



HAL
open science

Cognitive ageing in everyday life

Pierre Schweitzer

► **To cite this version:**

Pierre Schweitzer. Cognitive ageing in everyday life. Cognitive Sciences. Université Paris sciences et lettres, 2017. English. NNT: 2017PSLEP061 . tel-02106981

HAL Id: tel-02106981

<https://theses.hal.science/tel-02106981>

Submitted on 23 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'École Pratique des Hautes Études

Cognitive ageing in everyday life

École doctorale de l'EPHE – ED 472

Spécialité - Neurosciences

Soutenue par :

Pierre SCHWEITZER

le 12 décembre 2017

Dirigée par :

Joel SWENDSEN

COMPOSITION DU JURY :

Pr. Stacey CALLAHAN
Université Toulouse 2
Présidente du jury
Rapporteur

Pr. Henri CHABROL
Université Toulouse 2
Rapporteur

Mme Bérengère GUILLERY-GIRARD
Université de Caen
Membre du jury

Pr. Mathilde HUSKY
Université de Bordeaux
Membre du jury

Pr. Joel SWENDSEN
Institut de Neurosciences Cognitives
et Intégratives d'Aquitaine
Directeur de thèse



École Pratique
des Hautes Études



Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisor Pr. Joel Swendsen, for his continuous support of my Ph.D study, for his patience and his trust despite my atypical record, and Pr. Mathilde Husky for the support she provided during the writing of this thesis.

I would also like to thank everyone else who accepted to be a member of my thesis committee and review this work: Pr. Stacey Callahan, Pr. Henri Chabrol and Mrs Bérengère Guillery-Girard.

Additionally, I would like to thank the research staff that worked on the field and made this study possible in practice, especially Caroline Buisson and Martine Bordessoules.

I am also grateful to Mickael Ludwig, Caroline Bordes and Céline Stinus who helped me with the heavy task of processing the data, and all the Neuroimaging and Human Cognition team for their indefectible kindness.

I also thank the ISPED team, in particular Pr. Hélène Amieva, for sharing their data, without which this project would not have been possible.

My thanks also go to all my friends for their support, especially Vanessa Locussol for always encouraging me when I lacked confidence.

Finally, I want to express my profound gratitude to my parents for providing me with the best conditions one could hope for, despite the many years this work required, and to add a special mention to my cats for their purrs when I was in doubt, even if they didn't understand why (but who knows?).

Table of contents

Résumé.....	6
Introduction	21
Chapter 1. Human cognition	25
1.1. Memory	25
1.1.1. Memory processes	28
1.2. Episodic memory	30
1.2.1. Definition	30
1.2.2. Assessment.....	32
1.3. Semantic memory.....	36
1.3.1. Definition	36
1.3.2. Assessment.....	37
1.4. Executive functions	39
1.4.1. Definition	39
1.4.2. Assessment.....	41
Chapter 2. Benefits of EMA for cognitive assessment.....	46
2.1. Barriers and limitations of traditional neuropsychological testing	46
2.2. Self-administered mobile cognitive tests	52
2.2.1. Data capture methods and cognitive functions tested.....	52
2.2.2. Psychometrics	53
2.2.3. Usefulness above and beyond traditional laboratory-based testing.....	54
2.2.4. Decisions for researchers	55
Chapter 3. EMA application development	57
3.1. Existing applications	57
3.1.1. Purdue Momentary Assessment Tool (PMAT)	57
3.1.2. MyExperience	58
3.1.3. Testing conclusions.....	59
3.2. Platform choice	59
3.3. Device choice	60
3.4. Hacking	61
3.5. Development	65
3.5.1. Participant experience.....	65

3.5.2. Research staff experience	66
3.5.3. Project structure	68
3.6. Summary	84
Chapter 4. Feasibility and validity of EMA cognitive testing in the elderly	85
4.1. Methodology	85
4.1.1. Sample.....	85
4.1.2. Procedure	86
4.1.3. AMI neuropsychological instruments.....	87
4.1.4. EMA cognitive assessments	89
4.1.5. Data analysis	95
4.2. Results	96
4.2.1. Acceptance and compliance.....	96
4.2.2. Effects associated with duration of ambulatory monitoring.....	98
4.2.3. Convergent validity of neuropsychological and mobile cognitive tests	99
Chapter 5. Discussion of feasibility and validity of EMA cognitive testing in the elderly	110
Chapter 6. Global discussion and conclusion	114
References	119

List of illustrations

Figure 1. Number of publications per year using EMA for cognitive and psychiatric disorders	22
Figure 2. The multi-store memory model	26
Figure 3. The working memory model.....	27
Figure 4. Memory processes	30
Figure 5. Cognitive decline in dementia and matching assessment methods	48
Figure 6. Nature of clinical and epidemiologic data demonstrating a correlation between physical activity and cognitive functioning	50
Figure 7. Assessment of lifestyles and cognitive functioning.....	51
Figure 8. The Samsung Galaxy Wifi 4.2" tablet	61
Figure 9. Idle screen showing different states.....	69
Figure 10. Alarm silencing interface.....	70
Figure 11. Busy button interface	72
Figure 12. Notification screen	73
Figure 13. Instructions screens	78
Figure 14. Single-choice question.....	79
Figure 15. Multiple-choice questions.....	80
Figure 16. Recording tasks.....	81
Figure 17. Time question	82
Figure 18. Text question.....	83
Figure 19. Screenshots of the verbal fluency task.....	89
Figure 20. Screenshots of the category generation test.....	91
Figure 21. Screenshots of the episodic memory test.....	93
Figure 22. Screenshots of the phonemic fluency test.....	94
Figure 23. Screenshots of the color naming task	95
Figure 24. Number of completed cognitive tests by age.....	97
Figure 25. Distribution of scores for mobile verbal fluency tests.....	100
Figure 26. Distribution of scores for semantic category generation	101
Figure 27. Distribution of task duration for semantic category generation	102
Figure 28. Distribution of episodic memory free recall	103
Figure 29. Distribution of episodic memory recognition scores.....	104
Figure 30. Distribution of episodic memory delayed free recall.....	105
Figure 31. Distribution of delayed recognition scores	106
Figure 32. Distribution of phonemic fluency scores	107
Figure 33. Distribution of color naming test scores	108
Figure 34. Distribution of color naming test score-per-time ratio	109

List of tables

Table 1. Nine properties of episodic memories.....	31
Table 2. Busy button parameters.....	72
Table 3. Notification parameters.....	73
Table 4. Question parameters.....	74
Table 5. Variable types.....	75
Table 6. Time types.....	76
Table 7. Event types.....	76
Table 8. Survey organization.....	86
Table 9. Pairs of words used in the mobile category generation test.....	90
Table 10. List of the words used in EMA episodic memory assessments.....	92
Table 11. Socio-demographic characteristics of the final sample of participants.....	96
Table 12. Neuropsychological and mobile cognitive performance of the final sample.....	98
Table 13. Fatigue and practice effects observed for mobile cognitive tests.....	99
Table 14. Association of Isaacs Set Test scores with mobile verbal fluency scores.....	100
Table 15. Association of Isaacs Set Test scores with semantic category generation duration.....	102
Table 16. Association of Grober and Buschke scores with episodic memory free recall.....	103
Table 17. Association of Grober and Buschke scores with episodic memory recognition....	104
Table 18. Association of Grober and Buschke scores with episodic memory delayed free recall.....	105
Table 19. Association of Grober and Buschke scores with episodic memory delayed recognition.....	106
Table 20. Association of Wechsler coding test scores with phonemic fluency scores.....	107
Table 21. Association of Wechsler coding test scores with color naming test scores.....	109
Table 22. Association of Wechsler coding test scores with color naming test time.....	109

Résumé

1. Introduction

Il est estimé que 6% à 8% de la population mondiale de plus de 60 ans développe une démence, et la maladie d'Alzheimer (MA) représente environ entre 60 à 70% de ces cas. L'apparition insidieuse de la MA implique que les patients éprouvent des difficultés cognitives subtiles qui augmentent en gravité sur plusieurs années, les premiers signes de déclin des fonctions cognitives survenant plus d'une décennie avant que le diagnostic clinique soit possible. D'un point de vue méthodologique, la détection des premiers stades du déclin cognitif conduisant à la MA est rendue compliquée non seulement par la difficulté à distinguer le déclin pathologique de celui observé dans le vieillissement normal, mais aussi par la marge d'erreur relativement importante associée aux tests neuropsychologiques standards. En particulier, les scores obtenus par des tests neuropsychologiques ponctuels peuvent varier en fonction de diverses influences spécifiques aux contextes, aux patients ou aux cliniciens.

Les méthodes reposant sur des technologies mobiles telles que l'« Ecological Momentary Assessment » (EMA, également appelée la méthode d'échantillonnage des expériences) sont prometteuses pour surmonter ces obstacles en fournissant des évaluations répétées plusieurs fois par jour et pendant plusieurs jours consécutifs, dans des contextes écologiquement valides. L'EMA a été largement validée dans diverses populations normales ou cliniques et elle a été utilisée avec succès chez les personnes âgées. L'utilisation de l'EMA a également évolué en fonction de la technologie disponible et des nouvelles questions scientifiques. En particulier, il est de plus en plus clair que l'expansion de l'EMA pour fournir des évaluations répétées et *in vivo* des fonctions cognitives peut réduire la marge d'erreur

typiquement associée aux tests neuropsychologiques et ainsi caractériser plus précisément le déclin cognitif précoce chez les personnes âgées. Cependant, aucune étude publiée à ce jour n'a examiné la faisabilité de base ou encore la validité des tests cognitifs mobiles par EMA dans cette population.

Cette thèse de doctorat aborde ces questions à travers deux objectifs distincts. Le premier est d'origine purement technique, cet objectif étant de développer une nouvelle application EMA capable d'administrer des tests cognitifs mobiles d'une gamme de fonctions cognitives. Le deuxième objectif de cette thèse est d'ordre clinique, et vise à fournir le premier examen détaillé de la faisabilité et de la validité des tests cognitifs mobiles spécifiquement chez les personnes âgées.

2. La cognition humaine

Parmi les différentes fonctions cognitives, la mémoire et les capacités exécutives sont particulièrement importantes dans le contexte du vieillissement cognitif, de nombreuses études ayant montré que leur altération pouvait potentiellement prédire la démence plusieurs décennies avant le diagnostic. La mémoire a été définie comme le processus de conservation de l'information au fil du temps, mais la mémoire humaine peut difficilement être décrite comme une fonction cérébrale unique plutôt que comme un système complexe de composants en interaction, chacun ayant des rôles et des propriétés spécifiques. Bien que notre compréhension de la façon dont la mémoire est organisée soit encore incomplète, beaucoup de ses composantes ont été intensivement étudiés au cours des dernières décennies. La mémoire épisodique est celle qui permet de revivre des expériences passées, grâce à la conscience auto-noétique (la capacité de se placer mentalement dans le passé et de se souvenir de ses propres expériences antérieures). Comme ce système de mémoire est connu pour être parmi les premiers à présenter un déclin lors de la phase prodromique de la MA (plus de dix

ans avant le diagnostic clinique), il est donc capital de disposer de moyens fiables pour son évaluation. La mémoire sémantique fait référence à la connaissance générale du monde que nous avons accumulée tout au long de notre vie. C'est un sous-type de mémoire déclarative qui, contrairement à la mémoire épisodique, est indépendant de l'expérience personnelle et du contexte dans lequel la connaissance a été acquise. Le déclin de la mémoire sémantique peut également être l'un des premiers symptômes de la MA. Les fonctions exécutives font référence à une famille de processus cognitifs qui sont mobilisés dans des situations nécessitant un effort mental et pour lesquelles des réponses automatiques, intuitives ou instinctives seraient inappropriées, insuffisantes ou impossibles. De telles situations peuvent inclure par exemple traiter de la nouveauté, anticiper un danger, prendre des décisions ou surmonter des réponses habituelles fortes (situations qui nécessitent la capacité d'organiser mentalement, initier ou maintenir d'autres processus cognitifs).

Des tests neuropsychologiques validés accompagnés des données normatives établies sont essentiels pour évaluer les capacités cognitives notées précédemment et pour détecter d'éventuelles zones de difficulté ou de déclin. Une limitation majeure des tests neuropsychologiques traditionnels, cependant, est qu'ils sont le plus souvent administrés ponctuellement et évalués par un clinicien un jour donné et à un moment précis. Le problème potentiel avec de telles évaluations ponctuelles est que les scores obtenus peuvent ne pas être représentatifs du fonctionnement habituel ni de la manière dont fonctionnent les personnes « la plupart du temps ». Les scores de performance cognitive d'un individu fluctuent également en fonction des rythmes quotidiens, du stress et de nombreuses autres influences quotidiennes. Pour cette raison, les instruments cliniques traditionnels utilisés pour caractériser le fonctionnement cognitif ou pour détecter les déficits sont caractérisés par une « marge d'erreur » pouvant être très petite ou très grande. En raison de cette variance, un dysfonctionnement subtil mais cliniquement pertinent peut être indétectable au moment de

l'examen. L'impact de cette marge d'erreur pour la recherche est que la détection de certains signes de déclin ou de déficits cognitifs, en particulier ceux associés à l'apparition précoce de troubles, nécessiterait souvent le recrutement de très grands échantillons d'individus. Les études à grande échelle sont très coûteuses et la réalité de la littérature scientifique est qu'elles sont peu nombreuses comparées aux enquêtes de taille plus modérée, dont la puissance statistique est limitée pour détecter les déficits cognitifs mineurs.

Une solution alternative à l'augmentation du nombre de participants inscrits dans une étude consiste à augmenter le nombre d'observations acquises auprès de chaque participant. Cette approche n'est souvent pas réalisable dans les protocoles neuropsychologiques traditionnels en raison des contraintes de temps et des dépenses associées aux tests administrés par les cliniciens. Cependant, les technologies mobiles peuvent facilement fournir des évaluations brèves et répétées des fonctions cognitives, du comportement ou des émotions en temps réel et dans les différents contextes de la vie quotidienne. L'évaluation quotidienne répétée de ces variables peut réduire la marge d'erreur associée aux outils cliniques traditionnels et fournir des données à haute résolution concernant les fonctions cognitives dans les contextes de la vie quotidienne. Compte tenu de cette possibilité, nous avons développé le programme Samplex pour la recherche EMA dans le cadre du premier objectif de thèse. Ce programme a été développé spécifiquement pour le système Android OS car celui-ci répondait à tous les critères essentiels concernant la durabilité, la disponibilité, le prix et la portabilité. Samplex offre une flexibilité considérable dans la conception des études EMA, y compris les types de questions qui peuvent être posées aux participants, la manière dont les participants peuvent répondre et, surtout, sa capacité à administrer des tests mobiles de fonctions cognitives spécifiques.

Comme mentionné précédemment, le second objectif majeur de cette thèse était d'appliquer les avancées techniques (Samplex) à l'étude du déclin cognitif chez les personnes

âgées. Ce deuxième objectif a nécessité une étude clinique approfondie de la faisabilité et de la validité de ces nouveaux outils EMA.

3. Méthode

3.1. Echantillon

L'échantillon fait partie de la cohorte AMI qui est basée sur la population de 1002 travailleurs agricoles âgés dans les zones rurales. La cohorte AMI a sélectionné au hasard parmi les participants âgés de 65 ans et plus faisant partie du registre de la sécurité sociale des agriculteurs. Les procédures d'étude ont été approuvées par le comité éthique régional et tous les participants ont fourni un consentement éclairé par écrit. Après leur inclusion initiale dans la cohorte, les participants ont complété une batterie de tests neuropsychologiques et, pour un sous-échantillon, les participants ont accepté un examen de neuro-imagerie administré environ tous les deux ans. Une première évaluation de suivi du sous-groupe d'imagerie a eu lieu entre 2009 et 2012. Lors du second examen de suivi (entre 2012 et 2014), un sous-échantillon aléatoire de ces participants a également été invité à participer à une étude EMA d'une semaine avec une tablette Android (Samsung Galaxy Wifi 4''2) programmée pour administrer les évaluations cognitives via Samplex. Les critères d'inclusion pour la présente étude étaient un niveau de lecture de CM2, aucune déficience visuelle ou motrice, aucune déficience cognitive significative ou diagnostic de démence, et participation au projet de neuro-imagerie. Un total de 172 participants ont été contactés pour participer à la présente étude EMA (44% de femmes) avec un âge moyen de 72 ans (écart-type = 4,6).

3.2. Procédure

L'entretien EMA a interrogé les participants au sujet de leur environnement physique, de leurs interactions sociales et de leurs comportements spécifiques. Un bref test cognitif a également été administré à la fin des entretiens électroniques pour évaluer la mémoire sémantique et la fluence verbale, la mémoire épisodique ou le fonctionnement exécutif. Ces tests cognitifs mobiles ont été développés spécifiquement pour l'auto-administration par le participant. Afin d'éviter les biais liés au moment de la journée ou à la répétition des tests, chaque test a été administré cinq fois par semaine dans un ordre contrebalancé pendant la journée, et un contenu unique a été développé pour les cinq versions de chaque test. Toutes les fonctions de l'appareil ont été désactivées de sorte qu'il ne pouvait être utilisé qu'aux fins de l'étude. Tous les participants ont été formés à l'utilisation de l'appareil et deux évaluations tests ont été administrées, l'une guidée par le personnel de recherche et l'autre par réalisée par l'individu de manière autonome. Les participants ayant des difficultés à comprendre ou à compléter les évaluations ont reçu un entraînement supplémentaire. Les entretiens électroniques de l'EMA ont été administrés à des intervalles fixes, randomisés entre individus. Les premiers et derniers entretiens de la journée ont été ajustés pour correspondre aux horaires de sommeil typiques des participants. Les évaluations cognitives en EMA utilisaient la fonction d'enregistrement vocal du smartphone ou la sélection manuelle sur l'écran du smartphone. Les réponses verbales enregistrées ont été codées par des membres formés de l'équipe de recherche. La fiabilité inter-évaluateur a été calculée sur la base des corrélations de Pearson pour un sous-échantillon de 30 participants et variait de 0,90 à 0,98 pour tous les tests.

3.3. Tests neuropsychologiques de la cohorte AMI

3.3.1. Mémoire sémantique

Le « Isaacs Set Test » a été utilisé pour évaluer la mémoire sémantique, par la présentation d'une catégorie sémantique, telle que « Couleurs » ou « Villes ». Les participants ont été invités à lister autant de mots appartenant à la catégorie donnée que possible dans une période de 60 secondes. Le nombre de bonnes réponses a été utilisé pour noter le test.

3.3.2. Mémoire épisodique

Les sujets ont reçu le test de rappel sélectif libre et différé. Ce test commence par une phase d'étude où 4 feuilles de papier affichant chacune 4 mots ont été successivement présentées. Les 16 mots appartiennent à 16 catégories sémantiques distinctes. Pour chaque liste, il est demandé au participant de faire correspondre la catégorie avec l'élément correspondant (par exemple « Parmi ces mots, pouvez-vous me dire quel est le poisson? »). Les participants sont ensuite testés pour un rappel immédiat: chaque catégorie est présentée et on leur demande de donner l'item correspondant; quand ils ne parviennent pas à se souvenir, ils reçoivent la bonne réponse. Une fois la phase d'étude terminée, la phase de rappel commence par un rappel libre suivi d'un rappel indicé pour les éléments qui n'ont pas été mémorisés. Trois essais libres et indicés sont successivement complétés, séparés par une tâche d'interférence (compter à rebours). Après un délai de 20 minutes, un essai de rappel libre et indicé est administré. La somme des réponses correctes pour les trois tests de rappel et le nombre de bonnes réponses pour la période de rappel différé ont tous deux été utilisés pour noter le test.

3.3.3. Fonctionnement exécutif

Le test des symboles de la WAIS-IV a été utilisé pour évaluer le fonctionnement de l'exécutif. Dans ce test, on présente aux participants une liste de 9 symboles appariés avec des chiffres de 1 à 9. On leur présente ensuite un tableau de chiffres qu'ils doivent remplir avec le symbole correspondant dans une période de 90 secondes. Le nombre de symboles corrects a été utilisé pour noter le test.

3.4. Tests cognitifs mobiles

3.4.1. Mémoire sémantique: fluence verbale

Un test mobile de fluence verbale sémantique a été développé. Ce test présente aux participants une catégorie sémantique (à savoir « Instruments de musique », « Meubles », « Outils », « Légumes » et « Véhicules »), puis ils sont invités à dire autant de mots que possible appartenant à cette catégorie pendant 60 secondes. Les réponses verbales ont été enregistrées sur le smartphone et codées plus tard par le personnel de recherche. Le nombre total de mots corrects a été utilisé comme score de performance principal. De plus, le nombre d'erreurs, de répétitions et de réponses pour les catégories associées (les réponses qui sont elles-mêmes des catégories, par exemple « Plantes » lorsqu'on leur a demandé « Légumes ») a été évalué.

3.4.2. Mémoire sémantique: génération de catégories

Un test mobile a été développé pour évaluer la capacité à reconnaître la catégorie commune à plusieurs mots. Une paire de mots appartenant à la même catégorie sémantique a été présentée. Les participants ont été invités à nommer leur catégorie en 30 secondes. Les réponses ont été enregistrées verbalement et codées plus tard par le personnel de recherche. Pour ce test, les participants ont été autorisés à arrêter l'enregistrement une fois qu'ils ont

répondu. Le score principal pour ce test était 0 pour les erreurs, 1 pour une catégorie associée et 2 pour une réponse correcte. Le temps requis pour répondre a également été codé comme un sous-score.

3.4.3. Mémoire épisodique: rappel et reconnaissance immédiate et différée

Un test mobile d'apprentissage de liste de mots a été élaboré afin d'évaluer le rappel libre et la reconnaissance, de manière immédiate et différée. Des listes de mots représentant des objets matériels ont été sélectionnées à partir de la base de données BRULEX, avec une fréquence d'apparition comprise entre 200 et 1000 mots parmi un million de mots apparaissant dans la langue française. Une liste de 10 mots a été affichée pendant 30 secondes, après quoi les participants ont été immédiatement invités à rappeler librement tous les mots en une minute. À la suite de la tâche de rappel libre, le test a présenté aux participants une liste de 20 mots qui comprenaient les 10 mots de la liste précédemment affichée ainsi que 10 mots supplémentaires ne figurant pas dans la liste et qui avaient une fréquence similaire d'apparition dans la langue française. Les participants ont ensuite été invités à identifier les mots qu'ils reconnaissaient comme appartenant à la liste originale en sélectionnant chaque mot directement sur l'appareil. Les nombres de mots corrects pour la tâche de rappel libre (enregistrée verbalement) et la tâche de reconnaissance (enregistrée par sélection sur l'écran du dispositif EMA) ont servi de scores primaires pour la performance de mémoire épisodique immédiate. À l'évaluation EMA suivante, environ 3 heures plus tard, les participants ont été invités à effectuer les mêmes tâches (rappel libre et reconnaissance) sans que la liste originale ne soit affichée. Les nombres de mots corrects pour les deux tâches ont servi de scores primaires pour les performances de mémoire épisodique différée. Le nombre d'erreurs et les répétitions lors du rappel libre ont servi de sous-scores.

3.4.4. Fonctions exécutives: fluence phonémique

Un test mobile de fluence phonémique a été utilisé pour évaluer les fonctions exécutives. On a présenté aux participants une lettre et on leur a demandé de nommer autant de mots commençant par cette lettre que possible dans un délai d'une minute et sans nommer les noms propres. Nous avons évité les lettres « difficiles », c'est-à-dire les lettres pour lesquelles trop peu de mots existent; les lettres utilisées pour ce test comprenaient 'P', 'F', 'T', 'M' et 'S'. Les réponses verbales ont été enregistrées par le dispositif EMA et le nombre total de mots corrects a servi de score de performance primaire. Le nombre d'erreurs et de répétitions a également été évalué.

3.4.5. Fonctions exécutives: dénomination de couleur

Une tâche mobile de dénomination des couleurs a été développée, inspirée de la phase d'interférence du Stroop. Dans cette tâche, 16 noms de couleurs (à savoir « Bleu », « Jaune », « Vert » et « Rouge ») sont affichés au participant. Cependant, chaque mot est présenté dans une autre couleur que celle que le mot désigne. Les participants sont invités à nommer les couleurs dans lesquelles les mots sont affichés (enregistrés verbalement) en 60 secondes (les participants sont autorisés à arrêter l'enregistrement lorsque la tâche est terminée). Le nombre de bonnes réponses a servi de score principal pour ce test. Les sous-scores ont été évalués pour les erreurs, les mots sautés et la durée totale de la tâche de dénomination.

3.5. Analyses des données

Les données EMA ont été analysées à l'aide de la modélisation hiérarchique linéaire et non-linéaire (HLM version 7). Des analyses de réactivité aux tests EMA répétés ont été effectuées afin d'identifier les changements de fréquence ou d'intensité des variables

cognitives en fonction de la durée de l'étude. La validité convergente a été examinée en utilisant les scores de tests neuropsychologiques obtenus lors de l'évaluation de référence comme prédicteurs de la performance cognitive mobile.

4. Résultats

4.1. Acceptabilité et faisabilité

Dans l'ensemble, 114 (66%) des 172 participants contactés ont accepté de participer à la partie EMA de l'étude. Les individus qui ont refusé la participation étaient plus âgés, $t(170) = 1,971$, $p < 0,05$, et avaient des scores MMSE inférieurs, $t(170) = -4,417$, $p < 0,001$, à ceux des personnes qui ont accepté. En comparaison avec la cohorte globale, les 114 participants à l'étude étaient plus souvent des femmes, $X^2(1) = 4,41$, $p < 0,05$, plus jeunes, $t(1000) = 7,96$, $p < 0,001$, et avaient des scores globaux MMSE plus élevés, $t(984) = -6,83$, $p < 0,001$. Soixante-quinze de ces participants ont répondu à au moins un tiers des évaluations programmées et ont été considérés comme remplissant le critère d'observance minimale des procédures. Comparativement à la cohorte globale, les participants satisfaisant au critère d'observance minimale étaient plus souvent des femmes, $X^2(1) = 7,20$, $p < 0,05$, plus jeunes, $t(961) = 7,37$, $p < 0,001$, et avaient des scores MMSE globaux plus élevés, $t(945) = -6,10$, $p < 0,001$. Le nombre de tests cognitifs mobiles complétés diminuait avec l'âge ($\gamma = -0,4963$, $t = -2,380$, $p < 0,05$).

Les questionnaires EMA complétés étaient nombreux, 82% des évaluations possibles étant réalisées par les participants dans le contexte de leur vie quotidienne (résultant en 2158 observations). L'examen de l'observance aux tests cognitifs mobiles enregistrés verbalement par l'appareil a indiqué qu'une petite partie des participants a peut-être reçu de l'aide d'un tiers pour compléter les évaluations, n'aurait pas compris les instructions ou aurait

utilisé des moyens pour contourner la difficulté (p.ex., écrire la liste des mots au lieu de la mémoriser). Lorsque ces évaluations ont été considérées comme nulles, il restait 1646 observations valides, ce qui représente environ 63% du nombre maximal potentiel d'observations.

Aucun effet de fatigue n'a été observé dans la mesure où le nombre d'observations manquantes a diminué en fonction de la durée de l'étude. Concernant les effets d'entraînement, les performances aux tests de mémoire sémantique, les scores de dénomination des couleurs, les scores de reconnaissance immédiate et les scores de rappel libre différé se sont améliorés en fonction de la durée de l'étude. Cependant, les performances de la reconnaissance différée de l'apprentissage de listes ont diminué avec le temps. Les scores obtenus à partir des autres tests cognitifs mobiles n'ont pas été significativement affectés par la durée de l'étude.

Les scores pour les tests cognitifs mobiles étaient souvent significativement corrélés avec les résultats des tests neuropsychologiques de fonctions cognitives similaires, y compris après ajustement statistique pour les effets d'entraînement associés au nombre de tests mobiles administrés. Cela était vrai pour le test de fluence sémantique ($\gamma = 0,104$, $t = 7,095$, $p < 0,001$), le test de fluence phonémique ($\gamma = 0,162$, $t = 4,576$, $p < 0,001$), le test de dénomination des couleurs ($\gamma = 0,071$, $t = 2,900$, $p = 0,006$), le rappel de liste de mots libres ($\gamma = 0,084$, $t = 4,411$, $p < 0,001$), la reconnaissance de liste de mots ($\gamma = 0,328$, $t = 3,997$, $p < 0,001$) et la reconnaissance différée ($\gamma = 0,315$, $t = 2,785$, $p = 0,007$).

5. Discussion

Des travaux très récents ont confirmé la valeur des tests cognitifs mobiles chez les personnes âgées comme moyen de réduire la marge d'erreur associée aux scores dérivés des évaluations neuropsychologiques traditionnelles, fournissant ainsi des outils plus sensibles au

fonctionnement cognitif. Malgré ces progrès, la faisabilité et la validité des tests cognitifs mobiles n'ont été examinées que par très peu d'études et aucune à ce jour n'a estimé les effets de fatigue ou de réactivité chez les personnes âgées. Dans une cohorte de personnes âgées, cette étude a examiné l'acceptabilité et la faisabilité des tests cognitifs mobiles administrés par EMA, ainsi que l'observance avec ses évaluations quotidiennes multiples, les biais potentiels associés aux tests répétés et sa validité convergente avec les évaluations neuropsychologiques traditionnelles.

En ce qui concerne la faisabilité fondamentale des tests cognitifs mobiles, les résultats indiquent que l'obstacle le plus important à leur utilisation est probablement l'acceptation initiale et le respect des procédures d'étude. Dans cet échantillon de personnes âgées non atteintes de démence, nous avons constaté que seule une majorité modérée de participants (66%) acceptait d'utiliser les évaluations mobiles. Ceux qui ont refusé la participation étaient plus âgés et avaient des scores plus faibles pour le fonctionnement cognitif général. Ces taux d'acceptation sont inférieurs à ceux obtenus dans des échantillons psychiatriques ou neurologiques lorsque l'EMA est utilisé sans tests cognitifs, mais peuvent aussi être inférieurs en raison de l'acceptation par ce sous-échantillon d'évaluations supplémentaires incluses dans un protocole de recherche plus large (y compris l'IRM). Il est également important de noter qu'environ un tiers (34%) des personnes qui ont accepté de participer n'ont pas répondu à suffisamment d'évaluations de l'EMA pour être considérées comme observant les procédures de l'étude. Il est possible que ces taux d'observance aient pu être améliorés si des procédures spécifiques avaient été mises en place pour encourager la participation (comme une compensation financière par exemple). Néanmoins, ces taux d'acceptation et de conformité indiquent que l'utilisation des évaluations cognitives mobiles chez les personnes âgées peut être limitée à une sous-population d'individus avec un certain

fonctionnement cognitif, bien que ces échantillons puissent inclure certains individus avec des déficits cognitifs très légers.

Alors que les tests cognitifs mobiles peuvent être réalisables chez la majorité des personnes âgées, l'évaluation de l'observance des participants à ces évaluations cognitives nécessite une attention du fait qu'aucun clinicien n'était présent pour administrer les tests ou interpréter le contexte dans lequel le test était effectué. En particulier, une minorité des réponses verbales enregistrées par les participants ont fourni une preuve potentielle d'assistance non autorisée (comme un conjoint qui rappelle au participant les mots d'une liste). Lorsque ces observations ont été exclues des analyses, l'échantillon a néanmoins répondu à 63% de tous les tests cognitifs programmés. Ce taux de réponse est inférieur au taux de réponse global (82%) pour les autres questions de l'EMA dans ce même échantillon et généralement inférieur aux évaluations EMA non cognitives dans d'autres populations adultes saines ou cliniques.

La validité des données recueillies a également été examinée en ce qui concerne les biais potentiels associés à la méthodologie des tests répétés et aux associations attendues entre les scores de performance cognitive mobile et les tests neuropsychologiques traditionnels. Aucun effet de fatigue n'a été trouvé selon lequel les individus pourraient réduire le nombre d'évaluation complétées en fonction du temps dans l'étude; au contraire, l'observance a augmenté au fil du temps en ce qui concerne le nombre de tests cognitifs mobiles effectués. Cependant, les résultats de certains tests variaient en fonction du nombre de tests administrés ce qui indiquait donc des effets d'entraînement potentiels. Comme le contenu de chaque test était unique à chaque administration, ces effets sont probablement attribuables à l'apprentissage du participant concernant la façon de répondre plutôt que relative à la mémorisation du contenu déjà présenté. Dans tous les cas, de tels effets peuvent être facilement maîtrisés par les chercheurs en contrôlant dans les analyses statistiques le nombre

de fois qu'un test particulier a été administré aux participants. Enfin, les corrélations significatives généralement observées avec les tests neuropsychologiques traditionnels de mémoire sémantique, de mémoire épisodique et des fonctions exécutives étayent la validité convergente des évaluations cognitives mobiles développées pour cette étude.

L'enthousiasme légitime pour les évaluations cognitives mobiles chez les personnes âgées doit cependant être modéré par la connaissance de ses limites et de ses apports spécifiques par rapport aux autres instruments et mesures du fonctionnement cognitif. La présente étude soutient donc l'idée de la faisabilité et la validité de ces nouveaux outils d'évaluation. En conclusion, bien que leur utilisation puisse ne pas être faisable chez tous les individus âgés, les tests cognitifs mobiles peuvent néanmoins fournir des données précieuses pour l'identification de signes précoces de déclin cognitif chez les personnes âgées.

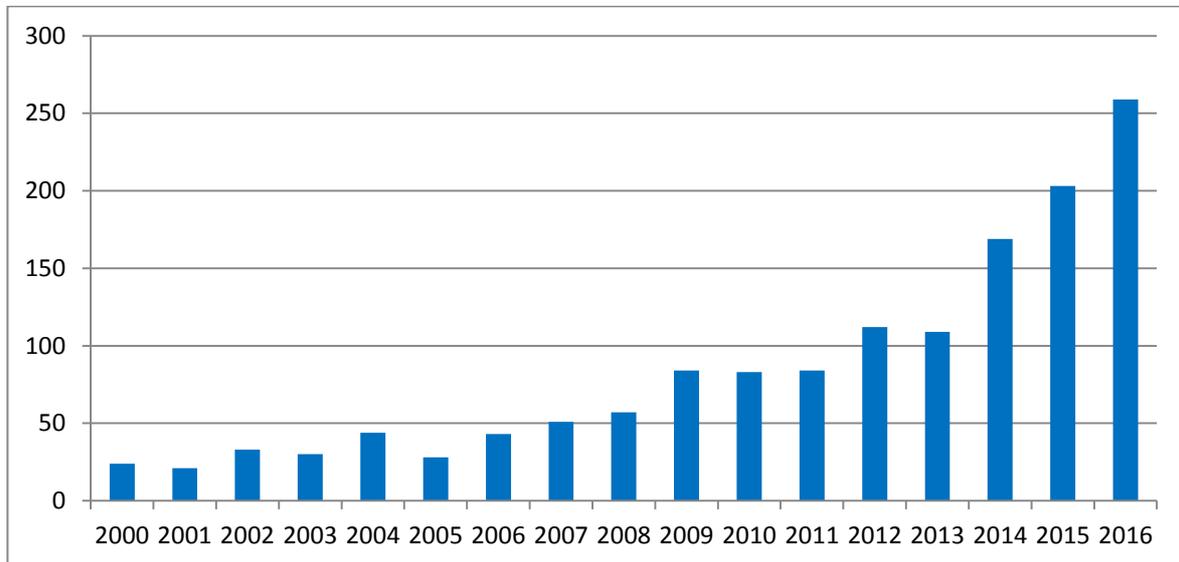
Introduction

It is estimated worldwide that 6% to 8% of the population over 60 years of age has developed dementia, and Alzheimer's Disease (AD) represents approximately 60 to 70% of these cases (World Health Organization, 2015). The insidious onset of AD implies that patients experience subtle cognitive difficulties that increase in severity over a number of years, with the first decline in any cognitive functions occurring over a decade before clinical diagnosis is possible (Amieva et al., 2008). From a methodological point of view, the detection of the earliest stages of cognitive decline leading to AD is not only complicated by the difficulty in separating pathological decline from cognitive decline observed in healthy aging, but also by the relatively large margin of error associated with standard neuropsychological testing (Hess et al., 2012; Metternich et al., 2009; Schmidt et al., 2007; Tollenaar et al., 2009). In particular, the scores derived through punctual neuropsychological testing may vary as a function of diverse context-specific, patient-specific or clinician-specific biases, and therefore such assessments often cannot differentiate variance attributable to measurement error from the subtle cognitive decline that occurs in patients who will eventually develop AD.

Methods relying on mobile technologies such as Ecological Momentary Assessment (EMA; also referred to as the Experience Sampling Method or ESM) hold promise for overcoming these barriers by providing repeated assessments multiple times a day and over several consecutive days in ecologically-valid contexts. EMA has been extensively validated in a range of healthy and clinical populations (Johnson et al., 2009a), and it has also been used successfully in elderly samples (Cain et al., 2009). Figure 1 presents the evolution of EMA investigations for cognitive and psychiatric conditions over time. While only a handful of such publications appeared before 2000, an increasingly rapid acceleration of investigations

using this method has been observed in the international literature (and particularly over the last 10 years).

Figure 1. Number of publications per year using EMA for cognitive and psychiatric disorders



Source: Pubmed

The use of EMA has also evolved as a function of both technology and new scientific questions. In particular, it is increasingly clear that the expansion of EMA to provide repeated and *in vivo* assessments of cognitive functions may reduce the margin of error typically associated with neuropsychological testing and thereby more reliably characterize early cognitive decline among the elderly (as well as daily-life cognitive functioning in other populations). In one recent study of non-demented elderly participants, intellectual activities of daily life were shown to improve momentary EMA measures of semantic memory, and these mobile memory scores were shown to be more sensitive than traditional neuropsychological measures in the identification of hippocampal atrophy (Allard et al., 2015). The repetition of such tests through EMA can also provide information concerning variability of cognitive functions (distinct from average functioning scores) which has been shown to be strongly predictive of cognitive decline (Holtzer et al., 2008). Despite

these encouraging findings, however, no published studies to date have examined the basic feasibility or validity of mobile cognitive testing through EMA among elderly individuals.

The present doctoral thesis addresses these issues by presenting and expanding on the following publications:

- Schweitzer P, Husky M, Allard M, Amieva H, Pérès K, Foubert-Samier A, Dartigues JF, Swendsen J. Feasibility and validity of mobile cognitive testing in the investigation of age-related cognitive decline. *International Journal of Methods in Psychiatric Research*. 2017; [Epub ahead of print]
- Swendsen J, Schweitzer P, Moore R. Mobile cognitive testing using experience sampling methodology. In J. Palmier-Claus (Ed). *Experience Sampling in Mental Health*. 2017 in press. Oxford: Routledge.

The first objective of this work is purely technical in origin, as its goal was to develop a new EMA platform capable of administering mobile cognitive tests covering a range of cognitive functions. While this is essential for responding to specific questions concerning age-related cognitive decline, it is important to note that such technical advances are also readily applicable to a range of mental and neurological disorders. Indeed, portions of this technical work have already been applied to numerous additional projects of our laboratory, including those addressing schizophrenia, mood and anxiety disorders, suicide risk, addiction, stroke, and HIV. The second objective of this thesis is clinically-oriented, as it provides the first detailed examination of the feasibility and validity of mobile cognitive testing specifically in the elderly.

Following this Introduction, different models of human cognition will first be described followed by a focus on several key functions of particular interest among the elderly (semantic memory, episodic memory and executive functioning). The second chapter will present the methodological challenges involved in the assessment of these functions

within the context of daily life, the benefits of EMA for this purpose, and the state of the literature concerning mobile cognitive testing. The subsequent chapters will then present the original scientific contributions of this thesis: Chapter 3 describes the fundamental EMA development work completed for this dissertation, while Chapter 4 presents a feasibility and validity study of EMA cognitive testing among the elderly. Following the presentation of these contributions, this document will conclude with a discussion of the findings of the feasibility and validity investigation, as well as by a general discussion of EMA cognitive testing.

Chapter 1. Human cognition

Cognition is the mental action or process of acquiring knowledge and understanding through thought, experience, and the senses (oxforddictionaries.com, 2016). It is achieved through a complex interaction of many neural networks involved in specific roles such as control, memory or reasoning. Even simple tasks usually involve subtle cooperation between several of these networks, making it hard to isolate and study them separately. While there is no clear categorization of elementary cognitive functions, the very active research in this field has led to identify some distinct functions and their underlying neural networks. Some of these cognitive functions are arousal, consciousness, decision making, language, motor coordination, perception, thought, executive functions and memory. Memory and executive functions are particularly important in the context of healthy vs. pathological cognitive aging, as numerous studies have shown that their alteration may predict dementia decades before diagnosis (Amieva et al., 2008; Rajan et al., 2015). The following section presents a brief description of each of these functions.

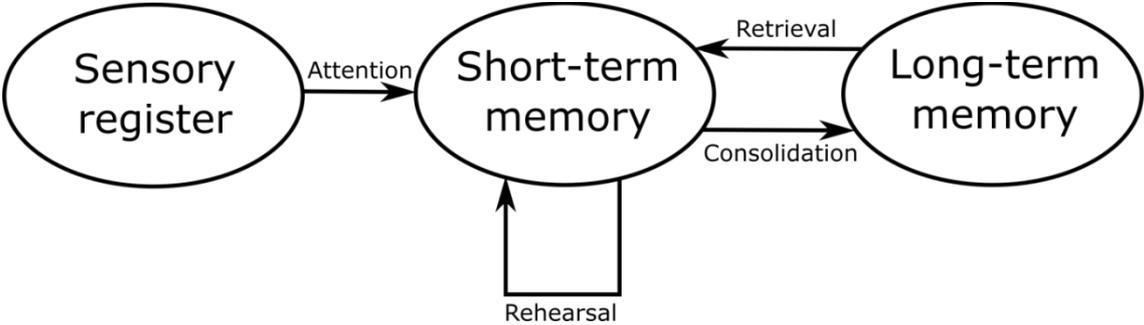
1.1. Memory

Memory has been defined as the process of maintaining information over time (Matlin, 2005). However, human memory can hardly be described as a single brain function, but rather as a complex system of interoperating components, each having specific roles and properties. While our understanding of how memory is organized is still incomplete, many of its components have been extensively investigated over the past decades. The discovery of new memory functions often follows a dichotomous pattern: a memory function that was once thought to be monolithic becomes conceptualized by researchers as being composed of

complementary sub-functions. Many of the most significant steps that have improved our understanding of how memory is organized have followed this scheme.

In the 60s, there was considerable debate about whether memory was a single, unified brain function, or rather composed of several components. A multi-store memory model was proposed where memory was assumed to be composed of a sensory register, short-term memory and long-term memory (Atkinson and Shiffrin, 1968). An illustration of this model is presented in Figure 2. In this model, the sensory register is described as an unconscious mechanism holding perceptual information for less than a second. Through attention, this information can be carried to short-term memory, a temporary storage holding 7 +/- 2 elements or “chunks” (Miller, 1956) for approximately 20 seconds (Peterson and Peterson, 1959). Rehearsal (mentally repeating information to strengthen its holding) is then assumed to facilitate the creation of durable traces in long-term memory, a virtually unlimited storage both in terms of capacity and duration, from which information can be later retrieved for processing. While this model is considered obsolete nowadays, it led to a tremendous amount of new research at the time, and many findings on memory organization derive from it.

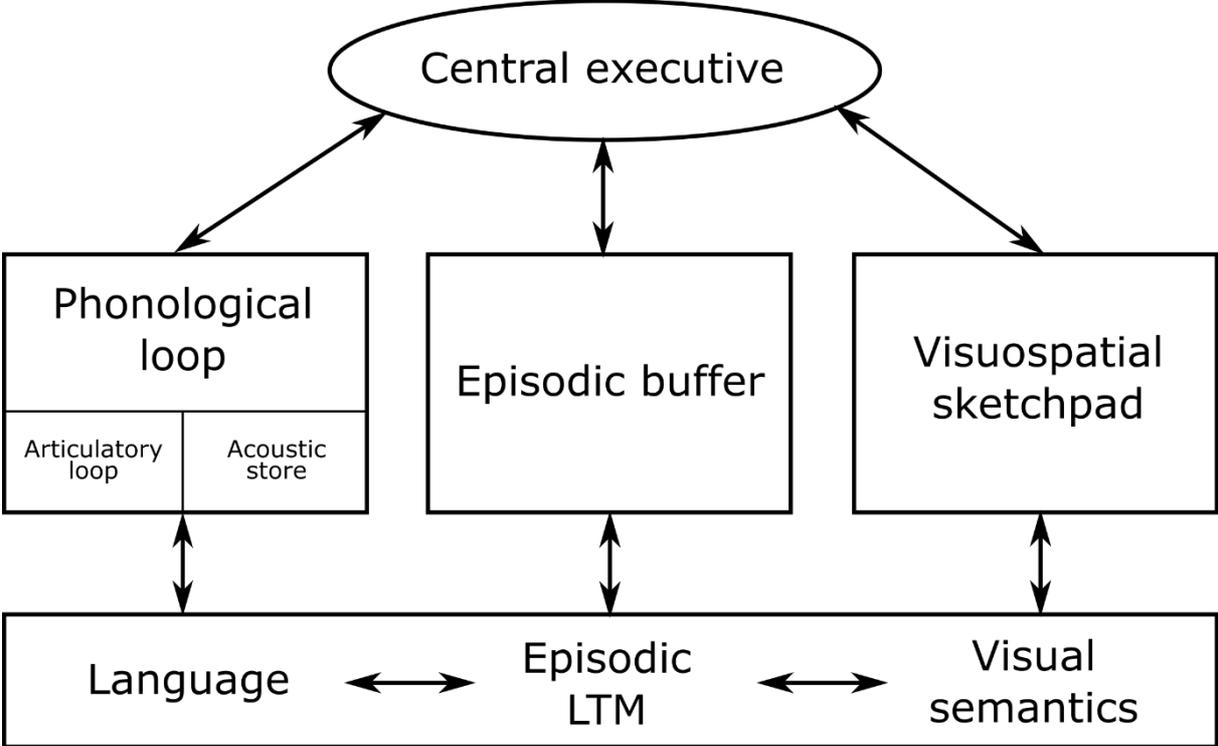
Figure 2. The multi-store memory model



The identification of each of these memory stores led to examine them more closely. The study of short-term memory revealed the existence of *working memory* (Figure 3), an alliance of separate but interacting temporary storage systems, possibly coordinated by a

single central executive component that enables the information not only to be maintained like with short-term memory, but also to be manipulated (Baddeley, 1983; Baddeley and Hitch, 1974). However the separation between short-term and working memory is not always clear: while they are conceptually different, there is evidence of a large overlap between them, though there is currently no consensus as to their relationship (Aben et al., 2012). The core component of working memory is assumed to be a *central executive* that supervises and coordinates three “slave systems”: the *phonological loop* comprised of an *acoustic store* retaining the sound of language and an *articulatory loop* used for speech perception and production, the *visuospatial sketchpad* dealing with visual and spatial information and the *episodic buffer* assumed to temporarily hold episodic representations of information and act as a bridge capable of retrieving from and feeding into long-term memory (Baddeley, 2000). Working memory is itself categorized as a component of *executive functions*.

Figure 3. The working memory model



Recently, there has been a growing interest regarding intermediate-term memory, a memory layer that was not present in the multi-store model (Rosenzweig et al., 1993). Intermediate-term memory is assumed to persist for up to three hours, and can be produced without the formation of long-term memory (Lukowiak et al., 2000). Intermediate-term memory declines completely before the onset of long-term memory (Sutton et al., 2001).

The long-term memory component of the Atkinson-Shiffrin model also went under close investigation and was found to be composed of distinct sub-functions (Tulving, 1972). One of these sub-functions, the *procedural memory*, was identified as an unconscious process (sometimes called *implicit memory*) holding the knowledge of actions to perform or of specific concepts (e.g. knowing how to ride a bicycle). The consciously controlled counterpart, the *declarative memory* (also known as *explicit memory*), was further split between *semantic memory* that reflects our knowledge of the world (e.g. describing properties of an object or concept), and *episodic memory* that is our capacity to recall past events in a spatial-temporal context (e.g. remembering a childhood birthday) (Tulving, 1972, 1983, 2002). Episodic memory and semantic memory are thought to be interdependent while being anatomically distinct (Greenberg and Verfaellie, 2010). In this thesis, particular attention is given to these two types of memory as they are known to decline early during the preclinical phase of AD, decades before a diagnosis can be established (Amieva et al., 2008; Rajan et al., 2015).

1.1.1. Memory processes

Remembering information involves several processes: *encoding*, *storage*, *retrieval* and *consolidation* of the learned material (Figure 4). Each of these processes is vulnerable to specific errors and may result in false memories (Straube, 2012).

Encoding is the initial process of acquiring information and transferring it into long-term memory (Goldstein et al., 2011). The perceived material is converted into a form that can be stored in memory brain structures (McLeod, 2007). Examples of types of encoded information are visual, acoustic or semantic in nature. Efficient encoding is thought to facilitate later retrieval, and it has been reported that when the encoded material relates to preexisting information, a deeper level of processing takes place that results in more durable traces in long-term memory (Craik and Lockhart, 1972).

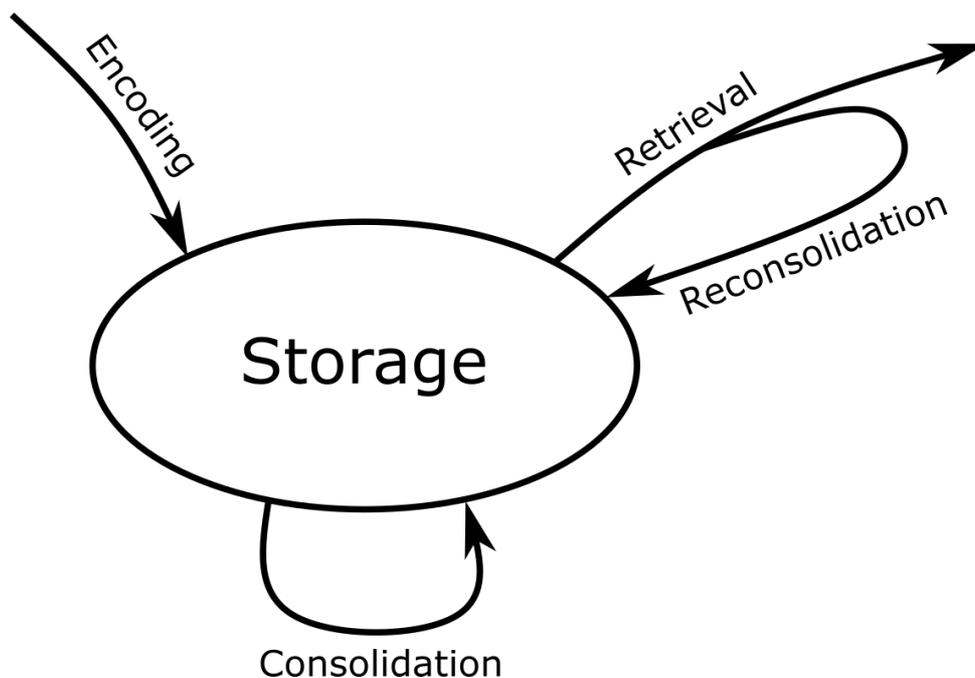
Storage refers to the capacity to hold encoded information. The memory systems that were defined in the previous section are supported by distinct storages with specific properties, including the nature of the information they hold, their capacity, the duration they retain information and the amount of control or effort they require. The overlaps in the durations these systems retain information support the idea that these systems work concurrently rather than sequentially (Lukowiak et al., 2000).

Retrieval or recall corresponds to the process of re-accessing information that was previously encoded and stored. The three main types of recall used in memory tasks are free recall (our capacity to remember items from a set in any order), cued recall (remembering items that cannot be recalled without a cue) and serial recall (remembering items in a certain order).

Consolidation is the process of stabilizing a memory trace after its initial acquisition. It comprises two phases: *synaptic consolidation*, an early phase taking place a few hours after encoding in which the hippocampus is believed to play an administrator role and *system consolidation* where memories progressively become independent of the hippocampus over a period of weeks to years. However, these consolidated memories can be destabilized at the time of their retrieval, allowing their content to be modified or erased through *reconsolidation* (Agren, 2014), a recently discovered process thought to be present in

many species (Reichelt and Lee, 2013). The mechanisms underlying reconsolidation are still under debate (Gisquet-Verrier et al., 2015).

Figure 4. Memory processes



1.2. Episodic memory

1.2.1. Definition

The existence of episodic memory was first assumed in the late 50s (Nielson, 1958; cited in Tulving, 2002), but went mostly unnoticed until the early 70s which saw the first attempts at its rigorous definition and distinction from semantic memory (Tulving, 1972). This controversial concept (at the time) has then been thoroughly studied and refined, and the distinction between episodic and semantic memory as the two components of declarative memory is now widely accepted.

Episodic memory makes it possible to re-experience life, through auto-noetic awareness, defined as the ability to mentally place one’s self in the past and to remember one's own previous experiences (Tulving, 2002). Episodic memories can involve words as well as shapes, and are characterized by nine properties shown in Table 1 (Conway, 2009).

Table 1. Nine properties of episodic memories

1. Contain summary records of sensory-perceptual-conceptual-affective processing
2. Retain patterns of activation/inhibition over long periods
3. Often represented in the form of (visual) images
4. They always have a perspective (field or observer)
5. Represent short time slices of experience
6. They are represented on a temporal dimension roughly in order of occurrence
7. They are subject to rapid forgetting
8. They make autobiographical remembering specific
9. They are recollectively experienced when accessed

Episodic memory is an early-deteriorating, past-oriented memory system, and it is more vulnerable than other memory systems to neuronal dysfunction. It was considered at first as a late-developing system, thought to be operational in children of at least 4 years of age (Tulving, 2002), but recent experiments tend to demonstrate that it might be already functional in 3-month old infants (Mullally and Maguire, 2014). Episodic memory requires but goes beyond the semantic memory system, and is subserved by a widely-distributed network of cortical and subcortical brain regions that overlap with, but also extend beyond, the networks subserving other memory systems. The essence of episodic memory lies in the conjunction of three concepts: the self, auto-noetic awareness, and subjectively-sensed time. It is sometimes believed to be unique to the human species, and to be the possible reason for the

appearance of culture and civilization (Tulving, 2002), though recent studies show that fundamental features of episodic memory capacity are present in mammals and birds and that the major brain regions responsible for episodic memory in humans have anatomical and functional homologs in other species (Allen and Fortin, 2013).

1.2.2. Assessment

As this memory system is known to be among the first to show decline in prodromal AD more than a decade before clinical diagnosis (Amieva et al., 2008; Traykov et al., 2007), it is crucial to have reliable means for its assessment. However, there is no consensus regarding the definition of what episodic memory is, and even slightly different definitions lead to large differences in assessments. Furthermore, episodic memory tests also differ in their specificity, some of them assessing multiple cognitive functions at once whereas others are more specifically targeted in the episodic memory system, leading to scores that have quadratic relationships or even no correlation at all (Cheke and Clayton, 2013).

1.2.2.1. Verbal paired associates

The verbal paired associates test consists of remembering words by using other stimuli words which are semantically close. After orally presenting all words, participants are given stimuli words and asked to remember which word was associated with it. This task allows for the investigation of associative memory, but is also considered not to be sensitive enough because it is easier than the other tests assessing episodic memory. It is used in tests such as the Wechsler memory scale “WMS-III” (Wechsler, 1997), the Mnemonic Efficiency Battery “BEM 144” (Signoret, 1991) and the Clinical memory scale MEM-III (Wechsler, 2001).

1.2.2.2. Logical memory

In the logical memory task, two short stories are read out loud by the investigator, and the participant is asked to recall these stories both in an immediate and in a delayed manner. In addition to episodic memory, the logical memory task also assesses oral understanding, short-term holding of information and verbal fluency, and as such is not suited in the case of multiple cognitive deficiencies. It also does not allow for testing of encoding and recollection. In order to compensate for these limitations, some variants include a delayed repeat of the story to optimize encoding, or a delayed recognition task to enable distinguishing storage processes from recollection. This task has been used by tests such as the WMS-III, MEM-III, Basso screening battery “BSB” (Basso et al., 1996) and the Story recall test (Caine et al., 1986).

1.2.2.3. Selective reminding test (SRT)

The selective reminding test consists of the presentation of a list of 10 to 12 unrelated words. After the list has been presented, the participant is instructed to recall as many words as possible from the list. Words that are not recalled are repeated by the examiner on the next trial. Trials continue until all words are remembered three times in a row, or until 12 iterations have been completed. This allows for the assessment of short-term storage over the first iterations and of long-term storage as the test progresses. Long-term recall can then be assessed using delayed recall occurring 20 to 30 minutes later. However, this paradigm does not allow for cued recall, and sensitivity to interference and learning strategies cannot be easily assessed.

Since its initial creation (Buschke and Fuld, 1974), the SRT has been widely used in the study of multiple sclerosis, and has been included in multiple test batteries:

- verbal selective reminding test “VSRT” (Hannay and Levin, 1985; Levin et al., 1982)
- the brief repeatable battery for neuropsychological examination "BRB-N" (Rao, 1990)
- neuropsychological screening battery for multiple sclerosis "NPSBMS" (Rao et al., 1991)
- short cognitive assessment battery for multiple sclerosis affected subjects "BCcogSEP" (Dujardin et al., 2004)
- open trial-SRT (DeLuca et al., 1994)

1.2.2.4. Rey auditory verbal learning test (RAVLT)

In the Rey auditory verbal learning test (Rey, 1964), a list of 15 words is orally presented, and the participant is asked to recall as many words from the list as possible. This process is repeated a total of five times. Then, an interference task consisting of another list of 15 words unrelated to those of the first list is presented once, and the participant is asked to list all the words he or she can remember. The participant is then asked to name as many words as possible from the first list. Thirty minutes later, the participant is once again asked to remember as many words as possible from the original list.

1.2.2.5. California verbal learning test (CVLT)

The California verbal learning test (Delis, 2000) follows the same scheme as the RAVLT mentioned above, except that the word lists are constituted of 16 words consisting of four categories of four semantically-related words. These semantic categories are “tools”, “fruits”, “clothing” and “spices and herbs”. The words are always said in the same fixed order

so that no consecutive words belong to the same semantic category. The interference list shares two categories with the first and two unshared categories, however no word is common to both lists. The semantic categories allow for the addition of a cued recall task after the free recall that follows the interference task and the delayed recall. The final part of the test is a recognition task where a list of 44 words is displayed and the participant is asked to state whether the words are distractors or whether they belong to the original list; distractors either share the same semantic categories of the first list, or whether they sound like them. This test has the advantage of offering means to assess learning strategy, e.g. if participants cluster words using semantic categories, as the order is specifically crafted to alternate these categories at presentation time.

1.2.2.6. Grober & Buschke (Free and cued recall, 16 items)

The Grober & Buschke test (Grober and Buschke, 1987) begins with a study phase where four sheets of paper each displaying four words are successively presented. The 16 words belong to 16 distinct semantic categories. For each list, the participant is asked to match the category with the corresponding item (e.g. “Among these words, can you tell me which one is a fish?”). Participants are then tested for immediate recall: each category is presented and they are asked to give the corresponding item; when they fail to remember, they are given the correct answer. Once the study phase has been completed, the recall phase begins using free recall followed by cued recall for the items that were not remembered. Three free and cued trials are successively completed, separated by an interference task (counting backwards). After a 20-minute delay, a delayed free and cued recall trial is administered. The sum of correct answers for the three recall tests and the number of correct answers for the delayed recall period are both used to score the test.

1.3. Semantic memory

1.3.1. Definition

Semantic memory refers to general world knowledge we have accumulated throughout our lives (McRae and Jones, 2013). It is a subtype of declarative memory that, unlike episodic memory, is independent of personal experience and the context in which knowledge was acquired. For example we know what a cat is (semantic memory), but we also might have memories about a particular cat (episodic memory). Semantic memories can derive from episodic memories when our experience gradually reinforces our understanding, effectively generalizing multiple episodes into abstracted concepts (Binder and Desai, 2011; human-memory.net, 2017), however semantic memory can also be formed without an episodic base. As with episodic memory, the formation of new memories of world knowledge is dependent on the hippocampus (Manns et al., 2003), however the long-term function of episodic and semantic memory is supported by distinct brain regions (Martin-Ordas et al., 2014).

Semantic memory impairment may be one of the earliest symptoms of Alzheimer's disease (AD), declining as early as 12 years before diagnosis (Amieva et al., 2008). It has been reported (Nebes, 1989) that while semantic production is impaired (giving words from a category, characteristics from objects), comprehension usually remains intact (recognizing which words belong to categories, if a characteristic is related to an object or concept). This fact argues in favor of a retrieval issue; however multiple studies have shown that in the context of AD, semantic memory impairment is mainly due to storage degradation rather than failure of access (Henry et al., 2004; Hodges et al., 1992; Mårdh et al., 2013; Salmon and Bondi, 2009). It has also been observed that object naming is relatively less impaired than the naming of famous faces or buildings (Ahmed et al., 2008).

1.3.2. Assessment

1.3.2.1. Isaacs set test

The Isaacs set test (Isaacs and Kennie, 1973) aims at evaluating the verbal fluency performance of participants, who are asked to give as many words as possible belonging to a given category within a limited amount of time. Initial versions of the test allowed very limited time and also put a limit to the number of correct words (e.g. a limit of 10 words or 15 seconds, whichever comes first). The risk of observing a ceiling effect increases as more time is given to participants, and usually a delay between 15 seconds to one minute is used. Commonly proposed categories include colors, animals, plants, and cities. Scoring is based on the number of correct words given. This test also indirectly involves short-term memory as participants need to keep track of the previous responses to avoid repeats. Sub-scores may include the number of such repeats, and also the number of intrusions (words that do not belong to a given category).

1.3.2.2. Similarities test

The Similarities test is a subtest of the Wechsler Adult Intelligence Scale (Wechsler, 1955) measuring a participant's ability to form new concepts. Participants are asked to determine the common attribute or category that is shared by two items presented to them. Traditionally, five pairs of items (e.g. "orange" and "banana") are presented in an order of increasing difficulty. Participants score two points for correct responses (e.g. "fruits") or one point if the response is correct but not fully accurate or insufficiently precise (e.g. "food").

1.3.2.3. Graded naming test

The Graded naming test (McKenna and Warrington, 1983) is used to assess object naming ability. Participants are shown images of objects or animals, and are asked to say

what the image represents. To avoid ceiling effects, the presented images are graded in difficulty, which also enables detecting word-finding difficulties even in participants that have an extensive naming vocabulary. This test is considered to have a cultural bias and is available only in English, and it has been re-standardized to match the evolution of overall performance (Warrington, 1997). Similar tests have been developed to assess word-finding of famous faces (Graded Faces Naming Test) and buildings (Graded Buildings Test).

1.3.2.4. Pyramids and palm trees test

The purpose of the Pyramids and palm trees test (Howard et al., 1992) is to assess the participant's ability to access detailed semantic representations from words and pictures. Subjects are shown a simple drawing or word (e.g. a pyramid) that they are asked to semantically match with one of two drawings or words, a target (e.g. palm trees) and a distracter (e.g. pine trees). By combining drawings, written and spoken words, six different versions of the test can be administered. Participants are usually trained with three or four examples before the test begins, then take the full test consisting of 52 triads with no time limit. Correct responses grant 1 point; responses for which the subject remains hesitant are scored 0.5 points. Producing more than five incorrect answers is considered clinically significant, although a rationale for this cut-point was not provided (Gudayol-Ferré et al., 2008). Because this test involves only concrete items, it does not permit to draw conclusions about comprehension of abstract words and concepts. However, unlike verbal tests, it is less sensitive to purely linguistic disorders. In very early dementia, the semantic deficit may not be sufficiently severe to cause difficulty with this task, and a more difficult version of this test, the Camel and Cactus test (Bozeat et al., 2000), was developed as a part of the Cambridge Semantic Memory Test Battery (Adlam et al., 2010), featuring a larger number of items and three foils for each trial.

1.4. Executive functions

1.4.1. Definition

Executive functions (EFs) refer to a group of cognitive processes that are used when dealing with situations requiring a mental *effort* and for which automatic, intuitive or instinctive responses would be inappropriate, insufficient or impossible. Such situations may include for example dealing with novelty, anticipating danger, making decisions or overcoming strong habitual responses (situations that require the ability to mentally organize, initiate or maintain *other* cognitive processes). While these properties characterize EFs, defining precisely what they are or describing their structure remains a topic of much debate, not only because of the various roles they are responsible for, but also because they often involve other non-executive processes, which render the task of isolating purely executive processes even more difficult. Despite this complexity, it is commonly admitted that EFs are not a single unitary construct, but are fractionated into distinct processes which share a common base. Indeed, it has been consistently observed that when confronted with two different EF tasks, some participants that were proficient at one task failed at the other. These observations raised the question of how many ‘core’ components constitute EFs, how do they relate to one another and how to characterize them. To this end, many studies used a latent variable approach in order to identify the factors that contribute to EF performance. This approach was partially successful as most studies were able to identify common factors, however these studies also found different numbers of factors (usually from two to five), or factors that have a slightly different meaning. It is argued that the difference in the participants’ age and the variations in the EF tasks that were used may be responsible for these differences. A brief description of these factors follows, namely ‘Shifting’, ‘Updating’, ‘Inhibition’, ‘Access’ and a mediating factor, ‘Speed’.

Shifting, also referred to as ‘cognitive flexibility’, concerns the ability to switch attention between different sub-tasks, operations or mental sets, that is, to disengage from an irrelevant task and subsequently engage into another relevant task set. It involves being flexible enough to adjust to changing demands or priorities, to admit being wrong, to take advantage of sudden or unexpected opportunities, or to change the way one thinks about something using a new perspective. This function is thought to build on updating and inhibition and to be fully operational only after early adolescence (Davidson et al., 2006).

Updating is closely related to working memory, as it involves evaluating incoming information and revising the existing content of working memory by deleting what is no longer relevant and incorporating more recent information (Adrover-Roig et al., 2012; Fisk and Sharp, 2004; Lehto et al., 2003; Miyake et al., 2000). It is critical for making sense of anything that unfolds over time, for that requires holding in mind what happened earlier and relating that to what comes later (Diamond, 2013).

Inhibition concerns the ability to deliberately withhold dominant, automatic or pre-potent responses when they are inappropriate (Miyake et al., 2000). Given that the term inhibition is commonly used to describe a wide variety of functions (Kok, 1999), it is important to note that this definition applies to the specific concept of executive inhibition, which should not be confused with inhibition of attention or motivational inhibition (Miyake et al., 2000; Nigg, 2000). However, inhibition has not yet been reliably identified as a core EF. Several studies found a high correlation between updating and inhibition (Adrover-Roig et al., 2012; Miyake et al., 2000), and it is suspected that resistance to interference and the maintenance of internal representations in working memory are closely related. One hypothesis is that older adults may exhibit ‘dedifferentiation’ of these cognitive functions (Adrover-Roig et al., 2012).

Access refers to the efficiency of accessing long-term memory, and is thought to play an important role in executive tasks that involve mechanisms such as lexical access or semantic fluency. It has also been reported that this factor is less sensitive to age than shifting, updating or inhibition (Fisk and Sharp, 2004).

Speed corresponds to the rate at which information can be processed. Several studies have shown that the relationship between speed and the other EF factors is strong and significant (Adrover-Roig et al., 2012; Fisk and Sharp, 2004). Information processing speed has been proposed as a modulator of executive capacity among the elderly (Salthouse, 1996; Salthouse et al., 1998) given that, once controlled for age differences in processing speed, many age-related deficits in EF performance are reduced to below statistical significance (Adrover-Roig et al., 2012; Salthouse and Babcock, 1991).

1.4.2. Assessment

Assessing EF performance raises a number of issues, as executive measures might not load on a single executive construct, and might overlap with one another (Miyake et al., 2000; Royall et al., 2002). Moreover, because EFs operate on other cognitive processes, it is virtually impossible to assess EF without incidentally assessing other nonexecutive processes (Suchy, 2009) and, as a result, a low score on a single executive test does not necessarily mean inefficient or impaired EF (Miyake et al., 2000). Finally, executive tasks tend to suffer from relatively low internal and/or test-retest reliability, because EF is inherently linked to novelty, and because participants may adopt different strategies on different occasions, or even within a single session (Miyake et al., 2000). Consequently, the task impurity, low reliability and construct validity issues make assessing EFs a difficult process.

1.4.2.1. Wisconsin Card Sorting Test (WCST)

In the Wisconsin Card Sorting Test (Berg, 1948), participants are presented with a stimulus card and four reference cards displaying three attributes: color (red, green, blue or yellow), number (from 1 to 4) or shape (circle, cross, star or square). Participants are required to match the stimulus card with one of the four reference cards, but are not told about which criteria must be used (color, number or shape). Participants have to figure out which attribute is the relevant one following the feedback that is given to them by the investigator (right or wrong). Moreover, participants are told that the matching criteria may suddenly change at any time, but they are not aware of when it will happen, the only clue being the feedback they get and their knowledge of the previous context. Once a stimulus card has been successfully matched, the next stimulus card is presented until there are no cards left in the deck. In the traditional version of this test, there are $4 \times 4 \times 4 = 64$ possible cards, four of which are used as reference cards, which leave a total of 60 cards for participants to match. Scores for this test consist of numbers, percentages or the percentile of categories achieved, as well as trials, errors and perseverative errors.

This task aims at assessing the participants' ability to shift mental sets when the matching attributes change, a skill which is thought to be reflected by the number of perseverative errors (that is, whether they keep matching cards using the previous attribute even after he or she has been told it was wrong). However, this task also involves the ability to remember which attribute is currently relevant or irrelevant, and maintaining this information in working memory falls under the definition of updating. Moreover, when the relevant attribute changes, participants are also required to inhibit the way they were matching cards prior to the change. Consequently, while the main target of this task is to assess the shifting ability of participants, it also indirectly assesses updating and inhibition.

1.4.2.2. Tower of Hanoi

The Tower of Hanoi was originally a mathematical puzzle popularized by the French mathematician Edouard Lucas in 1883. In this task, participants are presented with three rods, and a number of disks of varying diameter which can slide onto any rod. At the beginning, all disks form a stack in ascending order on the left-most rod. The objective is to move the entire stack of disks to the right-most rod following three rules: 1. only one disk may be moved at a time; 2. each move consists of taking the upper disk from one stack and placing it on top of another; 3. no disk may be placed on top of a smaller disk. There is usually no time limit on this task, and the score corresponds to the number of moves required to reach the target disposition. Variants of this task feature custom starting and target ending positions as in the Tower of London task (with or without the disk size constraint or additional constraints on legal moves such as in the Tower of Toronto task which uses colored disks). However, these variants seem to measure different abilities (Goel and Grafman, 1995) and have yielded low correlations with the original Hanoi test results (Humes et al., 1997).

Despite the fact that this task was originally believed to assess planning ability, it was later thought to be more dependent on inhibition given the goal/sub-goal conflict (Goel and Grafman, 1995), an observation that was confirmed in a subsequent latent variable analysis (Miyake et al., 2000). Working memory was also found to contribute to performance in complex variants of the task (using a higher number of disks), as the participant is required to keep in mind more sub-goal information (Goel et al., 2001; Morris et al., 1997).

1.4.2.3. Stroop

Since its invention by Stroop in 1935 (Stroop, 1935), many variants of the Stroop test have been developed (Strauss et al., 2006), usually consisting of three to four phases. Depending on the variant, the number of elements displayed varies between 24 and 100, and the number of colors used from three to six. The first steps usually include presenting

participants with a series of colored dots or squares and asking them to name the color of the shapes. Other common initial steps involve presenting participants with a series of words representing color names that are printed either in black ink, in the same color as designated by the word, in a different color, or a mix of congruent and different colors, and asking participants to read the words. Following these initial steps, the last step consists of presenting the participants with a series of words representing color names but printed in a different color than that designated by the words, and participants are asked to name the color in which these words are printed rather than reading the words. Depending on the variant of the test, there may be a time limitation for each trial (usually 45 seconds), and participants may be asked to correct errors as well as to do the task as quickly as possible. Raw scoring usually utilizes the number of items completed, but the main interest of this task is to obtain an interference score derived from the difference between normal color-naming and the color-naming of the mismatching colored words from the last step.

This measure was originally designed to assess inhibition, more specifically the ability to suppress a habitual response (word reading) in favor of a less familiar one (color naming). However, it is also believed that working memory contributes to performance in this task, as being subject to interference may reflect a failure to maintain the task goal of ignoring the word (Kane and Engle, 2003). Processing speed has also been consistently reported to play an important role in Stroop performance (Strauss et al., 2006).

1.4.2.4. Verbal fluency

The verbal fluency test (Thurstone, 1938) consists of asking participants to produce as many words as possible beginning with a specified letter for a period of time, usually one minute. Participants are also required to exclude proper names, and to avoid giving several words that share the same root (e.g. ‘shop’, ‘shopping’). Letters do not have the same difficulty (word frequencies are not evenly spread), and their difficulty also depends on the

language context. For instance, in the English versions of this test, the most commonly used letters are 'F', 'A' and 'S'. The score is the number of all valid words that are given during the allowed time, discounting any repetitions, proper names, wrong words and variations of the same word.

As with other measures of executive functions, verbal fluency performance requires the use of distinct cognitive abilities, as it involves self-monitoring to prevent errors and suppress previous responses while maintaining the objective. Generating cues to access new items is taxing on attention resources and during this operation other attention-demanding components (e.g. monitoring and suppression) may be negatively affected, resulting in increases in perseverative errors. A number of factor-analytic findings have supported the idea that working memory plays an important role in this task (Brocki and Bohlin, 2004; Elias et al., 1997; Anderson et al., 2001; cited in Strauss et al., 2006). Speed of processing has also been reported to contribute to verbal fluency performance (Boone et al., 1998; Salthouse et al., 2003).

1.4.2.5. Digit-Symbol Coding

In the digit-symbol coding test (Wechsler, 1939), participants are presented a list of nine symbols paired with digits from 1 to 9. Participants are given a matrix of digits which they are asked to complete with the corresponding symbol in a 90-second span. The number of correct symbols is used to score the test.

This test sometimes includes supplementary procedures that assess perceptual-motor speed and incidental learning as a way of determining what might eventually cause a low Digit-Symbol Coding score. While processing speed seems to be the main component associated to this task performance, the contribution of memory (albeit small) has also been observed (Joy et al., 2004; Kreiner and Ryan, 2001).

Chapter 2. Benefits of EMA for cognitive assessment

While the technological revolution assures that progress will continue to provide many innovative applications in the years to come, the newest of these developments involves *in vivo* assessments of human cognition. EMA tests are often administered through smartphones and are capable of assessing memory, attention, executive functions, and a wide range of other cognitive abilities that are necessary for everything from basic autonomous living to performance in more challenging circumstances. The capacity of applications to train and enhance these cognitive functions is also being increasingly explored, and there is a growing commercial market for ‘serious games’ and other applications in the objective of improving cognitive capacities or performance.

In light of increasing interest for EMA among researchers, clinicians and the general public, this chapter provides a review of the state of mobile cognitive assessments as they are currently applied in the fields of mental health and neurology. First, in order to fully understand the unmet needs to which mobile cognitive assessments respond, it is important to examine the main limitations of traditional neuropsychological testing that can at least partially be overcome by mobile assessments. A review of the literature will then be provided to describe the research to date using different forms of mobile cognitive testing, followed by a discussion of decisions for researchers using this approach.

2.1. Barriers and limitations of traditional neuropsychological testing

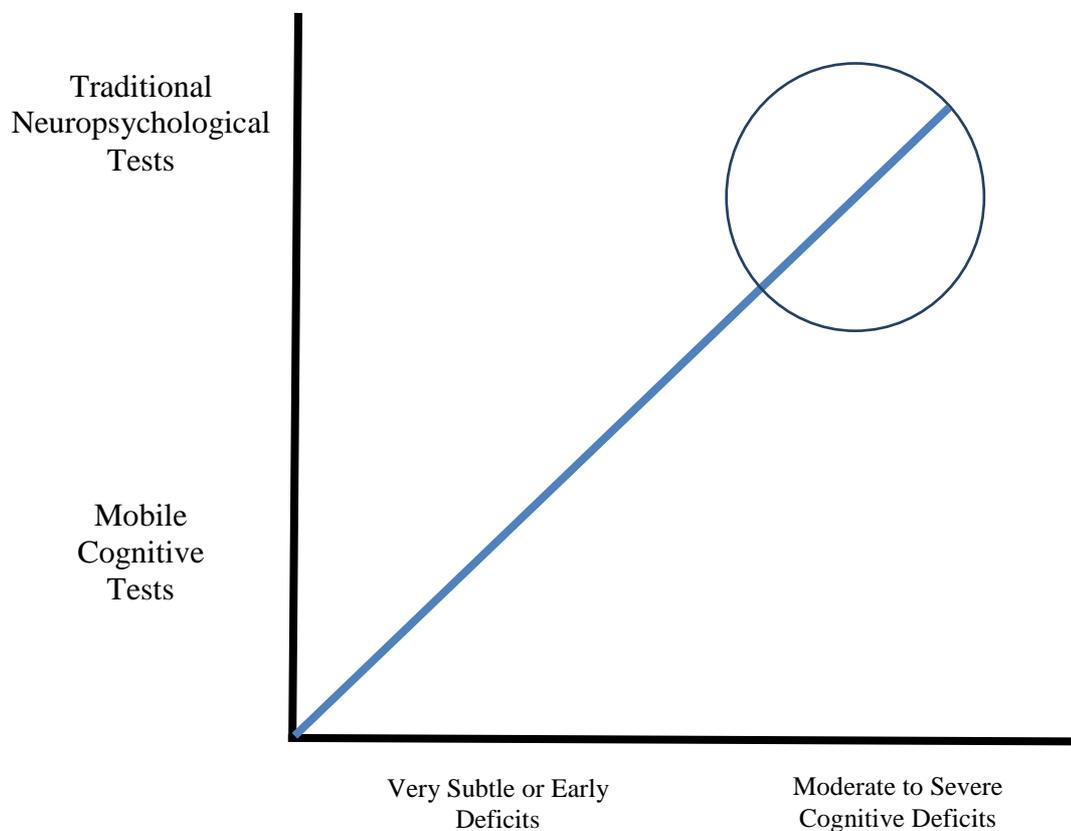
Validated neuropsychological tests with established normative data are essential to evaluating an individual’s cognitive capacities and to detect eventual areas of difficulty or

decline. A major limitation of traditional neuropsychological tests, however, is that they are most often administered only punctually, in that the patient is assessed by a clinician on a single given day and at a specific time. The potential problem with such punctual assessments is that the obtained scores may not be representative of the individuals functioning on ‘most days’ or ‘most times’. It is possible that on the day of testing, for example, the patient did not have his or her usual two cups of coffee, or he or she is distracted by thoughts of an upcoming medical procedure. Such circumstances may reasonably be thought to detrimentally impact the patient’s performance at the moment of testing. In the same way, it can also be argued that other events or circumstances may help to improve performance momentarily relative to what is “typical” for that patient, such as being particularly well-rested or having recently received very positive news. Such variance is quite common, as an individual’s cognitive performance scores fluctuate as a function of daily rhythms, stress and many other state-dependent influences. For this reason, traditional clinical instruments used for characterizing cognitive functioning or for detecting deficits are characterized by a “margin of error” that may range from very small to very large. Due to this variance, subtle but clinically-relevant dysfunction may go undetected at the time of examination. The impact that this margin of error has on research is that the detection of some signs of cognitive decline or deficits, especially those associated with the early onset of disorders, would often require the recruitment of very large samples of individuals. Large-scale studies are very expensive and the reality of the scientific literature is that they are few in number compared to investigations of more moderate sizes, which are limited in their statistical power to detect minor cognitive deficits.

Fortunately, an important alternative to increasing the number of participants enrolled in a study is to increase the number of observations acquired from each participant. This approach is often not feasible for traditional neuropsychological protocols due to the time and expense associated with single clinician-administered tests. However, mobile

technologies can readily provide brief and repeated assessments of cognitive functions, behavior and emotion in real-time and across the different contexts of daily life. The repeated daily assessment of these variables may reduce the margin of error associated with traditional clinical tools and provide high-resolution data concerning cognitive functions in daily life contexts. This point is illustrated in Figure 5, where the circle represents the hypothetical point at which cognitive decline can be detected using traditional instruments. The remaining blue line represents those cases that remain “under the radar” of traditional neuropsychological tests but that may nonetheless represent the earliest stages of dementia. Yet, one of the most coveted goals in aging research is to identify the earliest stages of cognitive decline in the hope that the overall trajectory might be altered or slowed if interventions can be provided long before the development of dementia.

Figure 5. Cognitive decline in dementia and matching assessment methods

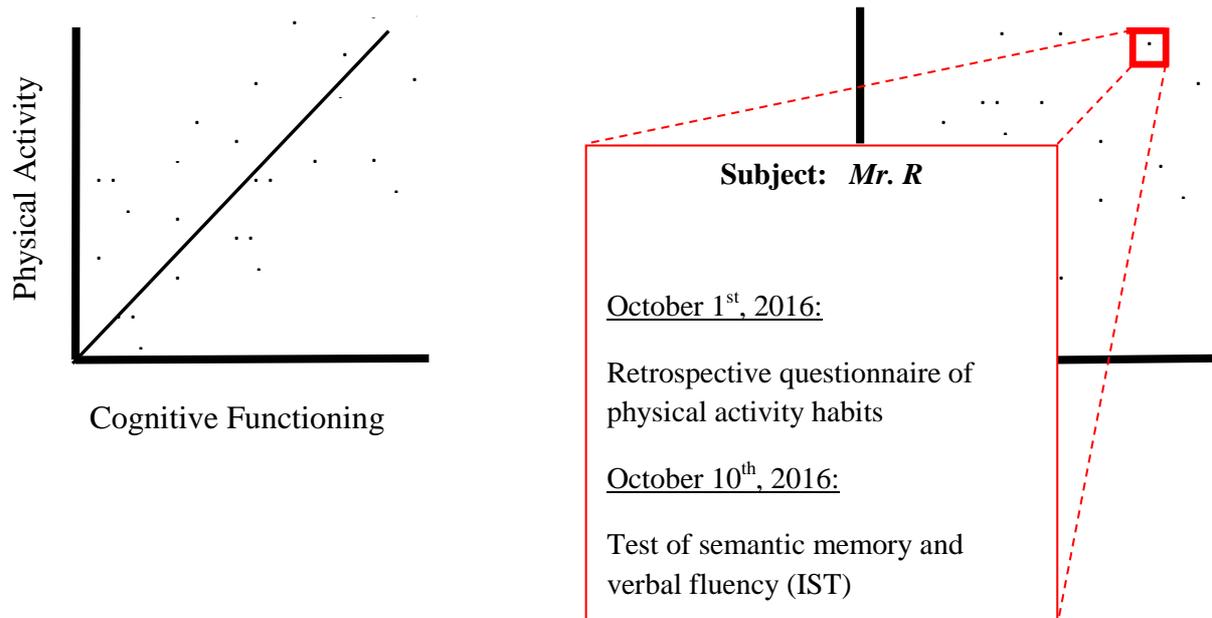


As an example of research using mobile tests to examine this question, Allard and colleagues (Allard et al., 2014) used both the Isaac Set Test (IST), a traditional neuropsychological test of semantic memory and verbal fluency, as well as repeated mobile assessments of these cognitive capacities. Despite the fact that poorer IST performance was previously shown to be associated with hippocampal atrophy in a sample of more than 300 individuals (Allard et al., 2014; Bernard et al., 2014), Allard and colleagues found no statistically significant association between these variables in their sample of 60 participants. This is not surprising given the link between the size of the margin of error and reduced statistical power associated with their smaller sample. However, when using repeated test scores administered through EMA in this study, a highly significant association was observed between semantic memory performance and hippocampal atrophy. Moreover, had EMA been used in the sample of 300 subjects (e.g. Bernard et al., 2014), particularly subtle associations may have been identified, thereby facilitating detection of the earliest stages of the cognitive decline and dementia.

A second major limitation of neuropsychological testing concerns the fixed environmental contexts in which the assessments occur. The provision of testing in one environment can provide information only about the overall state of the individual or, at best, only suggest potential correlations for which the causes cannot be verified. This second issue is illustrated in Figure 6. In the left panel of the figure, we present the general type of correlations frequently reported in the literature that have linked the risk of Alzheimer's disease and other forms of dementia to specific variables including physical activity, diet and other lifestyle characteristics. However, in the right panel one can observe that the type of data collected for each participant was based on a general and retrospective assessment of activity preferences as well as on a one-time administration of the IST. Based on such data, we are therefore unable to conclude whether increased physical activity actually improves

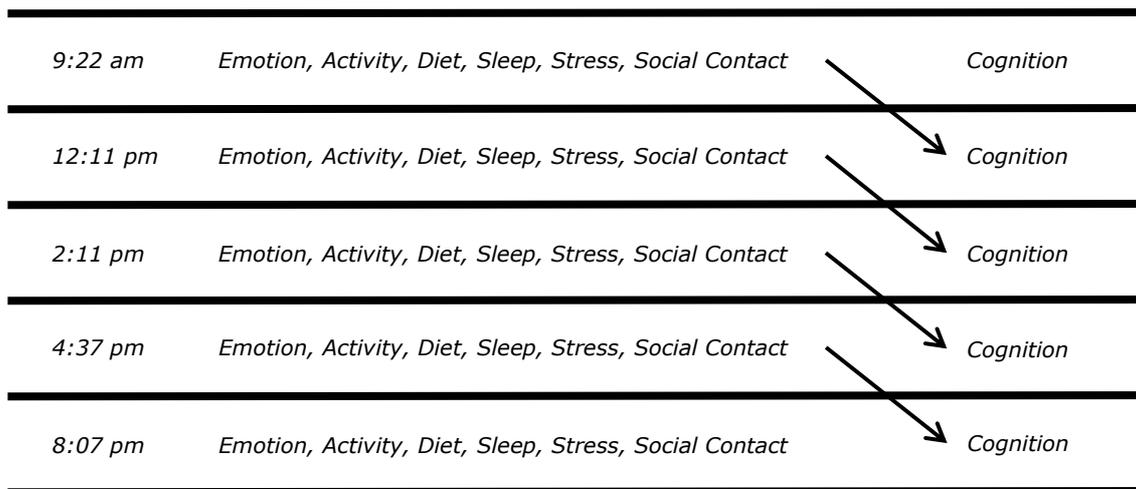
semantic memory, or if those with declining semantic memory perform less exercise, or again if additional factors may explain the link between these two variables in the sample.

Figure 6. Nature of clinical and epidemiologic data demonstrating a correlation between physical activity and cognitive functioning



In order to examine such daily-life correlations, as well as to determine their directionality, highly dynamic data, adapted to the natural life cycle of each variable, are needed. An example of one such protocol is presented in Figure 7, based on five electronic interviews per day over a week-long period as used by the Allard and colleagues study (Allard et al., 2014).

Figure 7. Assessment of lifestyles and cognitive functioning



In this investigation, a range of daily life experiences, behaviors and events were assessed at each electronic interview and a mobile test of semantic memory performance was administered. Allard and colleagues were able to demonstrate that intellectually-stimulating activities (such as playing scrabble, reading or crossword puzzles) increased semantic memory performance in the subsequent hours of the same day, while no association was found for other candidate variables (i.e., physical activity, social interaction, daily chores). Moreover, each arrow represented in Figure 7 presents the prospective link of these activities with later-memory performance, controlling for the state of memory at the previous assessment and thereby allowing for directionality to be assessed. This method also allows for such relationships to be examined in the reverse direction (i.e. cognitive performance predicting intellectually-stimulating activities), thereby overcoming the circularity of many correlations that have been identified with traditional research methods.

2.2. Self-administered mobile cognitive tests

2.2.1. Data capture methods and cognitive functions tested

We (Swendsen, Schweitzer & Moore, 2017) identified twelve studies other than our study (Schweitzer et al., 2017) that utilized self-administered, mobile cognitive tests within an EMA paradigm (Allard et al., 2014; Brouillette et al., 2013; Frings et al., 2008; Keenan et al., 2014; Kennedy et al., 2011; Riediger et al., 2014; Schuster et al., 2015; Sliwinski et al., 2016; Timmers et al., 2014; Tiplady et al., 2009; Veasey et al., 2015; Waters and Li, 2008). At this time, mobile cognitive tests have mostly been used in non-clinical samples of healthy volunteers. One exception was a study by Frings and colleagues (Frings et al., 2008) in which real-time working memory performance of patients with epilepsy (who were being titrated with Levetiracetam) was compared to a control group of patients on stable treatment with antiepileptic drugs. Participants in these studies ranged in age from 14 to 83 years old (mean weighted by sample size: 47 years). Most studies cited above were administered on a mobile phone provided to the participants, with the exception of four studies employing Personal Digital Assistants (Allard et al., 2014; Frings et al., 2008; Schuster et al., 2015; Waters and Li, 2008). Timing and duration of studies ranged from one assessment per day for one day (Brouillette et al., 2013) up to five assessments per day for fourteen days (Sliwinski et al., 2016). On average, studies deployed 3.4 assessments per day for 6.6 days. This variance is attributable to differences in specific study objectives as well as to whether investigators chose to link mobile cognitive test scores to other momentary assessments versus to retrospective data collected over the previous several hours.

Concerning the diversity of functions tested, the majority of these studies administered a single type of mobile cognitive test (Allard et al., 2014; Brouillette et al., 2013; Frings et al., 2008; Riediger et al., 2014; Schuster et al., 2015; Timmers et al., 2014),

while a minority administered multiple types of tests over the duration of the study (Sliwinski et al., 2016; Tiplady et al., 2009; Veasey et al., 2015). Each research team in the identified investigations developed at least one unique test for their study. Overall, tests were developed in the following cognitive domains: working memory, attention/reaction time, processing speed, semantic memory, short-term memory, delayed memory, and executive functions. To our knowledge, none of these tests are currently publically available. Touch screens were the most common mode of response on the mobile device, with scoring completed automatically.

2.2.2. Psychometrics

The psychometric properties of studies completed to-date are encouraging. Several studies have examined the convergent validity between standard in-lab tests and the mobile versions of tests of similar constructs. While we were unable to concatenate the correlations between laboratory-based and the mobile versions of tests, concordance with in-lab tests generally had medium effects (Allard et al., 2014; Brouillette et al., 2013; Schuster et al., 2015; Sliwinski et al., 2016; Waters and Li, 2008). It should be noted, however, that these concordance rates may be reduced due to difficulties in fully duplicating neuropsychological tests with a corresponding mobile version, and that some studies (Allard et al., 2014) compared distinct cognitive tests presumed to assess the same cognitive capacity. Additionally, two studies directly examined whether significant associations existed between their mobile cognitive tests and in-lab measures of different constructs, and found evidence in support of the mobile test's discriminant validity (Brouillette et al., 2013; Schuster et al., 2015).

As previously discussed, a unique advantage of EMA cognitive tests is the ability for repeated sampling, both within and between days. Studies to-date have demonstrated good test-retest reliability (Brouillette et al., 2013; Timmers et al., 2014), adequate internal

reliability (Waters and Li, 2008), and excellent between-person reliability and within-person variability (Sliwinski et al., 2016) for various mobile cognitive tests.

2.2.3. Usefulness above and beyond traditional laboratory-based testing

Five studies have directly examined the usefulness of EMA cognitive assessments beyond traditional in-lab testing. In addition to the research previously described by Allard and colleagues (Allard et al., 2014), Riediger and colleagues (Riediger et al., 2014) examined the association between EMA assessed nervousness, momentary heart rate, and concurrent working memory performance. The authors found an interaction effect with increased nervousness and stress related to poorer working memory test performance in older, rather than younger, adults. A third study examined the relationship between self-reported alcohol consumption and real-world attention and working memory scores (Tiplady et al., 2009). Expected diurnal changes throughout the day on the EMA attention tasks were detected, and participants' performance on the EMA tests were worse on days when alcohol had been consumed.

The final two studies examined the lagged effect of caffeine and eating breakfast, respectively, on future real-world cognitive performance. Keenan and colleagues (Keenan et al., 2014) evaluated the effects of bedtime caffeine use on next-day working memory in a randomized controlled trial. Blind to treatment allocation, participants took one week of caffeine or placebo pills. Despite disrupted sleep when taking bedtime caffeine, next-day working memory performance was significantly higher throughout the caffeine week than the placebo week. Veasey and colleagues (Veasey et al., 2015) were interested in the effect of breakfast size prior to morning exercise on working memory performance later in the day. Participants included twenty-four young, healthy, and habitually active females. Results indicated that consuming a larger breakfast was associated with mid-afternoon working

memory decrements; these decrements were not observed when participants consumed a smaller breakfast.

2.2.4. Decisions for researchers

As previously mentioned, the choice of sampling schedules depends largely on the research questions of the given investigation. For example, a single cognitive assessment per day over multiple days may be sufficient if the researcher is interested mainly in acquiring repeated performance scores, such as in the goal of reducing variance associated with traditional neuropsychological test or eventually to increase statistical power. However, more intensive sampling schedules would be required in order to examine within-day associations of cognitive performance with mood or behavior, such as that used by Allard and colleagues (Allard et al., 2014).

As for number of days of assessment, many EMA researchers have applied a one-week design in order to acquire information across the natural rhythms of work and leisure that are most common in society, but multiple-week assessments have also been applied with success (Sliwinski et al., 2016). Despite these choices that depend on the specific scientific questions at hand, researchers should apply care in the timing of cognitive assessments and by test type. Unless they are used at every within-day assessment, mobile cognitive tests should be counterbalanced across the day in order to avoid systematic time biases (as would be the case when assessing memory or attention only in the morning or evening).

The investigations presented in the previous review also involved study-specific “in-house” programming that presents certain difficulties for the broader diffusion of mobile cognitive tests and that also require pilot testing on a study-to-study basis. While establishing the acceptability and feasibility of these tests is often possible with relatively small sample sizes, their convergent or discriminant validity can often be examined only after a larger

sample has collected. For this reason, the selection of neuropsychological tests to accompany a mobile testing protocol are of particular importance, and as careful matching should be achieved relative to the specific cognitive functions being tested. This objective often requires creativity on behalf of researchers in order to respect the copyright of any test that is not in the public domain, and while waiting for the field to permit the availability of mobile cognitive test “batteries” to its scientific community that would avoid the replication of unnecessary validation efforts.

Finally, a range of strategies have been applied with success in order to increase subject compliance. Most often, these approaches generally involve either financial or psychological incentives. Financial incentives may include a ‘bonus’ for a particular level of compliance (such as completion of 75% of all programmed assessments), and psychological incentives have included providing a progressive “counter” that shows subjects how many assessments were completed to date or that administer positive feedback through messages on the mobile device. However, these strategies cannot substitute for a motivated explanation of the importance of the project to the participant and why mobile technologies help the researcher to understand their daily experiences, or their personal ‘story’, in a way that is not possible with other methods.

Chapter 3. EMA application development

As this thesis is based on the creation of a new and unique EMA program, it is important to first review existing options that were considered by the CNRS team but that eventually were rejected, resulting in the decision to develop new programming. This work began at the cusp of a major transition in technologies, and therefore included consideration of pre-smartphone applications as well as the earliest publically-available programs for these devices, presented below.

3.1. Existing applications

3.1.1. Purdue Momentary Assessment Tool (PMAT)

The original PMAT was developed as a shareware EMA program for Personal Digital Assistance (or PDAs, for which we primarily used the Tungsten model manufactured by PALM, incorporated). However, changes in PDA technology quickly rendered the original version obsolete and required additional programming in order to make use of this program. While a convenient graphical user interface was available to design studies (which were saved in the XML format), PMAT was also limited to single-choice questions, multiple-choice questions, and sliders. A considerable amount of program code that was specific to PALM, making it difficult to port to another platform, especially for the database related operations (PDB format). Concerning other inconveniences, the original PMAT study data (containing questions and response items) was not included in data files, probably for performance reasons. Any added or edited item was assigned a new ID based on a global value equaling the newest ID. As a result, modifying an existing study, for example correcting a

typographical error in a response, generated a new ID for that response and made the resulting data files difficult to synchronize with the previous files. As study files were not included with participants' data, this resulted in files that could not be matched with their original studies. To make matters even worse, the PMAT desktop application silently failed in that case, which made importing values from studies that were edited several times a very difficult process. An application was developed exclusively for that purpose: it analyzed all existing IDs from the different studies and participants' data files, and extracted the data using the best match that was found. These developments nonetheless led to a large number of studies from the Bordeaux group with whom I worked (Allard et al., 2014; Ben-Zeev et al., 2011, 2012; Depp et al., 2016; Fatseas et al., 2015; Granholm et al., 2013; Husky et al., 2014, 2010; Jean et al., 2013; Lassalle-Lagadec et al., 2013; Mazure et al., 2014; Salamon et al., 2011, 2014; Serre et al., 2012; So et al., 2014; Swendsen et al., 2011; Villain et al., 2016). However, a fatal flaw of PMAT was the durability of its platform: PALM was progressively losing market shares in favor of Android, Apple and Microsoft, and replacing out-of-order devices became harder over time.

3.1.2. MyExperience

The MyExperience software was more modern than PMAT in many ways: not only the platform (Windows Mobile) and hardware were more recent, but the structure of studies also featured a built-in scripting language that allowed for a much improved flexibility when designing studies. While the absence of a graphical application for programming studies might discourage researchers with no programming experience, the power resulting from its flexibility made it a good candidate for EMA studies. Unfortunately, stress-testing the application revealed that the memory management was poor, especially for extensive electronic questionnaires: every question was held in memory until the entire questionnaire

was finished. Consequently, after about 50 questions the application systematically crashed because of out-of-memory errors. This stability issue alone was enough to drive researchers (and our team) away from the app, as an EMA application is typically given to participants for one to several weeks, and it is critical that it remains stable during the entire assessment period.

3.1.3. Testing conclusions

As noted above, both available applications had pro and cons, however their issues were considered too detrimental to make them a valid choice for our research (including this thesis). As a result, we decided to develop our own application that would make use of the advantages of both the tested programs while avoiding their weaknesses.

3.2. Platform choice

Our application was named *Simplex* to reflect its purpose: sampling experience. As technology evolves at a very quick rate, the first thing we considered was the platform upon which we would build the application. Several criteria had to be taken into account, among which the most important were durability, availability, price and portability. Three platforms were considered: Android from Google, iOS from Apple, and Windows Mobile from Microsoft. Our choice naturally went in favor of Android, as it was a clear winner on all four criterions.

Durability and availability proved to be a critical issue with PALMs; however, in the case of Android, we were assured that it would not cause any problem for many years. Android OS ensures backward compatibility, which means that applications developed for a specific version will be able to run on subsequent versions. The hardware availability also increased our confidence: thousands of different Android devices exist, and many new

devices are brought to the market every year, ensuring we would never be short on replacements. Indeed, over the course of the few years we used Samplex, we had to use five distinct Android devices which did not imply any reprogramming to the survey aspect of Samplex, which proved this platform was a relevant choice. However the device locking remained a difficult task for every new device (see the “Hacking” section below).

The pricing might not seem to be a critical criterion at first. However, in the case of EMA studies, one often needs to have dozens of devices at one’s disposal. In this case, small differences in device costs may have a huge impact on the feasibility of such studies. The very dynamic market of Android devices naturally resulted in much more choices and lower prices for these devices, especially compared to Apple devices which the few existing ones often costs several times their Android equivalent. Devices running Microsoft Mobile OS prices were between Apple and Android, both in term of variety and price.

Portability also needed to be considered for several reasons. A good portability would allow developing a PC desktop application counterpart without programming everything from scratch. It would also potentially enable the application to run on other platforms without modifying the core of the application. The programming language used to build Android applications is Java, and has the interesting property of being multiplatform-oriented at its core, as Java programs are translated into *Java bytecode*, a universal instruction format, which is later interpreted and translated into machine code by platform-specific Java virtual machines.

3.3. Device choice

After investigating several devices, we opted for an Android tablet, the Samsung Galaxy Wifi 4.2” (displayed in Figure 8). As a tablet it does not feature telephone capabilities. It comes with a 1,500 mAh battery, which ensures at least a day of operation without

recharging. We found the touch-screen to be sensitive and accurate enough for participants. The stock Android version is Android 2.3; while it was not the latest version at the time we acquired these devices, it was suited to their relatively low processing power (single-core 1GHz processor with 512MB of memory). The storage capacity of 8GB ensured the device would have enough space to save even the most demanding studies. The price for each device was approximately 130 euros. However, this product reached its end of life, and for subsequent studies, other Android devices had to be acquired.

Figure 8. The Samsung Galaxy Wifi 4.2" tablet



3.4. Hacking

“Hacking” was originally a slang term for programmer-based actions that take control of, or modify, a computerized device in a manner other than what was originally intended by its developer. Over time, the word “hacking” has entered into the mainstream of computer programming language, and it no longer is used only as a synonym for illegal pirating of such devices. In the context of this thesis, hacking refers to user-based (our)

modifications that are made possible by the broader community of developers. It is important to note that our EMA application project was not a traditional application. Usually, applications are just part of a multipurpose operating system, in which the user controls and chooses what applications suit his or her needs, and has the freedom to install, configure and uninstall them, change operating system settings, and so on. However in our case, the idea was that we would handle the device installation and configuration in the laboratory, so that study participants would receive these devices in a study-ready state, without needing any technical knowledge about how to operate them. To make the week-long interviews as reliable as possible, we also wanted the devices to look as if they were dedicated to EMA, effectively locking participants out of the underlying operating system. On the programming side, this has several implications: not only should the application start as soon as the device is turned on, but once it is started, it should also be virtually impossible for a participant to escape the EMA application and access the system. Indeed, if a participant were able to exit the EMA application, not only interviews could be missed, but if the date and time were changed, the entire week of data collection could be lost.

Running an application at startup is easy as this feature exists since early Android versions. However until very recently, device locking was inexistent (Android 5 and above now include the concept of device ownership). The reason was probably that in order to improve user security, no random application should be able to take over their phones or tablets, but this reasoning did not apply in our case as we were fully responsible for the device programming and only wanted participants to use a single application.

In order to lock participants from exiting the application, our intent was to disable the device buttons. A typical Android device (at the time of Android 2.x) has six buttons: power, volume up, volume down, back, home and menu. Among these, two buttons potentially allow to escape an application: the home button function is to redirect a user to the

home screen (the equivalent of the desktop in computer operating systems), and the back button is a means to show the previous screen. The effect of pressing the back button is also controllable by applications; however this is not the case for the home button, which is handled by the system and no mechanism in Android 2.x allows for a modification of its behavior.

However, Android is built on top of an underlying Linux operating system, and the definition of all button functions happens at this system level through a dedicated text file. This file defines one button per line, each line consisting of a button name, a number representing the button *scancode*, and an optional field defining whether the button should be able to “awaken” a sleeping device or not. This file resides on the *system* partition (which contains Android), which is a block of data protected from any modification (a *read-only* partition) except by the administrator, which is called the *root* user in Linux. However, Android devices do not grant any root access by default for security reasons: any application with root capabilities would be able to read and modify other application’s data, hence breaking the isolation principle (each application in Android is meant to be isolated from others) and resulting in potential privacy violations. In our case, there was no such risk, especially as our devices did not have any network connectivity. Therefore, the first step was to obtain root privileges on our devices.

We first tried all the existing *exploits* at that time, especially the popular *zerg-rush* exploit; these could potentially give a temporary root access which would be enough to modify the files we needed, but they did not work on our version of Android. As a second approach, we considered that for most devices, alternative operating systems with integrated root access already exist and can be downloaded online from communities dedicated to Android hacking such as the XDA developer community (<https://www.xda-developers.com>). Unfortunately, the device we chose was not popular enough to already have been hacked, so

no *rooted* operating system existed for this device. We concluded that we had to hack into the devices ourselves to create a rooted system.

In order to create such a modified operating system, we used a method that requires the use of two devices. This method takes advantage of the fact that even if the system partition cannot be modified while the system is running, it is still possible to *flash* a partition when the system is offline. Flashing a system is the method that vendors use to upgrade systems to a new version. It consists of transferring a complete system partition (a file that typically weights several hundred megabytes) onto the previous system (overwriting it in the process) via an USB cable, while the device is set into a specific mode for this operation. However, flashing only allows for the individual to *write* to the device, not to *read* from it, so we still needed a way to obtain the stock system image in order to modify it before it could be subsequently flashed on devices.

The stock system was obtained in two steps. First, a non-functional system partition was manually crafted so that it allowed at least root access on a device through the super-user file *su* which was absent from the original system. This system was flashed on a device. As expected, this device did not run Android properly, but it was enough to access the device as root via Android *USB debugging* feature, which provides a command shell through USB. Thanks to these administrator privileges, the unmodified *kernel* partition (where the Linux startup files reside) could be downloaded on a computer. We then injected the super-user file into it, and flashed this modified kernel partition on a second, yet unmodified device. Thanks to the root access granted by the kernel partition, we were able to download the unmodified system partition from this second device. We then injected the super-user file on this system partition, hence obtaining a functional rooted system for our devices.

After flashing this rooted system, we were able to use the required root access to modify the button definition file and disable them. However, one specific function could not

be disabled: holding the power button for 10 seconds has the effect of rebooting a device, and seems to be hardware-controlled. As this knowledge is not likely to be known by participants, and as the EMA application would start after reboot anyway, we considered it was an acceptable drawback and focused on the development phase.

3.5. Development

In order to develop an Android application, several frameworks can be used. The most basic of these frameworks would consist of manually *compiling* programs into Java bytecode, and converting them into Android package files (APK) by *dexing*, *aligning* and *signing* them. However, integrated development environments (IDEs) exist that would improve the overall productivity by automating all these processes while also offering additional services such as syntax highlighting, inline help, instant install and run, version control integration, and so on. We started the project using Eclipse, as this was the most popular IDE for Android projects at that time, and was featuring an Android-ready version called Android Developer Tools (ADT). However, we later switched to Android Studio, an IDE from Google dedicated to Android that did not exist at the beginning of our development phase.

3.5.1. Participant experience

The application is designed with ease-of-use in mind, as participants are assumed to have no technical knowledge about how to manipulate Android. Therefore, the underlying presence of Android should be completely transparent from the user. This means the application should be run as soon as device starts up, but also that participants should be locked inside the application once the application is running. Additionally, the application

should be tolerant to battery failure, which means that any survey that should have occurred while the device was powered off should be gracefully skipped.

A typical scenario for a participant starts at the lab, with a brief training on how the application looks like and how to answer a series of simple questions, and is then asked to carry the device with him for the entire survey period (usually one or two weeks). The participant then goes back to his everyday life. When a survey is about to take place, the device vibrates and emit an alarm sound while displaying an invitation to start the survey until the participants accepts it or the alarm timeout is reached (usually about 10 minutes). When no survey is running, the idle screen shows information about whether another survey is planned for that day, no more surveys are planned for this day and battery should be recharged, or the entire survey period has ended and the device should be returned.

3.5.2. Research staff experience

An investigative team will typically use a dedicated computer for installing and configuring mobile devices as well as for importing data input by participants. Preparing a computer is straightforward as it only needs to have the appropriate drivers installed, which are usually provided on the manufacturer's website. Configuring mobile devices is a more complex task, so our team made installers that automate this process, which includes two steps: the modified system first has to be flashed on devices; once this is done, the device is configured so the buttons are locked and Samplex can be installed. The flashing step transparently uses two underlying open-source programs: "Zadig" provides the driver that allows flashing (<http://zadig.akeo.ie/>), and "Heimdall" is a console program that performs the actual flashing operation (<https://github.com/Benjamin-Dobell/Heimdall>).

In practice it is useful to have one device used exclusively for participant's training before the survey period begins. To this end, Samplex features a demonstration mode which

allows the research staff to present participants with a predefined set of questionnaires with a single button press (no responses are saved in that mode). Once the training has been completed, the device has to be initialized, which usually consists in specifying at least a subject identifier and the minimum and maximum times during which the survey will happen each day, and optionally more information depending on studies (e.g. participant's group or gender). Once the device has been initialized, it can be handled to a participant along with a charger for the study period.

Once the research team receives devices that contain the participant's responses, the data first need to be saved on a safe storage. This phase is done in two steps. First, the data has to be exported from Samplex to the mobile device storage. This is due to the structure of Android in which each application runs in isolation from others. In Samplex, this is done by accessing the administrator menu which is protected by a password known only by the research team, and using the « Export results » option. Once the data has been exported, the data can simply be transferred to a computer using a USB cable. The data consists of at least three files: an XML file which is the raw database that was used by Samplex during the survey period; a CSV file that is generated during export that contains only the relevant data, and allows to conveniently examine and analyze the data as it is the most compatible sheet file format; a LOG file that contains the application's internal events and is traditionally used for debugging issues. Additionally, any sound sample that was recorded will also be present in the same folder as the three aforementioned files in the Android's native 3GPP format. Once the data has been saved, the device can be reset by choosing the « Reset study » option in the administrator menu, so it is ready to be used by the next participant.

3.5.3. Project structure

The project was divided in two parts from the start: the *core* package, which would contain the application logic and be platform-independent, and the *graphical user interface (GUI)* package that would be specifically designed for a particular platform, in this case the Android platform. EMA studies are stored as XML files. In order to manipulate XML structures programmatically, the open-source *Simple* library from Niall Gallagher was used (<http://simple.sourceforge.net>). This library is light on resources, fast and syntactically convenient; however while it was actively maintained at the time the project started, it is no longer the case and the project might be adjusted to use another library in the future.

Almost all of Simplex activity articulates around two concepts: a *study*, which is a fixed object that corresponds to the questionnaires, signal times, specific messages and others constraints designed by researchers, and a *progress*, which is dynamic in nature as it represents the progression of a participant throughout the interview period of a study. A study is defined in a XML file (which can be formally described as a hierarchy of *nodes* with optional *attributes*, child nodes being called *elements*), in which the top-most node is named *study*, and has two attributes, name and duration (in days). The elements that are required to define a study will be presented. These elements represent messages, settings, variables, questionnaires, times and events.

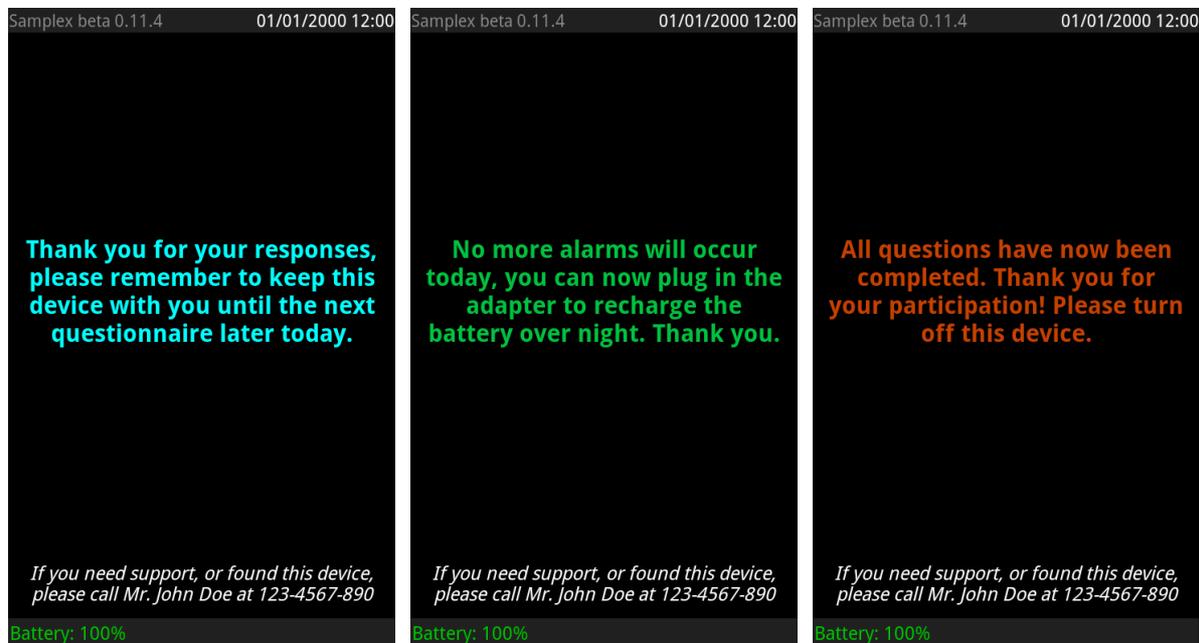
3.5.3.1. Messages

Various messages are displayed to participants over the course of an interview period. These can be customized for each study depending on the language and cultural context of participants.

Information and status messages

When Samplex is used by participants, it is running in user mode. In that mode, a default *idle screen* is displayed when no survey is running. This idle screen displays several points of information about the device such as battery level and the current date and time, as well as information about the study such as investigators contact information and the status of the study (shown in Figure 9).

Figure 9. Idle screen showing different states



The information message is constantly displayed on the bottom of the idle screen, and is meant to provide a way of contacting the investigative team in case a participant has questions, but also in the case a device is lost so that anyone who finds it would be able to retribute this device to the research staff.

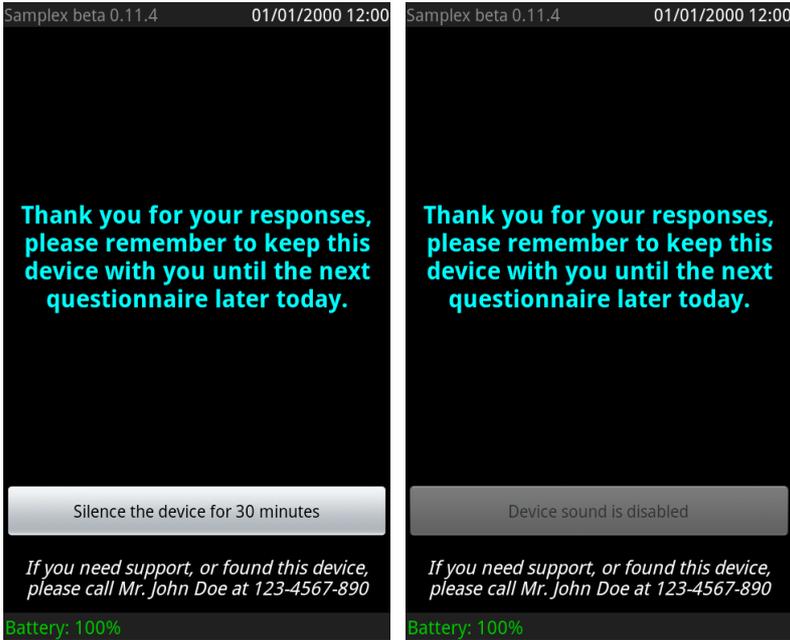
The study status informs the participant about the course of the survey. There are three different status messages that correspond to different study states. A specific message can be displayed when there is at least one pending survey in the current day; this message typically advises the participant to keep the device within reach as a survey is planned to

happen later in the day. When all planned surveys for the current day have been passed, another message tells the participant that no more surveys are planned this day, and advise the participant to recharge the device battery. Finally, when the entire interview period is complete, a specific message informs the participant that the study is finished and that the device may be sent back to the investigative team.

Interface messages

Sometimes, it is desirable to allow the participants to silence the device for a short period of time. In this case, the text displayed by the interface can be adjusted (Figure 10). Note that even when the device is silenced, it will still use vibrations to signal that a survey is ready.

Figure 10. Alarm silencing interface



3.5.3.2. Settings

A few settings can be set for the study. One is the alarm volume that will be used when triggering an audio signal to alert the participant that an electronic interview is ready for completion. Another important setting is the unlock code; this code will be used by the research staff to access Simplex administrator functions, such as exporting results, resetting the study, and so forth. This code is not directly stored in the study file: for security reasons its MD5 hash is stored. Even if MD5 is known to be easy to hack itself, it is enough to prevent direct exposure to a possible reader.

3.5.3.3. Parameters

Parameters are a more complex group of settings that can affect the behavior of question screens, such as their layout, timeout and so on. As different parameters might be desirable for distinct questions, they follow a hierarchical definition. For example, a question timeout can be defined for the study, but if another question timeout is defined for a questionnaire, it will take precedence for that specific questionnaire. If a specific parameter has not been defined at any level, its default will apply for the whole study. Parameters can be defined from the most global to the most specific level: study, event, questionnaire and question. Parameters defined in specific areas of the study will take precedence over those that are defined at a more global level.

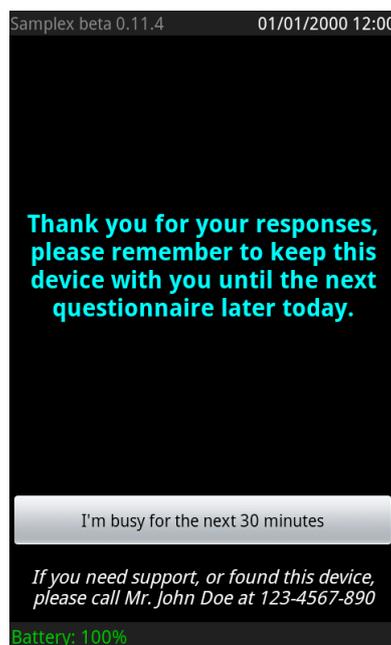
Busy button parameters

The busy button is an alternative to the alarm silencing feature (Figure 11). It allows for setting a small time interval during which no alarm will trigger; if a signal was planned at this interval, it will occur at the end of the ‘busy interval’. The parameters that control how this function behaves are described in Table 2.

Table 2. Busy button parameters

Parameter description	Default setting
Enables or disable the busy button	Disabled
Default button text	<i>I'm busy for the next 30 minutes</i>
Busy interval duration	30 minutes
Show during notification	False
Number of allowed presses for each interval	1

Figure 11. Busy button interface



Notification parameters

When the time of a planned survey is reached, the device will enter a notification state, where the participant is expected to respond that he or she is ready to answer questions (Figure 12). To this end, the device emits an audio signal, vibrates, and displays a message that typically invites the participant to press a large area to begin the survey. When a participant does not respond to a notification after a given time, the whole survey for this

signal is considered as being missed. Notification behavior may be adjusted through the parameters described in Table 3.

Table 3. Notification parameters

Parameter description	Default setting
Text displayed during notification	<i>Press here to start questions</i>
Duration after which the notification ends	10 minutes
Delay between audio and vibration signals	2 minutes
Custom signal sequence (replaces duration and delay)	Empty

Figure 12. Notification screen



Question parameters

Questions are what participants will be the most exposed to and constitute the main object of this application, as they are the very source of interest for researchers. As a result, many parameters modify the way questions are displayed: from layout to interface logic, several question behaviors can be adjusted to suit the researchers' and participants' needs. These parameters are described in Table 4.

Table 4. Question parameters

Parameter description	Default setting	Target question type
Demonstration mode: if enabled, responses are discarded	Disabled	All question types
Question text size	20 dip*	All question types
Inactivity timeout during a question, after which a survey is considered canceled	2 minutes	All question types
Smooth transition between questions	Disabled	All question types
Text displayed on the confirmation button	<i>OK</i>	All question types except record
Allow scrolling when the list of responses do not fit on a single screen	Enabled	Single-choice and multiple-choice
Response padding (inner space)	12 dip*	Single-choice and multiple-choice
Response margin (outer space)	3 dip*	Single-choice and multiple-choice
Response text size	20 dip*	Single-choice and multiple-choice
Quick single-choice mode that skips confirmation	Disabled	Single-choice
Text displayed for the special response in multiple-choice questions that means ‘nothing’	<i>None of the above</i>	Multiple-choice
Content text size	30 dip*	Record and instructions
Content layout	Centered	Record and instructions
Text displayed by the button to stop recording	<i>Stop</i>	Record
Text displayed by the button at the end of the record	<i>Done, thank you!</i>	Record
Initial delay during which confirmation button is disabled to prevent accidental presses	2 seconds	Instructions
Inactivity delay after which the confirmation button starts flashing	30 seconds	Instructions
Auto-completion (only used for testing/development)	Disabled	All question types

* dip: density-independent pixels, a unit that is used on mobile devices to homogenize output across different devices

3.5.3.4. Variables

Variables, like in most programming languages, hold typed values and can be tested or changed throughout the course of a survey. They are typically initialized at the beginning of a study by the research staff in order to set various constraints that will affect how Samplex behaves. For example, setting the gender or the group to which a participant belongs may affect which questions are asked. Several variable types are defined, as shown in Table 5.

Table 5. Variable types

Variable type	Description
Single	Holds a single value from a list of possible choices
Multi	Holds multiple values from a list of possible choices
Date	Holds a date
Number	Holds a number
Text	Holds a character string

3.5.3.5. Questionnaires

Questionnaires are simply defined as an ordered list of questions. Each questionnaire can feature a condition which will be checked every time the questionnaire is triggered; if this condition is not met, the entire questionnaire is skipped. Questionnaires also feature parameters which will override the default for all included questions (unless the questions themselves override the questionnaire parameters).

3.5.3.6. Times

Times represent the “time of a day”, independently of the date. They are referenced by *events* to determine when the alarms fire, signaling participants that a survey is ready. While it is possible to use the same fixed times for all participants, it is usually not desirable: not only participants have different habits and real-life constraints, but probing them at the same time every day may also introduce biases (e.g. a participant following a routine is likely

to be in the same state at a given hour of the day). In order to avoid these issues, Samplex allows defining times at initialization (typically wake up and sleep times), times relative to others, times randomly chosen between two others, or lists of predefined times. The available types of times are described in Table 6.

Table 6. Time types

Time type	Description
Raw	Defines a fixed time
User	Allows to define a customizable time during study initialization
Relative	Defines a time relatively to another, shifted by a given amount of minutes
Between	Set a time by as a fixed or random proportion between two others
List	Defines sets of predefined times

3.5.3.7. Events

Events define a list of questionnaires that will be asked given certain conditions. Typically, they are used in conjunction with times to trigger surveys, but can also be defined to trigger on a specific study day, weekend day, or week day. They also offer the possibility for participants to manually trigger surveys when researchers decide to allow that option, by pressing a button displayed in the *idle screen*. Events can define parameters that will apply to all questionnaires and questions that are triggered within that event, unless these questionnaire or questions override them. Event types are described in Table 7.

Table 7. Event types

Event type	Description
Time	A time event will trigger at the referenced time every day, unless a specific study day, week day or weekend day is specified
Manual	A manual event is triggered by a participant by pressing a button in the idle screen
Demo	Demonstration events are designed for the training phase of participants, and usually present them with the typical questionnaires without recording any information.

3.5.3.8. Questions

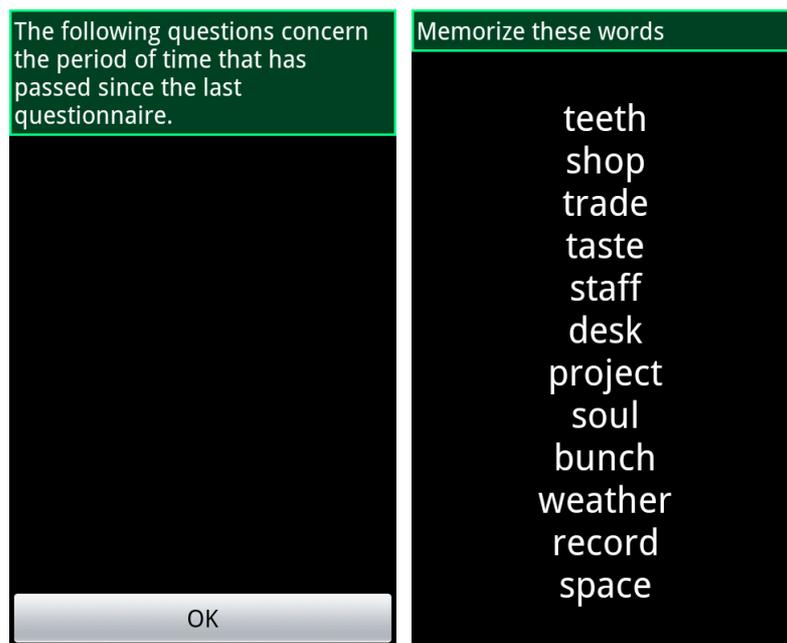
In order to flexibly adapt to researcher and participants needs, questions can be configured to behave in different ways (e.g. question, response and content text size can be adjusted, as well as their layout). Every question may feature a condition that needs to be true for the question to be asked, and parameters that will override the default behavior. Questions can also define *branches*, which define questionnaires that will be asked when specific responses are given. When a question has not been answered within a predefined timeout, responses to the previous questions are saved but the whole survey is marked as canceled, and the device goes back to the *idle screen*.

While all researcher needs cannot be anticipated, the most common types of questions are implemented in Samplex: instructions screens, single choice questions, multiple choice questions, recording tasks, questions asking for a time, and questions asking for text. When researchers need a type of task that is not implemented in Samplex, it is still possible to call on an external application through the use of the “custom” question type. The following sections present the different questions types that are predefined in Samplex.

Instruction screens

Instruction screens aim at guiding the participant throughout a survey, and usually consist of instructions text, a validation button, and optionally a visual content (Figure 13). By default, the validation button is disabled for two seconds to prevent accidentally pressing it, and can be set to flash after a given amount of time. The validation button may also be hidden, and in that case a fixed amount of time (adjustable in milliseconds) is set to determine when the survey should display the next question. This latter option is useful in studies that need to expose visual content for a predefined amount of time.

Figure 13. Instructions screens



Single-choice questions

Single-choice questions display the question text, followed by a list of possible responses, and a validation button (Figure 14). This type of question allows exactly one response, and the validation button will remain disabled as long as no response has been selected. If “quick mode” is enabled, no validation button is displayed and the first response that is selected will be instantly validated (this setting is not recommended as it is error prone). The order of responses can be randomized. The list of responses is scrollable by default (when they do not fit on one screen), however scrolling may be disabled if necessary: in that case, responses will be displayed screen by screen until one is selected.

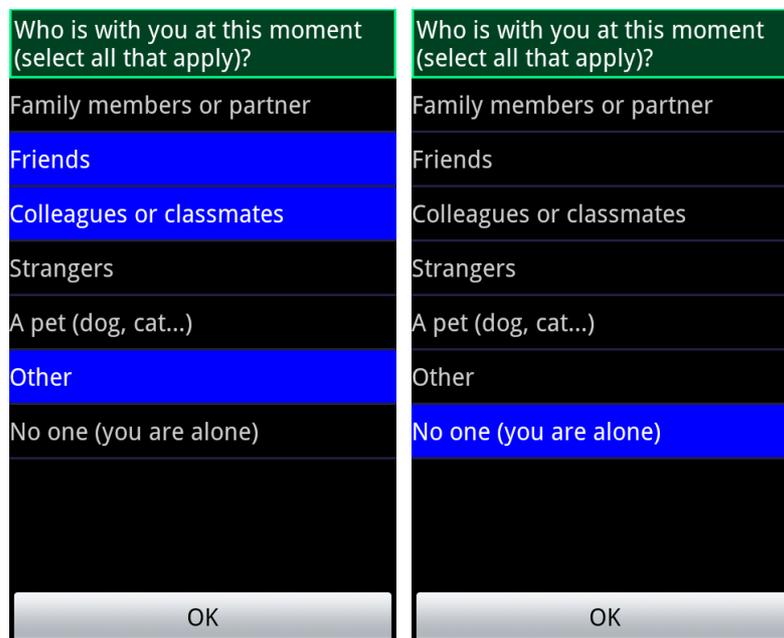
Figure 14. Single-choice question



Multiple-choice questions

Multiple-choice questions have the same appearance as single-choice questions, but differ in their behavior. They allow for multiple responses to be selected, and also feature a particular response that means “nothing” (Figure 15). In a multiple-choice question, every possible response corresponds to a binary item in the final data table (“1” if the response was selected, “0” otherwise). The order of the responses can be randomized. The special response meaning “nothing” does not have to be specified as it is a built-in mechanism of multiple-choice questions, but its text can be customized and pressing it will unselect any response that had been previously selected. This special response can also be disabled, forcing participants to choose at least one response.

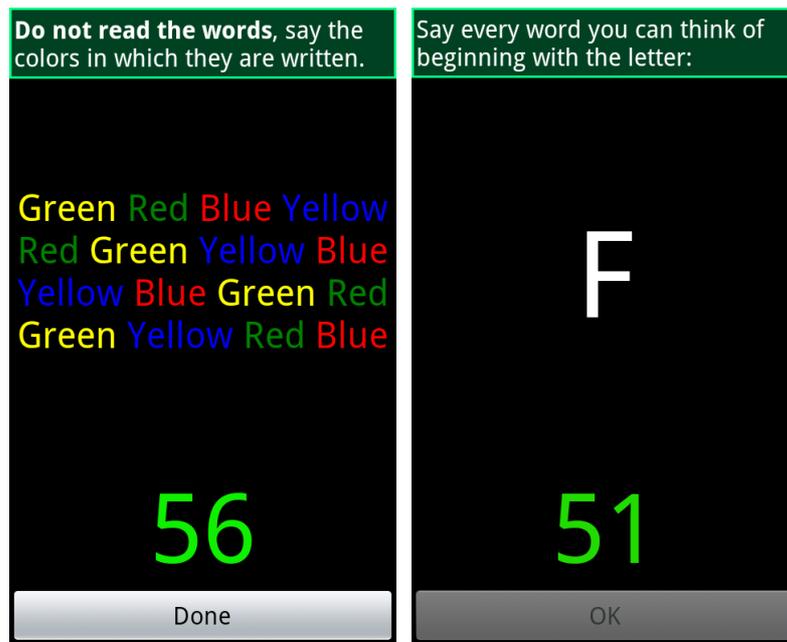
Figure 15. Multiple-choice questions



Recording tasks

Recording tasks can display instructions along with a visual content, and record the participant voice for a predetermined amount of time, using the device microphone (Figure 16). An option allows to set whether participants are able to interrupt the recording. When the recording is finished, another option controls whether a customizable message should be displayed, or if the next question should be asked right away. The recording is stored in a 3GPP formatted file, and the file name is stored in the final data table.

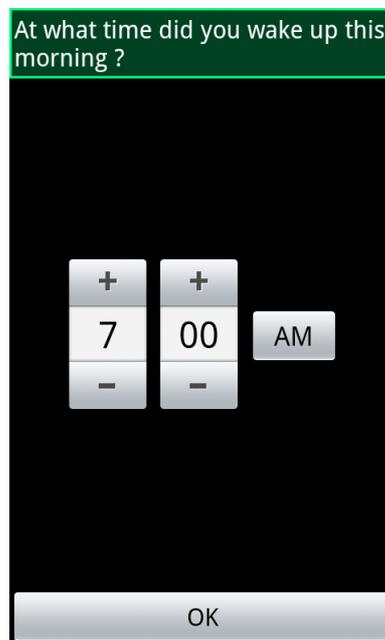
Figure 16. Recording tasks



Time questions

Time questions are useful when asking participants about a time: instead of displaying a long list of times, this question type takes advantage of the built-in “TimePicker” elements of Android, allowing participants to conveniently input a given time by pressing arrows (Figure 17). A minimum and a maximum time can optionally be defined.

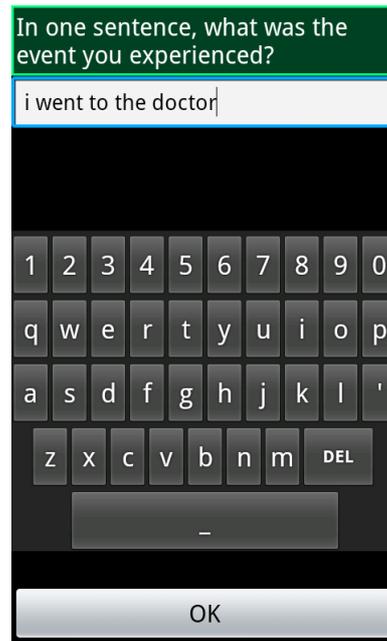
Figure 17. Time question



Text questions

Text questions allow participants to enter text through a minimalist software keyboard (Figure 18). Currently, only the QWERTY keyboard layout is available.

Figure 18. Text question



Custom questions

If a particular task is not available in Samplex, custom questions allow for the running of an external application by specifying its package name. However, this process is still experimental and lacks several features. When an external application is activated, Samplex will be unable to control it to detect a timeout, store results, or auto-complete the task: all these aspects are currently the external application's responsibility.

3.5.3.9. Conditions

Questionnaires and questions can use conditions to check whether they should be asked, and always return either "true" or "false". Boolean logic conditions include the

traditional “and”, “or” and “not” conditions, which are respectively true if all sub-conditions are true, if at least one sub-condition is true, or if the sub-condition is false. Other types of conditions allow for checking if a variable equals a given value, if a particular question was asked, or if the result to a particular question matches a certain value.

3.6. Summary

The Samplex application developed through this thesis provides considerably more flexibility than previous solutions, allows for dynamic surveys, adapts to individual participants with different life styles while being resilient to misuse and battery failures. For researchers it is also easy to deploy, maintain and use and provides convenient output that allows to focus on analysis.

Chapter 4. Feasibility and validity of EMA cognitive testing in the elderly

As previously mentioned, the second major objective of this thesis was to apply the technical advances described in Chapter 3 (Samplex) to the study of cognitive decline among the elderly. This second objective required an extensive clinical investigation of the feasibility and validity of these novel EMA tools, described in detail below.

4.1. Methodology

4.1.1. Sample

The sample was drawn from the AMI cohort, a population-based cohort of 1,002 French elderly agricultural workers in rural areas. The AMI cohort randomly selected participants aged 65 or older from the Farmers' Social Security Registry. The sampling strategy and methodology has been described in great detail elsewhere (Pérès et al., 2012). Study procedures were approved by the regional ethical review board and all participants provided written informed consent. Following baseline inclusion in the cohort from 2007 to 2009, participants were administered a battery of neuropsychological tests (see details below) and, in a sub-sample of the AMI cohort, participants completed a neuroimaging examination. The battery of neuropsychological tests and the neuroimaging exam were administered approximately every two years. The first follow-up assessment of the imaging subgroup occurred between 2009 and 2012. During the second follow-up examination (between 2012 and 2014), a random subsample of these participants was also invited to participate in a week-long ecological momentary assessment (EMA) study using an Android tablet (Samsung

Galaxy Wifi 4.2” with a 10.6 cm screen, default font size set to 12 points) programmed to administer electronic assessments through Samplex. Inclusion criteria for the present study were: having a sixth grade reading level, no visual or motor impairments, no significant cognitive impairment or diagnosis of dementia (based on a neuropsychological evaluation by a psychologist, a clinical examination by a geriatrician and confirmation in a case consensus conference by three dementia specialists) and participation in the ancillary neuroimaging project. Among the 316 subjects that received brain MRI, a total of 172 participants were contacted for participation in the present EMA study (44% female).

4.1.2. Procedure

The EMA electronic interview questioned participants regarding their physical environment, social interactions, and specific behaviors derived from previous EMA research (Johnson et al., 2009a). Table 8 presents a description of the survey organization.

Table 8. Survey organization

Triggering signal	Nature of the questions
First signal	Sleep quantity, quality and problems
Every signal	At the moment of interview: location, social context, level of happiness
Every signal	For the past hours: substance use, activities, difficulties remembering names

A brief cognitive test was also administered at the end of each electronic interview to assess semantic memory and verbal fluency, episodic memory or executive functioning. These cognitive functions were selected given their frequent decline among elderly individuals (Buczyłowska and Petermann, 2016; Harada et al., 2013) and the corresponding mobile cognitive tests were developed specifically for self-administration by the participant (see below). In order to avoid biases associated with time of day or test repetition, each test

was administered five times during the week in an order that was counterbalanced across the day and unique item content was developed for the five versions of each test (see details below). All functions of the device were deactivated so that the study smartphone could only be used for the purposes of the study. All participants were trained in how to operate the device and two practice assessments were administered, one guided by the research staff and one completed independently by the participant. Participants experiencing difficulty in understanding or completing the interviews were provided additional training. The general EMA electronic interviews occurred five times a day during seven consecutive days and were administered at fixed intervals, randomized across individuals. The first and last interviews of the day were adjusted so as not to disturb the typical sleep schedule of the participants. Cognitive assessments in EMA used the voice-recording function of the smartphone (for semantic memory, free-recall for episodic memory, and for executive functioning), or through manual selection on the smartphone screen (for episodic memory recognition). Recorded verbal responses were coded by trained members of the research team. Inter-rater reliability was computed based on Pearson correlations for the subsample of the 30 participants who completed the highest amount of cognitive tests, and ranged from .90 to .98 for all tests.

4.1.3. AMI neuropsychological instruments

4.1.3.1. Semantic memory

The Isaacs Set Test (Isaacs and Kennie, 1973) was used to assess semantic memory through the presentation of one semantic category, such as “Colors” or “Cities”. Participants were asked to list as many words belonging to a given category as possible in a 60-second span. The number of correct answers was used to score the test.

4.1.3.2. Episodic memory

Participants were administered the Free and Cued Selective Reminding Test (Grober and Buschke, 1987). This test starts with a study phase where four sheets of paper each displaying four words were successively presented. The 16 words belong to 16 distinct semantic categories. For each list, the participant is asked to match the category with the corresponding item (e.g. “Among these words, can you tell me which one is a fish?”). Participants are then tested for immediate recall: each category is presented and they are asked to give the corresponding item; when they fail to remember, they are given the correct answer. Once the study phase has been completed, the recall phase begins using free recall followed by cued recall for the items that were not remembered. Three free and cued trials are successively completed, separated by an interference task (counting backwards). After a 20-minute delay, a delayed free and cued recall trial is administered. The sum of correct answers for the three recall tests and the number of correct answers for the delayed recall period were both used to score the test.

4.1.3.3. Executive functioning

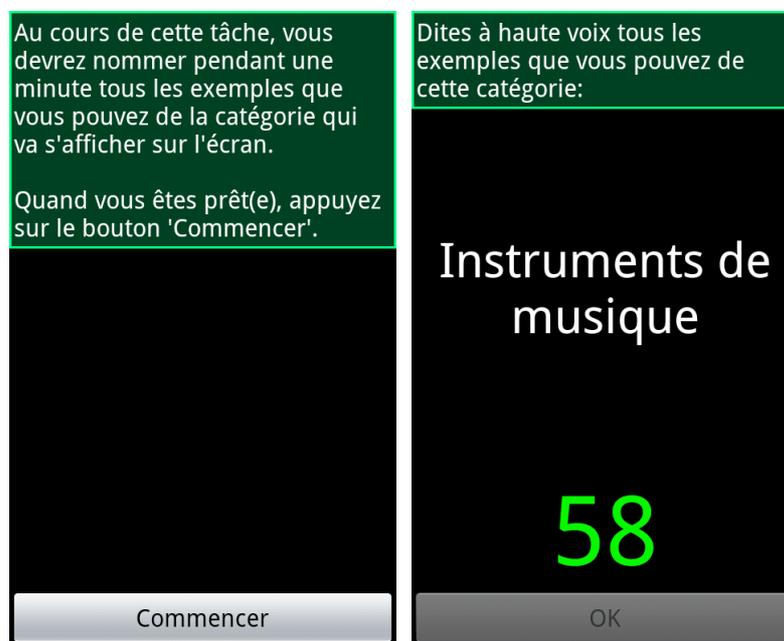
The digit-symbol test from the WAIS-IV (Wechsler et al., 2008) was used to assess executive functioning. In this test, participants are presented a list of nine symbols paired with digits from 1 to 9. Participants are given a matrix of digits which they have to complete with the corresponding symbol in a 90-second span. The number of correct symbols was used to score the test.

4.1.4. EMA cognitive assessments

4.1.4.1. Semantic memory: verbal fluency

A mobile test of semantic verbal fluency was developed to present participants with a semantic category (namely “Musical instruments”, “Furniture”, “Tools”, “Vegetables” and “Vehicles”). Participants are then asked to say as many words as possible belonging to this category within 60 seconds. Verbal responses were recorded on the smartphone and later coded by research staff. The total number of correct words was used as the primary performance score. Additionally, the number of errors, repeats and superordinate answers (answers that are categories, e.g. “Plants” when the correct answer was “Vegetables”) were assessed. Screenshots of this test are presented in Figure 19.

Figure 19. Screenshots of the verbal fluency task



4.1.4.2. Semantic memory: category generation

A mobile test was developed to assess the ability to recognize the category that several words have in common. A pair of words belonging to the same semantic category was presented (see details in Table 9). Participants were asked to name their category within 30 seconds. The responses were recorded verbally and later coded by the research staff. For this test, participants were allowed to stop the recording once they answered. The primary score for this test was 0 for errors, 1 for a superordinate category, and 2 for a correct answer. The time required to answer was also coded as a sub-score. Screenshots of this test are presented in Figure 20.

Table 9. Pairs of words used in the mobile category generation test

Words displayed	Expected answer
Pomme, Orange	Fruits
Oreille, Yeux	Parties du visage
Nord, Ouest	Points cardinaux
Pantalon, Chemise	Vêtements
Chien, Oiseau	Animaux

Figure 20. Screenshots of the category generation test



4.1.4.3. Episodic memory: immediate and delayed free-recall and recognition

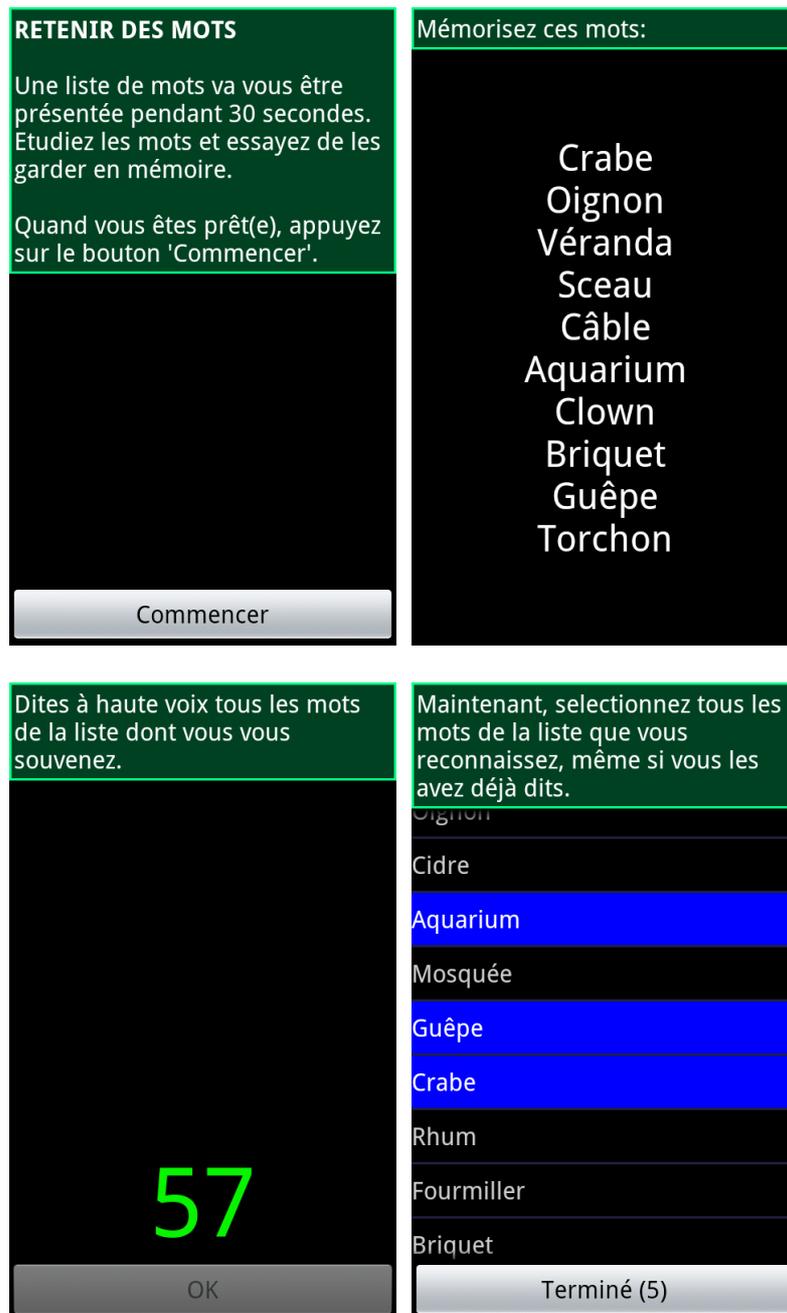
A mobile list-learning test was developed to assess immediate free recall and recognition as well as delayed recall and recognition. Lists of words representing material objects were selected from the BRULEX database (Content et al., 1990), and that had a frequency of appearance between 200 and 1,000 words among one million French words appearing in written text (the detailed list is presented in Table 10). A 10-word list was displayed for 30 seconds, after which participants were immediately asked to freely recall all words within one minute. Following the free-recall task, participants were presented a list of 20 words which included the 10 words in the previously-presented list as well as 10 additional words with matched frequency of appearance in the French language. Participants were then asked to identify the words they recognized as belonging to the original list by selecting each word directly on the device. The number of correct words for both the free-recall task (recorded verbally) and the recognition task (recorded by selection on the EMA device

screen) served as the primary scores for immediate episodic memory performance. At the following EMA assessment, approximately three hours later, participants were asked to complete the same tasks (free-recall and recognition) without the list being displayed. The number of correct words for both tasks served as the primary scores for delayed episodic memory performance. The number of errors, and repeats during free recall served as sub-scores. Screenshots of this test are presented in Figure 21.

Table 10. List of the words used in EMA episodic memory assessments

Test 1	Test 2	Test 3	Test 4	Test 5
Agenda	Stylo	Crabe	Pyramide	Piment
Yacht	Venin	Oignon	Sandwich	Camionnette
Dindon	Moutarde	Véranda	Casino	Tomate
Ouragan	Bagnole	Sceau	Chaussette	Guichet
Pâté	Chauve-souris	Câble	Orge	Glycine
Sardine	Tornade	Aquarium	Sirop	Peigne
Boisson	Peluche	Clown	Muséum	Citron
Crocodile	Canapé	Briquet	Hareng	Gratte-ciel
Phoque	Ecureuil	Guêpe	Grenouille	Méduse
Scie	Gazelle	Torchon	Andouille	Pétard

Figure 21. Screenshots of the episodic memory test

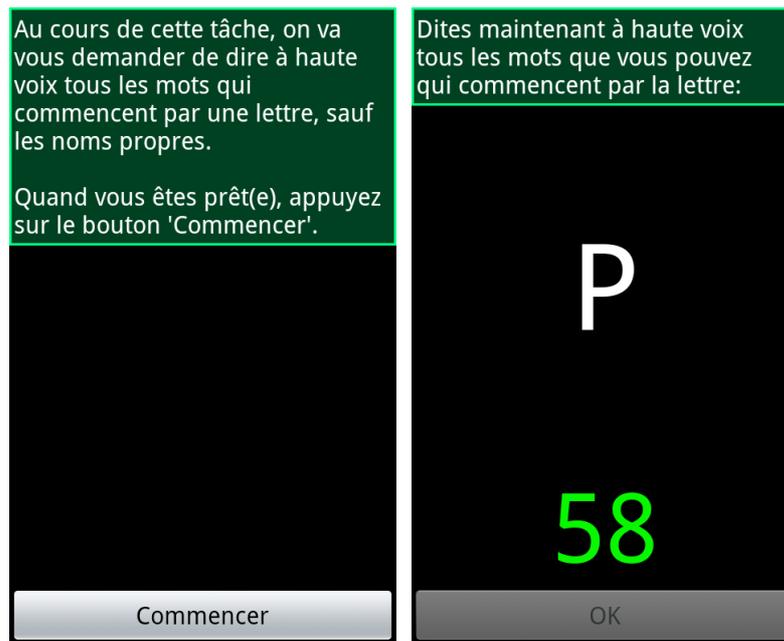


4.1.4.4. Executive functions: phonemic fluency

A mobile phonemic fluency test was used to assess executive functions. Participants were presented a letter and asked to name as many words beginning with that letter as possible within a one-minute span and without naming proper nouns. We avoided “difficult” letters, that is, words for which too few start with that letter in the French language. The

letters used for this test included “P”, “F”, “T”, “M” and “S”. Verbal responses were recorded by the EMA device and the total number of correct words served as the primary performance score. The number of errors and repetitions were also assessed. Screenshots of this test are presented in Figure 22.

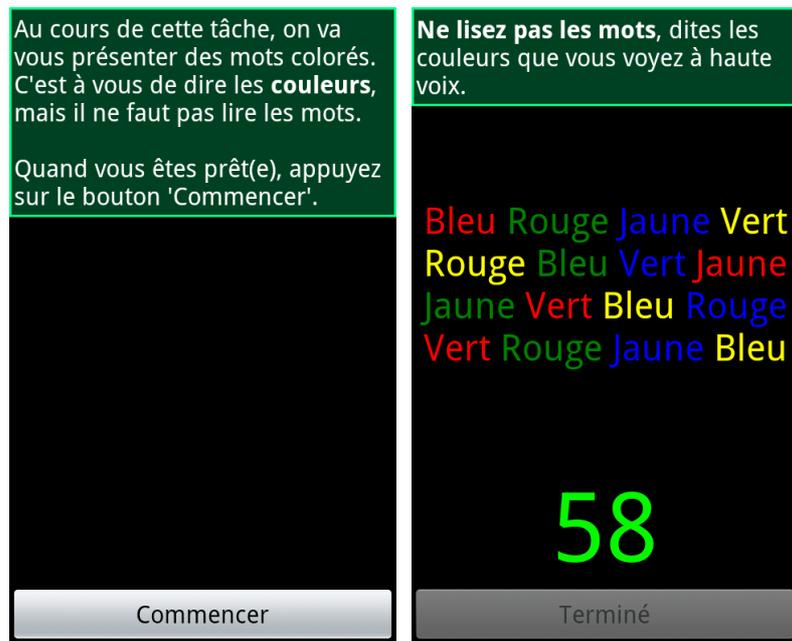
Figure 22. Screenshots of the phonemic fluency test



4.1.4.5. Executive functions: color naming

A mobile color naming task was developed, inspired by the interference phase of the Stroop test. In this task, 16 color names (namely “Blue”, “Yellow”, “Green” and “Red”) are displayed to the participant in a 4 x 4 grid. However, every word is presented in another color than that which the word designates. No word-color pair appears twice consecutively. Participants are asked to name the color in which words are displayed (recorded verbally), and reminded not to read the words themselves, within 60 seconds (participants are allowed to stop the recording when the task is finished). The number of correct answers served as the primary score for this test. Sub-scores were assessed for errors, skipped words and the total duration of the naming task. Screenshots of this test are presented in Figure 23.

Figure 23. Screenshots of the color naming task



4.1.5. Data analysis

EMA data were analyzed using Hierarchical Linear and Nonlinear Modeling (HLM) Version 7 (Raudenbush, 2004). Analyses of reactivity to repeated testing methodology were performed in order to identify significant changes in the frequency or intensity of cognitive variables as a function of study duration. Convergent validity was examined by using neuropsychological test scores acquired at the second AMI follow-up assessment as predictors of mobile cognitive performance assessed during the subsequent week. Analyses of reactivity and convergent validity were performed with participants meeting minimum compliance with the EMA methodology, defined as having completed at least one-third of programmed assessments (Johnson et al., 2009b). Means-as-outcomes models were used for continuous outcome variables and Bernoulli models for dichotomous outcomes.

4.2. Results

4.2.1. Acceptance and compliance

Overall, 114 (66%) of the 172 eligible participants agreed to take part in the EMA portion of the study. Individuals who refused participation were older, $t(170) = 1.971$, $p < 0.05$, and had lower MMSE scores, $t(170) = -4.417$, $p < 0.001$, than those who accepted. In comparison with the overall cohort ($N = 1,002$), the 114 participants enrolled in study were more often female, $X^2(1) = 4.41$, $p < 0.05$, younger, $t(1,000) = 7.96$, $p < 0.001$, and had higher global MMSE scores, $t(984) = -6.83$, $p < 0.001$. Seventy-five of these participants completed at least one third of programmed assessments and were considered as minimally compliant with the procedures. The socio-demographic characteristics of the sample of participants meeting minimal compliance criteria are presented in Table 11.

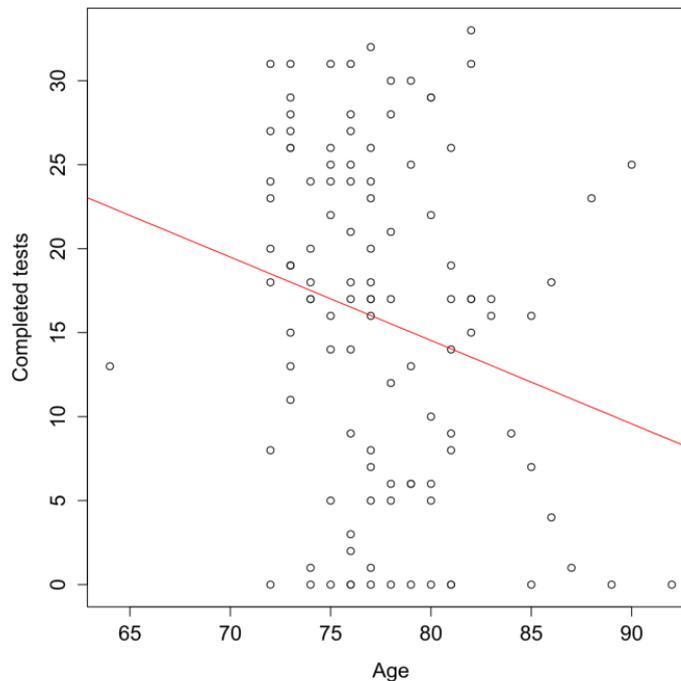
Table 11. Socio-demographic characteristics of the final sample of participants

Age	Mean	76.85
	SD	4.24
Sex	Female	56.8%
Education	Less than elementary school	19.4%
	Elementary school	31.3%
	More than elementary school	49.3%

In comparison with the overall cohort ($N = 1,002$), the 75 participants meeting the minimum compliance criterion were more often female, $X^2(1) = 7.20$, $p < 0.05$, younger, $t(961) = 7.37$, $p < 0.001$, and had higher global MMSE scores, $t(945) = -6.10$, $p < 0.001$. The number of completed mobile cognitive tests decreased with age ($\gamma = -0.4963$, $t = -2.380$,

$p < 0.05$), as illustrated in Figure 24. The cognitive performance scores for the sample of 75 participants meeting minimal compliance criteria are presented in Table 12.

Figure 24. Number of completed cognitive tests by age



Compliance with the self-report EMA interviews was high, with 82% of the possible assessments being completed by participants in the context of their daily lives resulting in 2,158 observations. Examination of compliance with mobile cognitive tests that were recorded verbally by the device indicated that a small portion of participants may have received help from a third-party to complete the assessments, did not understand the instructions, or used ways to circumvent the difficulty (e.g. writing down the list of words instead of memorizing it). When these assessments were considered as void, 1,646 valid observations remained, representing approximately 63% of the potential maximum number of observations.

Table 12. Neuropsychological and mobile cognitive performance of the final sample

	Cognitive function	Test name	Mean	SD
Neuropsychological Assessments	General	MMSE	27.39	2.17
	Semantic memory	IST-60	60.47	14.39
	Episodic memory	Free recall	25.95	6.70
		Cued recall	44.97	3.44
		Delayed free recall	10.45	2.59
		Delayed cued recall	15.37	1.07
	Executive functions	Digit-symbol test	29.30	10.07
Mobile Cognitive Tests	Semantic memory	Verbal fluency	10.02	2.90
		Category generation	1.73	0.63
	Episodic memory	Auto-biographic memory	7.12	0.99
		Free recall	4.92	1.52
		Recognition	7.18	2.18
		Delayed free recall	2.67	2.67
		Delayed recognition	3.06	2.10
	Executive functions	Phonemic fluency	8.86	3.40
		Color naming	15.35	1.42

MMSE (Mini-Mental Status Examination); IST-60 (Isaacs Set Test score at 60 seconds)

4.2.2. Effects associated with duration of ambulatory monitoring

No fatigue effect was observed in that the number of missing observations decreased as a function of study duration, rather than increased. Concerning practice effects, performance on verbal fluency scores, color naming scores, immediate recognition and delayed free-recall improved as a function of study duration. However, performance on the delayed recognition condition of list-learning decreased over time. Scores obtained from the other mobile cognitive tests were not significantly affected by study duration. These results are summarized in Table 13.

Table 13. Fatigue and practice effects observed for mobile cognitive tests

			Coef.	T-ratio	P-value
Fatigue effect			-0.360	-9.872	0.000
Practice effects	Semantic memory	Verbal fluency	0.420	2.851	0.005
		Category generation	0.430	1.425	0.155
	Episodic memory	Immediate free-recall	0.090	1.285	0.200
		Immediate recognition	0.266	2.331	0.021
		Delayed free-recall	0.328	2.339	0.020
		Delayed recognition	-0.958	-10.064	0.000
	Executive functioning	Phonemic fluency	-0.094	-0.592	0.554
		Color naming	0.235	2.870	0.005

4.2.3. Convergent validity of neuropsychological and mobile cognitive tests

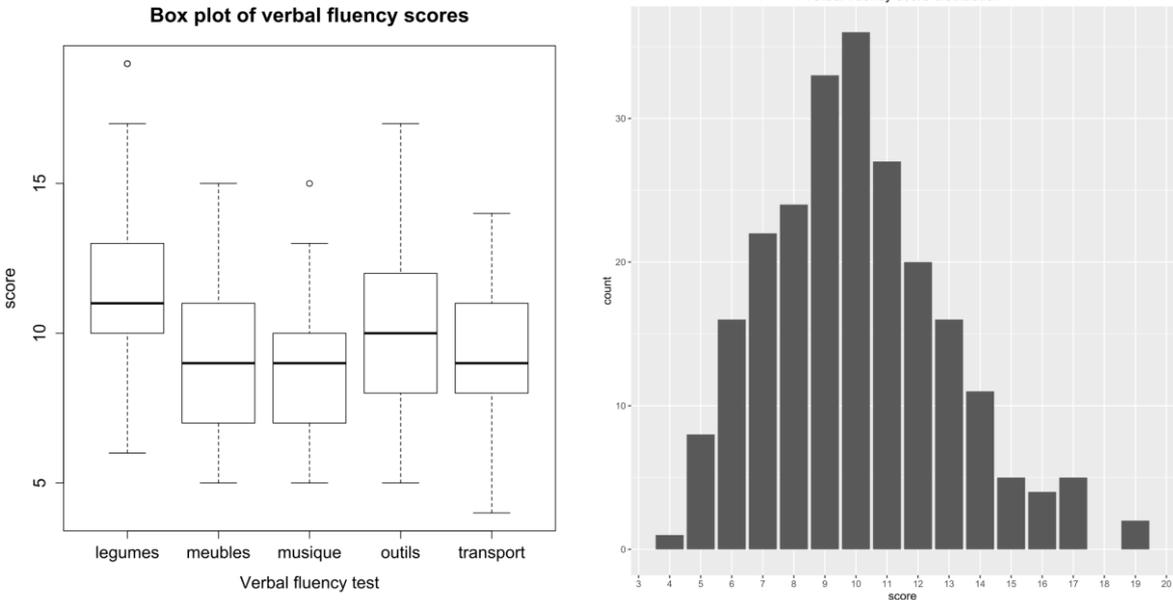
Scores for mobile cognitive tests were often significantly correlated with neuropsychological assessments of similar cognitive functions (see paired comparisons below), including after statistical adjustment for practice effects associated with the number of mobile tests administered. This was true for the semantic verbal fluency test ($\gamma = 0.104$, $t = 7.095$, $p < 0.001$), the phonemic fluency test ($\gamma = 0.162$, $t = 4.576$, $p < 0.001$), the color naming test ($\gamma = 0.071$, $t = 2.900$, $p = 0.006$), the word list free recall ($\gamma = 0.084$, $t = 4.411$, $p < 0.001$), the word list recognition ($\gamma = 0.328$, $t = 3.997$, $p < 0.001$) and the word list delayed recognition ($\gamma = 0.315$, $t = 2.785$, $p = 0.007$).

4.2.3.1. Semantic memory: Verbal fluency

The five variants of this mobile test were not of equal difficulty (e.g., the variant asking for “musical instruments” yielded fewer responses than the other tests, and the variant asking for “vegetables” yielded more responses), however the global scores were normally

distributed, a result also seen in clinical settings. In Figure 25, the left panel presents the box plots of scores for each of the five variants of this test, and the right panel shows the global distribution of these scores.

Figure 25. Distribution of scores for mobile verbal fluency tests



Scores for the mobile verbal fluency test were highly correlated with those of the Isaacs Set Test assessed in clinical settings, including when adjusted for age, sex and practice effects associated with the number of verbal fluency tests already administered (Table 14).

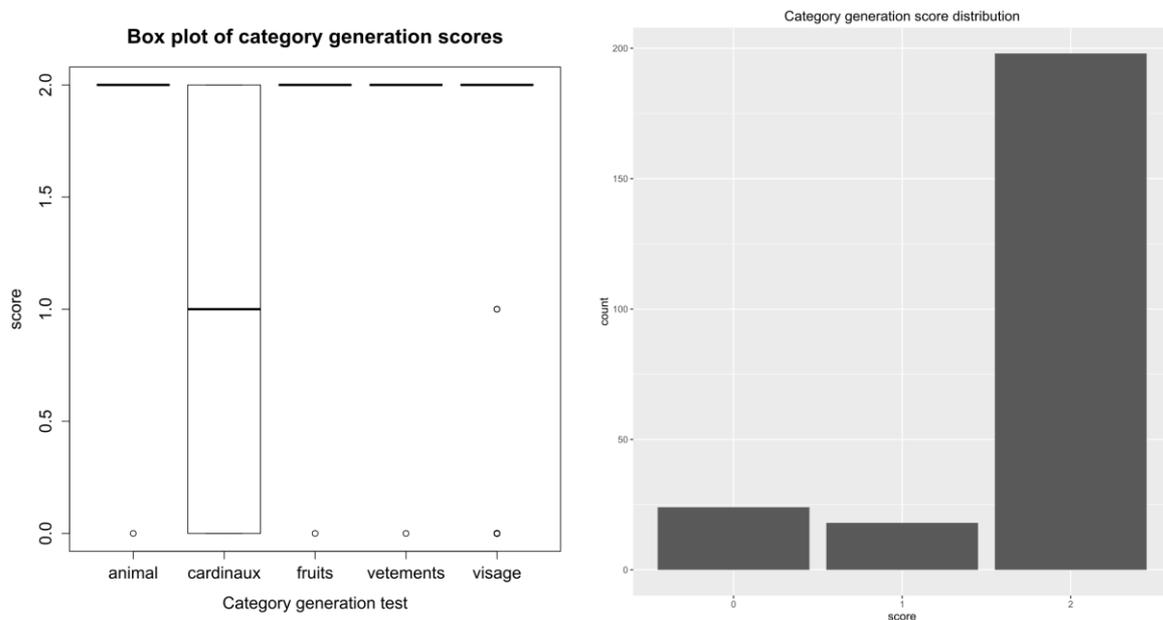
Table 14. Association of Isaacs Set Test scores with mobile verbal fluency scores

Covariates	Coefficient	T-ratio	P-value
no adjustment	0.100	6.451	0.000
test number	0.099	6.136	0.000
age, sex	0.104	7.095	0.000
test number, age, sex	0.102	6.669	0.000

4.2.3.2. Semantic memory: Category generation

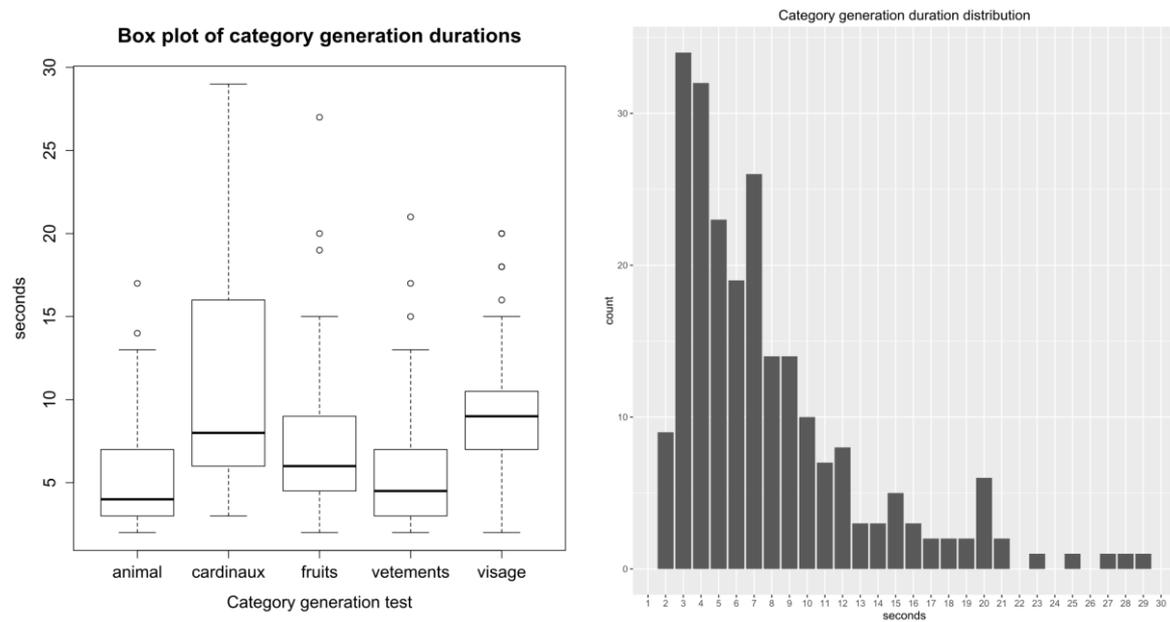
Due to the way this mobile test was scored (zero, one or two points), its distribution was not normal as most trials were likely to result in the maximum score. In Figure 26, the left panel presents the box plots of scores for each of the five variants of this test, and the right panel shows the global distribution of these scores.

Figure 26. Distribution of scores for semantic category generation



When considering the time participants required to complete the task, we observed that two of the variants took longer than the others, namely the “cardinal directions” and “parts of the face” variants, and the global duration of this test followed a positively skewed distribution. In Figure 27, the left panel presents the box plots of task durations for each of the five variants of this test, and the right panel shows the global distribution of these durations.

Figure 27. Distribution of task duration for semantic category generation



While the scores for this test were not correlated with those obtained from the Isaacs Set Test, a significant association was observed for time needed to complete the task (Table 15).

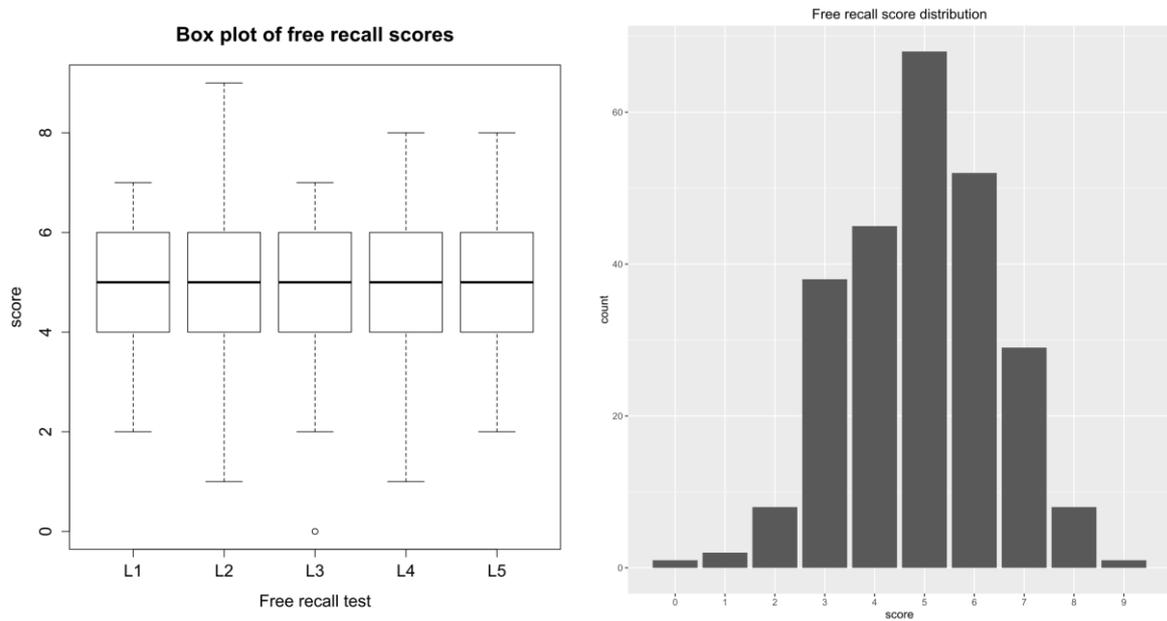
Table 15. Association of Isaacs Set Test scores with semantic category generation duration

Covariates	Coefficient	T-ratio	P-value
no adjustment	-0.108	-2.624	0.011
test number	-0.110	-2.743	0.008
age, sex	-0.098	-2.473	0.016
test number, age, sex	-0.910	-2.590	0.012

4.2.3.3. Episodic memory: Free recall

For the mobile free recall test, all variants showed similar difficulty, and the distribution of scores was normal. In Figure 28, the left panel presents the box plots of scores for each of the five variants of this test, and the right panel shows the global distribution of these scores.

Figure 28. Distribution of episodic memory free recall scores



The free recall scores were significantly correlated with those of the Grober and Buschke free recall test (Table 16).

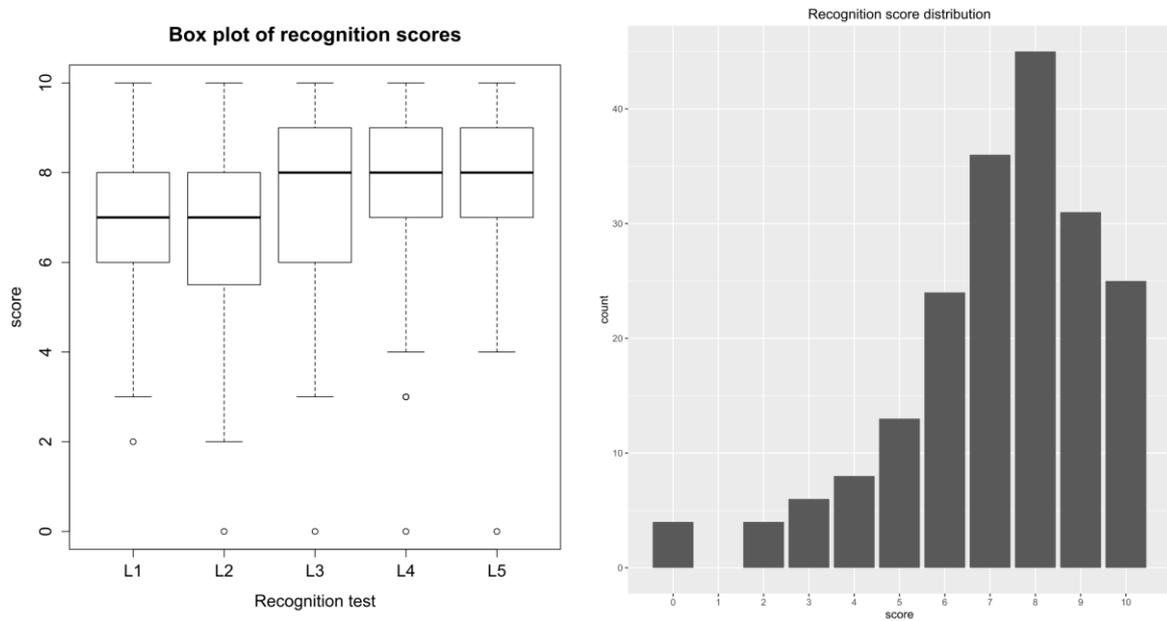
Table 16. Association of Grober and Buschke scores with episodic memory free recall

Covariates	Coefficient	T-ratio	P-value
no adjustment	0.085	4.351	0.000
test number	0.084	4.411	0.000
age, sex	0.077	3.703	0.000
test number, age, sex	0.077	3.758	0.000

4.2.3.4. Episodic memory: Recognition

All variants of the recognition tests showed somewhat similar difficulty, and the scores followed a negatively skewed distribution. In Figure 29, the left panel presents the box plots of scores for each of the five variants of this test, and the right panel shows the global distribution of these scores.

Figure 29. Distribution of episodic memory recognition scores



These scores were also significantly correlated with those of the Grober and Buschke cued recall test assessed in clinical settings (Table 17).

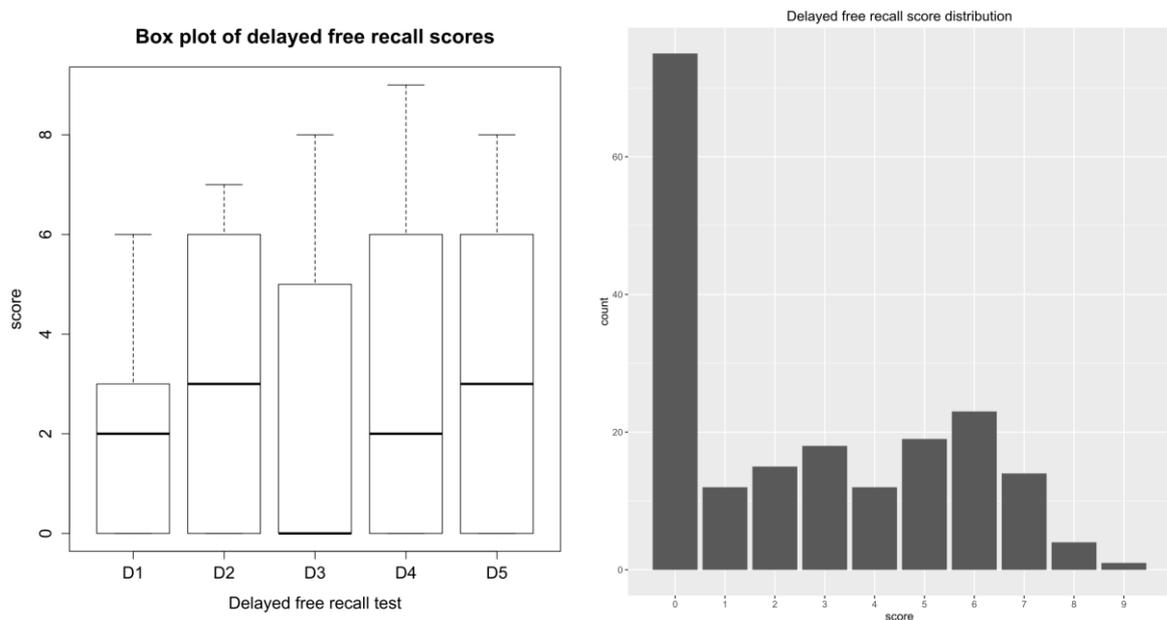
Table 17. Association of Grober and Buschke scores with episodic memory recognition

Covariates	Coefficient	T-ratio	P-value
no adjustment	0.336	3.811	0.000
test number	0.328	3.997	0.000
age, sex	0.332	3.724	0.000
test number, age, sex	0.323	3.892	0.000

4.2.3.5. Episodic memory: Delayed free recall

The variants of the delayed free recall test showed somewhat similar difficulty, and the scores approximated an exponential distribution. In Figure 30, the left panel presents the box plots of scores for each of the five variants of this test, and the right panel shows the global distribution of these scores.

Figure 30. Distribution of episodic memory delayed free recall



These scores were not correlated with those of the Grober and Buschke delayed free recall test (Table 18).

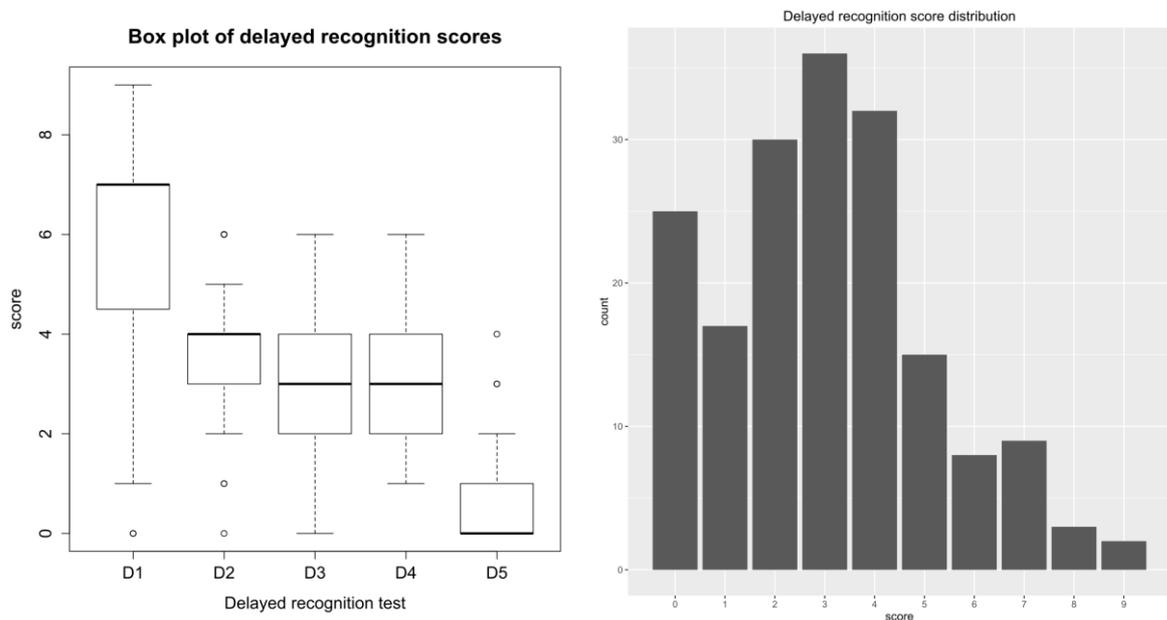
Table 18. Association of Grober and Buschke scores with episodic memory delayed free recall

Covariates	Coefficient	T-ratio	P-value
no adjustment	0.157	1.745	0.086
test number	0.154	1.767	0.083
age, sex	0.118	1.303	0.198
test number, age, sex	0.113	1.288	0.203

4.2.3.6. Episodic memory: Delayed recognition

Variants of the delayed recognition tests showed dissimilar difficulty, and the scores approximated a positively skewed distribution. In Figure 31, the left panel presents the box plots of scores for each of the five variants of this test, and the right panel shows the global distribution of these scores.

Figure 31. Distribution of delayed recognition scores



These scores were correlated with those of the Grober and Buschke delayed cued recall test when adjusted for test number, age and sex (Table 19).

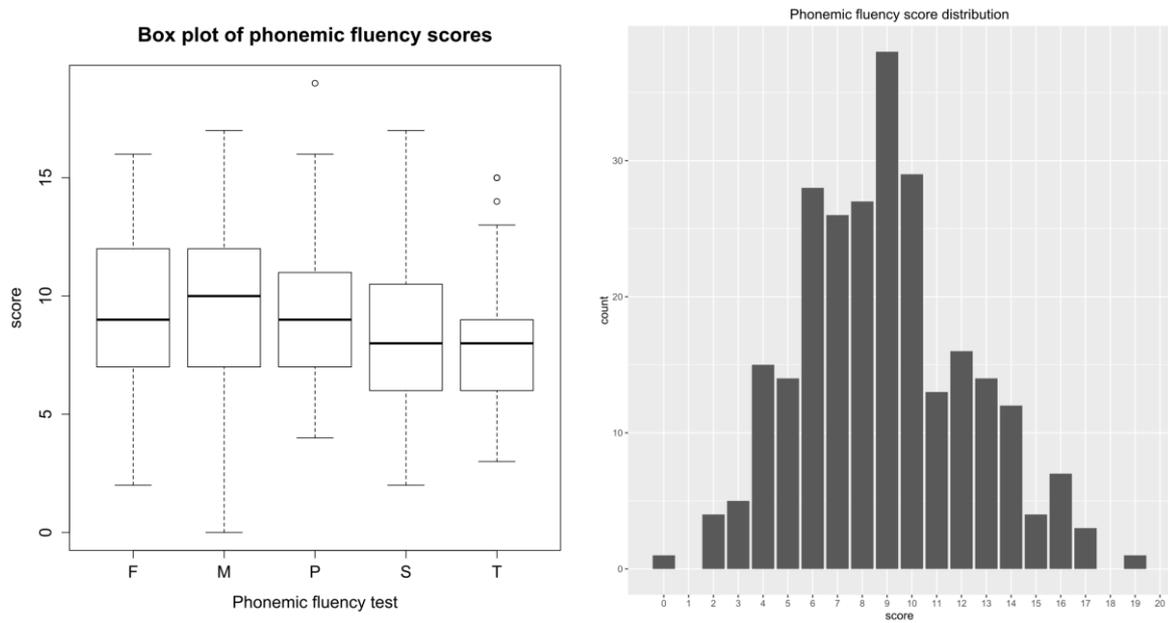
Table 19. Association of Grober and Buschke scores with episodic memory delayed recognition

Covariates	Coefficient	T-ratio	P-value
no adjustment	0.241	1.881	0.065
test number	0.315	2.785	0.007
age, sex	0.257	2.106	0.040
test number, age, sex	0.332	2.781	0.007

4.2.3.7. Executive functioning: Phonemic fluency

All variants of this mobile test showed a similar difficulty, and the scores were normally distributed. In Figure 32, the left panel presents the box plots of scores for each of the five variants of this test, and the right panel shows the global distribution of these scores.

Figure 32. Distribution of phonemic fluency scores



These scores were significantly correlated with the scores of the Wechsler coding test (Table 20).

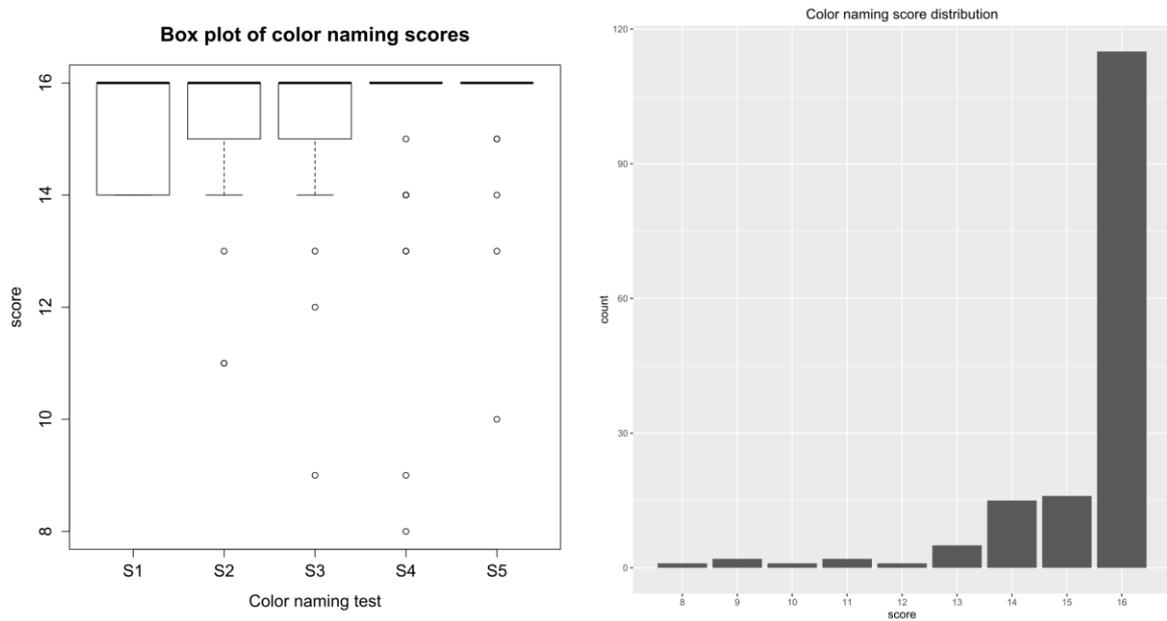
Table 20. Association of Wechsler coding test scores with phonemic fluency scores

Covariates	Coefficient	T-ratio	P-value
no adjustment	0.144	4.283	0.000
test number	0.145	4.308	0.000
age, sex	0.161	4.571	0.000
test number, age, sex	0.162	4.576	0.000

4.2.3.8. Executive functioning: Color naming test

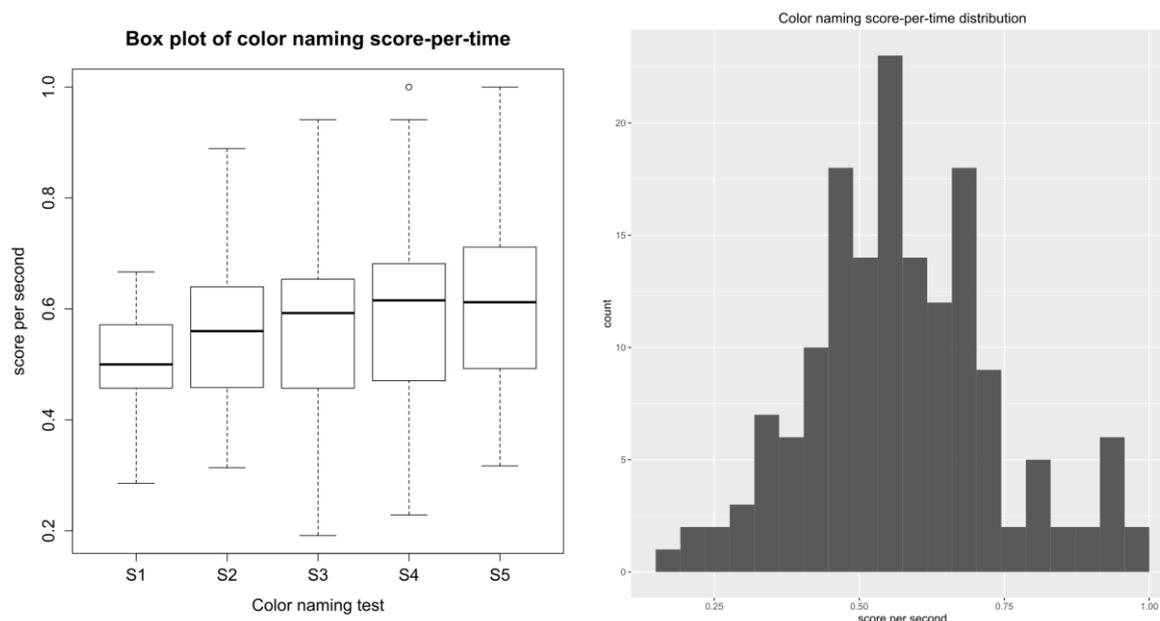
The scores of this mobile test were exponentially distributed, possibly indicating that the task was too easy, and all variants showed a somewhat similar difficulty. In Figure 33, the left panel presents the box plots of scores for each of the five variants of this test, and the right panel shows the global distribution of these scores.

Figure 33. Distribution of color naming test scores



However, when observing the ratio of the score divided by the time required to complete the task, we found that the distribution was normal. In Figure 34, the left panel presents the box plots of the score-per-time ratio for each of the five variants of this test, and the right panel shows the global distribution of these ratios.

Figure 34. Distribution of color naming test score-per-time ratio



The scores of the color naming task were correlated with those of the Wechsler coding test (Table 21). Similarly, the time required to complete the task was significantly correlated with the scores of the Wechsler coding test (Table 22).

Table 21. Association of Wechsler coding test scores with color naming test scores

Covariates	Coefficient	T-ratio	P-value
no adjustment	0.073	2.634	0.011
test number	0.074	2.735	0.009
age, sex	0.068	2.752	0.008
test number, age, sex	0.071	2.900	0.006

Table 22. Association of Wechsler coding test scores with color naming test time

Covariates	Coefficient	T-ratio	P-value
no adjustment	-0.390	-3.846	0.000
test number	-0.386	-3.697	0.001
age, sex	-0.366	-3.381	0.001
test number, age, sex	-0.375	-3.210	0.002

Chapter 5. Discussion of feasibility and validity of EMA cognitive testing in the elderly

Mobile technologies have been applied in clinical research for nearly three decades and its use is rapidly expanding among elderly populations (Cain et al., 2009). Very recent findings have confirmed the value of mobile cognitive testing in this population as a means of reducing the margin of error associated with scores derived from traditional neuropsychological assessments, thus providing more sensitive tools for the detection of corresponding brain markers as well as for the identification of the daily life activities that may improve cognitive functioning (Allard et al., 2014). Despite this progress, the feasibility and validity of mobile cognitive testing has been examined by very few investigations (Brouillette et al., 2013; Oliveira et al., 2014; Scholey et al., 2012; Schuster et al., 2015; Timmers et al., 2014) and no study to date has estimated fatigue or reactive effects in the elderly. In a cohort of elderly community residents, the present investigation examined the acceptability and feasibility of mobile cognitive testing through EMA, as well as compliance with its multiple daily assessments, potential biases associated with repeated testing, and its convergent validity with traditional neuropsychological assessments.

Concerning the basic feasibility of mobile cognitive testing in the elderly, the findings indicate that the most significant barrier to its use is likely to be initial acceptance and compliance with study procedures. In this sample of non-demented elderly persons, we found that only a moderate majority of participants (66%) accepted participation in the mobile assessments. Those who refused to participate were older and had lower scores of general cognitive functioning. These acceptance rates are lower than those found in either psychiatric

or neurologic samples when EMA is used without cognitive testing (Johnson et al., 2009b, 2009a), but may be lower due to the acceptance by this sub-sample of additional assessments involved in the larger research protocol (including MRI). It is also important to note that approximately one-third (34%) of individuals who accepted to participate did not respond to enough EMA assessments to be considered as minimally compliant with the study procedures. It is possible that these compliance rates could have been improved if specific procedures were implemented to encourage participation (such as financial compensation). Taken together, however, these acceptance and compliance rates indicate that the use of mobile cognitive assessments in the elderly may be limited to a subpopulation of higher-functioning individuals, though such samples may still include individuals with very mild cognitive deficits.

While mobile cognitive testing may be feasible in the majority of elderly individuals, evaluation of participant compliance with the cognitive assessments requires attention to essential limitations associated with the fact that no clinician was present to administer the tests or to interpret the context in which the test was completed. In particular, a minority of the verbal responses recorded by participants provided potential evidence of unauthorized assistance (such as a spouse who reminds the participant of the words on a list-learning task). When these items were excluded from analyses, the sample nonetheless responded to 63% of all programmed cognitive tests. This response rate is lower than the overall response rate for other EMA studies (Cain et al., 2009; Johnson et al., 2009b), however it represents a more rigorous selection of valid responses.

The validity of the data collected was also examined for potential biases associated with the repeated-testing methodology and for expected associations between mobile cognitive performance scores with traditional neuropsychological tests measuring similar cognitive functions. No evidence was found for a fatigue effect whereby individuals may

increasingly miss or skip assessments as a function of time in the study, but rather participants became even more compliant over time in completing the mobile cognitive tests. However, the scores on some tests were found to vary as a function of the number of tests administered and therefore indicate potential practice effects. As the actual item content of each test was unique to each administration, these effects are likely to be attributable to the participant's learning of how to respond rather than memorizing previously-presented content. In any case, such practice effects can be readily controlled by researchers through adjusting in statistical analyses for the number of times in which a particular test had been administered to participants. Finally, the significant correlations generally observed with traditional neuropsychological tests of semantic memory, episodic memory and executive functions provide support for the convergent validity of the mobile cognitive assessments developed for this study. In cases where no association was observed (category generation for semantic memory), it appears that a ceiling effect was observed in that most participants scored very highly on this test. This lack of sufficient variance may have therefore prevented detection of convergent validity. Nonetheless the time required to complete this test revealed a significant association with its corresponding neuropsychological test.

The legitimate enthusiasm for mobile cognitive assessments in the elderly should be balanced with knowledge of its limitations and comparative value relative to other instruments and measures of cognitive functioning. The present study provides support for the feasibility and validity of these novel assessment tools, though the findings should be interpreted in light of several methodological issues. Perhaps most importantly, the present sample of elderly individuals is characterized by a relatively low education level that, along with other aspects of study procedures (such as the inclusion of MRI), may have affected study acceptance or compliance rates. The development of a wider range of mobile tests would also allow for more comprehensive assessments of cognitive functions and their

examination relative to a range of socio-demographic characteristics should reinforce knowledge for the value of mobile testing. It is also important to note that mobile cognitive testing generates data that may be qualitatively different from that obtained from traditional neuropsychological testing, and therefore should be considered only as complementary information to such clinical information. Despite these limitations, the challenge of detecting subtle cognitive decline at the earliest stages of dementia requires the development of new methods and more sensitive assessment tools. Mobile technologies are an important contribution to this effort. In summary, while their use may not be feasible in all elderly individuals, the resulting data may nonetheless provide more broadly generalizable data for the identification of daily life activities associated with improved cognitive health (Allard et al., 2014) as well as for the examination of clinical markers associated with initial cognitive decline.

Chapter 6. Global discussion and conclusion

The present thesis may be considered atypical in the sense that it combined two distinct objectives: that of fundamental technical EMA development with that of the clinical testing of these new applications. As these goals were consecutive in nature, the time needed to complete this thesis was also atypical, nonetheless its findings may be considered of interest for several reasons.

From a technical point of view, the development of a new EMA platform has far-reaching implications for researchers using EMA or for those interested in brain-based disorders, as the scientific literature has underscored the need for new tools in order to examine enduring questions relative to the etiology, course and treatment of these disorders. Despite the conceptual challenges involved in its development, the resulting EMA platform known as “Simplex” nonetheless succeeded in its objectives of providing a means of collecting both standard EMA information (concerning behaviors, psychological states and symptoms) as well as objective data concerning cognitive performance in real-life circumstances. The juxtaposition of both sources of information provides a particularly powerful means for understanding the dynamic relation of daily life behaviors and experiences with cognitive functioning, as demonstrated by the recent study by Allard and colleagues (Allard et al., 2014). Using multilevel modeling techniques such as those provided by HLM, prospective analyses can be performed over periods of several hours that provide missing information concerning the directionality of correlated daily-life variables. A major purpose of this thesis was to provide the technical means through which cognitive functioning may be included among these daily life variables and to advance research in the field through methodological and technical development.

Concerning clinical applications of Samplex, my thesis work focused only on its application to elderly samples. The previous discussion of the results of this clinical investigation demonstrated the general feasibility and validity of EMA cognitive testing in this population, and highlighted specific caveats that should be considered by researchers wishing to apply these new tools of investigation. However, over and above its use in the elderly, the application of mobile cognitive testing will likely provide new insights into a wide range of mental and neurological disorders. Indeed, the Samplex program and its mobile cognitive tests have already been integrated into additional research projects conducted within our CNRS/EPHE laboratory. While a full discussion of these projects is beyond the scope of this thesis, one illustration from the field of addiction may be useful for the reader to better understand the expansion of these tools to other populations. For example, it has been well-documented that patients with different forms of addiction (e.g. alcohol, opiates, cannabis) perform poorly on neuropsychological tests of executive functioning as compared to healthy controls (Bechara, 2001; Bechara and Damasio, 2002; Bechara et al., 2002; Dom et al., 2006; Ernst et al., 2003; Grant et al., 2000; Hanson et al., 2008; Whitlow et al., 2004). However, these studies are incapable of verifying if executive deficits, such as those concerning decision-making, are the consequences of chronic substance abuse or if they may potentially contribute to the elevated rate of relapse seen in this population. Again, a primary reason why such information is unavailable concerns the methodological constraints that characterize traditional clinical research (as described in Chapter 2). Most research paradigms characterize neuropsychological deficits more or less as stable “traits”, however the reality is that they are highly variable (often fluctuating over periods spanning a few minutes to a few hours). For this reason, the use of EMA mobile cognitive assessments of executive functioning among individuals with addiction may provide key information concerning the potential causal relationships between moments of poor executive functioning and resistance to craving or

substance use. I have participated in the adaptation of Samplex for precisely this purpose, through a large ANR investigation known as ‘MobiCog’ that will complete full data collection this December. Additional projects using Samplex and mobile cognitive testing are also underway to examine other questions involving diverse clinical populations.

In addition to the many benefits of mobile cognitive testing, it is also important to appreciate their limitations that distinguish them from traditional neuropsychological instruments. First, at no point in this thesis was the phrase “mobile neuropsychological testing” used. This intentional caution in the use of vocabulary is necessary given the important differences between both types of testing. As briefly noted in my discussion of the findings of the clinical study, neuropsychological testing is completed by technicians or expert clinicians that are trained in the interpretation of test performance results elucidating the individual’s physical and psychological status at the moment of testing, as well as motivation in performing the test itself. Such information is essential for clinical diagnosis and it cannot be achieved through isolated self-administered mobile cognitive tests. For this reason, the latter can never replace the former, and the information provided by mobile tests must be considered a different source of data concerning cognitive functioning. This also applies to the nature of testing contexts, where neuropsychological tests are administered in quiet and distraction-free environments and after establishing a specific appointment time with the patient. Neuropsychological tests are therefore best adapted to characterizing the overall cognitive capacities of the patient. In turn, mobile cognitive tests may be completed at any time in daily life, whether or not the television is on, whether or not one has just awoken from a nap, or whether or not one is in the middle of gardening. Such data are therefore best adapted to characterizing the ‘typical’ cognitive functioning of the patient in daily life, with all of its distractions and constraints.

It is also important to note that mobile cognitive tests are limited by the technologies that are used in their administration and by the EMA methodology itself. A smartphone may be easily capable of administering certain tests such as the IST, but the small screen may render the administration of many other tests unfeasible. EMA also requires relatively brief assessments which mean that extensive cognitive testing may also be inappropriate for this methodology. As previously noted, while the repeated assessment of cognitive functions is a strength for many reasons, it introduces potential biases that must be acknowledged and controlled for in research investigations.

Finally, a word of caution is necessary concerning the proliferation of smartphone applications for clinical populations because they are not without risk. To use one salient illustration, there are currently hundreds of smartphone applications to aid individuals with different forms of addiction, but not all were developed using prudent clinical judgment. For example, among the numerous applications that exist for alcohol dependence, some encourage full and immediate abstinence. This approach fully ignores the danger that going ‘cold turkey’ off alcohol poses for some patients due to seizures or other medical complications. In the same line of reasoning, the mass diffusion of mobile cognitive tests may be counterproductive for some patients with cognitive difficulties who would otherwise have directly sought expert advice and diagnosis, but who do not do so due the assumption that their problem has been adequately assessed. The rapid advances in mobile technologies may therefore, to some extent, have progressed too quickly and it is the responsibility of researchers and clinicians alike to inform individuals of the appropriate place that such tools may have in their lives.

In conclusion, the caution that one may have about the use of mobile cognitive tests is balanced with the quite extraordinary promise they hold in advancing research as a complement to neuropsychological testing and other clinical tools. Their development is still

in its infancy, but the existing data underscores the feasibility of convergent validity of this form of testing. Ideally, a repository of validated tests could be constructed in order to provide their wider access to researchers and to reduce the heterogeneity associated with the development of unique tests by individual researchers for the purpose of specific investigations. It is now necessary to progress in the further development and refining of this novel approach through careful validation studies.

References

- Aben, B., Stapert, S., and Blokland, A. (2012). About the Distinction between Working Memory and Short-Term Memory. *Front. Psychol.* 3.
- Adlam, A.-L.R., Patterson, K., Bozeat, S., and Hodges, J.R. (2010). The Cambridge Semantic Memory Test Battery: detection of semantic deficits in semantic dementia and Alzheimer's disease. *Neurocase* 16, 193–207.
- Adrover-Roig, D., Sesé, A., Barceló, F., and Palmer, A. (2012). A latent variable approach to executive control in healthy ageing. *Brain Cogn.* 78, 284–299.
- Agren, T. (2014). Human reconsolidation: a reactivation and update. *Brain Res. Bull.* 105, 70–82.
- Ahmed, S., Arnold, R., Thompson, S.A., Graham, K.S., and Hodges, J.R. (2008). Naming of objects, faces and buildings in mild cognitive impairment. *Cortex* 44, 746–752.
- Allard, M., Husky, M., Catheline, G., Pelletier, A., Dilharreguy, B., Amieva, H., Pérès, K., Foubert-Samier, A., Dartigues, J.-F., and Swendsen, J. (2014). Mobile technologies in the early detection of cognitive decline. *PloS One* 9, e112197.
- Allen, T.A., and Fortin, N.J. (2013). The evolution of episodic memory. *Proc. Natl. Acad. Sci.* 110, 10379–10386.
- Amieva, H., Le Goff, M., Millet, X., Orgogozo, J.M., Pérès, K., Barberger-Gateau, P., Jacqmin-Gadda, H., and Dartigues, J.F. (2008). Prodromal Alzheimer's disease: successive emergence of the clinical symptoms. *Ann. Neurol.* 64, 492–498.
- Anderson, V.A., Anderson, P., Northam, E., Jacobs, R., and Catroppa, C. (2001). Development of executive functions through late childhood and adolescence in an Australian sample. *Dev. Neuropsychol.* 20, 385–406.
- Atkinson, R.C., and Shiffrin, R.M. (1968). Human Memory: A Proposed System and its Control Processes I. In *Psychology of Learning and Motivation*, K.W.S. and J.T. Spence, ed. (Academic Press), pp. 89–195.
- Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* 4, 417–423.
- Baddeley, A.D. (1983). Working Memory. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 302, 311–324.
- Baddeley, A.D., and Hitch, G. (1974). Working Memory. In *Psychology of Learning and Motivation*, G.H. Bower, ed. (Academic Press), pp. 47–89.
- Basso, M.R., Beason-Hazen, S., Lynn, J., Rammohan, K., and Bornstein, R.A. (1996). Screening for cognitive dysfunction in multiple sclerosis. *Arch. Neurol.* 53, 980–984.

- Bechara, A. (2001). Neurobiology of decision-making: risk and reward. *Semin. Clin. Neuropsychiatry* 6, 205–216.
- Bechara, A., and Damasio, H. (2002). Decision-making and addiction (part I): impaired activation of somatic states in substance dependent individuals when pondering decisions with negative future consequences. *Neuropsychologia* 40, 1675–1689.
- Bechara, A., Dolan, S., and Hinds, A. (2002). Decision-making and addiction (part II): myopia for the future or hypersensitivity to reward? *Neuropsychologia* 40, 1690–1705.
- Ben-Zeev, D., Ellington, K., Swendsen, J., and Granholm, E. (2011). Examining a cognitive model of persecutory ideation in the daily life of people with schizophrenia: a computerized experience sampling study. *Schizophr. Bull.* 37, 1248–1256.
- Ben-Zeev, D., Morris, S., Swendsen, J., and Granholm, E. (2012). Predicting the occurrence, conviction, distress, and disruption of different delusional experiences in the daily life of people with schizophrenia. *Schizophr. Bull.* 38, 826–837.
- Berg, E.A. (1948). A simple objective technique for measuring flexibility in thinking. *J. Gen. Psychol.* 39, 15–22.
- Bernard, C., Helmer, C., Dilharreguy, B., Amieva, H., Auriacombe, S., Dartigues, J.-F., Allard, M., and Catheline, G. (2014). Time course of brain volume changes in the preclinical phase of Alzheimer's disease. *Alzheimers Dement. J. Alzheimers Assoc.* 10, 143–151.e1.
- Binder, J.R., and Desai, R.H. (2011). The Neurobiology of Semantic Memory. *Trends Cogn. Sci.* 15, 527–536.
- Boone, K.B., Pontón, M.O., Gorsuch, R.L., González, J.J., and Miller, B.L. (1998). Factor analysis of four measures of prefrontal lobe functioning. *Arch. Clin. Neuropsychol. Off. J. Natl. Acad. Neuropsychol.* 13, 585–595.
- Bozeat, S., Lambon Ralph, M.A., Patterson, K., Garrard, P., and Hodges, J.R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia* 38, 1207–1215.
- Brocki, K.C., and Bohlin, G. (2004). Executive functions in children aged 6 to 13: a dimensional and developmental study. *Dev. Neuropsychol.* 26, 571–593.
- Brouillette, R.M., Foil, H., Fontenot, S., Corroero, A., Allen, R., Martin, C.K., Bruce-Keller, A.J., and Keller, J.N. (2013). Feasibility, reliability, and validity of a smartphone based application for the assessment of cognitive function in the elderly. *PloS One* 8, e65925.
- Buczyłowska, D., and Petermann, F. (2016). Age-Related Differences and Heterogeneity in Executive Functions: Analysis of NAB Executive Functions Module Scores. *Arch. Clin. Neuropsychol. Off. J. Natl. Acad. Neuropsychol.* 31, 254–262.
- Buschke, H., and Fuld, P.A. (1974). Evaluating storage, retention, and retrieval in disordered memory and learning. *Neurology* 24, 1019–1019.
- Cain, A.E., Depp, C.A., and Jeste, D.V. (2009). Ecological momentary assessment in aging research: a critical review. *J. Psychiatr. Res.* 43, 987–996.

Caine, E.D., Bamford, K.A., Schiffer, R.B., Shoulson, I., and Levy, S. (1986). A controlled neuropsychological comparison of Huntington's disease and multiple sclerosis. *Arch. Neurol.* *43*, 249–254.

Cheke, L.G., and Clayton, N.S. (2013). Do different tests of episodic memory produce consistent results in human adults? *Learn. Mem. Cold Spring Harb. N* *20*, 491–498.

Content, A., Mousty, P., and Radeau, M. (1990). Brulex. Une base de données lexicales informatisée pour le français écrit et parlé. *Année Psychol.* *90*, 551–566.

Conway, M.A. (2009). Episodic memories. *Neuropsychologia* *47*, 2305–2313.

Craik, F.I.M., and Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *J. Verbal Learn. Verbal Behav.* *11*, 671–684.

Davidson, M.C., Amso, D., Anderson, L.C., and Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia* *44*, 2037–2078.

Delis (2000). California Verbal Learning Test, Second Edition: CvLT-II ; Adult Version ; Manual (Pearson).

DeLuca, J., Barbieri-Berger, S., and Johnson, S.K. (1994). The nature of memory impairments in multiple sclerosis: acquisition versus retrieval. *J. Clin. Exp. Neuropsychol.* *16*, 183–189.

Depp, C.A., Moore, R.C., Perivoliotis, D., Holden, J.L., Swendsen, J., and Granholm, E.L. (2016). Social behavior, interaction appraisals, and suicidal ideation in schizophrenia: The dangers of being alone. *Schizophr. Res.* *172*, 195–200.

Diamond, A. (2013). Executive Functions. *Annu. Rev. Psychol.* *64*, 135–168.

Dom, G., De Wilde, B., Hulstijn, W., van den Brink, W., and Sabbe, B. (2006). Decision-making deficits in alcohol-dependent patients with and without comorbid personality disorder. *Alcohol. Clin. Exp. Res.* *30*, 1670–1677.

Dujardin, K., Sockeel, P., Cabaret, M., De Sèze, J., and Vermersch, P. (2004). La BCcogSEP : une batterie courte d'évaluation des fonctions cognitives destinées aux patients souffrant de sclérose en plaques. *Rev. Neurol. (Paris)* *160*, 51–62.

Elias, M.F., Elias, P.K., D'Agostino, R.B., Silbershatz, H., and Wolf, P.A. (1997). Role of age, education, and gender on cognitive performance in the Framingham Heart Study: community-based norms. *Exp. Aging Res.* *23*, 201–235.

Ernst, M., Grant, S.J., London, E.D., Contoreggi, C.S., Kimes, A.S., and Spurgeon, L. (2003). Decision making in adolescents with behavior disorders and adults with substance abuse. *Am. J. Psychiatry* *160*, 33–40.

Fatseas, M., Serre, F., Alexandre, J.-M., Debrabant, R., Auriacombe, M., and Swendsen, J. (2015). Craving and substance use among patients with alcohol, tobacco, cannabis or heroin addiction: a comparison of substance- and person-specific cues. *Addict. Abingdon Engl.* *110*, 1035–1042.

- Fisk, J.E., and Sharp, C.A. (2004). Age-related impairment in executive functioning: updating, inhibition, shifting, and access. *J. Clin. Exp. Neuropsychol.* 26, 874–890.
- Frings, L., Wagner, K., Maiwald, T., Carius, A., Schinkel, A., Lehmann, C., and Schulze-Bonhage, A. (2008). Early detection of behavioral side effects of antiepileptic treatment using handheld computers. *Epilepsy Behav.* EB 13, 402–406.
- Gisquet-Verrier, P., Lynch, J.F., Cutolo, P., Toledano, D., Ulmen, A., Jasnow, A.M., and Riccio, D.C. (2015). Integration of New Information with Active Memory Accounts for Retrograde Amnesia: A Challenge to the Consolidation/Reconsolidation Hypothesis? *J. Neurosci. Off. J. Soc. Neurosci.* 35, 11623–11633.
- Goel, V., and Grafman, J. (1995). Are the frontal lobes implicated in “planning” functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia* 33, 623–642.
- Goel, V., Pullara, D., and Grafman, J. (2001). A computational model of frontal lobe dysfunction: working memory and the Tower of Hanoi task. *Cogn. Sci.* 25, 287–313.
- Goldstein, E.B., Vanhorn, D., Francis, G., and Neath, I. (2011). *Cognitive psychology : connecting mind, research, and everyday experience* (Belmont, CA Wadsworth/Cengage Learning).
- Granholm, E., Ben-Zeev, D., Fulford, D., and Swendsen, J. (2013). Ecological Momentary Assessment of social functioning in schizophrenia: impact of performance appraisals and affect on social interactions. *Schizophr. Res.* 145, 120–124.
- Grant, S., Contoreggi, C., and London, E.D. (2000). Drug abusers show impaired performance in a laboratory test of decision making. *Neuropsychologia* 38, 1180–1187.
- Greenberg, D.L., and Verfaellie, M. (2010). Interdependence of episodic and semantic memory: Evidence from neuropsychology. *J. Int. Neuropsychol. Soc. JINS* 16, 748–753.
- Grober, E., and Buschke, H. (1987). Genuine memory deficits in dementia. *Dev. Neuropsychol.* 3, 13–36.
- Gudayol-Ferré, E., Lara, J.P., Herrera-Guzman, I., Böhm, P., Rodés, E., Ansaldo, A.I., and Peña-Casanova, J. (2008). Semantic memory as assessed by the Pyramids and Palm Trees Test: the impact of sociodemographic factors in a Spanish-speaking population. *J. Int. Neuropsychol. Soc. JINS* 14, 148–151.
- Hannay, H.J., and Levin, H.S. (1985). Selective reminding test: an examination of the equivalence of four forms. *J. Clin. Exp. Neuropsychol.* 7, 251–263.
- Hanson, K.L., Luciana, M., and Sullwold, K. (2008). Reward-related decision-making deficits and elevated impulsivity among MDMA and other drug users. *Drug Alcohol Depend.* 96, 99–110.
- Harada, C.N., Natelson Love, M.C., and Triebel, K.L. (2013). Normal cognitive aging. *Clin. Geriatr. Med.* 29, 737–752.
- Henry, J.D., Crawford, J.R., and Phillips, L.H. (2004). Verbal fluency performance in dementia of the Alzheimer’s type: a meta-analysis. *Neuropsychologia* 42, 1212–1222.

- Hess, T.M., Popham, L.E., Emery, L., and Elliott, T. (2012). Mood, motivation, and misinformation: Aging and affective state influences on memory. *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.* *19*, 13–34.
- Hodges, J.R., Salmon, D.P., and Butters, N. (1992). Semantic memory impairment in Alzheimer's disease: Failure of access or degraded knowledge? *Neuropsychologia* *30*, 301–314.
- Holtzer, R., Verghese, J., Wang, C., Hall, C.B., and Lipton, R.B. (2008). Within-person across-neuropsychological test variability and incident dementia. *JAMA* *300*, 823–830.
- Howard, D., Patterson, K.E., and Company, T.V.T. (1992). The pyramids and palm trees test : a test of semantic access from words and pictures (Bury St Edmunds : Thames Valley Test Company).
- human-memory.net (2017). Episodic Memory and Semantic Memory - Types of Memory - The Human Memory.
- Humes, G.E., Welsh, M.C., Retzlaff, P., and Cookson, N. (1997). Towers of Hanoi and London: Reliability and Validity of Two Executive Function Tasks. *Assessment* *4*, 249–257.
- Husky, M., Olié, E., Guillaume, S., Genty, C., Swendsen, J., and Courtet, P. (2014). Feasibility and validity of ecological momentary assessment in the investigation of suicide risk. *Psychiatry Res.* *220*, 564–570.
- Husky, M.M., Gindre, C., Mazure, C.M., Brebant, C., Nolen-Hoeksema, S., Sanacora, G., and Swendsen, J. (2010). Computerized ambulatory monitoring in mood disorders: feasibility, compliance, and reactivity. *Psychiatry Res.* *178*, 440–442.
- Isaacs, B., and Kennie, A.T. (1973). The Set Test as an Aid to the Detection of Dementia in Old People. *Br. J. Psychiatry* *123*, 467–470.
- Jean, F.A.M., Swendsen, J.D., Sibon, I., Fehér, K., and Husky, M. (2013). Daily life behaviors and depression risk following stroke: a preliminary study using ecological momentary assessment. *J. Geriatr. Psychiatry Neurol.* *26*, 138–143.
- Johnson, E.I., Sibon, I., Renou, P., Rouanet, F., Allard, M., and Swendsen, J. (2009a). Feasibility and validity of computerized ambulatory monitoring in stroke patients. *Neurology* *73*, 1579–1583.
- Johnson, E.I., Grondin, O., Barrault, M., Faytout, M., Helbig, S., Husky, M., Granholm, E.L., Loh, C., Nadeau, L., Wittchen, H.-U., et al. (2009b). Computerized ambulatory monitoring in psychiatry: a multi-site collaborative study of acceptability, compliance, and reactivity. *Int. J. Methods Psychiatr. Res.* *18*, 48–57.
- Joy, S., Kaplan, E., and Fein, D. (2004). Speed and memory in the WAIS-III Digit Symbol--Coding subtest across the adult lifespan. *Arch. Clin. Neuropsychol. Off. J. Natl. Acad. Neuropsychol.* *19*, 759–767.
- Kane, M.J., and Engle, R.W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *J. Exp. Psychol. Gen.* *132*, 47–70.

- Keenan, E.K., Tiplady, B., Priestley, C.M., and Rogers, P.J. (2014). Naturalistic Effects of Five Days of Bedtime Caffeine Use on Sleep, Next-Day Cognitive Performance, and Mood. *J. Caffeine Res.* 4, 13–20.
- Kennedy, D.O., Veasey, R.C., Watson, A.W., Dodd, F.L., Jones, E.K., Tiplady, B., and Haskell, C.F. (2011). Vitamins and psychological functioning: a mobile phone assessment of the effects of a B vitamin complex, vitamin C and minerals on cognitive performance and subjective mood and energy. *Hum. Psychopharmacol.* 26, 338–347.
- Kok, A. (1999). Varieties of inhibition: manifestations in cognition, event-related potentials and aging. *Acta Psychol. (Amst.)* 101, 129–158.
- Kreiner, D.S., and Ryan, J.J. (2001). Memory and motor skill components of the WAIS-III Digit Symbol-Coding subtest. *Clin. Neuropsychol.* 15, 109–113.
- Lassalle-Lagadec, S., Catheline, G., Mayo, W., Dilharreguy, B., Renou, P., Fleury, O., Allard, M., Swendsen, J., and Sibon, I. (2013). Cerebellum involvement in post-stroke mood: a combined ecological and MRI study. *Psychiatry Res.* 212, 158–160.
- Lehto, J.E., Juujärvi, P., Kooistra, L., and Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *Br. J. Dev. Psychol.* 21, 59–80.
- Levin, H.S., Benton, A.L., and Grossman, R.G. (1982). *Neurobehavioral Consequences of Closed Head Injury* (Oxford University Press).
- Lukowiak, K., Adatia, N., Krygier, D., and Syed, N. (2000). Operant Conditioning in *Lymnaea*: Evidence for Intermediate- and Long-term Memory. *Learn. Mem.* 7, 140–150.
- Manns, J.R., Hopkins, R.O., and Squire, L.R. (2003). Semantic memory and the human hippocampus. *Neuron* 38, 127–133.
- Mårdh, S., Nägga, K., and Samuelsson, S. (2013). A longitudinal study of semantic memory impairment in patients with Alzheimer's disease. *Cortex* 49, 528–533.
- Martin-Ordas, G., Atance, C.M., and Caza, J.S. (2014). How do episodic and semantic memory contribute to episodic foresight in young children? *Front. Psychol.* 5.
- Matlin, M.W. (2005). *Cognition*, 6th Edition (Crawfordsville: John Wiley & Sons, Inc.).
- Mazure, C.M., Weinberger, A.H., Pittman, B., Sibon, I., and Swendsen, J. (2014). Gender and stress in predicting depressive symptoms following stroke. *Cerebrovasc. Dis. Basel Switz.* 38, 240–246.
- McKenna, P., and Warrington, E.K. (1983). *Graded naming test* (Windsor: NFER-Nelson).
- McLeod, S.A. (2007). *Stages of Memory - Encoding Storage and Retrieval*.
- McRae, K., and Jones, M.N. (2013). *Semantic Memory*.
- Metternich, B., Schmidtke, K., and Hüll, M. (2009). How are memory complaints in functional memory disorder related to measures of affect, metamemory and cognition? *J. Psychosom. Res.* 66, 435–444.

- Miller, G.A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* *63*, 81–97.
- Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., and Wager, T.D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: a latent variable analysis. *Cognit. Psychol.* *41*, 49–100.
- Morris, R.G., Miotto, E.C., Feigenbaum, J.D., Bullock, P., and Polkey, C.E. (1997). The effect of goal-subgoal conflict on planning ability after frontal- and temporal-lobe lesions in humans. *Neuropsychologia* *35*, 1147–1157.
- Mullally, S.L., and Maguire, E.A. (2014). Learning to remember: The early ontogeny of episodic memory. *Dev. Cogn. Neurosci.* *9*, 12–29.
- Nebes, R.D. (1989). Semantic memory in Alzheimer’s disease. *Psychol. Bull.* *106*, 377–394.
- Nielson, J.M. (1958). *Memory and Amnesia* (San Lucas Press).
- Nigg, J.T. (2000). On inhibition/disinhibition in developmental psychopathology: views from cognitive and personality psychology and a working inhibition taxonomy. *Psychol. Bull.* *126*, 220–246.
- Oliveira, J., Gamito, P., Morais, D., Brito, R., Lopes, P., and Norberto, L. (2014). Cognitive assessment of stroke patients with mobile apps: a controlled study. *Stud. Health Technol. Inform.* *199*, 103–107.
- oxforddictionaries.com (2016). cognition - definition of cognition in English | Oxford Dictionaries.
- Pérès, K., Matharan, F., Allard, M., Amieva, H., Baldi, I., Barberger-Gateau, P., Bergua, V., Bourdel-Marchasson, I., Delcourt, C., Foubert-Samier, A., et al. (2012). Health and aging in elderly farmers: the AMI cohort. *BMC Public Health* *12*, 558.
- Peterson, L., and Peterson, M.J. (1959). Short-term retention of individual verbal items. *J. Exp. Psychol.* *58*, 193–198.
- Rajan, K.B., Wilson, R.S., Weuve, J., Barnes, L.L., and Evans, D.A. (2015). Cognitive impairment 18 years before clinical diagnosis of Alzheimer disease dementia. *Neurology*.
- Rao, S.M. (1990). *A manual for the brief repeatable battery of neuropsychological tests in multiple sclerosis*.
- Rao, S.M., Leo, G.J., Bernardin, L., and Unverzagt, F. (1991). Cognitive dysfunction in multiple sclerosis. I. Frequency, patterns, and prediction. *Neurology* *41*, 685–691.
- Raudenbush, S.W. (2004). *HLM 6: Hierarchical Linear and Nonlinear Modeling* (Scientific Software International).
- Reichelt, A.C., and Lee, J.L.C. (2013). Memory reconsolidation in aversive and appetitive settings. *Front. Behav. Neurosci.* *7*.
- Rey, A. (1964). *L’examen clinique en psychologie*. (Paris: Presses universitaires de France).

- Riediger, M., Wrzus, C., Klipker, K., Müller, V., Schmiedek, F., and Wagner, G.G. (2014). Outside of the laboratory: Associations of working-memory performance with psychological and physiological arousal vary with age. *Psychol. Aging* 29, 103–114.
- Rosenzweig, M.R., Bennett, E.L., Colombo, P.J., Lee, D.W., and Serrano, P.A. (1993). Short-term, intermediate-term, and long-term memories. *Behav. Brain Res.* 57, 193–198.
- Royall, D.R., Lauterbach, E.C., Cummings, J.L., Reeve, A., Rummans, T.A., Kaufer, D.I., LaFrance, W.C., and Coffey, C.E. (2002). Executive control function: a review of its promise and challenges for clinical research. A report from the Committee on Research of the American Neuropsychiatric Association. *J. Neuropsychiatry Clin. Neurosci.* 14, 377–405.
- Salamon, R., Johnson, E.I., and Swendsen, J. (2011). Daily life mechanisms of stress spillover among early adolescents experiencing academic difficulty. *Eur. J. Psychol. Educ.* 26, 453–463.
- Salamon, R., Swendsen, J.D., and Husky, M.M. (2014). Extracurricular behavior and activities: A daily life study of academic failure and success [French language]. [/data/revues/00034487/v172i4/S000344871300259X/](#).
- Salmon, D.P., and Bondi, M.W. (2009). Neuropsychological Assessment of Dementia. *Annu. Rev. Psychol.* 60, 257–282.
- Salthouse, T.A. (1996). The processing-speed theory of adult age differences in cognition. *Psychol. Rev.* 103, 403–428.
- Salthouse, T.A., and Babcock, R.L. (1991). Decomposing adult age differences in working memory. *Dev. Psychol.* 27, 763–776.
- Salthouse, T.A., Fristoe, N., McGuthry, K.E., and Hambrick, D.Z. (1998). Relation of task switching to speed, age, and fluid intelligence. *Psychol. Aging* 13, 445–461.
- Salthouse, T.A., Atkinson, T.M., and Berish, D.E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *J. Exp. Psychol. Gen.* 132, 566–594.
- Schmidt, C., Collette, F., Cajochen, C., and Peigneux, P. (2007). A time to think: circadian rhythms in human cognition. *Cogn. Neuropsychol.* 24, 755–789.
- Scholey, A.B., Benson, S., Neale, C., Owen, L., and Tiplady, B. (2012). Neurocognitive and mood effects of alcohol in a naturalistic setting. *Hum. Psychopharmacol.* 27, 514–516.
- Schuster, R.M., Mermelstein, R.J., and Hedeker, D. (2015). Acceptability and feasibility of a visual working memory task in an ecological momentary assessment paradigm. *Psychol. Assess.* 27, 1463–1470.
- Schweitzer, P., Husky, M., Allard, M., Amieva, H., Pérès, K., Foubert-Samier, A., Dartigues, J.-F., and Swendsen, J. (2017). Feasibility and validity of mobile cognitive testing in the investigation of age-related cognitive decline. *Int. J. Methods Psychiatr. Res.* 26.

- Serre, F., Fatseas, M., Debrabant, R., Alexandre, J.-M., Auriacombe, M., and Swendsen, J. (2012). Ecological momentary assessment in alcohol, tobacco, cannabis and opiate dependence: a comparison of feasibility and validity. *Drug Alcohol Depend.* *126*, 118–123.
- Signoret, J.-L. (1991). *Batterie d'efficence mnésique: BEM 144* (Elsevier).
- Sliwinski, M.J., Mogle, J.A., Hyun, J., Munoz, E., Smyth, J.M., and Lipton, R.B. (2016). Reliability and Validity of Ambulatory Cognitive Assessments. *Assessment*.
- So, S.H., Peters, E.R., Swendsen, J., Garety, P.A., and Kapur, S. (2014). Changes in delusions in the early phase of antipsychotic treatment - an experience sampling study. *Psychiatry Res.* *215*, 568–573.
- Straube, B. (2012). An overview of the neuro-cognitive processes involved in the encoding, consolidation, and retrieval of true and false memories. *Behav. Brain Funct. BBF* *8*, 35.
- Strauss, E., Sherman, E.M.S., and Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary* (Oxford University Press).
- Stroop, J.R. (1935). Studies of Interference in Serial Verbal Reactions. *J. Exp. Psychol.* *18*, 643.
- Suchy, Y. (2009). Executive Functioning: Overview, Assessment, and Research Issues for Non-Neuropsychologists. *Ann. Behav. Med.* *37*, 106–116.
- Sutton, M.A., Masters, S.E., Bagnall, M.W., and Carew, T.J. (2001). Molecular Mechanisms Underlying a Unique Intermediate Phase of Memory in Aplysia. *Neuron* *31*, 143–154.
- Swendsen, J., Ben-Zeev, D., and Granholm, E. (2011). Real-time electronic ambulatory monitoring of substance use and symptom expression in schizophrenia. *Am. J. Psychiatry* *168*, 202–209.
- Thurstone, L.L. (1938). *Primary mental abilities*. (Chicago: Univ. of Chicago Press).
- Timmers, C., Maeghs, A., Vestjens, M., Bonnemayer, C., Hamers, H., and Blokland, A. (2014). Ambulant cognitive assessment using a smartphone. *Appl. Neuropsychol. Adult* *21*, 136–142.
- Tiplady, B., Oshinowo, B., Thomson, J., and Drummond, G.B. (2009). Alcohol and cognitive function: assessment in everyday life and laboratory settings using mobile phones. *Alcohol. Clin. Exp. Res.* *33*, 2094–2102.
- Tollenaar, M.S., Elzinga, B.M., Spinhoven, P., and Everaerd, W. (2009). Immediate and prolonged effects of cortisol, but not propranolol, on memory retrieval in healthy young men. *Neurobiol. Learn. Mem.* *91*, 23–31.
- Traykov, L., Raoux, N., Latour, F., Gallo, L., Hanon, O., Baudic, S., Bayle, C., Wenisch, E., Remy, P., and Rigaud, A.-S. (2007). Executive functions deficit in mild cognitive impairment. *Cogn. Behav. Neurol. Off. J. Soc. Behav. Cogn. Neurol.* *20*, 219–224.
- Tulving, E. (1972). Episodic and semantic memory. In *Organization of Memory*, (New York: Academic Press), pp. 381–402.

- Tulving, E. (1983). *Elements of Episodic Memory* (Clarendon Press).
- Tulving, E. (2002). Episodic Memory: From Mind to Brain. *Annu. Rev. Psychol.* 53, 1–25.
- Veasey, R.C., Haskell-Ramsay, C.F., Kennedy, D.O., Tiplady, B., and Stevenson, E.J. (2015). The Effect of Breakfast Prior to Morning Exercise on Cognitive Performance, Mood and Appetite Later in the Day in Habitually Active Women. *Nutrients* 7, 5712–5732.
- Villain, M., Cosin, C., Glize, B., Berthoz, S., Swendsen, J., Sibon, I., and Mayo, W. (2016). Affective Prosody and Depression After Stroke: A Pilot Study. *Stroke* 47, 2397–2400.
- Warrington, E.K. (1997). The Graded Naming Test: A Restandardisation. *Neuropsychol. Rehabil.* 7, 143–146.
- Waters, A.J., and Li, Y. (2008). Evaluating the utility of administering a reaction time task in an ecological momentary assessment study. *Psychopharmacology (Berl.)* 197, 25–35.
- Wechsler, D. (1939). *The measurement of adult intelligence* (Baltimore, MD, US: Williams & Wilkins Co).
- Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale* (Oxford, England: Psychological Corp.).
- Wechsler, D. (1997). *Wechsler Memory Scale* (Psychological Corporation).
- Wechsler, D. (2001). *MEM-III, Échelle clinique de mémoire de Wechsler: manuel* (les Éd. du Centre de psychologie appliquée).
- Wechsler, D., Psychological Corporation, and PsychCorp (Firm) (2008). *WAIS-IV technical and interpretive manual* (San Antonio, Tex.: Pearson).
- Whitlow, C.T., Liguori, A., Livengood, L.B., Hart, S.L., Mussat-Whitlow, B.J., Lamborn, C.M., Laurienti, P.J., and Porrino, L.J. (2004). Long-term heavy marijuana users make costly decisions on a gambling task. *Drug Alcohol Depend.* 76, 107–111.
- World Health Organization (2015). *Dementia*.

Résumé

Contexte. La détection des premiers stades du déclin cognitif menant à la maladie d'Alzheimer est rendue compliquée par la marge d'erreur associée aux tests neuropsychologiques traditionnels. Les technologies mobiles sont prometteuses pour surmonter cet obstacle en fournissant des évaluations répétées dans des contextes écologiquement valables. Le premier objectif de cette thèse était donc de développer des tests cognitifs mobiles pour diverses fonctions cognitives. Son deuxième objectif était de fournir la première étude détaillée de la faisabilité et de la validité des tests cognitifs mobiles dans la population âgée.

Méthode. Un programme pour les smartphones Android, "Samplex", a été développé pour administrer des tests mobiles de mémoire sémantique, de mémoire épisodique et de fonctionnement exécutif. La faisabilité et la validité de cet outil ont ensuite été examinées chez un échantillon de 114 sujets âgés non démentés. L'échantillon a été sollicité cinq fois par jour pendant une période d'une semaine pour fournir des informations sur leurs comportements, leurs états psychologiques et leurs contextes environnementaux, suivis de brefs tests mobiles de fonctionnement cognitif.

Résultats. De l'échantillon initial, 75 personnes ont montré un bon niveau d'observance aux évaluations répétées. Aucun effet de fatigue n'a été constaté, mais des effets d'entraînement ont été observés pour de nombreux tests mobiles. Les évaluations neuropsychologiques traditionnelles étaient corrélées avec la majorité des tests mobiles correspondants.

Conclusions. Les tests cognitifs mobiles peuvent être utilisés avec une partie substantielle des personnes âgées et sans effets de fatigue lorsqu'ils sont effectués sur une période d'une semaine. Les changements de performance liés à la répétition des tests mobiles peuvent être contrôlés statistiquement, réduisant ainsi les biais potentiels dans ces données. Leur validité convergente avec les évaluations neuropsychologiques traditionnelles indique que les tests cognitifs mobiles peuvent fournir des données valides et complémentaires pour détecter les premiers stades du déclin cognitif.

Mots Clés

Vieillesse cognitive, Vie quotidienne, Méthode d'échantillonnage des expériences, Maladie d'Alzheimer

Abstract

Background. The detection of the earliest stages of cognitive decline leading to Alzheimer's Disease is complicated by the margin of error associated with traditional neuropsychological tests. Mobile technologies hold promise for overcoming this barrier by providing repeated assessments in ecologically-valid contexts. The first objective of this doctoral thesis was therefore to develop mobile cognitive tests for a range of cognitive functions. Its second objective was to provide the first detailed examination of the feasibility and validity of mobile cognitive testing in the elderly population.

Methods. A program for Android smartphones, "Samplex", was developed to administer mobile tests of semantic memory, episodic memory and executive functioning. The feasibility and validity of this tool was then examined in a sample of 114 non-demented elderly individuals. The sample was signaled five times a day for a one-week period to provide information about their behaviors, psychological states and environmental contexts, followed by brief mobile tests of cognitive functioning.

Results. Of the initial sample, 75 individuals achieved moderate to high compliance with the repeated assessments. No evidence was found for fatigue effects, but training effects were observed for many of the mobile tests. Traditional neuropsychological assessments were correlated with the majority of corresponding mobile tests.

Conclusions. Mobile cognitive testing may be used with a substantial portion of elderly individuals and without fatigue effects when conducted over a one-week period. Time-dependent changes in mobile test scores may be statistically controlled, thereby reducing potential biases in these data. Their convergent validity with traditional neuropsychological assessments indicates that mobile cognitive tests may provide valid and complementary data for detecting the earliest stages of cognitive decline.

Keywords

Cognitive ageing, Everyday life, Experience sampling method, ESM, Ecological momentary assessment, EMA, Alzheimer's disease