et al. 2010). Dynamic metrics are used to analyze the students' visual shifts along the time and across different areas of interest. However, a main issue is to relate these gaze metrics to students' learning.

Pachman and his colleagues (2016) used fixation-based metrics to detect learners' confusion in digital learning environments. They found that confusion ratings were positively correlated with the fixation duration on non-relevant areas, which helped them detect potential confusion areas with the future goal of setting self-regulatory techniques for students to manage their confusion during learning (Pachman et al. 2016). In (Ben Khedher and Frasson 2016b), the authors used eye movements in order to predict learners' level of performance during their interaction with a narrative-centred learning game. They examined the relationship between static metrics such as mean fixation duration and learners' outcomes. They were able to discriminate between two groups of learners in term of scores (high and low) in different

multiple-choice quizzes. Lum et al. (2016) investigated the relationship between the number of fixations and learners' mental workload during the completion of a computer-based training program. The results showed that there were statistically significant correlations implying that when the task was not challenging enough, learners' visual attention decreased (Lum et al. 2016).

Despite that static metrics are being used as a main measure for most studies, they do not glean enough information to show that focusing on a particular area for a certain period of time is the cause of performance differences among learners. Hence the use of dynamic metrics, which provide a deeper understanding of the learners' visual behaviour. In fact, scan paths are useful tools for providing differences in individuals eye movement patterns in the context of visual learning (Eisenbarth and Alpers 2011; Krupinski et al. 2006; Xu et al. 2018). For instance, in the work of Byun et al. (2014), the authors analyzed learners' in-situ data to investigate how eye tracking can be used to evaluate students' performance while they are interacting with a serious game. Scan path dynamic metrics were used to differentiate between novice and expert players. Results showed that experts have a more organized visual path and tend to focus more on the relevant areas in the game compared to the novices who have a randomly visual trajectory (Byun et al. 2014). In the same context, Susac et al. (2014) also used scan path measures to discriminate between expert and non-expert learners while solving mathematical equations. Results indicated that expert students executed a well-defined and organized visual strategy as compared to non-expert students who focused more on the presented solutions (Susac et al. 2014).

In this paper, we propose to study both static and dynamic fixation-based metrics to assess students' performance within problem-solving tasks. Our objective is to use both metrics to monitor the analytical reasoning process of novice medical students as they interact with a learning environment and explore a series of observations to resolve some clinical cases. First, we seek to analyze the students' visual behaviour across different areas of interest and check whether there are particular areas that contribute to the students' success. Second, we aim to examine whether static and dynamic eye gaze metrics can have an impact on students' reasoning outcomes.

5.3 Experimental design

An experimental study was conducted to track students' visual behaviour while interacting with a learning environment called Amnesia. Amnesia is a medical serious game developed to assess novice undergraduate medical students' cognitive abilities through clinical problem-solving tasks. The design of the environment was examined and validated by a medical professional. The game features a virtual hospital where the user is playing the role of a doctor who is mistakenly diagnosed with amnesia and trapped within this hospital. The players need to prove that they do not suffer from amnesia by resolving first some cognitive tasks such as simple visual memory exercises and logic tests (e.g. number series and analogies). Then, they have to demonstrate their clinical skills by resolving six medical cases.

For each case, the students are instructed to identify the correct diagnosis and the appropriate treatment. For that purpose, they have to consider a series of observations such as patients' demographic information, symptoms and antecedents. The players can make initial hypotheses from the demographic information and the symptoms exposed by the patient and establish a first diagnosis. The different diseases they have to discover are respectively: flu, bacterial pneumonia, measles, Ebola, mumps and whooping cough. For each diagnosis, different response alternatives are given and students have three attempts to find out the correct response. They can also collect additional clinical data such as analyses and antecedents until reaching the right diagnosis. Once the diagnosis is established, the students have also three attempts to identify the adequate treatment and after three errors made either in the diagnosis or the treatment, the game is over. Between the three attempts, the students can adjust their hypothesis and correct their answer.

5.3.1 Protocol

Upon arrival at the experimental room, students were asked to sign an ethic form explaining the study and the material. Then, they were placed in front of the eye tracker and were informed that free head movements were allowed. Following the calibration process, the game was displayed. An introductory scene informs the players about how to use the system and the main objectives of the learning environment. During the session, guidance messages were shown to help them move forward in the hospital. After 30-45 minutes of interaction, the session ends.

Participants were then asked to fill in a post-game questionnaire about the game design and usability in order to collect feedbacks for further improvement.

5.3.2 Participants and apparatus

Fifteen undergraduate medical students were recruited for the experiment. Their eye movements were recorded using a Tobii Tx300 eye tracker at a sampling rate of 300 Hz. Infrared sensors and a camera were used to record the session on a 23-inch computer monitor with a (1920 x 1080) resolution. The Amnesia environment was presented on this eye tracking system, which was approximately placed at a distance of 65 cm from the participant. A nine-point calibration grid was used to evaluate the eye gaze quality of each participant. The calibration process consists of evaluating the quality of the measured gaze points. It is a mandatory step in eye tracking in order to keep in the study only reliable data. All the participants who took part in the study successfully passed the calibration step (see table 5.1 for participants' statistics).

Table 5.1 Participants' statistics.

	<i>Male</i>	<i>Female</i>
Tested Respondents	8	7
Valid Respondents	8	7
Average Age	23.2	20.1
Data Quality	92%	91%

In order to obtain a detailed analysis of the students' eye movements, specific areas of interest (AOIs) representing task-relevant regions of the screen were created in each medical case (see figure 5.1). Six AOIs were defined as follows: Information (*I*), Antecedents (*A*), Symptoms (*S*) Analysis (*N*), Diagnosis (*D*) and Treatment (*T*). The *I* area of interest includes the demographic information of the patient (e.g. first and last name, age, weight and height). In the antecedent AOI, we introduce the eventual allergies and prior diseases that the patient could have. We present in the *S* region all the symptoms related to the specific disease. The analysis AOI includes data on the temperature, heart rate and the blood pressure of the patient. Concerning

the diagnosis area, it shows different suggestions of answers among which the participant may select one. Finally, the *T* area of interest presents different suggestions for the potential treatments of the selected disease. All the medical cases have the same AOIs. Gaze data relative to each area were recorded. From the obtained data three metrics were computed: two static metrics namely fixation duration and time to first fixation, and a dynamic metric namely scan path. Fixation duration is the average time spent on a particular AOI. Time to first fixation indicates the amount of time it takes to look at a specific AOI from stimulus onset. Scan path is a series of fixations and saccades representing the visual trajectory of a student’s eye movements across the AOIs.

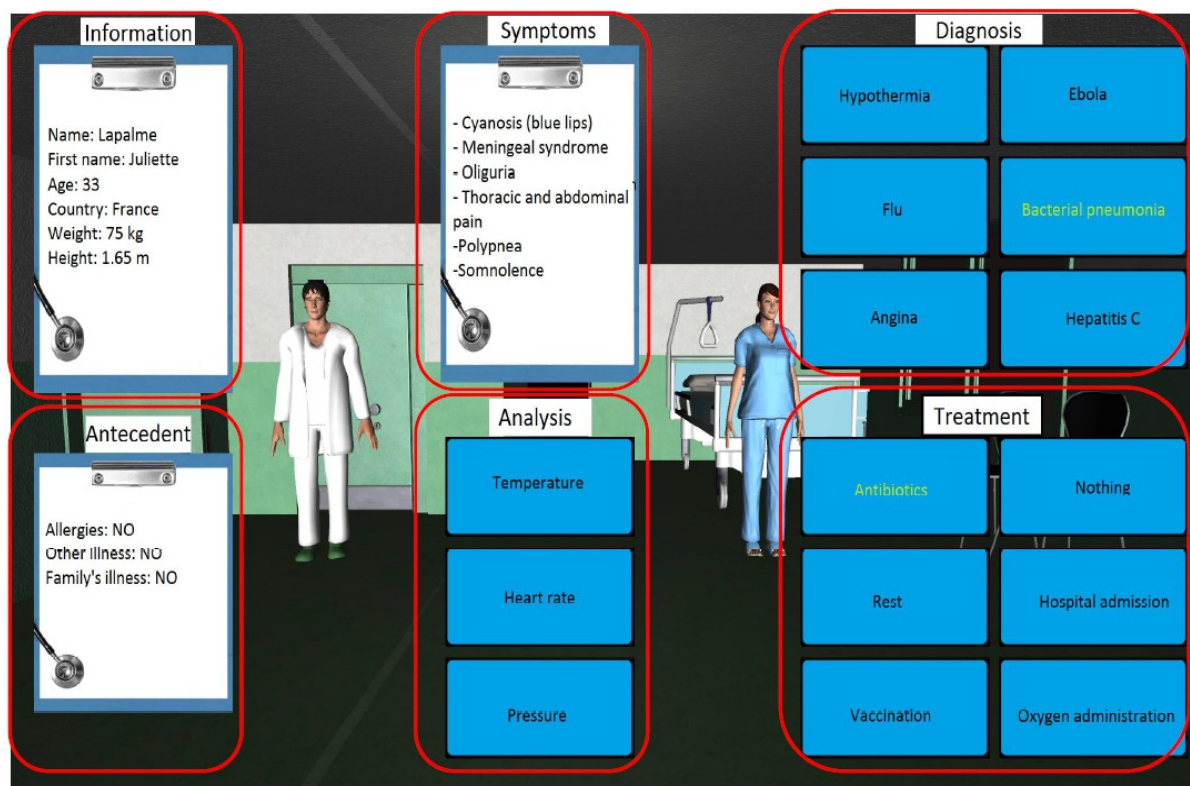


Figure 5.1 The six areas of interest identified in all the medical cases.

5.4 Results and discussion

In this study, we had two main objectives. First, we aimed to track students’ eye movements while they were resolving medical cases. We wanted to have a general overview of their visual behaviour across the different areas of interest in each medical case. The goal was to identify

which areas of interest caught the most the students' attention and whether such area(s) contributed to the students' success using static metrics. Second, we sought to investigate students' visual patterns using the scan path dynamic metric. That is, can the scan path actually reveal the students' thought process? Can this metric accurately reflect the learners' performance?

5.4.1 Fixation-based metrics and students' performance assessment across the AOIs

In each medical case, participants have the ability to inspect different regions of the screen as they attempted to make a diagnosis based on Symptoms (S), Information (I), Antecedents (A) and Analyses (N) as shown in figure 5.2. In each region, the fixation metrics were recorded to find out the most visited areas and to identify whether they could possibly lead to successful responses.

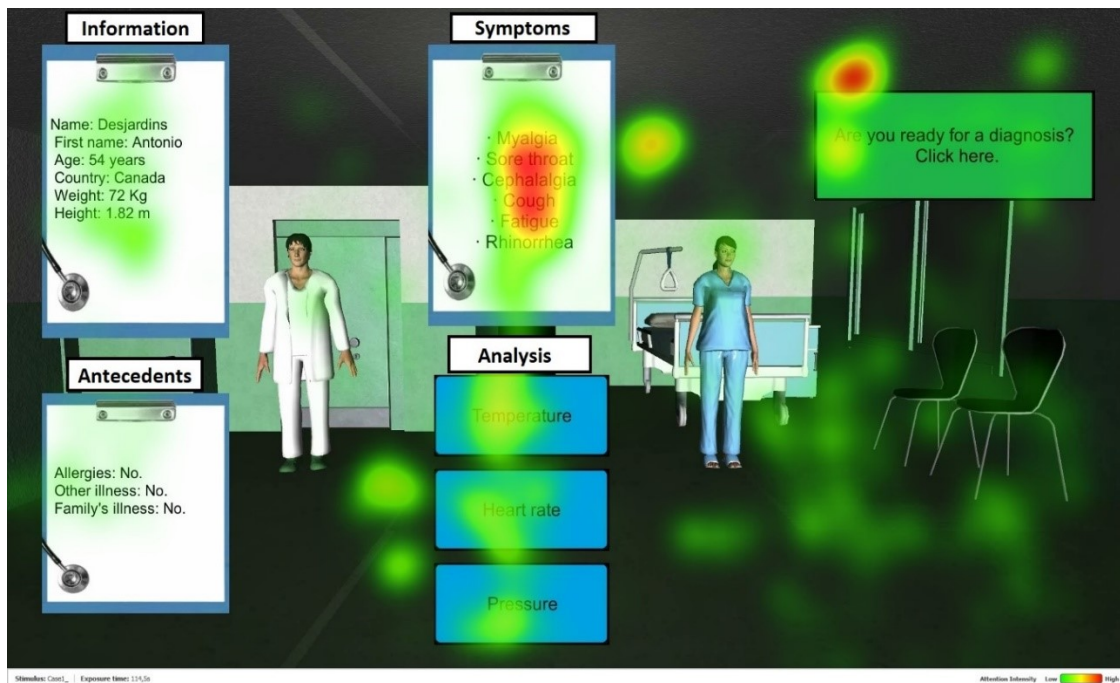


Figure 5.2 A screenshot of one medical case. The red colour represents the most fixated areas of interest within the screen, followed by yellow and then green for the areas with less fixation intensity.

5.4.1.1 How students explore the learning environment?

A preliminary statistical comparison between the different areas of interest using the mean values (M) and the standard deviations (SD) was performed. The results showed that the Symptoms' area was by far the most fixated zone by all the participants in all cases in terms of fixation duration. The most significant fixation rate was found in the last case (M = 28.72, SD = 17.48) and the least was during the resolution of case 1 (M = 13.90 SD = 11.48). Indeed, in the first case, participants discovered the interface for the first time. For this reason, the time allocated for the symptoms was not as different as the other areas, as compared for example to the analysis area of interest (M = 10.37, SD = 7.05).

In accordance with these findings, a one-way ANOVA was performed to examine whether there were significant differences among all the AOIs in term of fixation duration. Table 5.2 shows statistically significant results ($p < 0.001$) among all cases. Tukey post-hoc tests were conducted where all areas of interest were compared by pairs in order to find out which area of interest attracted the most the students' interest. The results showed that the Symptoms' AOI differed significantly ($p < 0.05$) from the other areas in almost all cases. Based on fixation duration, the time spent looking at the Symptoms' area far exceeded the time dedicated to the other areas ($p < .001$). As a matter of fact, these results are obvious since the analysis of a patient's symptoms is a paramount step in any clinical diagnosis process.

Table 5.2 Descriptive statistics for the fixation duration metric across all the AOIs, Means (Standard Deviations). ANOVAs are reported in italic ($p < 0.001$).

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
	<i>F (3, 11) = 9.83</i>	<i>F (3, 6) = 28.52</i>	<i>F (3,6) = 43.95</i>	<i>F (3, 3) = 7.15</i>	<i>F (3, 3) = 15.73</i>	<i>F (3, 1) = 15.89</i>
Information	4.35 (2.65)	2.85 (1.75)	2.56 (1.79)	1.22 (1.01)	1.60 (1.29)	1.26 (1.39)
Antecedents	1.79 (1.35)	0.69 (0.61)	0.71 (0.69)	1.68 (1.18)	0.70 (0.69)	0.78 (0.72)
Symptoms	16.78 (14.64)	25.89 (12.56)	18.69 (7.30)	13.90 (11.48)	20.72 (10.48)	23.72 (11.87)
Analyses	10.37 (7.05)	7.25 (4.82)	6.50 (8.15)	4.27 (1.46)	2.92 (1.87)	5.69 (3.07)

5.4.1.2 Do fixation-based metrics have an impact on students' response?

In this study, several Multivariate Analysis Of Variance (MANOVA) were performed to check in each case, whether looking to the “S” AOI can predict students' responses and if not, which area can then better predict it.

Symptoms' area. The results of the MANOVAs are shown in table 5.3. For the first case, the analysis was not conducted since the number of subjects was not fair between the two groups (group 1: success; group 2: failure). For cases 2, 3, 4 and 5, the results did not show a statistically significant relationship between the fixation duration and the identification of the correct diagnosis. Finally, for the last case, a significant effect was found showing that focusing on the “S” area may potentially have an impact on the success or failure.

Table 5.3 MANOVAs' results summary between the symptoms' fixation duration and the diagnosis success/failure.

		<i>Case 1</i>	<i>Case 2</i>	<i>Case 3</i>	<i>Case 4</i>	<i>Case 5</i>	<i>Case 6</i>
		F (3, 11)	F (3, 6)	F (3, 6)	F (3, 3)	F (3, 3)	F (3, 1)
<i>F.D</i>	<i>p</i>	.	.14	.45	.45	.77	.002
	<i>F</i>	.	2.6	.62	.66	.09	121.08

Remaining areas. In the previous section, the results demonstrated that one cannot conclude that longer fixation on the Symptoms' AOI impacts the students' performance. In this section we present additional case-by-case MANOVAs to investigate whether there is such area that leads to the students' success. Both fixation duration and time to first fixation were analyzed.

Case 1 was discarded since all participants succeeded in identifying the diagnosis. Case 2 ($F(1, 8) = 4.946$) and case 5 ($F(1, 3) = 0.084$) showed no significant differences ($p = n.s$) between the areas of interest in terms of both fixation metrics. For the remaining three cases, the results were statistically significant in terms of time to first fixation. Post hoc tests were

performed for separate correlational analyses to check which AOI could potentially abet students' success. Results showed that there was not a unique AOI that is related to participants' performance, but in each case there were different areas. For case 3 and 6 a significant effect was found for all areas ($p < 0.05$). In case 4, a unique significant result was found for the Antecedents' region ($p < 0.001$). In fact in this case, the diagnosis the students should identify is Ebola, and they needed to pay particular attention to what was listed in the antecedents' area, which included an important clue ("recently travelled to Guinea") that may help students give the correct answer.

In summary, we were able to identify a relationship between some areas of interest and students' performance using fixation-based static metrics. However, we cannot state that these static metrics can reliably assess students' performance. For instance, as stated in different studies (Djamasbi 2014), longer fixations do not always reflect users' engagement; fixation metrics and learners' performance may be negatively correlated.

5.4.2 Performance assessment using scan path metric

In this section, we investigate the relationship between students' saccadic eye movement sequences and their performance. We use students' scan paths and a sequence alignment algorithm namely, Smith-Waterman, to evaluate learners' analytical reasoning. Smith-Waterman algorithm is based on dynamic programming for pairwise sequence comparison (Setubal and Meidanis 1997). Our objective is to use this method to discriminate between the learners in a way that identifies those who follow a correct analytical reasoning process using students' visual scan paths. The alignment process is divided into three parts.

First, students' scan paths are recoded as a sequence of characters representing the student's followed steps to solve each case (e.g. "IIIAASSSSNNNNSSNDDSSDDTTDTT"). Each character of the sequence is a visited AOI with a fixation duration above a threshold of 250 ms. Redundant characters are collapsed to retain a unique occurrence for each visited area in the scan path (e.g. "IASNSNDSDTDT").

Second, a *reference sequence* is used to represent the hypothetico-deductive clinical reasoning process. This target sequence is a sequential representation of the different steps a

novice student has to perform to accomplish an optimal reasoning process yielding to this following sequence: "ISANDT". Indeed, in this analytical process, the clinician starts by collecting the patient's information, then he can make initial hypotheses from the symptoms exposed by the patient. Additional clinical data such as analyses or medical antecedents are then collected to validate or discard his hypotheses until he reaches the correct diagnosis and treatment. In order to have deeper information about the hypothetico-deductive clinical reasoning process, please refer to (Nendaz et al. 2005).

Third, both obtained sequences (i.e. the student's scan path and the reference sequence) are compared using the local sequence alignment Smith-Waterman algorithm. Aligning two sequences using the Smith-Waterman algorithm implies identifying regions of similarity between both sequences. The first step of the Smith-Waterman algorithm is to create a similarity-based matrix where the first column and row are initialized with zeros. The remainder cells are filled using the local alignment recursion relation. Once the matrix is completely filled, the algorithm selects the cell with the highest score. Then, from that cell, it backtracks through the array until reaching the first zero. The constructed path represents one of the possible optimal local alignments for the two sequences.

In order to quantify the optimal alignment found, a final score is computed based on three scoring metrics namely; match, mismatch and gap. A match means that two characters in both sequences are identical, a mismatch requires a mutation operation that is substituting one character to another, and a gap implies an insertion or a deletion in one of the two sequences. The higher the computed alignment score, the more the two sequences are similar. We refer to (Ben Khedher et al. 2018a) for more details about this procedure.

In order to highlight the contribution of the scan path dynamic metric in assessing students' performance, we are interested, in the next section, in analyzing the relationship between the alignment metrics and participants' progression through the three attempts of the tasks' resolution. Table 5.4 presents descriptive statistics about the different response attempts in each medical case: the Average Exposure Time (A.E.T) in seconds represents the average of the time spent solving the case, the Total Attempts is equal to the total number of attempts made per case and the Total Error Attempts shows the total number of failed attempts. According to these statistics, the first case is where the students spent more time to resolve as well as the

highest number of attempts because it is the first medical case in the game. In fact, the students were not familiar with the environment, it takes time to understand the interface they were shown. During the first case exposure time, almost half of the time is spent exploring the environment and reading the different symptoms and analysis (A.E.T between the beginning of the case and the click on the diagnostic button = 699 sec instead of 258 sec in the third case).

Table 5.4 Descriptive statistics for the response attempts in each medical case.

	<i>Case 1</i>	<i>Case 2</i>	<i>Case 3</i>	<i>Case 4</i>	<i>Case 5</i>	<i>Case 6</i>
Average Exposure Time	1614.7	880	803.4	510.7	326.7	285.6
Total Attempts	47	36	39	31	13	18
Total Error Attempts	22	16	21	19	3	9

5.4.2.1 Performance assessment across the different attempts

A one-way ANOVA was conducted in order to analyze the alignment metrics (score, gap, match and mismatch) across the three attempts allowed to resolve each medical case.

First, a significant contribution was found across the three attempts in terms of alignment score ($F(3, 52) = 5.193, p < 0.05$). This indicates that at each failure in one attempt, the alignment score decreases which suggests a reasoning sequence that deviates from the reference sequence. Similarly, for the number of gaps and mismatches, statistically significant differences were reported respectively ($F(3, 52) = 3.644, p < 0.05$) and ($F(3, 52) = 24.057, p < 0.001$). These results point out that the alignment metrics are strongly linked to students' performance when reasoning.

We performed a scheffe post hoc test to present detailed group-based statistics for each alignment metric. Results indicated that the alignment score significantly decreases in the third attempt compared to the first and second ones as shown in figure 5.3. These findings clearly show that after two failures, the student cannot produce effective results anymore, especially if he did not yet find the right diagnosis. In terms of sequence alignment, it can be said that the

similarity between the reference sequence and the students' reasoning sequence tend to decrease as the number of attempts increases since the number of gaps and mismatches has almost doubled in the third attempt compared to the first one. In fact, a low alignment score means a low number of matches and a high number of gaps.

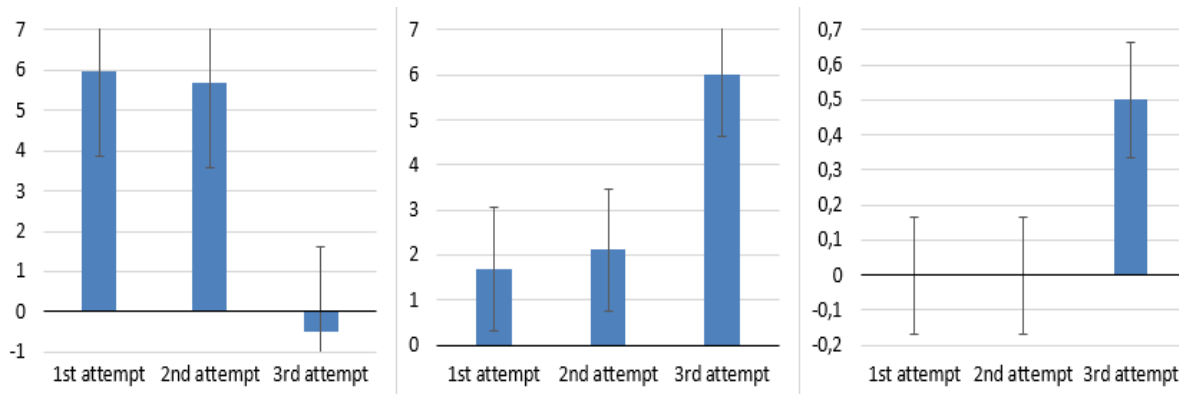


Figure 5.3 Comparison of alignment metrics for the three different attempts. (a) Score (b) Gap (c) Mismatch

To sum up, two methods that assessed the performance of the students as they were resolving medical cases were presented. In the first approach, we used static eye gaze metrics namely, fixation duration and time to first fixation in order to investigate how the participants visually explore the learning environment and specifically what can these metrics reveal about students' visual behaviour. The analysis demonstrated that almost all students have a similar visual pattern revealing that the symptoms' area of interest was fixated the most. Yet, contradictory results have been found showing that focusing on that region has no effect on learners' outcomes. In the second approach, we showed how to model students' eye movements using the dynamic eye gaze metric (i.e. scan path) and differentiate between them in a way to identify the learners who executed correct vs. incorrect reasoning process. Through these experiments, we were able to answer our two research questions by highlighting the importance of using scan paths when analyzing students' eye movements since they enable to monitor moment-to-moment changes in the participants' visual attention.

5.5 Conclusion

In this paper, we proposed a two-fold approach to assess students' reasoning process using first static metrics namely, fixation duration and time to first fixation and second a dynamic metric namely, scan path. An experimental protocol was established by recruiting 15 medical students to record their eye movements. Participants interacted with a medical serious game designed to elicit their analytical reasoning abilities when solving six clinical cases. Results showed that using static metrics, we were able to have a general overview of participants' visual experience within the learning environment. On one hand, we observed that the symptoms area of interest was by far the most fixated area regarding its importance in the reasoning process. However, on the other hand, fixating this area for a long period of time does not always lead to good performance. Then, we used the scan path metric in order to assess students' analytical abilities. This enabled us to discriminate between students in a way that identifies those who performed a correct/incorrect reasoning during medical case resolution.

These findings have potential implications particularly in high-risk medical environments where it is paramount to constantly monitor students' performances. As future works, we aim to propose a multimodal approach to assess students' reasoning performance on the basis of their cerebral and visual behaviour using jointly electroencephalography (EEG) and eye tracking. In the short term, we are planning to study two EEG indices namely, engagement and workload in order to monitor participants' brain activity through the different game phases to investigate whether there are differences in terms of performance. In the long term, we will be interested in developing help strategies that will be used in real time during the interaction with the environment to provide support for the students with the goal of improving their reasoning process.

Acknowledgements

This work was supported by SSHRC and NSERC (National Science and Engineering Research Council). We also thank Issam Tanoubi from the University of Montreal for his collaboration in the design of the serious game.

Chapitre 6.

Une approche d'évaluation du raisonnement de l'apprenant basée sur l'alignement de séquences

Dans le chapitre précédent, nous avons réalisé une étude exploratoire afin de distinguer entre les deux types de mesures statiques et dynamiques quant à leur efficacité comme sources d'information pour analyser les mouvements des yeux de l'apprenant durant son processus de raisonnement. En se basant sur les mesures statiques, nous avons pu obtenir un aperçu global du comportement visuel de l'apprenant en identifiant les zones d'intérêt qui ont le plus attiré son attention. Toutefois, bien que ces mesures soient les plus utilisées en littérature, elles ne sont pas capables de refléter la dynamique des mouvements des yeux dont nous avons besoin pour évaluer le raisonnement. D'où le recours au « scan path » comme mesure dynamique qui est capable de fournir des informations instantanées sur le chemin d'analyse visuelle exécuté par l'apprenant lors de son processus de raisonnement.

Ce chapitre est une continuation des travaux entrepris dans le chapitre précédent dans lequel nous avons montré la faisabilité d'utiliser le « scan path » pour évaluer le raisonnement d'un apprenant. Dans ce chapitre, nous décrivons plus en détail notre nouvelle approche d'évaluation basée sur l'alignement de séquences. Cette approche est fondée essentiellement sur des études menées en bio-informatique qui visent à aligner deux ou plusieurs séquences afin d'identifier des régions de similarité entre elles et représenter les évolutions qui peuvent se produire d'une séquence à l'autre. Ces études sont issues des travaux de recherches menés sur le système génétique humain et sur la conservation des séquences d'ADN.

Aligner deux séquences correspond à calculer le nombre minimal d'opérations d'édition (insertion, suppression et mutation) nécessaires pour transformer une séquence en une autre. En d'autres termes, il s'agit de faire correspondre les caractères qui composent chacune des séquences pour identifier s'il y a une similarité entre elles. Trois types d'alignement de

séquences existent dans la littérature à savoir : l'alignement global, l'alignement local et la recherche de motif (« *pattern searching* »). L'alignement global est approprié dans le cas de deux séquences homologues (appartenant à la même famille). Il s'agit d'évaluer la ressemblance globale entre elles, c'est-à-dire une comparaison sur l'ensemble de la longueur des deux séquences. L'alignement local quant à lui est approprié dans le cas de deux séquences différentes mais ayant des ressemblances locales entre elles. Il s'agit de chercher dans la première séquence des fragments semblables à la deuxième séquence (ou des parties). Finalement la recherche de motif consiste à rechercher dans une séquence donnée toutes les occurrences possibles d'un motif donnée.

Dans ce travail de recherche nous avons en premier lieu écarté la possibilité d'utiliser l'alignement global vu que nous manipulons des séquences hétérogènes non homologues. En deuxième lieu, nous avons commencé par la recherche de motif dans une première étude. Le travail a été accepté en 2017 dans la conférence *International Conference in Intelligent Computing*. Cependant les résultats n'étaient pas satisfaisants. L'algorithme utilisé était seulement capable de détecter les apprenants qui ont accompli un processus de raisonnement correct. C'est pour cette raison que nous avons entrepris une deuxième étude en utilisant l'alignement local qui semble le mieux approprié pour nos recherches.

Notre contribution dans cette étude est d'analyser le « scan path » de l'apprenant dans le but d'évaluer son processus de raisonnement. Cette étude est composée de trois étapes primordiales nécessaires dans l'approche d'analyse de similarité : la première étape consiste à modéliser le processus de raisonnement de l'apprenant en utilisant la méthode de codage à base de caractères (« *string-coding based method* »). Ensuite, la deuxième étape consiste à modéliser la séquence optimale de raisonnement (appelée « séquence référence ») inspirée du raisonnement clinique hypothéticodéductif. Finalement, la troisième étape est le « tracing » où nous utilisons un algorithme d'alignement de séquences pour comparer les deux séquences obtenues dans les étapes précédentes. L'algorithme d'alignement local choisi est le Smith-Waterman qui se base sur la programmation dynamique. A la fin de cette étape, quatre mesures d'alignement sont calculé à savoir le nombre de similarité, le nombre d'insertion/suppression, le nombre de mutation et le score de similarité pour permettre d'évaluer si l'apprenant est en train de suivre le bon processus de raisonnement.

En premier lieu, nous avons corrélé les différentes mesures d'alignement entre elles en utilisant le teste de Pearson pour examiner s'il existe des relations significatives. Les résultats ont montré deux relations statistiquement significatives : d'un côté une relation positive entre le score de similarité et le nombre de similarité et d'un autre côté une relation négative entre le score de similarité et le nombre d'insertion/suppression. Par exemple, si le nombre de similarité augmente et le nombre d'insertion/suppression diminue, ceci suggère que le score de similarité entre les deux séquences est élevé (par rapport à un seuil). En d'autres termes, cela explique que la « séquence référence » est parfaitement alignée avec la séquence du participant impliquant que ce dernier a suivi le bon raisonnement. Toutefois, l'apprenant peut s'égarer et ainsi sa séquence de raisonnement diffère de la « séquence référence » ce qui peut se traduire par un mauvais score de similarité (un nombre de similarité faible et un nombre d'insertion/suppression élevé).

Dans ce travail, nous nous sommes basés sur une 'séquence de référence' adaptée des travaux de (Nendaz 2005) pour évaluer le processus de raisonnement de chaque participant. Par ailleurs, il aurait été souhaitable de se baser sur l'avis de plusieurs experts afin de générer la séquence référence. Ce point représente une limite de notre recherche que nous avons détaillée dans la section travaux futurs du dernier chapitre de ce manuscrit. De plus, il aurait été intéressant de comparer les « scan path » des apprenants entre eux afin de détecter des différences possibles d'un apprenant à l'autre ou au contraire obtenir une tendance générale de raisonnement qui représente les étudiants novices.

La suite de ce chapitre est constituée de l'article de recherche intitulé « *Local Sequence Alignment for Scan Path Similarity Assessment* », publié dans le journal *International Journal of Information and Education Technology* en 2018. Le travail a été supervisé par mon directeur de recherche Claude Frasson. L'article a été révisé et corrigé par Imène Jraïdi.

Local Sequence Alignment for Scan Path Similarity Assessment

Asma Ben Khedher, Jraidi Imène, Claude Frasson

Published in : *International Journal of Information and Education Technology*, 8(7), 482-490 (2018).

<https://doi.org/10.18178/ijiet.2018.8.7.1086>

Abstract

It has long been shown that there is a close relationship between eye movement, human cognition and brain activity. The present work seeks to explore this relationship by investigating the students' saccadic eye movement sequences in a problem-solving task. We aim to assess students' reasoning process in a clinical problem-solving task using students' visual trajectories. We use students' scan path, followed while resolving medical cases, and a local sequence alignment algorithm, to evaluate their analytical reasoning during medical case resolution. An experimental protocol was conducted with 15 participants. Eye movements were recorded while they were interacting with our learning environment. The proposed approach, based on gaze data, can be reliably applied to eye movement sequence comparison. Our findings have implications for improving novice clinicians' reasoning abilities in particular and ultimately enhancing learning outcomes.

Keywords: Cognitive tasks, eye movements, local sequence alignment, medical reasoning, scan path similarity.

6.1 Introduction

Assessing students' learning experience is taken as a compulsory component in virtual learning environments. Indeed, it is extremely difficult to evaluate students' level of understanding during a learning session. Many questions arise: are they assimilating the materials? Do they pay enough attention on the learning targets? Is the reasoning process they perform is correct or incorrect? In order to overcome these issues, several studies tended to measure students' learning outcomes to evaluate their performance. However, learning performance and especially how students learn are not easily measured.

Learning environments are limited in that they do not really focus on the reasons behind the learners' outcomes, meaning, why there was a failure not a success (Westera 2016). Therefore, endowing computer-based learning environments with the faculty of assessing learners' behavior during the learning process is of primary interest, to provide a more valuable learning experience.

Sensing technology (e.g. wearable cameras, bio sensors, eye tracking, etc.) has shown considerable promising results in analyzing learners' behavior and improving their learning performance (S. D'Mello et al. 2007; Jraidi et al. 2012, 2014; Jraidi and Frasson 2010). Chen and his colleagues (2017) developed an attention aware system based on EEG signals to assess students' level of attention within an e-learning environment. The results proved that their system was able to identify low-attention periods that were significantly correlated with post-test scores (C.-M. Chen et al. 2017). Sawyer et al. (2017) introduced facial expression tools to their game-based learning environment in order to predict students learning gains. The findings suggest that facial expression tracking provide additive value compared to baseline predictive student models (Sawyer et al. 2017). Some other studies proposed multimodal sensor-based methods for learning assessment (Chaouachi et al. 2010, 2011; Jraidi et al. 2013; Jraidi and Frasson 2013; Krithika L.B and Lakshmi Priya GG 2016; Taub et al. 2017). Krithika and Priya (2016) proposed a system with three data sources namely, head rotation, eyelid status and facial recognition to detect learners' concentration in order to enhance learning experience. The measured level of concentration was then sent in real-time to instructors to provide students with feedbacks. Such techniques provide unobtrusive measurement of students' behavior within the

learning environment. However, there is a need to provide an in-depth assessment by employing a more efficient tool such as eye tracking. Tracking students' visual behavior is importance in learning environments as a natural source of information to provide pertinent real time data about the location and duration of an individual's eye fixation (Frutos-Pascual and Garcia-Zapirain 2015).

Eye tracking systems have flourished since the past few years due to its ease of use, high sensitivity and especially non-intrusiveness (C.-. Y. Wang et al. 2016; Yeo et al. 2015; Zhang et al. 2017). It has been a very useful tool in many research domains including visualization (Toker and Conati 2014), activity recognition (Courtemanche et al. 2011), affect detection (Jaques et al. 2014) and especially, learning (Ghali et al. 2016; Kardan and Conati 2012; Martínez-Gómez and Aizawa 2014; Yi et al. 2015), to understand how students evolve and progress while learning. In fact, how students are reasoning while solving problems is a fundamental issue, regarding the technical and methodological challenges due to the complexity of the reasoning process. Researches had demonstrated that a relationship exists between *eye movements and cognitive processes* (Chakraborty and McGuire 2016; Hunter et al. 2010). As a result, the use of eye tracking could be beneficial as a mean to identify the link between the task-relevant information and the learner's knowledge state (Ben Khedher et al. 2016).

Current eye tracking studies are using mostly static metrics (Takeuchi and Habuchi 2007). These metrics are measured from the observed data, such as fixation counts within a particular area of interest (AOI). Ben Khedher et al. (2016b) explored eye movements in order to predict learners' level of performance during an interaction with a narrative-centered learning game. They assessed the relationship between static metrics (i.e. mean fixation duration, number of fixations, number of revisits and time to first fixation) and learners' performance. They were able to discriminate between two groups of learners in term of their scores by achieving a good accuracy using the random forest classifier (Ben Khedher and Frasson 2016b). Lum and his colleagues (2016) examined also the effect of the number of fixations on learners' performance. Results indicated that they were positively correlated with the mental workload meaning that when the task was not challenging enough, participants' visual attention decreased (Lum et al. 2016). Yet, despite the increased use of this type of metrics, they lack moment-to-moment analysis of students' visual behavior. As a result, another type of eye tracking metrics has been

used to study the sequential properties of eye movements. Dynamic metrics are measured according to spatiotemporal dimensions, such as the scan path, which is a series of fixations and saccades representing the visual trajectory of a users' eye movement. These metrics provide richer information as they reveal the dynamics of the visual behavior.

This paper proposes to use eye tracking metrics to model the students' reasoning process. In particular, we use the scan path to represent the sequential visual path a student follows while resolving medical cases. We seek to assess the analytical reasoning process the learners use to yield a general conclusion from existing observations. The objective is to discriminate between the learners in a way that identifies those who follow a correct analytical reasoning that will be later referred to as *reference process*. In this approach, we use a well-known method applied in bioinformatics to quantitatively measure the similarity between two biological sequences.

An experimental protocol was conducted in order to record learners' eye movements while interacting with our virtual environment, trying to identify the matching diagnosis and treatment of six medical cases. We describe the methodology used to extract the scan paths and evaluate the students' analytical reasoning. The rest of the paper is ordered as follows: section 2 relates some existing work, section 3 presents the algorithm we will use. In section 4, we present our learning environment and the experimental protocol. In section 5, we explain how we model students' reasoning process. Section 6 discusses the obtained results and finally in section 7 we conclude the paper with some future work directions.

6.2 State of the art

Although most of the current literature use fixation-based and saccade-based metrics for visual analysis, there is a growing interest in using the scan path as a primary metric, as it enables to monitor moment-to-moment changes in the individual's attention and focus (Cutrell and Guan 2007; Hayhoe and Ballard 2005). It has thus far been used in diverse domains such as medicine (Hood and Raza 2014; Tien et al. 2012), visualization (Eraslan et al. 2015; Kurzhals et al. 2014; Raschke et al. 2014) and affect detection (Jaques et al. 2014). Since then, the use of scan path in educational research has gained much interest, since it is closely related to information processing according to (Noton and Stark 1971). Ho et al. (2014) examined students' visual behaviors when reading a web-based scientific report including textual and graphical contents.

They aimed to investigate whether students' level of prior knowledge (high and low) has an impact on their reading process. Results showed that both types of students spent more time reading text than graphics with a higher fixation duration on the graphic for the high prior knowledge students, which implies that the other students were struggling when reading a report that integrates both text and graphic information (Ho et al. 2014). In the same context, Lee and Wu (2018) explored students' scan paths when reading geometric descriptions. They found the same results which imply that more than 80% of the students were text-directed (Lee and Wu 2018). In a problem-solving task, the scan path of 40 students was recorded while solving simple algebraic equations in order to provide insight into how they rearrange them. The results indicated that expert students perform a well-defined and organized visual strategy compared to non-expert students who fixated more the presented solutions than the equations themselves (Susac et al. 2014).

It is obvious that the scan path provides precise and temporal information about the users' visual attention. However, some crucial questions arise: how to analyze a students' scan path? What is the most suitable approach? Can the recorded scan paths be applied to assess students' reasoning process? We will be answering the first question by presenting thereafter, the main existing approaches for scan paths' analysis. Concerning, the second question, there has been a long discussion going on but no answer really exists. Although, we describe in the next section, the method we use in this paper to analyze the students' scan path. As for the last question, it is examined in the results and discussion section.

Scan path analysis methods can be classified into four categories: similarity/dissimilarity calculation, transition probability calculation, pattern detection and common scan path identification (Eraslan et al. 2015). The first method compares two scan paths. The second determines the probability of a visual element to be before or after a particular element in the scan path. The third method identifies a specific visual pattern among the observed scan paths or detects the patterns with a high degree of matching. The fourth method detects a common visual pattern within all the identified scan paths. In the present work, we deal with the scan path similarity. Three main methods for scan path comparisons are described below.

String-coding based methods. These methods are the most common used techniques known in the literature. The scan path is recoded as a string of characters. The recoding consists in assigning a letter to each region of interest. The obtained sequences are then compared using string comparison methods. The Levenshtein Distance is based on the string-edit algorithm which computes the minimum number of editing operations (insertion, deletion and substitution) required to transform one scan path to another (Levenshtein 1966). Eraslan and his colleague (2015) developed the eMine scan path algorithm to identify a common scan path among the users when visualizing web pages in order to improve users' experience. They use the string-edit algorithm as a step in their approach to find in a list of scan paths, the two most similar ones (Eraslan and Yesilada 2015). Another popular algorithm is ScanMatch which computes the similarity between two scan paths rather than computing the distance, using the Needleman-Wunch algorithm utilized in the bioinformatics domain for comparing two DNA sequences (Cristino et al. 2010).

Visual methods. Eye tracking offers a visual attention tool namely, the heat map. It highlights the frequently fixated regions by the user. The areas with a high number of fixations and/or fixations duration are called hotspots (Sundstedt 2012). Apart from the visual aspects, in such methods, comparing two scan paths remains difficult, hence the use of another graphical method. The Dotplot is used in bioinformatics for visualizing similarities in genetic sequences (Huang and Zhang 2004). Goldberg and Helfman (2010) applied the dotplot technique to represent and identify similar scan paths. As for the string-edit methods, scan paths are coded as a string of characters and then aligned on the axis of a two-dimensional matrix. Each shared character is marked with a dot and finally, the algorithm consists in finding the longest straight line in the matrix representing the common scan path between the two sequences (Goldberg and Helfman 2010).

Geometric methods. Scan paths are coded as multidimensional vectors representing fixations and saccade properties (e.g. location, fixation durations, etc.) and the objective is to compute the minimal distance mapping between the vectors. Jarodzka et al. (2010) proposed MultiMatch, a geometric method. Using a vector with five properties namely, length, position, direction,

shape and duration, they search for the shortest route (i.e. scan path) between all the fixations based on the Dijkstra algorithm (Jarodzka et al. 2010).

In this paper, we propose a **string-coding based method** to assess the students' reasoning process while resolving clinical cases. First, we use the scan path to model the student's analytical reasoning. Then, we seek to discriminate between learners in a way to identify those who follow a correct reasoning using a local sequence alignment algorithm namely, Smith-Waterman. We are not the first to perform scan path alignment. There is the work of Cristino and his colleagues (2010) who applied a global alignment based on the Needleman-Wunch algorithm within a web-based visual search task (Cristino et al. 2010).

6.3 Sequence alignment algorithm

Comparing two nucleotides or amino acid sequences to quantify the degree of similarity between them is a commonly used technique in bioinformatics. In the human genetic system, it is fundamental to discover if a DNA sequence is conserved from one species to another or how many similar residues genes they share. Sequence alignment algorithms are intended to identify regions of similarity between the sequences and represent their evolutionary relationships (Pearson 2013).

Aligning two sequences (pairwise alignment) or multiple sequences (multiple alignment) refers to compute the minimum number of editing operations (insertion, deletion and mutation) required to transform one sequence into another. That is, corresponding the sequences' characters to find out if they are similar. Three possible situations can occur when comparing two characters, namely, match, mismatch and gap. A match means that the two characters are identical (A-A, B-B and Y-Y). A mismatch requires a mutation, which is substituting one character for another (G-O, A-D and R-K) and a gap implies an insertion or a deletion in one of the two sequences (a gap in sequence one leads to an insertion in sequence two and vice versa). Given an alignment between Seq 1 and Seq 2, figure 6.1 shows these different situations.

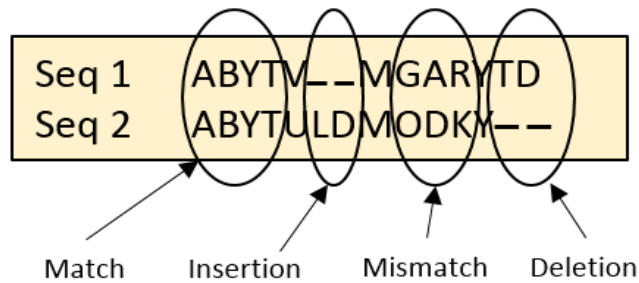


Figure 6.1 Editing operations for aligning two sequences.

To quantify the alignment between two sequences, a score is computed. It can measure either the similarity or the distance between the sequences. An elementary score is assigned for each position of the alignment where two characters are compared and the final score is the sum of these elementary scores [50]. Let us consider the two previous sequences Seq 1 and Seq 2, we associate for example $+3$ if match, -1 if mismatch and -2 if gap. Figure 6.2 shows the alignment of the two sequences with the associated scores. In this example, there are five matches, four mismatches and four gaps, given a total score of $3 = 5*(+3) + 4*(-1) + 4*(-2)$.

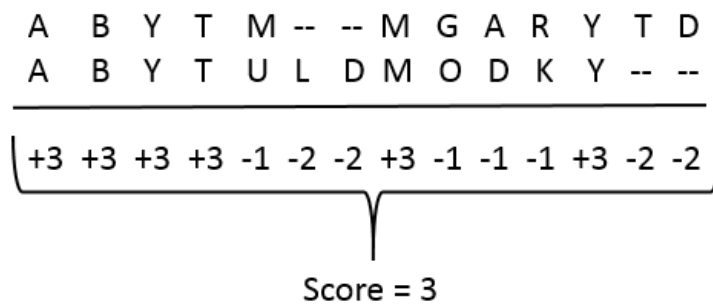


Figure 6.2 An alignment of two sequences with the elementary scores for each two characters.

6.3.1 Local alignment

Thus far, biologists attempt to perform an end-to-end alignment, meaning over the whole length of both sequences. However, years later, they found it more interesting to focus on isolated regions of similarity between the sequences, which is not feasible by a global alignment; hence, the new definition of local similarity, known as local alignment. A local alignment of two sequences s and t is an alignment of a subsequence of s and a subsequence of t . It is more

appropriate than the global and this for different reasons. The first reason is that global alignment compares two sequences that have approximately the same length and are quite similar, whereas in the local alignment arbitrary-length segments of each sequence can be aligned with no penalty for the rest of the sequence regions. In fact, the sequences we use in this study are different in size and are quite different from one another since no participant has the same eye movements as the others. Second, the global alignment, as we have presented earlier, performs an alignment over the entire length of the sequences whereas, what really matters for us, is the regions of the two sequences that coincide the best. That is to find the sub-sequences that have a high level of similarity. Finally, it is more relevant to use local alignment as it is faster and above all, it can be easily considered as global alignment if the two sub-sequences are similar.

There exists another sequence alignment method known as pattern matching which consists in searching for all the approximate occurrences of a particular pattern in a given sequence. In a previous work, we applied this algorithm to our experimental data (Ben Khedher et al. 2017a, 2017b). We ran bivariate correlations to assess the relationships between the alignment score and students' results in the medical cases. However, no statistically significant results were found. Hence the need to change the algorithm and to test the local alignment and more specifically the Smith Waterman algorithm, which is the most often used algorithm in biology when looking for local similarities between two sequences.

6.3.2 Smith-Waterman algorithm

The Smith-Waterman alignment algorithm is based on dynamic programming to compare two sequences. Dynamic programming is a method by which the original problem is divided into smaller sub-problems. In the alignment context, it can be used to compute the optimal alignment scores between two given sequences. Basically, the concept is that each elementary score in the alignment is computed using the previously calculated scores (i.e. the characters that were aligned before the current alignment).

The algorithm creates first a similarity-based matrix where a score is computed for each two characters' alignment, and second selects the optimal alignment from all the possible alignments found. Let s and t two sequences of length m and n respectively. We create a $(m+1)$

x ($n+1$) matrix where the first row and column are initialized with zeros. The remainder cells are filled using the following recursion relation shown in (1).

$$V(i, j) = \max(0, V(i-1, j) + g, V(i, j-1) + g, V(i-1, j-1) + p(s_i, t_j)) \quad (1)$$

where $V(i, j)$ represents the similarity score between the i^{th} element of sequence s and the j^{th} element of the sequence t , g is the gap penalty and $p(s_i, t_j)$ is a match if $s_i = t_j$ and a mismatch if $s_i \neq t_j$. The zero in the recursion allows to ignore possible negative alignment scores. If the score for a given position in the matrix is negative, it is replaced by the zero value and the alignment may be interrupted. It means that we can restart a local alignment at any position.

Once the matrix is completely filled, we find the cell with the highest score that represents where the alignment ends. Then we do a trace back through the matrix until reaching the first zero, which represents the beginning of the alignment. The constructed path represents one of the optimal local alignments since many alignments may exist for two sequences if we have the same maximum value throughout the matrix. Figure 6.3 illustrates an example of an optimal local alignment for two sequences $s = \text{TTCATA}$ and $t = \text{TGCTCGTA}$.

	T	G	C	T	C	G	T	A	
0	0	-6	-12	-18	-24	-30	-36	-42	-48
T	-6	5	-1	-7	-13	-19	-25	-31	-37
T	-12	-1	3	-3	-2	-8	-14	-20	-26
C	-18	-7	-3	8	2	3	-3	-9	-15
A	-24	-13	-9	2	6	0	1	-5	-4
T	-30	-19	-15	-4	7	4	-2	6	0
A	-36	-25	-20	-10	1	5	2	0	11

Figure 6.3 Similarity matrix using the local alignment (the red arrows represent the optimal alignment path).

6.4 Experimental protocol

We conducted an experiment to record the students' eye movement activity while interacting with our virtual environment called Amnesia. This system was specially designed for the experiment.

6.4.1 Amnesia

Amnesia is a medical serious game that assesses the cognitive abilities of novice medical students through clinical decision-making. The curriculum underlying the environment was designed by a medical professional and approved by a doctor. The game features a virtual hospital where the user plays the role of a doctor who was mistakenly diagnosed with amnesia and found himself trapped within the hospital. The player has to prove that he does not suffer from this disease, by resolving in a first step some cognitive tasks (such as memory, attention and logic tests). Memory and attention tests involve quite funny and simple exercises such as card matching, divided attention (simultaneously responding to multiple tasks) and visual memory. Logic tests include more complex tasks that involve inferential skills on numerical and verbal reasoning such as number series and analogies. Then, in a second step, the player has to demonstrate his clinical skills by resolving medical cases.

6.4.2 Medical cases

The game includes six medical cases. For each case, the students are asked to find out the correct diagnosis and the appropriate treatment. To this end, the players are instructed to analyze a series of observations including the patient's demographic information, antecedents, symptoms and clinical data.

The different diseases to be identified are respectively flu, bacterial pneumonia, measles, ebola, mumps and whooping cough. For each diagnosis and treatment, different response alternatives are given. For the diagnosis, there is only one correct answer, which is the current disease. The students are given three attempts to find out the correct answer. Once the diagnosis is established, the student has to identify the right treatment, the same scheme is used as for the diagnosis. The students are given three attempts to respond; after two errors the game is over

and there are up to three possible correct responses. Between the attempts, the student can re-analyze the patient in order to correct his answer.

6.4.3 Participants and apparatus

Fifteen participants (8 males) aged between 20 and 27 ($M = 21.8 \pm 2.73$) were recruited for the experiment. Participants were all undergraduate medicine students at the University of Montreal. Upon arrival at the laboratory, they were briefed about the experimental objectives and familiarized with the material. They were asked to sign a consent form and placed in front of the eye tracker. They were informed that free head movements were allowed.

A Tobii Tx300 eye tracker was used with a sampling rate of 300 Hz. It includes infrared sensors and a camera, all were integrated within a 23-inch computer monitor (1920 x 1080) resolution. The eye tracking system was placed at a distance of approximately 65 cm from the participants' eyes. A 9-point calibration grid was used to evaluate the quality of the students' measured gaze points. This calibration process is essential since it influences the reliability of the data. Once the calibration process is established, the session begins with the game's introductory scene where the players are instructed to discover the environment and meet the different cognitive tasks. Guidance messages were used to help them move forward in the hospital.

At the end of the game session, the participants were shown with the medical cases they were resolving and prompted to recall their reasoning process. They were asked to self-report their visual sequence (i.e. the order of the areas they looked at). They were also asked to give their opinion about the game design and usability in order to have feedbacks for potential corrections.

6.5 Data analysis

Our main objective was to demonstrate how a sequence alignment algorithm could be applied to eye movements in order to assess students' analytical reasoning while resolving medical cases. After completing the recording process as described in the previous section, participants' eye movement sequences were extracted from the collected eye gaze data. First, we aimed to

inspect whether we are able to visually model students' reasoning process. Second, given the reasoning sequences, we were interested in assessing their correctness.

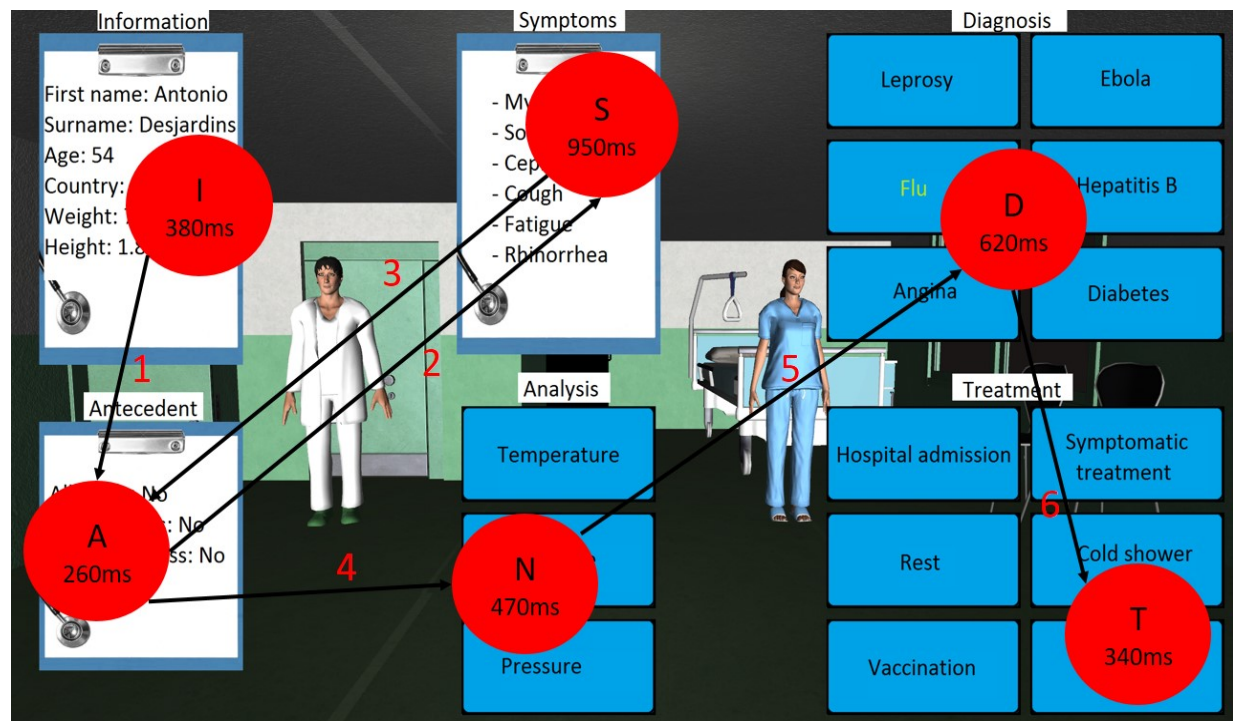


Figure 6.4 An example of scan path over the six AOIs.

6.5.1 Students' reasoning process

In this study, we need first to represent the visual steps the students use to resolve the medical cases as a sequence. To that end, we used the recorded eye movements to extract, for each medical case, the sequential visual path the students followed to yield their conclusions from the existing observations. First, the obtained data were filtered to fixations. We considered only the regions on which the learner focuses with a fixation duration above 250 ms; this threshold was chosen according to previous studies suggesting that eye fixations should last over 200 ms to be considered as meaningful (Liversedge and Findlay 2000) Then, to be able to represent the students' visual scan path as a sequence, we needed to divide our environment into areas of interest (AOI).

We defined six areas and assigned a letter to each one, namely (I) Information, (A) Antecedent, (S) Symptoms, (N) aNalysis, (D) Diagnosis and (T) Treatment (see Figure 6.4);

each fixation occurring within a specific area is then coded with the corresponding letter. In this way, we obtain a scan path recoded as a sequence of characters as follows “IIASSSSANDDDTT”. Each character corresponds to a visited AOI, and redundant characters refer to areas that have been visited many times successively. Finally, redundancies are removed to maintain only one occurrence of each visited area and create at the end this sequence of eye movement “IASABDT” representing students’ reasoning path.

Reference sequence.

Once we modelled the students’ reasoning process, our next investigation was to assess whether the learner is performing the correct reasoning. More precisely, we need to model this correct reasoning in order to check whether the student’s followed scan path, matches the correct sequence. To this end, we used the hypothetico-deductive clinical reasoning shown in figure 6.5 (Adapted from (Nendaz et al. 2005)) in order to represent the correct sequence, which will be called hereafter reference sequence.

In this type of analytical reasoning, the clinician generates initial hypotheses at the beginning of the medical interview based on the symptoms exposed by the patient. Then additional clinical data (analysis, radiography, medical antecedents, etc.) are collected and interpreted to evaluate these hypotheses that will be validated or discarded until reaching a correct diagnosis and treatment. From this analytical process, we draw a sequential representation of the different steps a student needs to perform to have a correct reasoning. We obtain thus, this scan path: $I \rightarrow S \rightarrow A \rightarrow N \rightarrow D \rightarrow T$, which yields the following reference sequence: “ISANDT”.

6.5.2 Sequence comparison

The aim of this paper is to investigate whether the reference sequence matches the student’s reasoning sequence by quantitatively comparing two eye movement sequences using the Smith-Waterman alignment algorithm. The algorithm uses a similarity-based substitution matrix seeking to find the best alignment with the highest score (Nakamaru 2011), which implies a strong similarity between the two sequences, meaning that the learner followed the correct reasoning process.

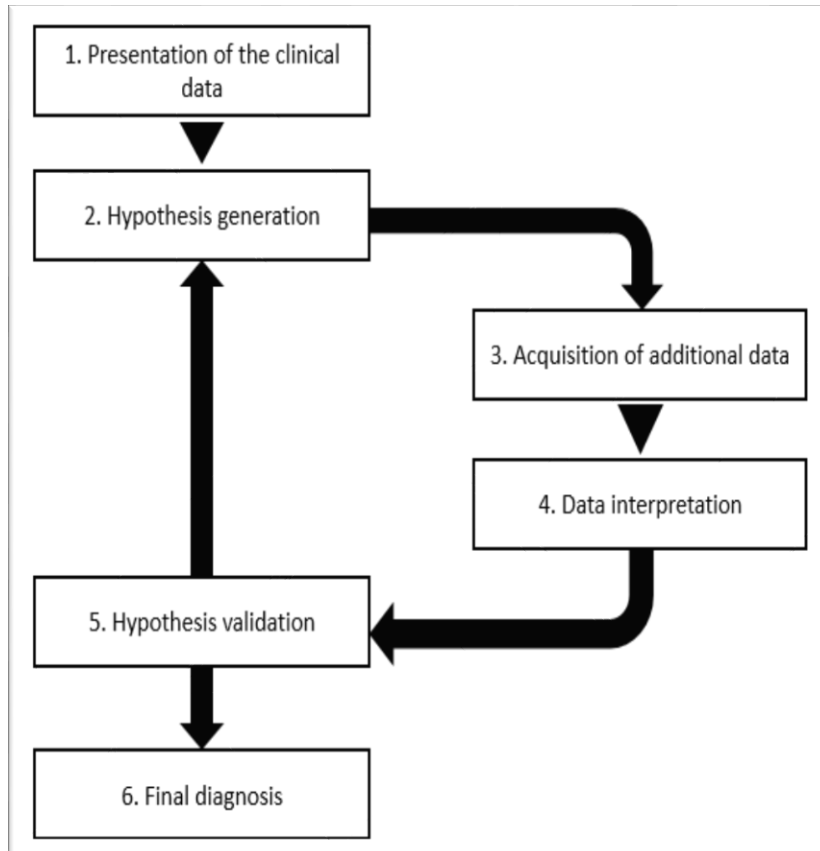


Figure 6.5 The hypothetico-deductive reasoning process.

Figure 6.6 presents two examples of the local alignment algorithm execution for two different participants. As it can be seen in the figure, we describe the similarity matrix as well as a possible optimal alignment we can obtain for the two sequences. The scoring metrics are set as follows: 2 for match and -1 for mismatch and gap, since we claim to decrease the number of observed differences between the two sequences and increase the alignment score. Once the scores in the array are computed, we search for the maximum value and where it comes from. The cell with the highest score will be used as a starting point to search for the optimal local alignment which is constructed by tracing back through the matrix until we reach an entry with a score of zero. On the left side (Figure 6.6 (a)), an example of a participant who performed a

correct reasoning process is shown, as we can notice from the obtained alignment score equal to 12. This means that the input reference sequence is perfectly aligned with the participant sequence. In other words, this implies that the student followed a good reasoning process. At the opposite, the learner’s reasoning process can deviate from the reference sequence and hence errors (mismatches and gaps) can occur during the alignment. Figure 6.6 (b) shows an alignment with two errors where the obtained score is equal to 4, due to the two obtained gaps.

We note in this last example that the alignment was not done over the entire length of the first sequence but only on the beginning. It is, in fact, a local alignment characteristic, which consists in selecting the best possible alignment with the subsequence offering the highest score. It can be easily noticed in the table that there is another maximum value equal to four, but the algorithm did not select it as a starting point for the alignment, as it gives a very low score with a lot of gaps which is not optimal.

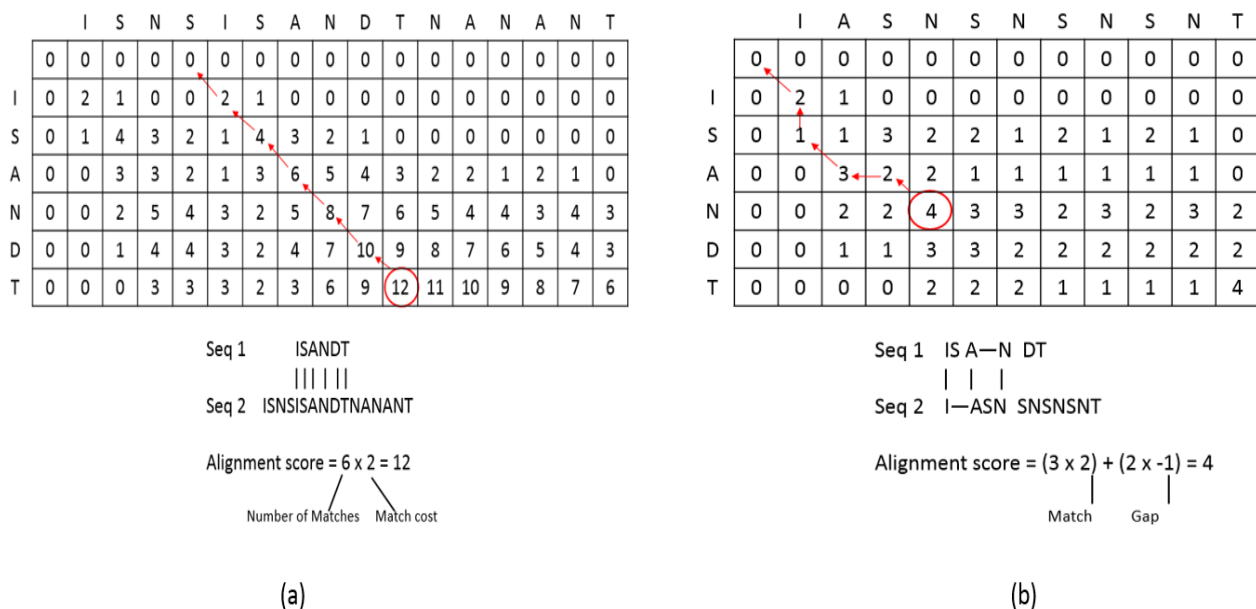


Figure 6.6 Two examples of sequence alignment with the corresponding similarity matrices. (a) A perfect match between the sequences. (b) An alignment with two errors “|” means a match and “—” means a gap.

Using this proposed approach, four alignment metrics were extracted: score, number of matches (NbMa), number of mismatches (NbMis) and number of gaps (NbGap). These metrics

were correlated with the performance-based features: success in the medical case (Yes or No), number of attempts in the diagnostic (NbAtmpDiag) and number of attempts in the treatment (NbAtmpTreat) to investigate the relationships between the learners' performance during medical cases resolution and eye movement alignment results. One-way analyses of variance (ANOVA) were also performed to check if these alignment metrics are significantly different among diagnostic and treatment attempts.

6.6 Results and discussion

We start first by considering the relationships between the alignment metrics and second their relationships according to students' performance recorded across the attempts in order to highlight the reliability of using a sequence alignment algorithm in assessing students' reasoning.

Statistically significant Pearson's correlation coefficients (r) were found between the score and the number of matches and gaps metrics respectively. We found a positive high correlation between the score and the NbMa ($r = 0.637$) and a high negative correlation ($r = -0.815$) between the score and the NbGap. These correlations were statistically significant ($p < 0.01$) suggesting that the alignment score depends on these two metrics and so the participants' reasoning. The more the alignment score is high (high number of matches and low number of gaps), the more the student performs a good reasoning.

6.6.1 Association of alignment metrics and the number of attempt features

We ran bivariate correlations to assess the relationships respectively between the alignment metrics and the NbAtmpDiag feature and the NbAtmpTreat feature as shown in table 6.1. No significant correlation was found with regards to the number of attempts in the treatment feature. However, for the number of attempts in the diagnostic, statistically significant Pearson's correlation coefficients were found.

Two alignment metrics showed a statistically moderate correlation: score ($r = -0.310$, $p = 0.024$) and NbMis ($r = 0.477$, $p = 0.000$) and a low positive correlation was also found ($r = 0.288$, $p = 0.036$) with the number of gaps. These results are consistent since a high number of

attempts in the diagnostic is likely to lead to a failure in the medical case resolution. In fact, a failure means a low alignment score and a high number of gaps.

6.6.2 Performance assessment through the attempts

In addition, we explored whether there were statistically significant differences between the alignment metrics when the students were reasoning across the different attempts of the medical case resolution. Statistical testing was performed using a one-way analysis of variance (ANOVA). We reported significant differences during the diagnostic attempts, however no differences were reported concerning the treatment.

First, a significant difference was observed across the three attempts in terms of the alignment score ($F(3, 52) = 5.193, p = 0.09$). This suggests that at each failure in an attempt, the alignment score decreases which implies a reasoning sequence that deviates from the reference sequence. Similarly, for the number of gaps and mismatches, statistically significant differences were found respectively ($F(3, 52) = 3.644, p = 0.036$) and ($F(3, 52) = 24.057, p = 0.000$). These findings highlight the fact that the alignment metrics are strongly linked with the students' performance when reasoning.

Table 6.1 Bivariate correlational results.

<i>Correlation between the NbAtmpDiag and the alignment metrics (N = 53)</i>		
Alignment metric	r	p
Score	-0.310*	0.024
NbMa	-0.111	0.429
NbGap	0.288*	0.036
NbMis	0.477**	0.000

*Correlation is significant at the 0.05 level.

**Correlation is significant at the 0.01 level.

A scheffe post hoc test was also performed to present detailed group-based statistics for each alignment metric. The results show only three groups corresponding to the three attempts given to students to find out the diagnosis, meaning that there is no failure in the diagnostic step. Figure 6.7 shows the results of the means of score, gap and mismatch metrics for the three different attempts. For the score metric, we found that the alignment score significantly decreases in the third attempt compared to the other attempts. These findings clearly show that at the end of the second attempt, the student can no longer reason properly, especially if he has not yet found the right diagnosis. This is reflected in the alignment score (Figure 6.7 (a)) where it decreased in the second attempt and it has even dropped below zero in the third attempt. The same results are reflected in the other metrics (i.e. gap and mismatch). One can notice the number of gaps which almost doubled as well as the number of mismatches.

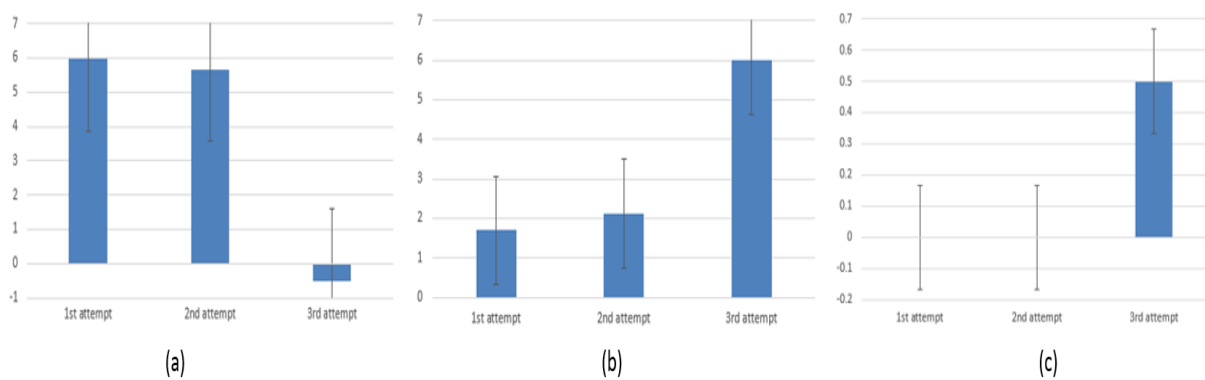


Figure 6.7 Comparison of alignment metrics for the three different attempts. (a) Score (b) Gap (c) Mismatch.

6.7 Conclusion

In this paper, we have presented our approach for comparing two scan paths extracted from students' eye gaze while interacting with the Amnesia learning environment. The method is based on a string representation of the scan path and a local sequence alignment algorithm, namely, Smith-Waterman. The algorithm is used to compute the similarity between two sequences. An experimental protocol was conducted with 15 participants during medical case resolution tasks. They had to identify in each case the correct diagnostic and treatment. An eye

tracker was used to record their eye movement and then the obtained scan paths were compared to a reference pattern to compute the similarity between them.

The objective was to evaluate the analytical reasoning process of novice medical students. The approach revealed that we were able to discriminate between the learners in a way to identify those who performed a correct/incorrect reasoning process. These findings could be applied in a more general context for human-computer applications that seek particularly to monitor the user's decision-making process within problem-solving tasks.

As future work, we have planned the classification of both alignment and eye tracking metrics to strengthen our approach and attempt to predict during the task resolution whether the student will perform a good or bad reasoning process. We also plan to integrate further physiological variables such as electroencephalography to assess both the student's visual attention and mental state.

Acknowledgement

We acknowledge SSHRC (Social Science and Human Research Council) through the LEADS project and NSERC (National Science and Engineering Research Council) for funding this research. Thanks to Issam Tanoubi from the University of Montreal for his collaboration in the experimental design.

Chapitre 7.

Analyse de l'état mental de l'apprenant lors d'une activité de raisonnement

Ce chapitre est dédié au deuxième volet de nos recherches qui porte sur l'analyse cérébrale de l'état mental d'un apprenant durant une activité d'apprentissage. Comme nous l'avons mentionné dans l'introduction de cette thèse, l'un de nos objectifs principaux est de suivre l'état mental de l'apprenant afin d'identifier des indicateurs cérébraux permettant d'évaluer le processus de raisonnement. Nous proposons dans ce chapitre une étude expérimentale dans laquelle nous mesurons un index d'engagement mental extrait à partir des signaux EEG.

Dans cette étude, nous avons exploré la faisabilité d'utiliser les EEG comme source d'information pour le suivi de l'état mental d'engagement des apprenants au cours d'un processus de raisonnement clinique. Pour cela, une étude a été réalisée en utilisant la deuxième expérimentation (II) faite avec Amnesia dans laquelle des étudiants novices en médecine doivent résoudre six cas médicaux. Ce travail de recherche est fait en trois étapes : tout d'abord, nous avons suivi l'évolution du niveau d'engagement des apprenants durant les trois phases de la tâche de résolution du problème clinique. Ensuite, nous avons analysé la distribution de l'indice d'engagement à travers les différentes zones d'intérêt qui composent l'environnement. Enfin, nous avons examiné la relation entre l'engagement mental de l'apprenant et ses performances.

Dans un premier temps, nous avons analysé comment le niveau d'engagement des apprenants évolue au cours des trois phases du processus de résolution des cas cliniques à savoir : l'exploration, le diagnostic et le traitement afin d'identifier s'il existe une période de temps particulière qui a suscité le plus d'attention. Les résultats ont montré que durant les trois phases le niveau d'engagement était le même chez tous les participants. Cependant, une seule variation particulière (un niveau d'engagement mental plus élevé) a été identifiée au niveau du

cas 1 et au cours de la phase d'exploration puisque c'est le premier cas médical à résoudre. En effet, l'interface qui est affichée diffère des autres scènes présentées dans l'environnement ce qui a tendance à pousser l'apprenant à être plus attentif et de fournir un effort mental supplémentaire pour comprendre la manière de résoudre le cas.

Dans un deuxième temps, nous avons examiné comment l'indice d'engagement est réparti entre les différentes zones d'intérêt pour déterminer les zones qui ont le plus attiré l'attention des apprenants. Des tendances similaires ont été observées dans la plupart des cas médicaux. D'un côté la zone nommée « traitement » est celle où les apprenants étaient le plus concentrés vu qu'elle nécessite plus d'effort mental que les autres. De l'autre côté, la zone nommée « antécédents » est la moins engageante de toutes. Ces résultats sont à l'opposé de ce que nous avons obtenu dans l'étude précédente où la zone « symptômes » était la plus fixée ce qui suggère qu'une durée de fixation plus longue n'implique pas forcément un niveau d'engagement élevé. De même un niveau d'engagement élevé au niveau d'une zone d'intérêt en particulier n'implique pas que cette attention particulière aura un impact sur les performances. Ce qui a été validé dans la troisième partie de cette étude.

Finalement, dans la dernière partie nous avons étudié la relation entre l'engagement mental et les performances en explorant la question suivante : est-ce que le fait d'être concentré sur des informations spécifiques pertinentes peut potentiellement amener les apprenants à résoudre correctement les cas médicaux ? L'analyse statistique des données a confirmé qu'il existe une relation statistiquement significative entre performances et indice d'engagement. Toutefois, il n'existe pas une zone unique en particulier qui a un impact direct sur la réussite ou l'échec.

Le reste de ce chapitre est constitué de l'article intitulé « Tracking Students' Mental Engagement Using EEG Signals during an Interaction with a Virtual Learning Environment » publié dans le journal *of Intelligent Learning Systems and Applications* en 2019. Ma contribution dans cet article était de surveiller l'état mental de l'apprenant dans une activité de raisonnement en mesurant son indice d'engagement. L'objectif est de pouvoir détecter les zones ou les phases dans l'environnement d'apprentissage qui suscitent le plus d'engagement mental et si tel est le cas est ce qu'elles ont un impact sur les performances ? Nous avons pu répondre à cette dernière question, toutefois nous avons voulu consolider nos résultats avec une autre

approche où nous demanderons à l'aide d'un posttest l'avis des participants afin de savoir d'un côté quelle AOI a le plus attirée leur attention et de l'autre côté quelle AOI les a aidée pour construire leur diagnostic. Nous n'avons pas pu malheureusement faire compléter le test par les participants pour deux raisons : la première est une contrainte temporelle vu que la session dure à peu près 1h30 de temps (20 à 25 minutes pour formulaire de consentement, prétest et installation du matériel en plus de 30 à 40 minutes de jeu) qui est beaucoup pour des étudiants en premières années de médecine avec un emploi du temps très chargé. La deuxième raison est qu'à la fin du jeu on leur demande de remplir un premier posttest. Nous estimons que la session est déjà assez remplie et qu'il est très difficile de garder les participants attentifs et surtout motivés plus qu'une heure ce qui risque d'affecter leurs réponses au deuxième posttest si nous avons décidé de le faire.

Ce travail de recherche a été supervisé par mon directeur de recherche. Imène Jraïdi a contribué avec ses conseils et ses corrections dans la rédaction de l'article. Par ailleurs, le système que j'ai utilisé pour le calcul de l'index d'engagement est celui développé par Maher Chaouachi (2015) dans le cadre de ses travaux de recherche (Chaouachi 2015).

Tracking Students' Mental Engagement using EEG Signals during an Interaction with a Virtual Learning Environment

Asma Ben Khedher, Imène Jraidi, Claude Frasson

Published in : *Journal of Intelligent Learning Systems and Applications*, 11(1), 1-14, (2019).

<https://doi.org/10.4236/jilsa.2019.111001>

Abstract

Monitoring students' level of engagement during learning activities is an important challenge in the development of tutoring interventions. In this paper, we explore the feasibility of using electroencephalographic signals (EEG) as a tool to monitor the mental engagement index of novice medicine students during a reasoning process. More precisely, the objectives were first, to track students' mental engagement evolution in order to investigate whether there were particular sections within the learning environment that aroused the highest engagement level among the students, and, if so, did these sections have an impact on learners' performance. Experimental analyses showed the same trends in the different resolution phases as well as across the different regions of the environments. However, we noticed a higher engagement index during the treatment identification phase since it aroused more mental effort. Moreover, statistically significant effects were found between mental engagement and students' performance.

Keywords: EEG, Engagement Index, Learners' Performance, Computer-based Learning Environments

7.1 Introduction

Endowing computer-based environments with the faculty of monitoring users' experience is an important step toward improving human-computer interactions and understanding users' needs (Ben Khedher et al. 2017a, 2018b; Jraidi et al. 2013). More precisely, the use of neurophysiological sensors such as heart rate, galvanic skin response, body temperature and electroencephalography (EEG) is gaining importance continuously as methods to assess users' behavior and mental state.

In particular, EEG data provide valuable quantitative and unbiased information on brain activity in a millisecond time-frame. It has become a low-cost and a non-invasive tool that is increasingly being used in a wide range of applications, including: medical diagnosis (e.g. patient coma monitoring and epilepsy detection) (Ben Hamida et al. 2015; Loo et al. 2016; Roach and Mathalon 2008), emotion recognition (e.g. boredom assessment) (Kim et al. 2018; Soleymani et al. 2016; Zhuang et al. 2017), affective modeling (e.g. uncertainty modeling, emotional state classification) (Jraidi et al. 2014; X.-W. Wang et al. 2014), performance assessment (e.g. outcome prediction, learners' classification) (Ben Khedher et al. 2018c; Jraidi et al. 2012; Jraidi and Frasson 2010; Lujan-Moreno et al. 2016; Van der Hiele et al. 2007) and users' mental activity assessment (Aricò et al. 2016; Keith et al. 2015; S. Wang et al. 2016).

Monitoring learners' mental state is of primary interest particularly in computer-based learning environments (C.-M. Chen et al. 2017; Jraidi and Frasson 2013; Mills et al. 2017). The ability to recognize and measure students' attention during the learning process is an important part of a successful knowledge acquisition since it impacts their cognitive performance. Therefore, obtaining an accurate view of learners' mental state may allow interactive learning systems to adjust the tutoring content, formulate the appropriate help strategies and enhance learning outcomes.

In this paper, we recorded learners' electrophysiological activity as they interact with a medical serious game in order to track their mental state and assess their learning performance.

We aim first to monitor EEG signals in order to explore how students' mental engagement evolve across the different phases of the learning environment. Second, we examine whether being engaged during particular phases of the game or paying attention to specific relevant areas of the environment have an impact on the learners' outcomes.

The remainder of the paper is organized as follows: Section 2 outlines some previous work concerning students' engagement assessment within computer-based learning environments. Section 3, describes our experimental setup Section 4, discusses the obtained results and section 5 presents a conclusion and future works.

7.2 Related work

The key to a successful learning experience is the ability for educators to provide adequate assistance to learners. In this context, many researchers are currently devoting a growing interest in assessing students' mental engagement in an effort to keep students engaged during computer-based learning interactions (Henrie et al. 2015; Jraidi et al. 2014). Engagement is a mental state that can be seen as attention, involvement, focus, or interest (Nakamaru 2011; Yang 2011). In the education context, the engagement state consists in deploying all the mental mechanisms involved in information processing to achieve optimal learning performance (Azevedo 2015). Both qualitative and quantitative methods are used to measure students' engagement.

The quantitative approaches are the most common measures. They are based on self-assessment tools such as questionnaires and surveys to be completed by the student. Self-report measures are used to assess students' emotional and cognitive engagement (P.-S. D. Chen et al. 2010; Fredricks and McColskey 2012; Jaafar et al. 2014). Pintrich and De Groot (1990) used a self-report questionnaire to measure students' self-regulated learning components, and how these are related to students' performance and cognitive engagement. The results demonstrated a positive correlation between self-efficacy and cognitive engagement. In fact, students who are self-regulating are more cognitively engaged and perform better than the others. These methods are practical, low cost and easy to use for large samples and distant learning. However, the obtained data are not entirely reliable since the responses may be biased. In fact, as the learners fill in the self-report questionnaires by themselves, the data will not be enough objective to draw

effective conclusions about students' engagement. One of the proposed solutions to face these issues is the use of qualitative approaches.

Qualitative methods measure students' engagement by means of various techniques such as observations, interviews and educators' ratings (Alford et al. 2016; Helme and Clarke 2001; Wigfield et al. 2008). Helme and Clarke (2001) used interview data to identify indicators of cognitive engagement during mathematics lessons in four classroom situations. Results revealed different patterns of cognitive engagement. For instance, in student-student interactions activity, the learners showed a higher level of engagement compared to student-teacher activity. Another study was based on the use of classroom observations in order to help conceptualize students' engagement and identify academic disengagement. The types of observations conducted in this work include discussions, projects and labs (Turner et al. 2014). Despite qualitative measures provide reliable indicators of the level of students' engagement, they remain time-consuming and non-scalable since the data are gathered and analyzed by humans.

Another alternative method to measure students' engagement was observed during these last years, which is the use of physiological sensors. The prime advantage of using such sensing technologies in learning environments is that they can provide valuable quantitative data about the cognitive behavior of the learner, which cannot be directly observable. The use of physiological sensors such as skin conductance, heart rate and electroencephalography proved its effectiveness in monitoring changes in learners' mental state (Benlamine et al. 2015; Berka et al. 2007; Boucheix et al. 2013; S. D'Mello et al. 2007; Pham and Wang 2015; Whitehill et al. 2014). Boucheix et al. (2013) used eye tracking to study how different graphic representations can have an impact on students' engagement and learning outcomes. Whitehill et al. (2014) used facial expressions for automatic detection of students' engagement. D'Mello, Chipman and Graesser (2007) used student's posture to discriminate between low engagement (boredom) and high engagement (flow).

Among all the existing sensors, researchers reported promising results when using EEG to measure students' engagement. In fact, EEG can be a viable indicator of moment-to-moment changes in learners' attention. In their tutoring system, Chaouachi and his colleagues (2015) automatically adapt the learning material according to learners' attention and workload, measured using EEG signals. For instance, when the learner is disengaged or

overloaded/underloaded, a worked example is given as a next activity in order to keep the learner engaged. In another study, EEG technology was used in an e-learning environment that detects individual mental effort using a mental state classification system (F.-R. Lin and Kao 2018). The authors used a supervised learning technique to identify video segments where students produce high mental effort. Both students and instructors can have access to these segments in order to improve learning outcomes: first, students can self-regulate their mental state and second, instructors can adapt the learning activities accordingly.

In this paper, we propose to use EEG to track learners' mental engagement while they are reasoning and solving different medical cases. We aim to assess how students' engagement varies between the three phases of the problem-solving task and across the different areas of interest within the environment. Then we investigate whether learners' engagement can have an impact on their reasoning outcomes.

7.3 Experimental design

Fifteen undergraduate medicine students (7 females) with a mean age of 21.8 ± 2.73 years ranging from 20 to 27 years participated in the experiment with a compensation of 20 dollars.

Upon arrival at the laboratory, participants were asked to sign a consent form explaining the experiment and the material. They were then outfitted with the Emotiv EEG headset and placed in front of the eye tracker to record respectively their brain activity and eye movements while interacting with our medical learning environment called Amnesia. During the session, 30-45 minutes were dedicated to the game and at the end, participants were invited to complete a post-game questionnaire in order to collect feedbacks about the ergonomics and usability of the game.

7.3.1 Amnesia: an educational video game

Amnesia is a learning environment developed for novice medicine students to assess their clinical abilities through different problem-solving tasks. The game features a virtual hospital where the player who plays the role of a doctor is mistakenly diagnosed with amnesia and trapped within the hospital. In order to get out of the trap, the students need to resolve first some cognitive tasks such as logic tests and second, they have to prove their clinical skills by resolving

six medical cases that were designed and validated with the help of a medical professional. Each medical case represents a different disease: flu, bacterial pneumonia, measles, Ebola, mumps and whooping cough.

The resolution task is divided into three phases: exploration, diagnostic and treatment and in each phase, they can also collect additional data such as analyses and antecedents to establish a diagnosis. The exploration phase represents the first part of the clinical case resolution process where the student should analyze the patient's demographic information as well as his/her clinical data (e.g. symptoms, antecedents, etc.). Once all the useful data gathered, the objective of the diagnostic phase is to identify the correct disease from a list of six proposed ones. The student has up to three trials to find out the correct response. After completing the diagnosis, the student is shown a list of different treatments and he/she is asked to find out the adequate one(s). Participants are also given up to three attempts to discover the appropriate treatment.

Furthermore, in each medical case, we divided the resolution environment into different sections in order to analyze the level of engagement according to different areas of interest (AOI). For that purpose, we defined six specific sections, representing six task-relevant regions of the screen, as follows: Information (I), Antecedents (A), Symptoms (S), Analyses (N), Diagnosis (D) and Treatment (T). The *I* area of interest includes the demographic information of the patient (e.g. name, origin, weight and height). In the antecedents AOI, we introduce the diseases that the patient has had before (e.g. allergies). The *S* region includes all the symptoms related to the specific disease. In the analyses AOI, we present other clinical data (e.g. temperature, heart rate and blood pressure). The *D* area shows the different diagnoses offered to the student among which he/she has to choose the correct answer. Finally, the treatment AOI presents different proposals among which the student should select the appropriate treatments.

7.3.2 EEG recordings

Participants' brain activity was recorded using the Emotiv headset that contains 16 electrodes placed according to the 10-20 international standard (Klem et al. 1999). It records simultaneously 14 regions of the brain (O1, O2, P7, P8, T7, T8, C5, FC6, F3, F4, F7, F8, AF3 and AF4). Two more electrodes are used as references corresponding to the P3 and P4 regions

called respectively Driven Right Leg (DRL) and Common Mode Sense (CMS). The EEG data were recorded at a sampling rate of 128 Hz. The methodology of Chaouachi and his colleagues (Chaouachi et al. 2011, 2015) was used to measure each second, an engagement index by establishing a ratio between the three EEG frequency bands namely, θ (4-8 Hz), α (8-13 Hz) and β (13-22 Hz) as follows: Engagement index = $\beta / \theta + \alpha$.

The three frequency bands were extracted by multiplying one second of the EEG signal by a Hamming window and applying a Fast Fourier Transform. Then, a combined value of the θ , α and β was computed by summing their values over all the 14 measured regions. Finally, as in (Chaouachi et al. 2010) the EEG engagement index at instant T is computed by averaging each engagement ratio within a 40-second sliding preceding instant T. This procedure is repeated every 2s and a new 40s sliding window is used to update the index.

7.4 Results and discussion

In this study, EEG data were gathered and analyzed among all participants to investigate students' mental state using the engagement brain index. The experimental results are mainly divided into three parts: 1) We assess the evolution of students' brain activity during the three phases of the clinical problem-solving task (exploration, diagnostic and treatment), 2) we analyze the distribution of the engagement index across the different AOIs and 3) we investigate the impact of the engagement level on performance.

7.4.1 Engagement index evolution through the resolution process

Our first objective was to analyze how the engagement index evolves in the different phases of the resolution process across all participants, in order to investigate whether there was a particular period of time within the game that aroused the most attention. From figure 7.1, we can clearly see that case 1 yielded the highest engagement index especially at the beginning of the case. This result is pretty obvious since the interface of the medical case resolution was shown for the first time to the participants. In fact, they needed to provide more mental effort to understand the different elements with which they had to interact in the environment. In case 3, we notice a decrease in the engagement index during a certain period of time during which the students must identify the correct disease. This decrease in the participants' attention has

affected their performance since the number of failed attempts, in this case, is equal to 21 which represents 54% of the total number of attempts (see table 7.1). The remaining cases have similar variations.

As to have detailed comparisons, we dressed a bar chart to study the variation of the engagement index across the exploration, diagnostic and treatment phases.

Figure 7.2 shows the distribution of the engagement index in each phase of each medical case. We notice that in almost all cases the exploration phase caught most attention followed second, by the treatment phase and third the diagnostic. However, in general, we can clearly distinguish that there is not an important difference between the cases or the phases in terms of engagement level. The highest variation is about 30% in the first case. To support this conclusion, statistical testing was performed using analyses of variance (ANOVA) with repeated measures. No statistically significant results ($p = n.s$) were found across the three phases suggesting that the engagement index is not related to a particular phase or a case. These findings underline the fact that the students were engaged in the same way through all the reasoning process from the exploration to the treatment identification.

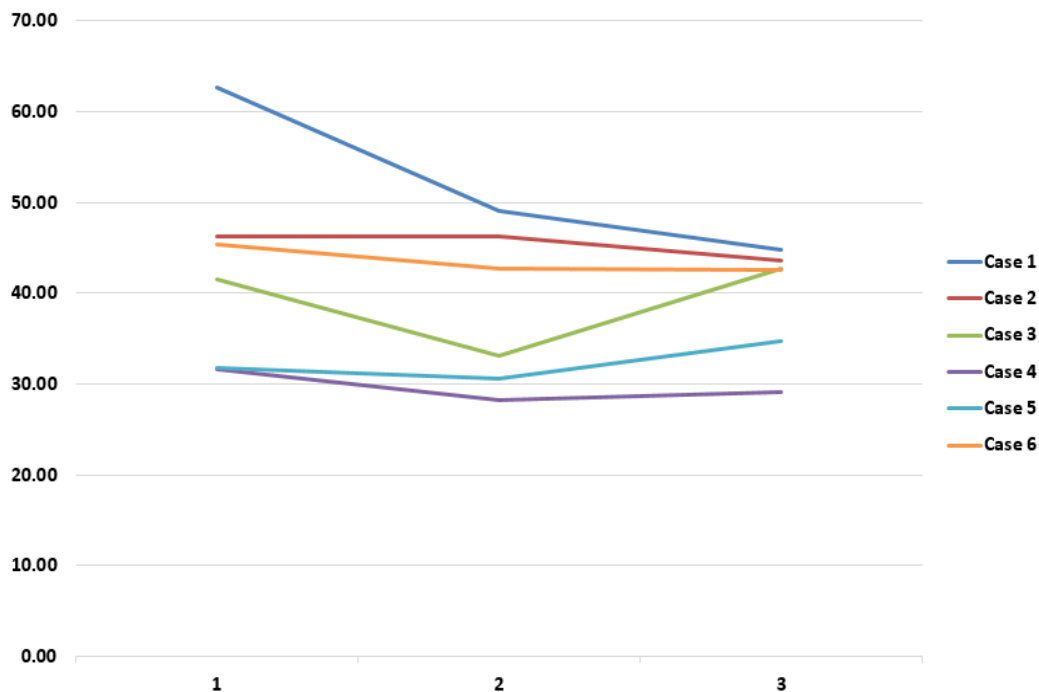


Figure 7.1 EEG engagement evolution in each medical case.

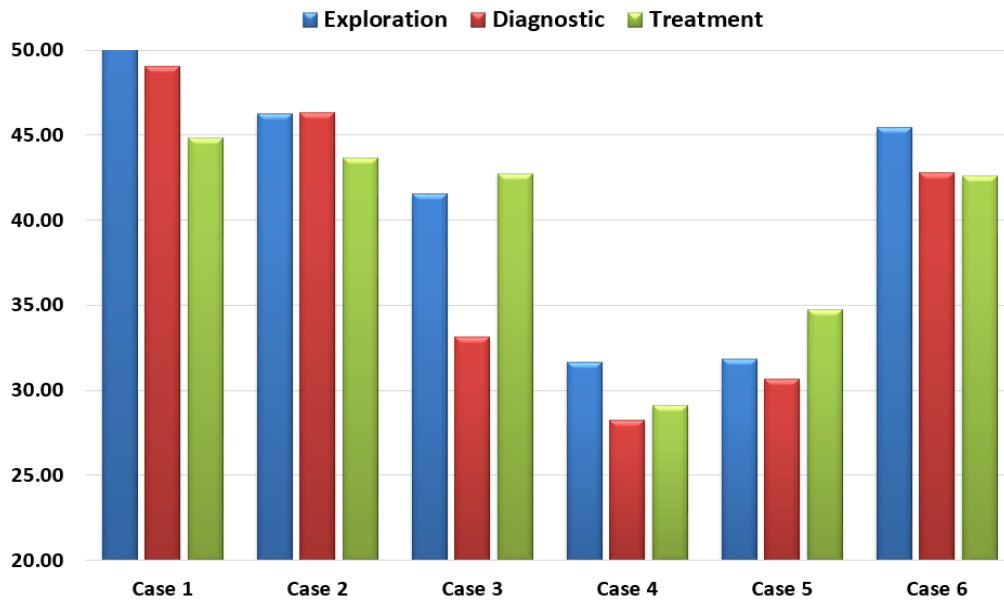


Figure 7.2 Distribution of students' engagement index across the different phases of the resolution process.

Table 7.1 Descriptive statistics for the medical case resolution.

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Number of participants who succeeded	15	10	10	8	5	5
Number of total attempts	47	36	39	31	13	18
Number of failed attempts	22	16	21	19	3	9

7.4.2 Engagement index distribution across the AOIs

In the previous analyses, it was found that the engagement index varies in a nearly similar way in all phases of the game all cases combined. In our next investigation, we performed further analyses to examine how the engagement index is distributed across the AOIs and similarly whether there is a specific area among those identified (i.e. Information, Antecedents,

Symptoms, Analyses, Diagnosis and Treatment) that may have an impact on students' engagement.

Preliminary statistical comparisons were made in terms of engagement index. Table 7.2 shows the descriptive statistics using mean values and standard deviations in order to see, in which areas there was a high level of engagement.

In the first case, the highest value was obtained for the Information area. Indeed, this corresponds to the first time the participants interacted with the clinical environment. That is why they were more engaged and focused to understand the tutoring material compared to the other regions. The Antecedents area was the least engaging in almost all cases since it did not contain specific relevant material, except in the fourth medical case where the engagement index was high. Indeed, in order to correctly resolve this case and to be able to identify the appropriate disease, the students needed to pay a particular attention to the important information contained in this area. As in this case, the diagnosis was Ebola, the learners needed to focus on the Antecedents area to successfully identify the disease, which includes an important clue (“recently travelled to Guinea”) that could help them provide the correct answer. In the remaining areas, the Treatment AOI has sparked the highest engagement level. Indeed, once the

Table 7.2 Descriptive statistics for the engagement index per medical case, Means (Standard Deviation).

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
<i>I</i>	67.26 (64.33)	29.80 (7.47)	46.03 (61.20)	27.08 (13.56)	34.02 (10.26)	34.10 (8345)
<i>A</i>	61.55 (56.31)	26.57 (4.53)	32.24 (12.37)	26.36 (8.69)	29.11 (7.18)	33.97 (13.99)
<i>S</i>	51.11 (43.88)	34.76 (19.92)	41.53 (33.92)	31.06 (0.6)	31.10 (9.62)	44.39 (15.40)
<i>N</i>	46.75 (40.14)	33.49 (22.96)	29.53 (11.35)	29.07 (10.97)	34.72 (11.29)	41.81 (20.79)
<i>D</i>	44.64 (34.37)	36.25 (23.11)	42.57 (35.37)	28.29 (7.30)	34.80 (14.58)	40.15 (18.49)
<i>T</i>	40.73 (28.75)	38.45 (19.86)	48.96 (32.03)	28.77 (7.94)	35.69 (11.08)	47.72 (15.36)

diagnosis is made, the students had to establish the appropriate treatments for the identified illness. A list of six different treatments was presented to the participants, and they were asked to choose the right ones (two or three depending on the case). Hence, the students had to pay attention to all the suggestions in order not to make mistakes. As a result, this step required more concentration than the others especially that the students were informed that the game will be over after three errors.

Based upon the obtained results, one-way ANOVAs were performed to examine whether there were significant differences among all AOIs in terms of engagement index. Three main effects were found respectively in the first case ($F(5, 1312) = 7.905, p < 0.01$), the third case ($F(5, 738) = 4.495, p < 0.01$) and the fifth one ($F(5, 292) = 2.559, p < 0.05$). This suggests that statistically significant differences exist between the six regions of interest in terms of mental engagement.

Post hoc tests with a Bonferroni adjustment were run in each case to show which specific regions raised the highest attention. Results indicate that in the first case, the engagement level was statistically significantly higher in the Information area than in the Symptoms ($p = .021$), Diagnosis ($p = .006$) and Treatment ($p = .000$) areas. Significant differences were also found in case 3 for the Treatment area compared to the Symptoms ($p = .010$) and Analyses ($p = .002$). Finally, in case 5 the pairwise comparisons demonstrated no significant effect indicating that there were no differences between the AOIs. Indeed, in this case, and as shown in table 7.2 all the values of the engagement index are relatively similar.

These results demonstrated that there was no particular AOIs that aroused the most the students' attention. These findings are contradictory with our previous work (Ben Khedher et al. 2018b) where the statistical analyses showed the opposite effect in terms of fixation duration. In fact, one-way ANOVAs showed that the S region was the most fixated one over all the medical cases. Thus, one can explain that longer fixations do not necessarily imply higher attention. Yet, does this imply that paying attention to specific relevant information can potentially lead students to correctly resolve the medical cases? We will explore this question in the next section. First, we will be interested in checking whether there exist statistical relationships between performance in the medical case resolution (success or failure) and

engagement. Second, we will check whether the engagement level in each AOI has an impact on students' performance.

7.4.3 Engagement index and performance

One way ANOVAs were first conducted to compare two groups of learners (group 1: success; group 2: failure) in terms of engagement level throughout the case resolution. For cases 2 and 5, the analysis was not conducted since the number of subjects was not fair between the two groups (all the participants succeeded in resolving the cases). For the remaining cases, the analysis yielded a main effect ($p < 0.05$) showing a significant difference of the students' engagement index in both groups as depicted in table 7.3. In cases 1, 4 and 6, we observed the highest engagement levels for group 1 compared to group 2. This suggests that being involved and concentrated in the task's resolution lead to a correct reasoning and thus a correct response. In fact, the more the learners were engaged the more they were able to reach the solution. Yet, in case 3 another trend was observed: the highest engagement index was identified in group 2 (i.e. the participants who failed in resolving the medical case). One can explain this differently; in that case, the learners were rather struggling in identifying the diagnosis/treatment. Hence, they tried to be more engaged and provided more mental effort to reach the solution.

Moreover, within-subject repeated measure ANOVAs were performed to investigate whether there is a specific region that may have an effect on students' performance. Yet, no significant differences ($p = n.s.$) were found between the AOIs in terms of success/failure over all cases suggesting that being engaged when focusing on a specific AOI did not have an impact on students' performance.

These results confirm our hypothesis in (Ben Khedher et al. 2018b) regarding fixation duration. Indeed, it was found that longer fixation duration on specific information areas has no effect on learners' performance. One can explain these results in different ways: the first reason can be that the fixated AOI may not contain important information, i.e. the student is admittedly engaged but not on the right direction, as is in case 4 where learners had to pay particular attention to the Antecedents area of interest. However, based on the statistical results, it is shown that this was the area with the lowest engagement index (see table 7.2). The second explanation is that the student can face difficulties to understand some knowledge. So, he/she tends to engage

Table 7.3 ANOVAs' results summary.

	<i>Case 1</i> F(1, 1319)	<i>Case 3</i> F(1, 742)	<i>Case 4</i> F(1, 475)	<i>Case 6</i> F(1, 222)
<i>p</i>	.025	.000	.000	.039
<i>F</i>	5.019	70.976	81.630	4.334

more and provide more mental effort when focusing on certain materials that can help her/him find the right answer such as the S or the N areas.

To sum up, the first experimental analyses are in line with our assumptions. In case 1 and during the exploration phase, students were more likely to experience high mental engagement since it is the first medical case they had to resolve. Also, the displayed interface differed from the other scenes of the game, which makes them more attentive in order to understand how they would solve the case. No particular variation was identified for the other cases in terms of engagement index: the three resolution phases aroused almost the same level of attention. When solving the medical cases, participants needed to focus on particular regions of the screen. Therefore, we were interested to examine whether there was a specific area(s) of interest whither the mental engagement was high. Similar trends have been observed in most cases: learners were mostly engaged when focusing on the Treatment AOI. In fact, this part requires more mental effort knowing that for each identified disease in each medical case, two treatments at least are needed. On the other side, the Antecedents area is the least engaging, as it does not contain significant information to support the students' reasoning process. Finally, in the second part of the experimental design, we developed the hypothesis that a change in the students' engagement level across the fixated AOIs can have an impact on learners' performance. Results showed statistically significant relationships between students' performance and engagement index. Nevertheless, based on the remaining analyses we cannot conclude that there is a unique area that has an effect on students' outcomes: i.e. being engaged when fixating an AOI, important or not, cannot necessarily lead to their success or failure in the medical case resolution. Indeed, in clinical problem-solving tasks, all the steps of the resolution

are mandatory. For instance, we cannot only be engaged when reading the symptoms and forget about the antecedents that may involve relevant information as well.

7.5 Conclusion

In this study, an experimental protocol was established to measure the mental engagement index of fifteen participants using EEG signals. Participants were shown six medical cases and asked to resolve them by identifying the correct diagnosis and treatment. The objective of this research was to draw a general overview of the students' brain activity changes when reasoning in order to subsequently develop appropriate help strategies. Indeed, tracking learners' mental engagement is very important especially in high-risk medical learning environments in order to constantly have an updated monitoring of their progress and level of knowledge acquisition.

First, we assessed the evolution of the engagement index across the different phases of the resolution process and the different areas of interest, with the aim to identify potential time frames or regions that could yield a high level of engagement. We found out that the first case aroused a high level of engagement especially during the exploration phase, which was the first scene exhibited to the participants. Then, we analyzed the relationship between mental engagement and learners' performance. The results showed that engagement has overall a significant positive impact on students' outcomes, however, we found that this relationship was not modulated by paying more attention to a particular section of the game or to a particular area of interest.

In our future research, we propose to incorporate eye tracking data with EEG in order to have a multimodal sensor-based assessment of students' learning behavior. In the long term, we plan to provide novice medicine students with timely interventions to foster their analytical reasoning process according to both their mental and visual behaviors.

Acknowledgements

This work was supported by NSERC (National Science and Engineering Research Council) and SSHRC (Social Science and Human Research Council) through the LEADS project.

Chapitre 8.

Une approche multimodale à base de senseurs physiologiques pour une analyse visuelle et cérébrale de l'état d'un apprenant durant un processus de raisonnement

Ce chapitre est une extension des travaux entrepris dans une étude exploratoire que nous avons réalisé pour examiner si la combinaison de deux senseurs physiologiques (ET et EEG) permet d'améliorer les performances de prédiction de deux classes d'apprenants. À partir des données collectés dans l'expérimentation (II) à savoir les mouvements des yeux et l'activité cérébrale, 15 variables oculaires et cérébrales ont été utilisées comme variables d'entrée pour l'entraînement des modèles prédictifs afin d'identifier automatiquement deux groupes d'apprenants : ceux qui ont réussi à résoudre les différents cas médicaux et ceux qui ont échoué. Six algorithmes de classification ont été testés et un taux de précision de 83% a été atteint. Nous avons aussi testé les performances de prédiction de chaque modalité séparément. Les résultats ont montré qu'en utilisant uniquement les mouvements des yeux un taux de précision faible a été obtenu comparé aux deux modalités et en utilisant les données cérébrales une seule classe a pu être détectée. Ce travail a été accepté en 2019 dans la conférence *FLorida Artificial Intelligence Research Society*.

Dans ce chapitre, nous avons étendu les travaux entrepris précédemment afin de proposer une approche multimodale capable d'analyser les indicateurs d'état mental et du comportement visuel des apprenants lors d'un processus de raisonnement. Dans cette étude, nous avons tout d'abord déterminé quelles mesures visuelles et cérébrales sont liées aux performances des apprenants. Ensuite, nous avons analysé la relation entre le comportement visuel des apprenants et leur activité cérébrale.

A partir des données oculaires collectées, différentes mesures ont été utilisées à savoir la durée de fixation et les variables reliées au « scan path » visuel qui ont été calculées en utilisant l'algorithme d'alignement local comme le nombre de similarité, d'insertion/suppression et de mutation et le score de similarité. L'objectif était d'examiner si ces variables ont un impact sur les performances des apprenants. Les résultats statistiques ont montré qu'aucune corrélation significative n'a été trouvée pour la durée de fixation contrairement aux autres variables où des résultats significatifs ont été obtenus par rapport à la réussite/l'échec dans le cas médical et aussi par rapport aux nombres de tentatives par cas. Par exemple, le score de similarité était significativement plus élevé dans les cas de réussites ($M = 5.8$) comparé aux cas d'échecs ($M = 3.6$). De plus, une corrélation négative a été trouvée entre le score de similarité et le nombre de tentatives signifiant que plus la séquence visuelle de l'apprenant s'approche de la séquence optimale, plus le nombre de tentatives diminue.

De la même manière, nous avons analysé la relation entre les variables EEG à savoir, l'état mental, la charge de travail et l'engagement et les mesures de performances. D'un côté la charge mentale de travail avait un impact sur la réussite/échec dans les cas médicaux, d'un autre côté l'engagement mental avait un impact sur le nombre de tentatives par cas. Concernant l'état mental (positif ou négatif), une relation statistiquement significative a été trouvée par rapport aux nombres de tentatives. Les résultats ont démontré qu'en cas d'état mental négatif c'est-à-dire désengagement, surcharge ou sous-charge cognitive le nombre de tentatives par cas est élevé.

Enfin, dans la dernière partie de cette étude, nous avons analysé la relation entre les différentes variables visuelles et cérébrales. Les résultats ont montré en premier lieu des relations significatives entre la durée de fixation et les trois niveaux de charge mentale avec une durée de fixation moyenne plus élevée en cas de surcharge cognitive. Pour l'indice d'engagement, nous avons aussi trouvé une relation significative avec le score de similarité qui était en moyenne plus élevé en cas d'engagement positif et avec le nombre de mutations qui était plus faible en cas de désengagement.

Notre contribution dans cet article était de proposer un modèle d'évaluation du processus de raisonnement en se basant sur le comportement visuel et cérébral de l'apprenant. L'objectif est d'étudier les mouvements des yeux et l'activité cérébrale pour fournir des indicateurs clé et

déterminer quelles caractéristiques visuelles et cérébrales ont un impact sur les performances en vue de les utiliser dans nos travaux futurs pour adapter nos stratégies d'aide.

La suite de ce chapitre est constituée de l'article de recherche intitulé « *Assessing Students' Clinical Reasoning using Gaze and EEG Features* », accepté à la conférence *Intelligent Tutoring System*. Le travail a été supervisé par mon directeur de recherche Claude Frasson et divisé entre les 3 premiers co-auteurs de cette façon : la collecte des données et le prétraitement des variables oculaires statiques et dynamique ont été faits par moi-même. Les deux indices d'engagement et de charge mentale ont été mesurés par moi-même en utilisant le système MENTOR développé par Maher Chaouachi dans le cadre de ses travaux de recherche (Chaouachi 2015). Les différentes analyses statistiques ont été réalisées par Imène Jraïdi. Finalement l'article a été rédigé par les deux premiers auteurs à titre équitable.

Assessing Students' Clinical Reasoning using Gaze and EEG features

Imène Jraidi, Asma Ben Khedher, Maher Chaouachi and Claude Frasson

Accepted in: *The 15th International Conference On Intelligent Tutoring Systems Conference, (ITS 2019).*

Abstract

The purpose of this work is to analyze the learners' visual and brain behaviors during clinical reasoning. An experimental study was conducted to record gaze and EEG data of 15 novice medical students as they interacted with a computer-based learning environment in order to treat medical cases. We describe our approach to track the learners' reasoning process using the visual scanpath followed during the clinical diagnosis and present our methodology to assess the learners' brain activity using the engagement and the workload cerebral indexes. We determine which visual and EEG features are related to the students' performance and analyze the relationship between the students' visual behavior and brain activity.

Keywords: Eye tracking, Scanpath, EEG, Engagement, Workload, Clinical reasoning, Learning Performance.

8.1 Introduction

The democratization of the use of affective computing techniques has enabled a multifaceted understanding of the learners' behavior (Hou 2015; Jraidi and Frasson 2013). Particularly, these techniques allowed computer-based learning environments and intelligent tutoring systems to

have access to a learner-centered data and to have a fine-grained analysis of the variables that impact learning performance (Ben Khedher et al. 2018b; S. K. D’Mello et al. 2008; Pardo et al. 2017).

Techniques such as electroencephalography (EEG) and eye tracking, which have been up to recent years mainly used in strict laboratory conditions, are being increasingly used in realistic learning settings (Ben Khedher et al. 2018d, 2019; Berka et al. 2007; Maynard et al. 2013). Their capacity to offer real-time qualitatively rich information about the learners’ state could be highly useful to optimize the learning strategies. This is particularly true in medical learning environments in which the learners’ interactivity with the system is very limited (Lajoie et al. 2013; Poitras et al. 2018). In fact, in such a clinical context, and throughout the diagnostic phases, the learners are generally engaged in various reading activities during which a complex clinical reasoning is performed with a very restrained action field. This consequently limits the learners’ data acquired by the system and hinders its capacity to assess the learning process. Therefore, the integration of EEG sensors to track the learners’ mental state, and eye tracking to track gaze patterns could be highly beneficial to assess learning, especially when the learner is just staring at a computer screen.

In this paper, we propose to use these techniques as a mean to analyze the learners’ clinical reasoning. More precisely, we use EEG data to assess two mental states, namely engagement and workload, and eye tracking to monitor learner’s visual focus and scanpath. We are interested in analyzing which features are related to the students’ reasoning and performance.

The remainder of the paper is organized as follows. We present previous work on the use of EEG and eye tracking in the learning context. Next, we describe the developed tutoring system, experimental setup and methodology. Then, we discuss the obtained results, conclude and present directions for future work.

8.2 Previous work

A wide range of sensing technologies is nowadays capable of providing highly accurate data on the users’ state in human computer interactions and particularly in ITS, where the behavioral and affective component is being more and more important.

Specifically, eye tracking and EEG devices are being increasingly used within the learning context (Ben Khedher et al. 2018a; Brouwer et al. 2017; D’Mello et al. 2012; Lallé et al. 2016; Slanzi et al. 2017). Slanzi and his colleagues predicted behavioral variables such as the learners’ click intentions within a web search task using gaze fixation, pupil dilation and EEG (Slanzi et al. 2017). Brouwer et al. used fixation duration, pupil size and EEG data to assess learners’ attention in a visual search task (Brouwer et al. 2017). El-Abbasy et al. developed an affective e-learning platform that changes the learning materials when the learner experiences a negative emotion. Emotions such as sadness and frustration were recognized using eye tracking and EEG (El-Abbasy et al. 2019). In (Alhasan et al. 2018), Alhassan and his colleagues used data mining techniques to classify the learners in terms of two different learning styles, namely visual and verbal, by taking the participants’ EEG and gaze data as input features. In the same context, in (Muldner and Burleson 2015) the authors used machine learning algorithms to discriminate between high and low creativity students using skin conductance bracelet, an eye tracker and EEG sensors. Two classifiers namely, Nearest Neighbor and Naïve Bayes achieved up to 80% of true positive rate. Makransky et al. analyzed EEG measures, such as alpha and theta frequency band powers, and eye tracking metrics, such as the percentage of time spent on relevant regions of the interface, to assess learners’ cognitive load during multimedia learning. Statistically significant relationships were found showing that eye tracking measures are indicators of extraneous cognitive load, and that EEG measures are indicators of intrinsic cognitive load (Makransky et al. 2019).

In this paper, we propose to analyze the learners’ visual behavior and brain activity within clinical reasoning using eye tracking and EEG. We use gaze and neural patterns to assess learners’ visual focus and cognitive state, and determine which features are related to students’ analytical process and performance.

8.3 Experimental design and methodology

An experimental study was conducted to record gaze and EEG data of novice medicine students as they interact with a computer-based learning environment (Amnesia). 15 participants (8 males) aged 20-27 years ($M = 21.8 \pm 2.73$) were recruited. Upon their arrival, they were briefed about the experimental procedure and asked to sign a consent form. They were then outfitted

with an Emotiv EEG headset and placed in front of a Tobii TX-300 eye tracker (at 65 cm approximately). A sampling rate of 128 Hz was used to collect the EEG data from 14 different regions of the scalp (O1, O2, P7, P8, T7, T8, FC5, FC6, F3, F4, F7, F8, AF3 and AF4). A sampling rate of 300 Hz was used with the eye tracker, and a nine-point calibration grid was used to calibrate the participants' point of gaze.

Amnesia. Amnesia is a realistic environment developed to assess undergraduate medical students' analytical skills through clinical problem-solving (Ben Khedher and Frasson 2016a). The system features a virtual hospital where the learner acts as a doctor who is mistakenly diagnosed with amnesia. The learner needs to prove that he does not suffer from this disease by resolving six different medical cases. In each case, the learner is instructed to identify both the correct diagnosis and the appropriate treatment through a series of observations including the patients' demographic information, symptoms, antecedents and clinical data. The diseases he must find are respectively: flu, bacterial pneumonia, measles, Ebola, mumps and whooping cough. For each diagnosis and treatment, different response alternatives are given, and the student has up to three attempts to find out the correct answer.

8.3.1 Visual behavior

Two different metrics were computed from the eye tracker, namely fixation duration and scanpath. Fixation duration is a temporal measure of the visual focus. It corresponds to the amount of time the fovea (center of gaze) is directed towards Areas of Interest (AOI) of the screen, i.e. the task-relevant elements in the system's interface. In total, six AOI were defined for each medical case (see Figure 1 (Ben Khedher et al. 2018c)) :

- Information (I) includes the demographic information of the patient (e.g. name, age and weight).
- Antecedents (A) introduces the allergies and the prior diseases.
- Symptoms (S) highlights the symptoms of the disease.
- Analysis (N) includes data on the patient's temperature, heart rate and blood pressure.
- Diagnosis (D) shows different response suggestions for the disease to be identified.

- Treatment (T) presents different suggestions for the potential treatments of the disease.

The scanpath is a spatiotemporal measure representing the dynamic visual trajectory of the eye movements. It uses both fixations and saccades across the AOI and is recorded as a sequence (e.g. IIIAASSSSNNNNSSNDDSSDDTTDTT); each letter denotes a visited AOI with a fixation duration above a threshold of 250 ms. Redundant characters are then collapsed to retain a unique occurrence of each area (e.g. IASNSNDSDTDT). Smith-Waterman algorithm is used to evaluate this scanpath. The algorithm, which is notably used in bioinformatics for DNA sequence alignment, compares the learner's visual scanpath with a target reference sequence. This sequence, denoted as ISANDT, represents the hypothetico-deductive analytical process a novice clinician should follow in clinical reasoning (Swanson et al. 1990). In this process, the clinician starts by collecting the patient's information, and then formulates an initial hypothesis according to the symptoms. Next, additional clinical data such as analysis or medical antecedents are collected to validate or discard this hypothesis until reaching a final diagnosis, and a treatment is prescribed.

The algorithm aligns both sequences (the learner's scanpath and the reference sequence) by optimizing a *similarity score* derived from the following features: the number of *matches*, number of *mismatches* and number of *gaps*. Matches are convergent elements (i.e. identical letters) in both sequences. Mismatches are the divergent elements (requiring mutations, i.e. substituting one letter for another). Gaps are the missing elements (implying an insertion or a deletion in one of the two sequences); we refer to (Ben Khedher et al. 2018a) for more details. The higher the aligned similarity score, the more the two sequences are similar; which means the closer the learner's visual sequence is to the optimal reference sequence.

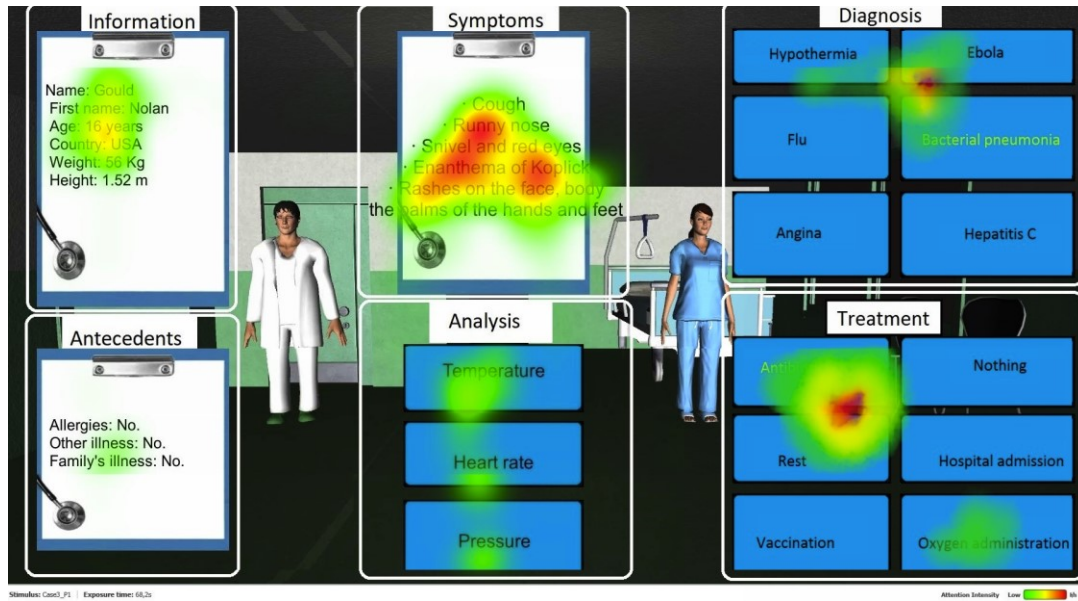


Figure 8.1 Screenshot of Amnesia with the AOI. The red color represents the most fixated areas, followed by yellow and then green with lower intensity.

8.3.2 Brain activity

Two brain indexes were computed from the EEG recordings, namely engagement and workload as depicted in Figure 2 (Chaouachi 2015). The engagement index is a neural indicator of the level of alertness and attention allocated during a task, and the workload index (also known as cognitive load index) measures the amount of information processing demands and mental effort induced during a task (Berka et al. 2007).

The engagement index is computed using three frequency bands, namely θ (4-8 Hz), α (8-13 Hz) and β (13-22 Hz) as follows: $\beta / (\theta + \alpha)$ (Pope et al. 1995). The extraction of these bands is performed by multiplying one second of the EEG signal by a Hamming window in order to reduce the spectral leakage, and applying a Fast Fourier Transform (FFT). As the EEG headset measures 14 different regions at the same time, the values of θ , α and β are summed over all these regions. An engagement index is computed each second. Then, a 40-second moving average mobile window is used in order to smooth the index and reduce its fluctuation.

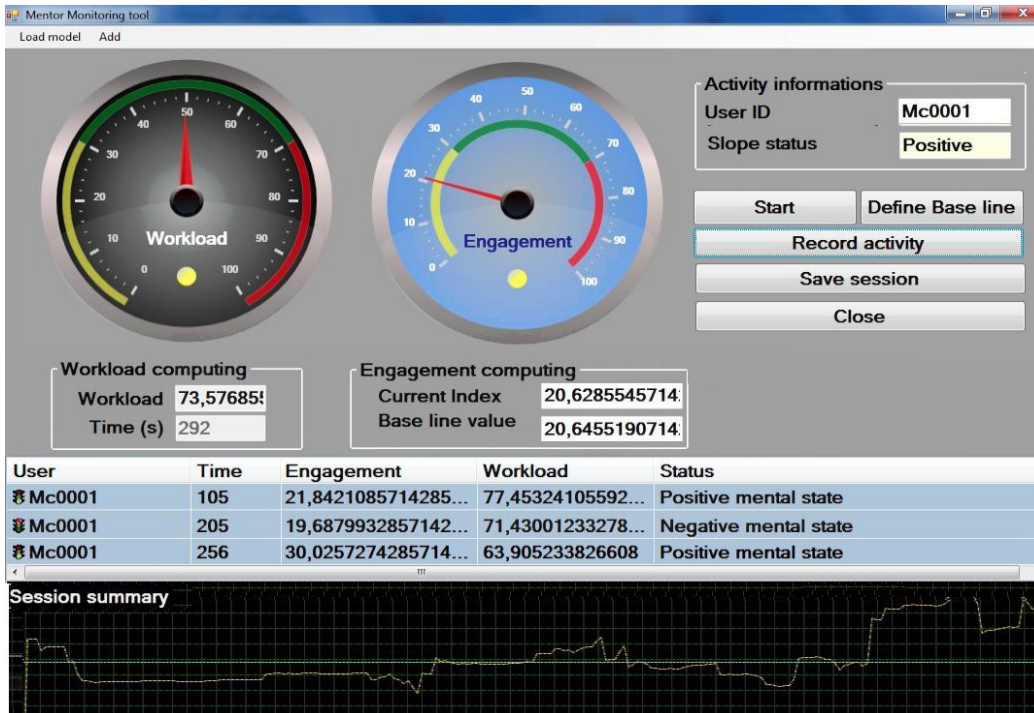


Figure 8.2 Screenshot from the acquisition module used for the EEG data

Unlike the engagement index, which is directly extracted from the EEG raw data, the workload index is computed using a pre-trained and validated predictive model (Chaouachi et al. 2011). EEG signal is cut into 1-second segments and multiplied by a Hamming window. A FFT is then applied to transform each segment into a spectral frequency and generate a set of 40 bins of 1 Hz ranging from 4 to 43 Hz. The dimensionality of the data is reduced using a Principal Component Analysis and then mean normalized. Next, a Gaussian Process Regression algorithm with an exponential squared kernel and a Gaussian noise is run in order to compute the EEG workload index from the normalized data (we refer to (Chaouachi et al. 2011) for more details about this procedure).

To assess the learner's mental state, a slope of each index is computed using the least squared error function of the indexes' values throughout each medical case. For the engagement index, if the slope value is positive, then the learner is considered as mentally engaged. Otherwise, the learner is considered as mentally disengaged. For the workload index, if the slope value is between -0.03 and +0.03, then the workload is considered as optimal. Otherwise, if the slope value is above 0.03, the learner is considered as overloaded, and if the slope is below -0.03 the

learner is considered as underloaded (Chaouachi et al. 2011). Moreover, if the learner is mentally engaged and has an optimal workload, then the learner's mental state is considered as positive. Otherwise, the mental state is negative.

8.4 Results and discussion

Results are presented in three sections: the first section analyzes the learners' visual behavior, the second section deals with the brain activity, and the third section describes the interplay between both measures.

8.4.1 Visual behavior

First, we wanted to analyze the relationship between fixation duration and the learners' outcomes (correctness and number of attempts in each medical case). That is whether focusing (visually) at the task-relevant elements of the screen has an impact on the learners' performance. The answer is no: there was no statistically significant correlation between fixation duration and performance measures ($p = n.s.$).

Then, we were interested in analyzing the dynamic visual behavior of the learner, namely assessing the relationship between the performance measures and the alignment metrics of the learners' visual scanpath: numbers of matches, mismatches and gaps, and the similarity score between the used scanpath and the optimal reference sequence. First, with regards to the success/failure (i.e. correct diagnosis and treatment), we found that the number of mismatches was significantly lower ($F(1, 75) = 13.585, p < 0.001$) if the case is successfully solved ($M = 0.00, SD = 0.00$) compared to failed cases ($M = 0.21, SD = 0.41$). Also, the similarity score was significantly higher ($F(1, 75) = 5.879, p < 0.05$) for the succeeded cases ($M = 5.79, SD = 2.42$) compared to the unsolved ones ($M = 3.58, SD = 5.61$). This implies that for the correct answers, the alignment scores between the students' scanpath and the optimal reference sequence were significantly higher and the number of mismatches smaller.

In addition, statistically significant results were found for the number of attempts per case. A significant positive correlation was found between the number of mismatches and the number of attempts ($r = 0.36, p < 0.001$). The more mismatches or deviations between the learner's visual sequence and the reference sequence, the more they were attempts per case; i.e. the more the

learners had trouble finding the right answer. A negative correlation with the score alignment was also found ($r = -0.26, p < 0.05$): the higher the alignment score, the lower was the number of attempts. This means that the more the learner's visual sequence was close to the optimal reasoning sequence, the lower was the number of attempts to answer. In other words, the more the learners' reasoning was 'optimal', the faster they could find the right answer.

Hence, the analysis of the gaze behavior showed that the fixation duration had no impact on the learning performance. A student can watch/fix the screen without being engaged in the task. This will be indeed confirmed through the analysis of the learners' brain data (EEG engagement index). The analysis of the dynamic visual behavior allowed us to track and evaluate the visual scanpath followed by the learner during his reasoning. The more the visual sequence was close to the hypothetico-deductive analytical process (the reference sequence), the better were the performance. This validates our approach to monitor the learners' analytical process through the analysis of the followed visual scanpath over the AOI of the game and its evaluation according to the optimal reference path using a sequence alignment algorithm.

8.4.2 Brain activity

Similarly, we analyzed the relationship between the EEG variables (mental state, workload and engagement) and the performance measures (correctness and number of attempts in each medical case). We found a statistically significant association between correctness and the workload variable (underload, optimal load and overload), $\chi(2) = 7.445, p < 0.05$. In particular, more than half of the correct answers (54.10% of the cases) were associated to the state of optimal load (against 26.20% to overload and 19.70% to underload).

For the number of attempts, we found a statistically reliable relationship with the mental state variable (positive/negative): $F(1, 74) = 13.725, p < 0.001$. The number of attempts was significantly higher when the mental state was negative ($M = 3.84, SD = 0.96$), against ($M = 2.96, SD = 1.02$) if the state was positive. We recall that a negative mental state is associated with one of these states: disengagement, or-and underload/overload. Two other ANOVAs were conducted to study each index separately. There was no significant relationship between workload and the number of attempts. However, a significant relationship between the engagement variable (engaged/disengaged) and the number of attempts was found: $F(1, 74) =$

17.27, $p < 0.001$. This number was significantly higher if the learners were disengaged ($M = 3.95$, $SD = 0.89$) against ($M = 3.03$, $SD = 1.04$) if the learners were mentally engaged. That is as the engagement level dropped, the learners had more difficulty in solving the medical cases. Or conversely when the engagement index was high, the number of attempts to resolve the cases decreased. In other words, the more attentive they were, the easier they found the right answers.

To summarize, the two cerebral metrics seem to have an impact on the performance measures. The cognitive load had an impact on the failure/success of the medical cases, and the engagement index had an impact on the number of attempts per case. On one hand, we have a significant association between the optimal load interval of the workload index and the succeeded medical cases. On the other hand, we have a lower number of attempts for the positive values of the engagement index.

8.4.3 Relationship between Gaze and EEG Data

We started by analyzing the relationship between the learners' mental state and the fixation duration in the medical cases. A first ANOVA showed that the fixation duration was on average significantly higher ($F(1, 74) = 5.99$, $p < 0.05$) when the mental state was negative ($M = 326.24$, $SD = 44.18$) vs. ($M = 302.82$, $SD = 38.62$) in case of positive mental state.

Two more ANOVAs were performed to analyze the engagement and the workload indices respectively. The first analysis showed that there was no statistically significant relationship between the engagement level and the fixation duration. On the other side, the analysis of the mental load index showed that there was a significant relation between fixation duration and (underload, optimal load and overload): $F(2, 73) = 4.275$, $p < 0.05$.

This shows that the fixation duration is not a good indicator of mental engagement. One can fix or look at an area of interest on the screen, without really lending focus or attention. This rather suggests that a high fixation time is a sign of overload and mental fatigue. Moreover, we found a significant negative correlation between the engagement and the workload indexes' values ($r = -0.295$, $p < 0.05$), which means that the higher the level of attention, the more the induced level of mental effort decreased. Inversely also, if the mental load increased, i.e. in case of mental fatigue, the level of attention decreased. This was observed not only during the

resolution of the medical cases, but throughout all the interaction with the game ($r = -0.204$, $p < 0.05$).

8.5 Conclusion

This paper presented an experimental study that analyzes eye tracking and EEG features of novice medical students while they interact with a hospital simulation game. The goal was to assess their visual and cerebral behaviors during clinical reasoning using different medical cases and sensors to record their brain and gaze data.

The analysis of these data first led us to the conclusion that the fixation duration on the screen or on relevant elements of the environment is not a good indicator of performance (outcomes) or attention (EEG engagement), but rather a sign of mental fatigue (overload). However, the analysis of the dynamic visual behavior using the learners' scanpath across the different AOI enabled us to evaluate the analytical process followed by the learners during diagnostic reasoning. Gaze data were then correlated with performance on one hand, and brain measurements (mental state and engagement/workload indexes) on the other hand. Furthermore, the analysis of the EEG data showed that there were also statistically significant correlations between the different brain measurements and learners' performance.

Therefore, this paper allowed us to confirm that both sensors (eye tracking and EEG) are highly valuable sources of information for the monitoring of the students' external gaze behavior in terms of visual scanpath and also their internal physiological state in terms of cerebral activity during problem solving. Our future work is directed towards using both sensors to actively adapt the game's tutoring interventions according to the learners' visual behavior and brain activity.

Acknowledgments

This work was supported by NSERC (National Science and Engineering Research Council) and SSHRC (Social and Human Research Council) through the LEADS project. We also thank Issam Tanoubi from the University of Montreal for his collaboration in the design of the Amnesia environment.

Chapitre 9.

Conclusion

9.1 Contributions

La première contribution de cette thèse a été de proposer une méthode d'évaluation des performances de l'apprenant en utilisant les mouvements des yeux. A cet effet, nous avons utilisé des mesures oculaires statiques et dynamiques permettant de nous renseigner sur le comportement visuel de l'apprenant durant son interaction avec un jeu sérieux dédié à l'apprentissage. Plus précisément, nous avons calculé des mesures statiques à partir des données observées selon une dimension temporelle, telles que la durée de fixation et le nombre de fixation et une mesure dynamique mesurée selon une dimension spatio-temporelle qui est le « scan path ». Deux études expérimentales ont été réalisées pour implanter notre approche en utilisant deux environnements d'apprentissage différents.

La première étude réalisée est une étude exploratoire dont l'objectif est d'examiner si les données oculaires sont capables de prédire les performances d'apprentissage. Nous avons enregistré les mouvements des yeux de 20 participants de l'Université de Montréal lorsqu'ils interagissaient avec le jeu sérieux Crystal Island (J. Rowe et al. 2009) en répondant à des questionnaires à choix multiples. Quatre variables ont été mesurées et utilisées par la suite comme variables d'entrée pour l'entraînement des modèles prédictifs en utilisant le logiciel de data mining WEKA afin d'identifier automatiquement deux classes d'apprenant : ceux qui ont réussi la résolution des questionnaires et ceux qui ont échoué. Quatre algorithmes d'apprentissage machine ont été testés à savoir : les classifieurs Bayésiens naïfs, les forêts aléatoires, le perceptron multicouches et les arbres de décision. Le meilleur taux de prédiction a été obtenu avec les arbres de décision (taux de précision : 70.10%).

Dans la deuxième étude, l'objectif était double : premièrement, nous avons analysé l'expérience visuelle de l'apprenant durant son interaction avec le jeu sérieux Amnesia qui met en évidence une activité de raisonnement clinique et deuxièmement, nous avons examiné si éventuellement le comportement visuel a un impact sur les performances d'apprentissage. Pour le deuxième volet de cette étude, nous nous sommes basés sur des travaux menés en bio-informatique concernant l'alignement de séquences et plus précisément deux méthodes le codage de caractères et l'alignement local. Les mouvements des yeux de 15 étudiants novices de la faculté de médecine de l'Université de Montréal ont été enregistrés durant l'expérimentation. A partir des données collectées, deux mesures statiques à savoir : la durée de fixation et le temps de la première fixation et une mesure dynamique à savoir : le « scan path » ont été mesurées. Deux apports ont été réalisés : dans un premier temps, nous avons détecté la même tendance visuelle chez tous les participants où la zone d'intérêt qui représente les symptômes était la plus fixée, toutefois les statistiques ont montré qu'il n'y a pas de relations statistiquement significatives entre les performances de l'apprenant et le fait de fixer une des zones clés du processus de raisonnement analytique. Dans un deuxième temps, l'analyse de la séquence visuelle de l'apprenant suivie au cours du raisonnement (c.-à-d. le scan path) a indiqué que les mesures dynamiques sont capables d'évaluer les performances d'apprentissage en identifiant les apprenants qui ont bien raisonné et ceux qui ne l'ont pas fait.

Notre deuxième contribution a été de suivre l'état mental d'un apprenant durant la résolution des cas médicaux afin de déterminer des indicateurs cérébraux permettant d'évaluer le processus de raisonnement. Pour ce faire, un index d'engagement mental basé sur les trois bandes de fréquences α , β et θ a été calculé en utilisant l'activité cérébrale. L'objectif est de suivre l'évolution de l'état mental de l'apprenant d'une part, et de déterminer s'il existe une relation entre état d'engagement et performances d'une autre part.

L'activité cérébrale des apprenants a été enregistrée en utilisant EEG au cours de leur interaction avec Amnesia. Afin d'obtenir des analyses approfondies, nous avons suivi le niveau d'engagement d'un apprenant tout d'abord, durant les trois phases du processus de résolution des cas médicaux. Ensuite, nous avons analysé la distribution de l'indice d'engagement mental à travers les différentes zones d'intérêt que nous avons défini au préalable au sein de l'interface de résolution. Enfin, nous avons essayé d'investiguer s'il existe une relation significative entre

engagement et performances. Les résultats ont démontré que globalement dans tous les cas, les trois phases de résolution à savoir : l'exploration, le diagnostic et le traitement ont suscité le même niveau d'engagement chez les participants sauf pour le premier cas dans lequel la phase d'exploration a généré un effort mental plus élevé. Pareillement au niveau des zones d'intérêt, la même tendance a été observée dans la plupart des cas médicaux. D'un côté, la région nommée « traitement » était celle qui a suscité plus d'engagement mental et de l'autre côté la région nommée « antécédents » était celle qui a attiré le moins d'attention. De plus, les analyses statistiques ont indiqué qu'il existe une relation statistiquement significative entre performances et indice d'engagement.

Notre dernière contribution a été de proposer une approche multimodale à base de senseurs physiologiques pour une analyse combinée du comportement visuel et de l'état mental de l'apprenant. Cette approche a pour objectif de déterminer quels sont les indicateurs cérébraux et visuels qui permettent d'évaluer le processus de raisonnement. Plusieurs chercheurs ont porté une attention particulière à l'utilisation simultanée de nombreuses sources d'information afin de surmonter les lacunes rencontrées en utilisant une seule modalité. En effet, cette approche multimodale permet de fournir un modèle d'évaluation plus robuste et fiable.

Une analyse des indicateurs d'état mental et de comportement visuel a été faite en trois étapes : tout d'abord, nous avons analysé le comportement visuel de l'apprenant en utilisant une mesure statique (la durée de fixation) et des mesures dynamiques reliées au « scan path » (nombres de similarité, d'insertion/suppression et de mutation) afin d'examiner si ces variables ont un impact sur les performances des participants. Aucune corrélation significative n'a été obtenue pour la durée de fixation, contrairement à la variable dynamique où des résultats significatifs ont été trouvés par rapport aux nombres de tentatives par cas médical. En effet, plus la séquence visuelle de l'apprenant se rapproche de la séquence de raisonnement optimal, plus le nombre de tentatives est faible. Ensuite, nous avons analysé l'activité cérébrale en étudiant la relation entre les variables EEG (état mental, charge mentale de travail et engagement) et les performances de l'apprenant. Les résultats ont démontré qu'en cas d'état mental négatif c'est-à-dire désengagement, surcharge ou sous-charge cognitive le nombre de tentatives par cas est élevé. Pareillement, une relation significative a été trouvée entre le nombre de tentatives et l'état d'engagement suggérant que plus l'apprenant est attentif plus il a de la facilité à répondre (le

nombre de tentatives diminue). Enfin pour la dernière étape de l'étude, nous avons analysé les deux modalités conjointement (mouvements des yeux et activité cérébrale). Les résultats statistiques ont pu montrer en premier lieu que l'état d'engagement est corrélé avec les mesures dynamiques. En effet, un engagement positif est corrélé avec le bon processus de raisonnement. En deuxième lieu, nous avons constaté que la mesure statique est corrélée aux deux niveaux de charge mentale de travail. Une analyse post-hoc a révélé qu'il y a une différence significative en termes de durée de fixation entre le niveau de charge optimale et la sous-charge.

9.2 Travaux futurs

Les résultats de nos travaux de recherche ont des implications pour les environnements d'apprentissage, en particulier dans les applications médicales à haut risque où il est primordial d'évaluer en permanence les performances des apprenants en surveillant leur raisonnement et leur processus décisionnel lors de la résolution de problèmes. Plus généralement, nos travaux ont aussi des implications pour les interactions homme-machine qui ont pour objectif de suivre l'expérience d'interaction de l'utilisateur afin de lui fournir l'intervention nécessaire au besoin. Nos travaux présentent toutefois quelques limitations qui ouvrent les horizons à de nombreuses perspectives de recherche. En voici quelques exemples :

Améliorer l'approche proposée : dans le chapitre 8, nous avons proposé une approche multimodale à base de deux senseurs physiologiques à savoir le traceur de regard et l'EEG qui permettent de déterminer quels sont les indicateurs cérébraux et visuels capables d'évaluer le processus de raisonnement. Nous pouvons améliorer cette approche de diverses manières. D'abord d'autres mesures statiques pourraient être calculées à partir des mouvements des yeux comme par exemple la taille de la pupille ou encore la longueur de saccade. D'autres senseurs physiologiques pourraient être intégrés tels que la conductance cutanée ou la pression sanguine.

Une autre limitation de nos travaux est que nous ne prenons pas en considération les émotions des apprenants lors de l'apprentissage. Une perspective intéressante serait de renforcer la composante émotionnelle dans l'approche proposée afin d'avoir un aperçu de l'état affectif de l'apprenant. Nous pouvons à titre d'exemple, mesurer le niveau de stress de l'apprenant, sa frustration ou sa confusion ce qui permet d'avoir d'autres types de données qui renforceraient la relation étroite qui existe entre émotion et apprentissage vu que les émotions ont une influence

sur les processus cognitifs de l'individu tels que la prise de décision et la mémorisation (Clore and Huntsinger 2007; Loewenstein and Lerner 2003).

Le nombre restreint des participants dans les deux expérimentations constitue aussi une limite. En effet, dans la deuxième expérimentation les participants étaient des étudiants en médecine qui n'étaient pas toujours disponibles sachant la rigidité de leur emploi du temps. La durée de cette expérimentation (presque 2 mois et demi) comparée à la première (1 mois) renseigne sur la difficulté que nous avons eu à les recruter. D'habitude, j'envoie un appel général à tous les étudiants du département concerné comme c'est le cas avec les étudiants en informatique mais pour ceux de la médecine je suis restée plus que deux semaines sans recevoir de réponses. Donc, j'étais obligée de faire du porte à porte et j'ai même demandé aux professeurs de bien vouloir transmettre le message à leurs étudiants durant les heures de cours. Aussi, en ce qui concerne l'impact potentiel du genre des participants sur leur performance d'apprentissage nous considérons que c'est une piste très intéressante à creuser surtout au niveau des différences possibles dans leur processus de raisonnement en termes de leur « scan path ».

Améliorer l'environnement d'apprentissage Amnesia : le jeu sérieux éducatif Amnesia a été développé spécifiquement pour des fins expérimentales dont le but est d'évaluer les capacités cognitives des étudiants novices en premières années de médecine. L'environnement contient actuellement juste six cas médicaux basiques et trois types d'exercices cognitifs incluant un exercice de mémorisation et deux exercices de logique. Nous envisageons d'ajouter d'autres cas cliniques un peu plus complexes qui mettent en situation par exemple un patient virtuel qui peut lui-même interagir avec l'étudiant en lui exposant ses symptômes. D'autres améliorations sont envisageables au niveau de l'aspect esthétique du jeu. Par exemple, après l'interaction avec l'environnement chaque participant est invité à remplir un posttest dans lequel nous lui avons demandé son avis concernant l'ergonomie et la convivialité du jeu et dans lequel il peut aussi donner quelques remarques et suggestions. Nous envisageons donc de prendre en compte les avis des participants dans l'étape de révision d'Amnesia.

Valider le raisonnement hypothético-déductif optimal par des experts du domaine : dans le chapitre 6, nous avons présenté le raisonnement analytique sur lequel nous nous sommes inspirés pour générer notre séquence de raisonnement optimale appelée « séquence référence ». Ce processus de raisonnement décrit les étapes proposées par (Nendaz et al. 2005) du premier

contact avec le patient jusqu'à l'établissement du diagnostic et du traitement approprié. Une des limitations de ces travaux serait que nous nous sommes basés sur un seul travail de recherche sans valider concrètement nos cas médicaux par des experts du domaine. Pour pallier cette lacune, nous proposons de ne plus se baser sur le schéma logique de Nendaz et ses collègues mais de tester directement nos cas médicaux par 10 experts ou plus. L'idée est d'enregistrer le « scan path » de chacun afin d'inférer une tendance générale de raisonnement qui elle sera utilisée comme le processus de raisonnement référence par rapport auquel nous comparerons les séquences des apprenants. Mais le problème est que dans le milieu médical, il est très difficile d'avoir accès aux opinions des experts, moi-même j'ai eu cette expérience lors de la conception des cas médicaux dans Amnesia. Sans l'intervention d'une personne du corps médical je n'aurais pas pu arriver à valider mes six cas par un expert. De plus, même si nous sommes capables d'obtenir l'avis de 10 experts nous nous ne sommes pas sûre de pouvoir consolider leur différents avis et pouvoir générer un 'raisonnement référence' sur lequel tous les experts se mettent d'accord.

Développer des stratégies d'aide dans la plateforme proposée : l'objectif de notre thèse est de proposer une approche unifiée d'analyse de l'état mental et du comportement visuel de l'apprenant lors d'une activité de raisonnement afin de déterminer des indicateurs clés visuels et cérébraux de performances. Un des travaux futurs importants que nous comptons élaborer est d'intégrer un module décisionnel à notre plateforme qui se base sur ces indicateurs pour prédire en temps réel le besoin d'aide et offrir l'intervention adéquate en conséquent.

Le but essentiel de notre environnement d'apprentissage est d'améliorer le processus de raisonnement des étudiants novices tout en surveillant leur état mental. Ce module décisionnel permettra alors d'offrir deux types de stratégies d'aide : des interventions simples qui ne risquent pas de perturber le déroulement de l'expérience et des interventions complexes en cas d'états mentaux défavorables à l'apprentissage tels que le décrochage ou le désengagement. Dans le premier cas, nous utiliserons principalement les mouvements des yeux pour identifier à un moment t où est portée l'attention de l'apprenant. Si nous détectons qu'il est en train de s'égarer du chemin de raisonnement optimal, nous pouvons juste l'éclairer et le remettre sur le bon chemin en mettant en surbrillance les éléments importants qui doivent attirer son attention. Dans le deuxième cas d'interventions, nous utiliserons plutôt une combinaison de deux

modalités à savoir : l'activité cérébrale et les mesures oculaires dynamiques. Nous envisageons pour cela avoir recours aux réseaux bayésiens dynamiques ou les arbres de décisions pour implanter cette partie du module décisionnel qui se base sur les indicateurs cérébraux et visuels pour générer l'intervention appropriée. Toutefois, ce module décisionnel est très complexe à développer vu la nature hautement risquée d'un diagnostic clinique. Les interventions à mettre en place doivent être bien étudiées pour ne pas risquer de perturber l'apprenant. En effet, dans leur raisonnement les cliniciens novices appliquent ce qu'ils ont appris à la lettre, ils ne sont pas assez expérimentés pour s'adapter à une situation particulière qui risque de perturber leurs fondements théoriques.

Expérimenter notre approche multimodale dans un autre environnement : les expérimentations de cette thèse sont principalement faites avec le jeu éducatif Amnesia qui met en évidence le raisonnement clinique qui compte parmi les types de raisonnement les plus complexes à analyser sachant les enjeux et les risques associés à un mauvais diagnostic de la part du clinicien. Nous nous sommes basés dans nos travaux à l'analyse du processus de raisonnement des étudiants novices. Une perspective serait d'analyser aussi le raisonnement des cliniciens experts qui n'utilisent pas le raisonnement hypothético-déductif. Toutefois, il est très difficile de généraliser vu que les cliniciens experts utilisent d'une façon intuitive soit le raisonnement inductif soit le déductif selon les circonstances. Le problème est que pour évaluer si un processus de raisonnement est correct ou non, dans notre approche nous le validons par rapport à un raisonnement référence et vu le caractère 'intuitif' du raisonnement des experts l'évaluation risque d'être plus compliquée.

Bibliographie

- Acee, T. W., Kim, H., Kim, H. J., Kim, J.-I., Chu, H.-N. R., Kim, M., et al. (2010). Academic boredom in under- and over-challenging situations. *Contemporary Educational Psychology*, *35*(1), 17–27. doi:10.1016/j.cedpsych.2009.08.002
- Alford, B. L., Rollins, K. B., Padrón, Y. N., & Waxman, H. C. (2016). Using Systematic Classroom Observation to Explore Student Engagement as a Function of Teachers' Developmentally Appropriate Instructional Practices (DAIP) in Ethnically Diverse Pre-kindergarten Through Second-Grade Classrooms. *Early Childhood Education Journal*, *44*(6), 623–635. doi:10.1007/s10643-015-0748-8
- Alhargan, A., Cooke, N., & Binjammaz, T. (2017). Affect recognition in an interactive gaming environment using eye tracking. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 285–291. doi:10.1109/ACII.2017.8273614
- Alhasan, K., Chen, L., & Chen, F. (2018). An Experimental Study of Learning Behaviour in an ELearning Environment. *The IEEE 20th International Conference on High Performance Computing and Communications*, pp. 1398–1403.
- Anderson, C. A., & Bushman, B. J. (2001). Effects of Violent Video Games on Aggressive Behavior, Aggressive Cognition, Aggressive Affect, Physiological Arousal, and Prosocial Behavior: A Meta-Analytic Review of the Scientific Literature. *Psychological Science*, *12*(5), 353–359.
- Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Pozzi, S., & Babiloni, F. (2016). Chapter 10 - A passive brain-computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks. In D. Coyle (Ed.), *Progress in Brain Research*, 295–328. doi:10.1016/bs.pbr.2016.04.021
- Atherton, M., Shah, M., Vazquez, J., Griffiths, Z., Jackson, B., & Burgess, C. (2017). Using learning analytics to assess student engagement and academic outcomes in open access enabling programmes. *Open Learning: The Journal of Open, Distance and e-Learning*, *32*(2), 119–136. doi:10.1080/02680513.2017.1309646

- Azevedo, R. (2015). Defining and Measuring Engagement and Learning in Science: Conceptual, Theoretical, Methodological, and Analytical Issues. *Educational Psychologist*, 50(1), 84–94. doi:10.1080/00461520.2015.1004069
- Ball, L. J., Phillips, P., Wade, C. N., & Quayle, J. D. (2006). Effects of Belief and Logic on Syllogistic Reasoning. *Experimental Psychology*, 53(1), 77–86. doi:10.1027/1618-3169.53.1.77
- Ben Hamida, S., Penzel, T., & Ahmed, B. (2015, August 25). EEG time and frequency domain analysis of Primary insomnia. *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6206–6209.
- Ben Khedher, A., & Frasson, C. (2016a). AMNESIA, a Dynamic Environment for Progressive Assessment of Cognitive Functions. In *Intelligent Tutoring Systems*, pp. 545–546.
- Ben Khedher, A., & Frasson, C. (2016b). Predicting User Learning Performance From Eye Movements During Interaction With a Serious Game. In *EdMedia: World Conference on Educational Media and Technology, Association for the Advancement of Computing in Education (AACE)*, pp. 1504–1511.
- Ben Khedher, A., Jraidi, I., & Frasson, C. (2016). Learners' performance tracking using eye gaze data. *1st International Workshop on Supporting Dynamic Cognitive, Affective, and Metacognitive Processes (SD-CAM), part of the 13th International Conference on Intelligent Tutoring Systems*, pp. 15–24.
- Ben Khedher, A., Jraidi, I., & Frasson, C. (2017a). Assessing Learners' Reasoning Using Eye Tracking and a Sequence Alignment Method. In *International Conference on Intelligent Computing*, pp. 47–57.
- Ben Khedher, A., Jraidi, I., & Frasson, C. (2017b). Tracking Students' Analytical Reasoning Using Visual Scan Paths. In *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, pp. 53–54. doi:10.1109/ICALT.2017.151
- Ben Khedher, A., Jraidi, I., & Frasson, C. (2018a). Local Sequence Alignment for Scan Path Similarity Assessment. *International Journal of Information and Education Technology*, 8(7).
- Ben Khedher, A., Jraidi, I., & Frasson, C. (2018b). Static and dynamic eye movement metrics for students' performance assessment. *Smart Learning Environments*, 5(1), 14. doi:10.1186/s40561-018-0065-y

- Ben Khedher, A., Jraidi, I., & Frasson, C. (2018c). Exploring Students' Eye Movements to Assess Learning Performance in a Serious Game. In *EdMedia + Innovate Learning: Association for the Advancement of Computing in Education (AACE)*, pp. 394–401.
- Ben Khedher, A., Jraidi, I., & Frasson, C. (2018d). What Can Eye Movement Patterns Reveal About Learners' Performance? In *14th International Conference on Intelligent Tutoring Systems*, pp. 415–417.
- Ben Khedher, A., Jraidi, I., & Frasson, C. (2019). Tracking Students' Mental Engagement Using EEG Signals during an Interaction with a Virtual Learning Environment. *Journal of Intelligent Learning Systems and Applications*, *11*(1), 1–14. doi:10.4236/jilsa.2019.111001
- Benbunan-Fich, R. (2001). Using protocol analysis to evaluate the usability of a commercial web site. *Information & Management*, *39*(2), 151–163. doi:10.1016/S0378-7206(01)00085-4
- Benlamine, S., Bouslimi, S., Harley, J., Frasson, C., & Dufresne, A. (2015). Toward Brain-based Gaming: Measuring Engagement During Gameplay. *EdMedia: World Conference on Educational Media and Technology*, pp. 717–722.
- Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., et al. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine*, *78*(5 Suppl), B231-244.
- Blascheck, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., & Ertl, T. (2014). *State-of-the-Art of Visualization for Eye Tracking Data*. In: EuroVis-STARs, pp. 63–82.
- Boucheix, J.-M., Lowe, R. K., Putri, D. K., & Groff, J. (2013). Cueing animations: Dynamic signaling aids information extraction and comprehension. *Learning and Instruction*, *25*, 71–84. doi:10.1016/j.learninstruc.2012.11.005
- Braunagel, C., Geisler, D., Rosenstiel, W., & Kasneci, E. (2017). Online Recognition of Driver-Activity Based on Visual Scanpath Classification. *IEEE Intelligent Transportation Systems Magazine*, *9*(4), 23–36. doi:10.1109/MITS.2017.2743171
- Brouwer, A.-M., Hogervorst, M. A., Oudejans, B., Ries, A. J., & Touryan, J. (2017). EEG and Eye Tracking Signatures of Target Encoding during Structured Visual Search. *Frontiers in Human Neuroscience*, *11*, 264. doi:10.3389/fnhum.2017.00264

- Bulling, A., Ward, J. A., Gellersen, H., & Troster, G. (2011). Eye Movement Analysis for Activity Recognition Using Electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4), 741–753. doi:10.1109/TPAMI.2010.86
- Burns, H. L., & Capps, C. G. (1988). Foundations of intelligent tutoring systems: an introduction. In J. J. R. Martha C. Polson (Ed.), *Foundations of Intelligent Tutoring Systems*, pp. 1–19.
- Byrne, B., Freebody, P., & Gates, A. (1992). Longitudinal Data on the Relations of Word-Reading Strategies to Comprehension, Reading Time, and Phonemic Awareness. *Reading Research Quarterly*, 27(2), 141–151. doi:10.2307/747683
- Byun, J. H., Loh, C. S., & Zhou, T. (2014). Assessing play-learners' performance in serious game environments by using in situ data: Using eye tracking for Serious Game Analytics. In *Annual Conference of the Association for Educational Communications and Technology (AECT)*, Jacksonville, FL.
- Cao, Y., Theune, M., & Nijholt, A. (2009). Modality Effects on Cognitive Load and Performance in High-load Information Presentation. In *Proceedings of the 14th International Conference on Intelligent User Interfaces*, pp. 335–344. doi:10.1145/1502650.1502697
- Catellin, S. (2004). L'abduction: une pratique de la découverte scientifique et littéraire. *Hermès, La Revue*, n° 39(2), 179–185.
- Chakraborty, J., & McGuire, M. P. (2016, December). Directional scan path characterization of eye tracking sequences: A multi-scale approach. *Future Technologies Conference (FTC)*, pp. 51–61.
- Chang, C., Zhang, C., Chen, L., & Liu, Y. (2018). An Ensemble Model Using Face and Body Tracking for Engagement Detection. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, pp. 616–622. doi:10.1145/3242969.3264986
- Chaouachi, M. (2015, April 30). *Modélisation de l'engagement et de la charge mentale de travail dans les Systèmes Tutoriels Intelligents*. Retrieved from <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/11958>
- Chaouachi, M., Chalfoun, P., Jraid, I., & Frasson, C. (2010). Affect and Mental Engagement: Towards Adaptability for Intelligent Systems, 6.

- Chaouachi, M., Jraidi, I., & Frasson, C. (2011). Modeling Mental Workload Using EEG Features for Intelligent Systems. In *User Modeling, Adaption and Personalization*, pp. 50–61.
- Chaouachi, M., Jraidi, I., & Frasson, C. (2015, June 29). MENTOR: A Physiologically Controlled Tutoring System. *User Modeling, Adaptation and Personalization*, pp. 56–67.
- Chen, B., Chang, Y.-H., Ouyang, F., & Zhou, W. (2018). Fostering student engagement in online discussion through social learning analytics. *The Internet and Higher Education*, 37, 21–30. doi:10.1016/j.iheduc.2017.12.002
- Chen, C.-M., Wang, J.-Y., & Yu, C.-M. (2017). Assessing the attention levels of students by using a novel attention aware system based on brainwave signals. *British Journal of Educational Technology*, 48(2), 348–369. doi:10.1111/bjet.12359
- Chen, P.-S. D., Lambert, A. D., & Guidry, K. R. (2010). Engaging online learners: The impact of Web-based learning technology on college student engagement. *Computers & Education*, 54(4), 1222–1232. doi:10.1016/j.compedu.2009.11.008
- Chen, S., Epps, J., Ruiz, N., & Chen, F. (2011). Eye Activity As a Measure of Human Mental Effort in HCI. In *Proceedings of the 16th International Conference on Intelligent User Interfaces*, pp. 315–318. doi:10.1145/1943403.1943454
- Chen, S.-C., She, H.-C., Chuang, M.-H., Wu, J.-Y., Tsai, J.-L., & Jung, T.-P. (2014). Eye movements predict students' computer-based assessment performance of physics concepts in different presentation modalities. *Computers & Education*, 74, 61–72.
- Clifton, C., Staub, A., & Rayner, K. (2007). Chapter 15 - Eye movements in reading words and sentences. In R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye Movements*, pp. 341–371. doi:10.1016/B978-008044980-7/50017-3
- Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences*, 11(9), 393–399. doi:10.1016/j.tics.2007.08.005
- Connell, J. P. (1990). Context, self, and action: A motivational analysis of self-system processes across the life span. In *The self in transition: Infancy to childhood*, pp. 61–97.
- Cordier, A., & Fuchs, B. (2005). Un assistant pour la conception et le développement des systèmes de RàPC. In S. Després (Ed.), *Plate-Forme AFIA 2005 - Atelier Raisonnement à Partir de Cas*, pp. 5–14.

- Courtemanche, F., Aïmeur, E., Dufresne, A., Najjar, M., & Mpondo, F. (2011). Activity recognition using eye-gaze movements and traditional interactions. *Interacting with Computers*, 23(3), 202–213. doi:10.1016/j.intcom.2011.02.008
- Cristino, F., Mathôt, S., Theeuwes, J., & Gilchrist, I. D. (2010). ScanMatch: A novel method for comparing fixation sequences. *Behavior Research Methods*, 42(3), 692–700. doi:10.3758/BRM.42.3.692
- Cutrell, E., & Guan, Z. (2007). What are you looking for?: an eye-tracking study of information usage in web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 407–416.
- Delécraz, J. (2017, April 29). Processus cognitifs : qu'est-ce que c'est et comment pouvons-nous améliorer nos processus mentaux. *CogniFit's Blog*. <https://blog.cognifit.com/fr/processus-cognitifs-processus-mentaux/>. Accessed 17 January 2019
- Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: analyses and review. *Journal of Experimental Psychology. General*, 140(4), 552–572. doi:10.1037/a0023885
- Djamasbi, S. (2014). Eye tracking and web experience. *AIS Transactions on Human-Computer Interaction*, 6. doi:10.17705/1thci.00060
- D'Mello, S., Chipman, P., & Graesser, A. (2007). Posture as a Predictor of Learner's Affective Engagement. *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, pp. 905–910.
- D'mello, S. K., Craig, S. D., Gholson, B., & Franklin, S. (2005). Integrating affect sensors in an intelligent tutoring system. In *In Affective Interactions: The Computer in the Affective Loop Workshop at International Conference on Intelligent User Interfaces*, pp. 7–13.
- D'Mello, S. K., Jackson, T., Craig, S. D., Morgan, B., Chip-man, P., White, H. L., et al. (2008). AutoTutor Detects and Responds to Learners Affective and Cognitive States. Presented at the Workshop on emotional and cognitive issues at the international conference on intelligent tutoring systems, pp. 306–308.
- D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies*, 70(5), 377–398. doi:10.1016/j.ijhcs.2012.01.004

- D'Mello, S., Taylor, R., Davidson, K., & Graesser, A. (2008). Self Versus Teacher Judgments of Learner Emotions During a Tutoring Session with AutoTutor. In B. P. Woolf, E. Aïmeur, R. Nkambou, & S. Lajoie (Eds.), *Intelligent Tutoring Systems*, pp. 9–18.
- Drachen, A., Nacke, L. E., Yannakakis, G., & Pedersen, A. L. (2010). Correlation Between Heart Rate, Electrodermal Activity and Player Experience in First-person Shooter Games. In *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*, pp. 49–54. doi:10.1145/1836135.1836143
- Duchesne, K. (2005, April). *Perception du soutien social et engagement affectif organisationnel* (masters). Université du Québec en Abitibi-Témiscamingue, Rouyn-Noranda. Retrieved from <http://depositum.uqat.ca/312/>
- Duchowski, A. T., Driver, J., Jolaoso, S., Tan, W., Ramey, B. N., & Robbins, A. (2010). Scanpath Comparison Revisited. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pp. 219–226.
- Eisenbarth, H., & Alpers, G. W. (2011). Happy mouth and sad eyes: Scanning emotional facial expressions. *Emotion, 11*(4), 860–865. doi:10.1037/a0022758
- Eivazi, S., & Bednarik, R. (2011, January 1). Predicting Problem-Solving Behavior and Performance Levels from Visual Attention Data. *Proceedings of the 2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction at IUI 2011*, pp. 9–16.
- El-Abbasy, K., Angelopoulou, A., & Towell, T. (2019). Measuring the Engagement of the Learner in a Controlled Environment using Three Different Biosensors. Presented at the 10th International Conference on Computer Supported Education, pp. 278–284.
- Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 8*(2), 97–120. doi:10.1016/j.trf.2005.04.012
- Eraslan, S., & Yesilada, Y. (2015). Patterns in Eyetracking Scanpaths and the Affecting Factors. *J. Web Eng., 14*(5–6), 363–385.
- Eraslan, S., Yesilada, Y., & Harper, S. (2015). Eye tracking scanpath analysis techniques on web pages: A survey, evaluation and comparison. *Journal of Eye Movement Research, 9*(1). doi:10.16910/jemr.9.1.2
- Evans, J. S. B. T. (2002). Logic and human reasoning: an assessment of the deduction paradigm. *Psychological Bulletin, 128*(6), 978–996.

- Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General*, 113(3), 464–486. doi:10.1037/0096-3445.113.3.464
- Festinger, L. (1962). *A Theory of Cognitive Dissonance*. Stanford University Press.
- Forsberg, E., Ziegert, K., Hult, H., & Fors, U. (2014). Clinical reasoning in nursing, a think-aloud study using virtual patients – A base for an innovative assessment. *Nurse Education Today*, 34(4), 538–542. doi:10.1016/j.nedt.2013.07.010
- Fredricks, J. A., & McColskey, W. (2012). The Measurement of Student Engagement: A Comparative Analysis of Various Methods and Student Self-report Instruments. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of Research on Student Engagement*, pp. 763–782. doi:10.1007/978-1-4614-2018-7_37
- Frutos-Pascual, M., & Garcia-Zapirain, B. (2015). Assessing Visual Attention Using Eye Tracking Sensors in Intelligent Cognitive Therapies Based on Serious Games. *Sensors (Basel, Switzerland)*, 15(5), 11092–11117. doi:10.3390/s150511092
- Fu, W.-T. (2001). ACT-PRO action protocol analyzer: A tool for analyzing discrete action protocols. *Behavior Research Methods, Instruments, & Computers*, 33(2), 149–158. doi:10.3758/BF03195360
- Ghali, R., Frasson, C., & Ouellet, S. (2016). Towards Real Time Detection of Learners' Need of Help in Serious Games. In *The Twenty-Ninth International Flairs Conference*.
- Glady, Y., Thibaut, J.-P., & French, R. M. (2013). Visual Strategies in Analogical Reasoning Development: A New Method for Classifying Scanpaths. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*, 35(35).
- Goldberg, J. H., & Helfman, J. I. (2010). Scanpath Clustering and Aggregation. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pp. 227–234.
- Gunes, H., & Pantic, M. (2010). Automatic, Dimensional and Continuous Emotion Recognition. *International Journal of Synthetic Emotions (IJSE)*, 1(1), 68–99. doi:10.4018/jse.2010101605
- Gutica, M., & Conati, C. (2013). Student Emotions with an Edu-game: A Detailed Analysis. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp. 534–539. doi:10.1109/ACII.2013.94

- Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R., & Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior*, *48*, 615–625. doi:10.1016/j.chb.2015.02.013
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology*, *52*, 139–183. doi:10.1016/S0166-4115(08)62386-9
- Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194. doi:10.1016/j.tics.2005.02.009
- Helme, S., & Clarke, D. (2001). Identifying cognitive engagement in the mathematics classroom. *Mathematics Education Research Journal*, *13*(2), 133–153. doi:10.1007/BF03217103
- Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, *90*, 36–53. doi:10.1016/j.compedu.2015.09.005
- Heraz, A., Razaki, R., & Frasson, C. (2007). Using machine learning to predict learner emotional state from brainwaves. In *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT)*, pp. 853–857. doi:10.1109/ICALT.2007.277
- Ho, H. N. J., Tsai, M.-J., Wang, C.-Y., & Tsai, C.-C. (2014). Prior knowledge and online inquiry-based science reading: Evidence from eye tracking. *International journal of science and mathematics education*, *12*(3), 525–554. doi:10.1007/s10763-013-9489-6
- Holland, C., & Komogortsev, O. V. (2011, October). Biometric identification via eye movement scanpaths in reading. *International Joint Conference on Biometrics (IJCB)*, pp. 1–8.
- Hood, D. C., & Raza, A. S. (2014). On improving the use of OCT imaging for detecting glaucomatous damage. *British Journal of Ophthalmology*, *98*(Suppl 2), ii1–ii9. doi:10.1136/bjophthalmol-2014-305156
- Hou, H.-T. (2015). Integrating cluster and sequential analysis to explore learners' flow and behavioral patterns in a simulation game with situated-learning context for science courses: A video-based process exploration. *Computers in Human Behavior*, *48*, 424–435. doi:10.1016/j.chb.2015.02.010

- Huang, Y., & Zhang, L. (2004). Rapid and sensitive dot-matrix methods for genome analysis. *Bioinformatics*, 20(4), 460–466. doi:10.1093/bioinformatics/btg429
- Hume, D. (2012). Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them. *Organizational Behavior*, 258–297.
- Hunter, M., Mach, Q. H., & Grewal, R. S. (2010). The Relationship Between Scan Path Direction and Cognitive Processing. *Proceedings of the Third C* Conference on Computer Science and Software Engineering*, pp. 97–100.
- Hwang, T., Kim, M., Hwangbo, M., & Oh, E. (2014). Comparative analysis of cognitive tasks for modeling mental workload with electroencephalogram. In *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2661–2665. doi:10.1109/EMBC.2014.6944170
- Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction*, 20. doi:10.1016/j.learninstruc.2009.02.013
- Ishimaru, S., Kunze, K., Kise, K., Weppner, J., Dengel, A., Lukowicz, P., & Bulling, A. (2014). In the Blink of an Eye: Combining Head Motion and Eye Blink Frequency for Activity Recognition with Google Glass. In *Proceedings of the 5th Augmented Human International Conference*, pp. 15:1–15:4. doi:10.1145/2582051.2582066
- Jaafar, S., Awaludin, N. S., & Bakar, N. S. (2014). MOTIVATIONAL AND SELF-REGULATED LEARNING COMPONENTS OF CLASSROOM ACADEMIC PERFORMANC. *Journal of educational psychology*, 8(1), 33.
- Jaques, N., Conati, C., Harley, J. M., & Azevedo, R. (2014). Predicting affect from gaze data during interaction with an intelligent tutoring system. *International Conference on Intelligent Tutoring Systems*, pp. 29–38.
- Jarodzka, H., Holmqvist, K., & Nyström, M. (2010). A Vector-based, Multidimensional Scanpath Similarity Measure. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pp. 211–218.
- JASPER, H. H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol*, 10, 370–375.

- Jo, I.-H., Yu, T., Lee, H., & Kim, Y. (2015). Relations between Student Online Learning Behavior and Academic Achievement in Higher Education: A Learning Analytics Approach. *Emerging Issues in Smart Learning*, pp. 275–287.
- Jr, P. T. C., & McCrae, R. R. (2013). *Personality in Adulthood: A Five-Factor Theory Perspective*. Routledge.
- Jraidi, I., Chalfoun, P., & Frasson, C. (2012). Implicit Strategies for Intelligent Tutoring Systems. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Intelligent Tutoring Systems*, pp. 1–10. doi:10.1007/978-3-642-30950-2_1
- Jraidi, I., Chaouachi, M., & Frasson, C. (2013). A dynamic multimodal approach for assessing learners' interaction experience. *International Conference on Multimodal Interaction*, pp. 271–278.
- Jraidi, I., Chaouachi, M., & Frasson, C. (2014). A Hierarchical Probabilistic Framework for Recognizing Learners' Interaction Experience Trends and Emotions. *Advances in Human-Computer Interaction, 2014(632630)*, 1–16. doi:10.1155/2014/632630
- Jraidi, I., & Frasson, C. (2010). Subliminally Enhancing Self-esteem: Impact on Learner Performance and Affective State. *Intelligent Tutoring Systems*, pp. 11–20.
- Jraidi, I., & Frasson, C. (2013). Student's Uncertainty Modeling through a Multimodal Sensor-Based Approach. *Journal of Educational Technology & Society, 16(1)*, 219–230.
- Jung, Y. J., Zimmerman, H. T., & Pérez-Edgar, K. (2018). A Methodological Case Study with Mobile Eye-Tracking of Child Interaction in a Science Museum. *TechTrends, 62(5)*, 509–517. doi:10.1007/s11528-018-0310-9
- Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8(4)*, 441–480. doi:10.1016/0010-0285(76)90015-3
- Kardan, S., & Conati, C. (2012, July 16). Exploring Gaze Data for Determining User Learning with an Interactive Simulation. *User Modeling, Adaptation, and Personalization*, pp. 126–138.
- Keith, J. R., Rapgay, L., Theodore, D., Schwartz, J. M., & Ross, J. L. (2015). An assessment of an automated EEG biofeedback system for attention deficits in a substance use disorders residential treatment setting. *Psychology of Addictive Behaviors, 29(1)*, 17–25. doi:10.1037/adb0000016

- Kiili, K., Ketamo, H., & Kickmeier-Rust, M. D. (2014). Evaluating the usefulness of Eye Tracking in Game-based Learning. *International Journal of Serious Games*, 1(2). doi:10.17083/ijsg.v1i2.15
- Kim, J., Seo, J., & Laine, T. H. (2018). Detecting boredom from eye gaze and EEG. *Biomedical Signal Processing and Control*, 46, 302–313. doi:10.1016/j.bspc.2018.05.034
- Klem, G. H., Lüders, H. O., Jasper, H. H., & Elger, C. (1999). The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalography and Clinical Neurophysiology. Supplement*, 52, 3–6.
- Knoepfle, D. T., Wang, J. T., & Camerer, C. F. (2009). Studying Learning in Games Using Eye-Tracking. *Journal of the European Economic Association*, 7(2–3), 388–398. doi:10.1162/JEEA.2009.7.2-3.388
- Kobrin, J. L., Larson, S., Cromwell, A., & Garza, P. (2015). A Framework for Evaluating Learning Progressions on Features Related to Their Intended Uses. *Journal of Educational Research and Practice*, 5(1), 58–73.
- Krithika L.B, & Lakshmi Priya GG. (2016). Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric. *Procedia Computer Science*, 85, 767–776. doi:10.1016/j.procs.2016.05.264
- Krupinski, E. A., Tillack, A. A., Richter, L., Henderson, J. T., Bhattacharyya, A. K., Scott, K. M., & Weinstein, R. S. (2006). Eye-movement study and human performance using telepathology virtual slides. Implications for medical education and differences with experience. *Human Pathology*, 37. doi:10.1016/j.humpath.2006.08.024
- Kurzhals, K., Heimerl, F., & Weiskopf, D. (2014). ISeeCube: Visual Analysis of Gaze Data for Video. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 43–50. doi:10.1145/2578153.2578158
- Lajoie, S. P., Naismith, L., Poitras, E., Hong, Y.-J., Cruz-Panesso, I., Ranellucci, J., et al. (2013). Technology-Rich Tools to Support Self-Regulated Learning and Performance in Medicine. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies*, pp. 229–242. doi:10.1007/978-1-4419-5546-3_16

- Lallé, S., Conati, C., & Carenini, G. (2016). Predicting Confusion in Information Visualization from Eye Tracking and Interaction Data. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp. 2529–2535.
- Lallé, S., Conati, C., & Carenini, G. (2017). Impact of Individual Differences on User Experience with a Visualization Interface for Public Engagement. *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, pp. 247–252.
- Lane, A. M., Whyte, G. P., Terry, P. C., & Nevill, A. M. (2005). Mood, self-set goals and examination performance: the moderating effect of depressed mood. *Personality and Individual Differences*, 39(1), 143–153. doi:10.1016/j.paid.2004.12.015
- Lee, W.-K., & Wu, C.-J. (2018). Eye Movements in Integrating Geometric Text and Figure: Scanpaths and Given-New Effects. *International Journal of Science and Mathematics Education*, 16(4), 699–714. doi:10.1007/s10763-016-9790-2
- Lemaire, P. (1999). *Psychologie cognitive. Ouvertures psychologiques, De Boeck Université.*
- Levenshtein, V. I. (1966). Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*, 10, 707.
- Leyman, E. L., Mirka, G. A., Kaber, D. B., & Sommerich, C. M. (2004). Cervicobrachial muscle response to cognitive load in a dual-task scenario. *Ergonomics*, 47(6), 625–645. doi:10.1080/00140130310001629766
- Lin, F.-R., & Kao, C.-M. (2018). Mental effort detection using EEG data in E-learning contexts. *Computers & Education*, 122, 63–79. doi:10.1016/j.compedu.2018.03.020
- Lin, H.-C. K., Su, S.-H., Chao, C.-J., Hsieh, C.-Y., & Tsai, S.-C. (2016). Construction of Multi-mode Affective Learning System: Taking Affective Design as an Example. *Journal of Educational Technology & Society*, 19(2), 132–147.
- Lin, J. J.-H., & Lin, S. S. J. (2014). COGNITIVE LOAD FOR CONFIGURATION COMPREHENSION IN COMPUTER-SUPPORTED GEOMETRY PROBLEM SOLVING: AN EYE MOVEMENT PERSPECTIVE. *International Journal of Science and Mathematics Education*, 12(3), 605–627. doi:10.1007/s10763-013-9479-8
- Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 4(1), 6–14. doi:10.1016/S1364-6613(99)01418-7

- Loboda, T. D., Brusilovsky, P., & Brunstein, J. (2011). Inferring Word Relevance from Eye-movements of Readers. *Proceedings of the 16th International Conference on Intelligent User Interfaces*, pp. 175–184.
- Loewenstein, G., & Lerner, J. S. (2003). The role of affect in decision making. In *Handbook of affective sciences*, pp. 619–642.
- Loo, S. K., Lenartowicz, A., & Makeig, S. (2016). Use of EEG biomarkers in child psychiatry research: current state and future directions. *Journal of child psychology and psychiatry, and allied disciplines*, 57(1), 4–17. doi:10.1111/jcpp.12435
- Lu, O. H. T., Huang, J. C. H., Huang, A. Y. Q., & Yang, S. J. H. (2017). Applying learning analytics for improving students engagement and learning outcomes in an MOOCs enabled collaborative programming course. *Interactive Learning Environments*, 25(2), 220–234. doi:10.1080/10494820.2016.1278391
- Lujan-Moreno, G. A., Atkinson, R. K., & Runger, G. (2016, January 1). EEG-based user performance prediction using random forest in a dynamic learning environment. *Intelligent Tutoring Systems: Structure, Applications and Challenges*, pp. 105–128.
- Lum, H. C., Greatbatch, R. L., Waldfogle, G. E., Benedict, J. D., & Nembhard, D. A. (2016). The Relationship of Eye Movement, Workload, and Attention on Learning in a Computer-Based Training Program. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60(1), 1477–1481. doi:10.1177/1541931213601339
- Luo, L., Kiewra, K. A., Peteranetz, M. S., & Flanigan, A. E. (2017). Using Eye-Tracking Technology to Understand How Graphic Organizers Aid Student Learning. *Eye-Tracking Technology Applications in Educational Research*, 220–238. doi:10.4018/978-1-5225-1005-5.ch011
- Ma, J., Han, X., Yang, J., & Cheng, J. (2015). Examining the necessary condition for engagement in an online learning environment based on learning analytics approach: The role of the instructor. *The Internet and Higher Education*, 24, 26–34. doi:10.1016/j.iheduc.2014.09.005
- Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Role of subjective and objective measures of cognitive processing during learning in explaining the spatial contiguity effect. *Learning and Instruction*, 61, 23–34. doi:10.1016/j.learninstruc.2018.12.001

- Martínez-Gómez, P., & Aizawa, A. (2014). Recognition of Understanding Level and Language Skill Using Measurements of Reading Behavior. *Proceedings of the 19th International Conference on Intelligent User Interfaces*, pp. 95–104.
- Maynard, O. M., Munafò, M. R., & Leonards, U. (2013). Visual attention to health warnings on plain tobacco packaging in adolescent smokers and non-smokers. *Addiction (Abingdon, England)*, *108*(2), 413–419. doi:10.1111/j.1360-0443.2012.04028.x
- Mega, C., Ronconi, L., & De Beni, R. (2014). What Makes a Good Student? How Emotions, Self-Regulated Learning, and Motivation Contribute to Academic Achievement. *Journal of Educational Psychology*, *106*(1), 121–131. doi:10.1037/a0033546
- Miller, B. W. (2015). Using Reading Times and Eye-Movements to Measure Cognitive Engagement. *Educational Psychologist*, *50*(1), 31–42. doi:10.1080/00461520.2015.1004068
- Mills, C., Fridman, I., Soussou, W., Waghray, D., Olney, A. M., & D’Mello, S. K. (2017). Put Your Thinking Cap on: Detecting Cognitive Load Using EEG During Learning. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, pp. 80–89. doi:10.1145/3027385.3027431
- Muldner, K., & Bursleson, W. (2015). Utilizing sensor data to model students’ creativity in a digital environment. *Computers in Human Behavior*, *42*, 127–137. doi:10.1016/j.chb.2013.10.060
- Muldner, K., Christopherson, R., Atkinson, R., & Bursleson, W. (2009). Investigating the Utility of Eye-Tracking Information on Affect and Reasoning for User Modeling. In G.-J. Houben, G. McCalla, F. Pianesi, & M. Zancanaro (Eds.), *User Modeling, Adaptation, and Personalization*, pp. 138–149.
- Nakamaru, S. (2011). Investment and Return. *Journal of Research on Technology in Education*, *44*(4), 273–291. doi:10.1080/15391523.2012.10782591
- Nasoz, F., Alvarez, K., Lisetti, C. L., & Finkelstein, N. (2004). Emotion recognition from physiological signals using wireless sensors for presence technologies. *Cognition, Technology & Work*, *6*(1), 4–14. doi:10.1007/s10111-003-0143-x
- Nendaz, M., Charlin, B., Leblanc, V., & Bordage, G. (2005). Le raisonnement clinique: données issues de la recherche et implications pour l’enseignement. *Pédagogie Médicale*, *6*. doi:10.1051/pmed:2005028

- Newble, D. I., Van der Vleuten, C., & Norman, G. (1995). Assessing clinical reasoning. In: *Higgs J, Jones M (eds) Clinical reasoning and the health professions*, pp. 168-178.
- Ninaus, M., Kober, S. E., Friedrich, E. V. C., Neuper, C., & Wood, G. (2014, September). The Potential Use of Neurophysiological Signals for Learning Analytics. *6th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES)*, pp. 1–5.
- Noordzij, M. L., Dorrestijn, S. M., & Berg, I. A. van den. (2017). An idiographic study into the physiology and selfreported mental workload of learning to drive a car. *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2016 Annual Conference*.
- Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, *11*(9), 929-IN8. doi:10.1016/0042-6989(71)90213-6
- Nourbakhsh, N., Chen, F., Wang, Y., & Calvo, R. A. (2017). Detecting Users' Cognitive Load by Galvanic Skin Response with Affective Interference. *ACM Trans. Interact. Intell. Syst.*, *7*(3), 12:1–12:20. doi:10.1145/2960413
- Nourbakhsh, N., Wang, Y., & Chen, F. (2013). GSR and Blink Features for Cognitive Load Classification. In P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2013*, pp. 159–166.
- Nourbakhsh, N., Wang, Y., Chen, F., & Calvo, R. A. (2012). Using Galvanic Skin Response for Cognitive Load Measurement in Arithmetic and Reading Tasks. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*, pp. 420–423. doi:10.1145/2414536.2414602
- Ochs, M., & Frasson, C. (2004). Emotionally Intelligent Tutoring Systems (EITS). *FLAIRS Conference*, pp. 251–256.
- Olafson, K. M., & Ferraro, F. R. (2001). Effects of Emotional State on Lexical Decision Performance. *Brain and Cognition*, *45*(1), 15–20. doi:10.1006/brcg.2000.1248
- Pachman, M., Arguel, A., Lockyer, L., Kennedy, G., & Lodge, J. M. (2016). Eye tracking and early detection of confusion in digital learning environments: Proof of concept. *Australasian Journal of Educational Technology*, *32*(6).
- Pardo, A., Han, F., & Ellis, R. A. (2017). Combining University Student Self-Regulated Learning Indicators and Engagement with Online Learning Events to Predict Academic

- Performance. *IEEE Transactions on Learning Technologies*, 10(1), 82–92. doi:10.1109/TLT.2016.2639508
- Pearson, W. R. (2013). An Introduction to Sequence Similarity (“Homology”) Searching. *Current Protocols in Bioinformatics*, 42(1), 3.1.1-3.1.8. doi:10.1002/0471250953.bi0301s42
- Pekrun, R., Goetz, T., Anne, Frenzel, Barchfeld, P., & Perry, V. H. (2012). Measuring emotions in students’ learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary educational psychology*, 36(1), 36–48.
- Pham, P., & Wang, J. (2015). AttentiveLearner: Improving Mobile MOOC Learning via Implicit Heart Rate Tracking. *Artificial Intelligence in Education*, pp. 367–376.
- Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33–40. doi:10.1037/0022-0663.82.1.33
- Poitras, E. G., Doleck, T., & Lajoie, S. P. (2018). Towards detection of learner misconceptions in a medical learning environment: a subgroup discovery approach. *Educational Technology Research and Development*, 66(1), 129–145. doi:10.1007/s11423-017-9555-9
- Politzer, G. (1993). *La psychologie du raisonnement : lois de la pragmatique et logique formelle* (thesis). Paris 8. Retrieved from <http://www.theses.fr/1993PA080793>
- Poole, A., & Ball, L. J. (2005). Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future. In *Prospects*”, Chapter in C. Ghaoui (Ed.): *Encyclopedia of Human-Computer Interaction*. Pennsylvania: Idea Group, Inc.
- Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology*, 40(1–2), 187–195.
- Ranellucci, J., Hall, N. C., & Goetz, T. (2015). Achievement goals, emotions, learning, and performance: A process model. *Motivation Science*, 1(2), 98–120. doi:10.1037/mot0000014
- Raptis, G. E., Katsini, C., Belk, M., Fidas, C., Samaras, G., & Avouris, N. (2017). Using eye gaze data and visual activities to infer human cognitive styles: method and feasibility studies. *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, pp. 164–173.

- Raschke, M., Blascheck, T., Richter, M., Agapkin, T., & Ertl, T. (2014, January). Visual analysis of perceptual and cognitive processes. *International Conference on Information Visualization Theory and Applications (IVAPP)*, pp. 284–291.
- Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye Movements as Reflections of Comprehension Processes in Reading. *Scientific Studies of Reading*, *10*(3), 241–255. doi:10.1207/s1532799xssr1003_3
- Reid, G. B., & Nygren, T. E. (1988). The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology*, *52*, 185–218. doi:10.1016/S0166-4115(08)62387-0
- Renkewitz, F., & Jahn, G. (2012). Memory indexing: a novel method for tracing memory processes in complex cognitive tasks. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *38*(6), 1622–1639. doi:10.1037/a0028073
- Rienties, B., Lewis, T., McFarlane, R., Nguyen, Q., & Toetenel, L. (2018). Analytics in online and offline language learning environments: the role of learning design to understand student online engagement. *Computer Assisted Language Learning*, *31*(3), 273–293. doi:10.1080/09588221.2017.1401548
- Roach, B. J., & Mathalon, D. H. (2008). Event-Related EEG Time-Frequency Analysis: An Overview of Measures and An Analysis of Early Gamma Band Phase Locking in Schizophrenia. *Schizophrenia Bulletin*, *34*(5), 907–926. doi:10.1093/schbul/sbn093
- Rodrigue, M., Son, J., Giesbrecht, B., Turk, M., & Höllerer, T. (2015). Spatio-Temporal Detection of Divided Attention in Reading Applications Using EEG and Eye Tracking. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pp. 121–125. doi:10.1145/2678025.2701382
- Rowe, J., Mott, B., W. MCQUIGGAN, S., Sabourin, J., LEE, S., & C. LESTER, J. (2009, January 1). Crystal Island: A Narrative-Centered Learning Environment for Eighth Grade Microbiology. *14th International Conference on AI in Education Workshops Proceedings*, pp.11–20.
- Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2011). Integrating Learning, Problem Solving, and Engagement in Narrative-Centered Learning Environments. *International Journal of Artificial Intelligence in Education*, *21*(1–2), 115–133. doi:10.3233/JAI-2011-019

- Salvucci, D. D., & Anderson, J. R. (2001). Automated Eye-Movement Protocol Analysis. *Human-Computer Interaction, 16*(1), 39–86. doi:10.1207/S15327051HCI1601_2
- Sawyer, R., Smith, A., Rowe, J., Azevedo, R., & Lester, J. (2017). Enhancing Student Models in Game-based Learning with Facial Expression Recognition. *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, pp. 192–201.
- Scholz, A., von Helversen, B., & Rieskamp, J. (2015). Eye movements reveal memory processes during similarity- and rule-based decision making. *Cognition, 136*, 228–246. doi:10.1016/j.cognition.2014.11.019
- Setubal, J. C., & Meidanis, J. (1997). *Introduction to computational molecular biology*. Boston: PWS Pub.
- Setz, C., Arnrich, B., Schumm, J., Marca, R. L., Tröster, G., & Ehlert, U. (2010). Discriminating Stress From Cognitive Load Using a Wearable EDA Device. *IEEE Transactions on Information Technology in Biomedicine, 14*(2), 410–417. doi:10.1109/TITB.2009.2036164
- Sharma, K., Jermann, P., & Dillenbourg, P. (2014). How Students Learn using MOOCs: An Eye-tracking Insight. *Proceedings of the 2nd MOOC European Stakeholders Summit*, pp. 147–154.
- Shen, L., Wang, M., & Shen, R. (2009). Affective e-Learning: Using “Emotional” Data to Improve Learning in Pervasive Learning Environment. *Educational Technology & Society, 12*, 176–189.
- Siegle, G. J., Ichikawa, N., & Steinhauer, S. (2008). Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology, 45*(5), 679–687. doi:10.1111/j.1469-8986.2008.00681.x
- Slanzi, G., Balazs, J., & Velasquez, J. (2017). Combining eye tracking, pupil dilation and EEG analysis for predicting web users click intention. *Information Fusion, 35*, 51–57. doi:10.1016/j.inffus.2016.09.003
- Soleymani, M., Asghari-Esfeden, S., Fu, Y., & Pantic, M. (2016, March). Analysis of EEG Signals and Facial Expressions for Continuous Emotion Detection. *IEEE Transactions on Affective Computing*, pp. 17–28.
- Sperandio, J.-C. A. du texte. (1980). *La psychologie en ergonomie / Jean-Claude Sperandio,...* <https://gallica.bnf.fr/ark:/12148/bpt6k4806560k>.

- Su, Y.-C., & Grauman, K. (2016). Detecting Engagement in Egocentric Video. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision – ECCV 2016*, pp. 454–471.
- Sundstedt, V. (2012). Gazing at Games: An Introduction to Eye Tracking Control. *Synthesis Lectures on Computer Graphics and Animation*, 5(1), 1–113. doi:10.2200/S00395ED1V01Y201111CGR014
- Susac, A., Bubic, A., Kaponja, J., Planinic, M., & Palmovic, M. (2014). EYE MOVEMENTS REVEAL STUDENTS’ STRATEGIES IN SIMPLE EQUATION SOLVING. *International Journal of Science and Mathematics Education*, 12(3), 555–577. doi:10.1007/s10763-014-9514-4
- Swanson, H. L., O’Connor, J. E., & Cooney, J. B. (1990). An Information Processing Analysis of Expert and Novice Teachers’ Problem Solving. *American Educational Research Journal*, 27(3), 533–556. doi:10.3102/00028312027003533
- Tai, R. H., Loehr, J. F., & Brigham, F. J. (2006). An exploration of the use of eye-gaze tracking to study problem-solving on standardized science assessments. *International Journal of Research & Method in Education*, 29(2), 185–208. doi:10.1080/17437270600891614
- Takeuchi, H., & Habuchi, Y. (2007, March). A Quantitative Method for Analyzing Scan Path Data Obtained by Eye Tracker. *IEEE Symposium on Computational Intelligence and Data Mining*, pp. 283–286.
- Taub, M., Mudrick, N. V., Azevedo, R., Millar, G. C., Rowe, J., & Lester, J. (2017). Using multi-channel data with multi-level modeling to assess in-game performance during gameplay with Crystal Island. *Computers in Human Behavior*, 76, 641–655. doi:10.1016/j.chb.2017.01.038
- Tchounikine, P. (2002). Quelques éléments sur la conception et l’ingénierie des EIAH.
- Tien, G., Atkins, M. S., & Zheng, B. (2012). Measuring Gaze Overlap on Videos Between Multiple Observers. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 309–312. doi:10.1145/2168556.2168623
- Toker, D., & Conati, C. (2014). Eye Tracking to Understand User Differences in Visualization Processing with Highlighting Interventions. *User Modeling, Adaptation, and Personalization*, pp. 219–230.

- Toker, D., Conati, C., Steichen, B., & Carenini, G. (2013). Individual User Characteristics and Information Visualization: Connecting the Dots Through Eye Tracking. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 295–304.
- Trevors, G., Feyzi-Behnagh, R., Azevedo, R., & Bouchet, F. (2016). Self-regulated learning processes vary as a function of epistemic beliefs and contexts: Mixed method evidence from eye tracking and concurrent and retrospective reports. *Learning and Instruction*, 42. doi:10.1016/j.learninstruc.2015.11.003
- Tricot, A. (1998). Charge cognitive et apprentissage. Une présentation des travaux de John Sweller. *Revue de Psychologie de l'éducation*, 3, 37–64.
- Tsai, M.-J., Hou, H.-T., Lai, M.-L., Liu, W.-Y., & Yang, F.-Y. (2012). Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers & Education*, 58(1), 375–385. doi:10.1016/j.compedu.2011.07.012
- Turner, J. C., Christensen, A., Kackar-Cam, H. Z., Trucano, M., & Fulmer, S. M. (2014). Enhancing Students' Engagement: Report of a 3-Year Intervention With Middle School Teachers. *American Educational Research Journal*, 51(6), 1195–1226. doi:10.3102/0002831214532515
- Vachon, F., & Tremblay, S. (2014). What Eye Tracking Can Reveal about Dynamic Decision-Making. *Advances in cognitive engineering and neuroergonomics*, pp. 157–165.
- Van der Hiele, K., Vein, A. A., Reijntjes, R. H. A. M., Westendorp, R. G. J., Bollen, E. L. E. M., van Buchem, M. A., et al. (2007). EEG correlates in the spectrum of cognitive decline. *Clinical Neurophysiology*, 118(9), 1931–1939. doi:10.1016/j.clinph.2007.05.070
- Varin, C. (2007). *L'enseignement du raisonnement conditionnel: de la logique aux neurosciences*.
- Wang, C.-Y., Tsai, M.-J., & Tsai, C.-C. (2016). Multimedia recipe reading: Predicting learning outcomes and diagnosing cooking interest using eye-tracking measures. *Computers in Human Behavior*, 62. doi:10.1016/j.chb.2016.03.064
- Wang, F., Li, W., Mayer, R. E., & Liu, H. (2018). Animated pedagogical agents as aids in multimedia learning: Effects on eye-fixations during learning and learning outcomes. *Journal of Educational Psychology*, 110(2), 250–268. doi:10.1037/edu0000221

- Wang, S., Gwizdka, J., & Chaovalitwongse, W. A. (2016). Using Wireless EEG Signals to Assess Memory Workload in the n-Back Task. *IEEE Transactions on Human-Machine Systems*, 46(3), 424–435. doi:10.1109/THMS.2015.2476818
- Wang, X.-W., Nie, D., & Lu, B.-L. (2014). Emotional state classification from EEG data using machine learning approach. *Neurocomputing*, 129, 94–106. doi:10.1016/j.neucom.2013.06.046
- Weil-Barais, A., & Dubois, D. (1994). *L'homme cognitif*. Presses universitaires de France.
- Weill-Tessier, P., Turner, J., & Gellersen, H. (2016). How do you look at what you touch?: a study of touch interaction and gaze correlation on tablets. *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, pp. 329–330.
- Wenger, E. (1987). *Artificial intelligence and tutoring systems: computational and cognitive approaches to the communication of knowledge*. Morgan Kaufmann Publishers Inc.
- Westera, W. (2016). Performance assessment in serious games: Compensating for the effects of randomness. *Education and Information Technologies*, 21(3), 681–697. doi:10.1007/s10639-014-9347-3
- Whitehill, J., Serpell, Z., Lin, Y., Foster, A., & Movellan, J. R. (2014). The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions. *IEEE Transactions on Affective Computing*, 5(1), 86–98. doi:10.1109/TAFFC.2014.2316163
- Wigfield, A., Guthrie, J. T., Perencevich, K. C., Taboada, A., Klauda, S. L., McRae, A., & Barbosa, P. (2008). Role of reading engagement in mediating effects of reading comprehension instruction on reading outcomes. *Psychology in the Schools*, 45(5), 432–445. doi:10.1002/pits.20307
- Xu, M., Song, Y., Wang, J., Qiao, M., Huo, L., & Wang, Z. (2018). Predicting Head Movement in Panoramic Video: A Deep Reinforcement Learning Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1. doi:10.1109/TPAMI.2018.2858783
- Yang, Y.-F. (2011). Engaging Students in an Online Situated Language Learning Environment. *Computer Assisted Language Learning*, 24(2), 181–198.
- Yeo, H.-S., Lee, B.-G., & Lim, H. (2015). Hand Tracking and Gesture Recognition System for Human-computer Interaction Using Low-cost Hardware. *Multimedia Tools Appl.*, 74(8), 2687–2715. doi:10.1007/s11042-013-1501-1

- Yi, J., Sheng, B., Shen, R., Lin, W., & Wu, E. (2015, August). Real Time Learning Evaluation Based on Gaze Tracking. *14th International Conference on Computer-Aided Design and Computer Graphics (CAD/Graphics)*, pp. 157–164.
- Yoshimura, K., Kise, K., & Kunze, K. (2015, August). The eye as the window of the language ability: Estimation of English skills by analyzing eye movement while reading documents. *13th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 251–255.
- Yu, T., & Jo, I.-H. (2014). Educational Technology Approach Toward Learning Analytics: Relationship Between Student Online Behavior and Learning Performance in Higher Education. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, pp. 269–270.
- Zacharis, N. Z. (2015). A multivariate approach to predicting student outcomes in web-enabled blended learning courses. *The Internet and Higher Education*, 27, 44–53. doi:10.1016/j.iheduc.2015.05.002
- Zeidner, M. (2007). Test Anxiety in Educational Contexts (pp. 165–184). doi:10.1016/B978-012372545-5/50011-3
- Zhang, X., Liu, X., Yuan, S.-M., & Lin, S.-F. (2017). Eye Tracking Based Control System for Natural Human-Computer Interaction. *Computational Intelligence and Neuroscience*. Research article. doi:10.1155/2017/5739301
- Zhuang, N., Zeng, Y., Tong, L., Zhang, C., Zhang, H., & Yan, B. (2017). Emotion Recognition from EEG Signals Using Multidimensional Information in EMD Domain. *BioMed Research International*, 2017(8317357). doi:10.1155/2017/8317357
- Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer*, 38(9), 25–32. doi:10.1109/MC.2005.297

Publications

Articles de revue

- Ben Khedher, A., Jraidi, I., & Frasson, C. (2019). Tracking Students' Mental Engagement using EEG Signals during an Interaction with a Virtual Learning Environment. *Journal of Intelligent Learning Systems and Applications*, 11(1), 1-14.
- Ben Khedher, A., Jraidi, I., & Frasson, C. (2018). Static and Dynamic Eye Movement Metrics for Students' Performance Assessment. *Smart Learning Environments*, 5(1).
- Ben Khedher, A., Jraidi, I., & Frasson, C. (2018). Local Sequence Alignment for Scan Path Similarity Assessment. *International Journal of Information and Education Technology*, 8(7), 482-490.

Articles de conférence

- Ben Khedher, A., Jraidi, I., Chaouachi, M., & Frasson, C. (2019). Predicting Learners' Performance using EEG and Eye Tracking Features. The 15th International Conference on Intelligent Tutoring Systems, June 03-07, 2019, Jamaica [Accepted as a full paper].
- Ben Khedher, A., Jraidi, I., & Frasson, C. (2019). Predicting Learners' Performance using EEG and Eye Tracking Features. The 32nd International FLAIRS Conference, May 19-22, 2019, Florida, USA [Accepted as a full paper].
- Ben Khedher, A., Jraidi, I., & Frasson, C. (2018). Exploring Students' Eye Movements to Assess Learning Performance in a Serious Game. *Proceedings of EdmMedia: World Conference on Educational Media and Technology International Conference*, June 25-29, 2018, Amsterdam, Netherlands, pp 394-401.
- Ben Khedher, A., Jraidi, I., & Frasson, C. (2018). What Can Eye Movement Patterns Reveal about Learners' Performance? The 14th International Conference on Intelligent Tutoring Systems, June 11-15, 2018, Montreal, Canada [Poster].
- Ben Khedher, A., Jraidi, I., & Frasson, C. (2017). Tracking Students' Analytical Reasoning Using Visual Scan Paths. The 17th IEEE International Conference on

Advanced Learning Technologies, pp. 53-54, July 3-7, 20017, Timisoara, Romania [Poster].

- Ben Khedher, A., Jraidi, I., & Frasson, C. (2017). Assessing Learners' Reasoning Using Eye Tracking and a Sequence Alignment Method. The 13th International Conference on Intelligent Computing, pp. 47-57, August 7-10, 2017, Liverpool, United Kingdom.
- Ben Khedher, A., & Frasson, C. (2016). AMNESIA, a dynamic environment for progressive assessment of cognitive functions. The 13th International Conference on Intelligent Tutoring Systems, June 6-10, 2016, Zagreb, Croatia, pp 545 [Young Research Track].
- Ben Khedher, A., & Frasson, C. (2016). Exploring Students' Eye Movements to Assess Learning Performance in a Serious Game. The Educational Media and Technology International Conference, June 28-30, 2016, Vancouver, British Columbia, Canada, pp 1504-1511.
- Ben Khedher, A., & Frasson, C. (2016). Recommending Regulation Strategies. The 28th International FLAIRS Conference, May 18-20, 2015, Hollywood, Florida, USA [Short Paper].

Articles de workshops

- Ben Khedher, A., Jraidi, I., & Frasson, C. (2016). Learners' Performance Tracking using Eye Gaze Data. The 1st International Workshop on Supporting Dynamic Cognitive and Metacognitive Processes, ITS Conference, June 6-7, 2016, Zagreb, Croatia.