HAL
open science

Quels apports de l'approche contextualisée à l'évaluation de l'intelligence chez les adolescent×e×s ?

Katia Terriot

To cite this version:

Katia Terriot. Quels apports de l'approche contextualisée à l'évaluation de l'intelligence chez les adolescent×e×s ?. Psychologie. Conservatoire national des arts et metiers - CNAM, 2018. Français. NNT : 2018CNAM1220 . tel-04317025

HAL Id: tel-04317025

https://theses.hal.science/tel-04317025

Submitted on 1 Dec 2023

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.

le cnam

ÉCOLE DOCTORALE Abbé Grégoire

Centre de recherche sur le travail et le développement (CRTD)

THÈSE présentée par :

Katia TERRIOT née LEAL

soutenue le : **18 décembre 2018**

pour obtenir le grade de : **Docteur du Conservatoire National des Arts et Métiers**

Discipline/ Spécialité : Psychologie, psychologie clinique, psychologie sociale / Psychologie

Quels apports de l'approche contextualisée à l'évaluation de l'intelligence chez les adolescent.e.s ?

THÈSE dirigée par :

M. LOARER Even	Professeur titulaire de Chaire, Cnam

THÈSE co-dirigée par :

M. GRÉGOIRE Jacques	Professeur ordinaire, Université Louvain-la-Neuve

RAPPORTEURS :

Mme BESANÇON Maud	Professeure des Universités, Rennes 2
Mme JEANRIE Chantale	Professeure titulaire, Université de Laval, Québec

---

JURY :

M. HOUSSEMAND Claude	Professeur, Université du Luxembourg
Mme BESANÇON Maud	Professeure des Universités, Rennes 2
Mme JEANRIE Chantale	Professeure titulaire, Université de Laval, Québec
M. CHARTIER Philippe	Maitre de conférences, Cnam
M. LOARER Even	Professeur titulaire de Chaire, Cnam
M. GRÉGOIRE Jacques	Professeur ordinaire, Université Louvain-la-Neuve

Remerciements

Mes premiers remerciements vont à ma famille : Jean, mon époux qui a su, avec patience, s'accommoder de la situation et du manque physique et psychique ; à mes enfants, sources d'inspiration ; à mes parents, sans qui rien n'aurait pu se faire, toujours présents malgré les années qui passent.

Je tiens à remercier chaleureusement mes directeurs de thèse : M. Grégoire et M. Loarer pour leur soutien, conseils avisés, le temps accordé et leur expertise et M. Houssemand pour ses riches recommandations qui m'ont permis d'améliorer mes analyses et mes compétences en statistiques notamment.

Je n'oublie pas toutes et tous les collègues de l'Inetop qui m'ont soutenue moralement et proposé leur aide tout au long de ce travail, en particulier Laurence Thiénot pour ses relectures, ainsi que Ghislaine Labouret pour son aide sur le logiciel R.

Merci également à plusieurs stagiaires Psychologues sans qui ce travail n'aurait pu se faire : Mmes Brax, Brémont, Cazeneuve, Cusset, Diaz, Dubost, Khettar, Lafitte, Lamazière, Petel, Roux et Vaz pour leur implication dans le recueil de données. Et aux psychologues de l'Education Nationale en Centres d'information et d'Orientation. Ainsi qu'à Catherine Denier pour son aide dans la diffusion de l'étude et le recueil de données.
Sans oublier les jeunes et leur famille pour leur participation ainsi que Laura Berson et Serge De Almeida pour leur contribution à la construction du matériel.

Et enfin les expert.e.s qui ont permis une amélioration notable de la qualité de certains articles qui composent cette thèse.

Résumé

Dans le domaine de l'évaluation de l'intelligence, les tests constitués de tâches décontextualisées ont une position hégémonique (Huteau & Lautrey, 1999). Ils reposent sur des modèles, issus de la psychologie classique, centrés sur la mesure de capacités générales supposées fonctionner également avec tous les types de contenus (Bartholomew, 2004 ; Binet, 1911). Or, d'autres modèles insistent sur l'importance des répertoires de connaissances spécifiques et de la contextualisation des processus cognitifs (McLellan, 1996). L'évaluation contextualisée place le sujet dans les conditions les plus proches des situations de la vie quotidienne. L'objectif principal de la thèse est de montrer les éventuels apports de cette approche contextualisée à l'évaluation de l'intelligence chez les adolescent.e.s, en termes de preuves de validité (liens à la performance scolaire et/ou adaptation sociale) à travers cinq articles. Le premier fait l'état des lieux de l'évaluation contextualisée de l'intelligence, sa définition, sa mesure et les relations entretenues avec l'évaluation classique. Le second développe plus particulièrement les évaluations auto et hétéro rapportées de l'intelligence. Les outils recensés restent très peu nombreux, et presque exclusivement pour un public adulte. Les trois articles suivants rendent compte du développement de trois outils d'évaluation contextualisée de l'intelligence auprès d'un public adolescent et de leurs apports notamment en termes de validité : un questionnaire d'auto-évaluation, deux questionnaires d'hétéro-évaluation (parent et enseignant.e) et un test de performance. Les quatre outils présentent des qualités psychométriques acceptables. Seul le facteur de raisonnement fluide du questionnaire d'auto-évaluation corrèle avec un test de performance classique (WISC-IV) et avec les résultats scolaires. Les liens avec le comportement adaptatif (mesuré par la Vineland-II) existent, notamment avec le facteur de vitesse de traitement du questionnaire (r=.69). Les questionnaires d'hétéro-évaluation entretiennent des liens modérés à forts avec les résultats scolaires. Enfin, pour le test de performance composé de sept subtests, seule une épreuve (intelligence cristallisée) corrèle positivement et significativement avec la moyenne scolaire générale (r=.42) ; et trois avec le comportement adaptatif (corrélations néanmoins faibles). Les corrélations avec le WISC-IV sont, quant à elles, élevées. Les résultats et implications de ces trois études sont discutés dans la conclusion et une ouverture sur la réalité virtuelle est proposée.

Mots clés : Cattell-Horn-Carroll (CHC), évaluation 360°, auto-évaluation, hétéro-évaluation, comportement adaptatif, test cognitif, WISC

Résumé en anglais

In the field of intelligence assessment, tests consisting of decontextualized tasks have a hegemonic position (Huteau & Lautrey, 1999). They are based on models from classical psychology, centered on the measurement of general abilities supposed to work equally with all types of content (Bartholomew, 2004; Binet, 1911). Other models emphasize the importance of specific knowledge repertoires and the contextualization of cognitive processes (McLellan, 1996). Contextualized evaluation places the subject in the conditions closest to the situations of everyday life. The main objective of the thesis is to show the contributions of this contextualized approach to the assessment of intelligence in adolescents, through five articles. The first takes stock of the contextualized evaluation of intelligence, its definition, its measurement and the relations maintained with classical evaluation. The second one develops more particularly the self and hetero assessment of intelligence. The tools identified remain very few and almost exclusively for an adult audience. The following three articles report on the development of four contextualized intelligence assessment tools for an adolescent audience and their contributions in particular in terms of validity: a self-assessment questionnaire, two hetero-assessment questionnaires (parent and teacher), and a performance test. All three tools have acceptable psychometric qualities. Only the fluid reasoning factor of the self-assessment questionnaire correlates with a classic performance test (WISC-IV) and with academic performance. Links to adaptive behavior (measured by Vineland-II) exist, especially with the processing speed factor (r = .69). Hetero-assessment questionnaires have moderate to strong links to academic performance. Finally, for the performance test composed of seven subtests, only one subtest (crystallized intelligence) correlates positively and significantly with the academic performance (r = .42); and three with adaptive behavior (yet weak correlations). Correlations with WISC-IV are high. The results and implications of these three studies are discussed in the conclusion and an opening on virtual reality is proposed.

Key words : intelligence, assessment, CHC, context

Table des matières

Remerciements....2
Résumé....3
Résumé en anglais....4
Table des matières....5
Liste des tableaux....8
Liste des annexes....10
Introduction....11
Première partie Cadre théorique....16
1. L'évaluation contextualisée de l'intelligence....17
Introduction....17
1.1 Qu'entend-on par une évaluation contextualisée ?....18
1.2 Les formes d'évaluations contextualisées....20
1.2.1 Les tests papier-crayon....20
1.2.2 Simulations et mises en situation....28
1.2.3 Evaluation en situation naturelle....34
Discussion....37
2. Les évaluations auto et hétéro-rapportées sont-elles des alternatives à la mesure psychométrique classique de l'intelligence ? Revue de 25 ans de recherche....42
Introduction....42
2.1 Les mesures auto-rapportées de l'intelligence....46
2.1.1 L'auto-estimation....48
2.1.2 L'auto-évaluation....55
2.2 Les évaluations hétéro-rapportées....60
2.2.1 L'hétéro-estimation....61
2.2.2 L'hétéro-évaluation....62
Discussion....65
Synthèse du cadre théorique....70
Seconde partie Contributions empiriques....74

3. Etude 1 : Le développement d'un outil d'auto-évaluation des capacités cognitives pour adolescent.e.s : le QAE 75
3.1 Introduction 75
3.1.1 Le modèle Cattell-Horn-Carroll (CHC) 77
3.1.2 L'auto-évaluation des capacités cognitives 79
3.2 Objectifs 82
3.3 Méthodologie 83
3.3.1 Développement du questionnaire 83
3.3.2 Participant.e.s 85
3.3.3 Mesure 85
3.4 Résultats 88
3.4.1 Les relations entre dimensions du QAE 93
3.4.2 Preuves de validité concourante et différentielle. 94
3.5 Discussion 100
4. Etude 2 : Evaluation multiple ou multi-source : apports pour la prédiction de la réussite scolaire ? 111
4.1 Introduction 111
4.1.1 De l'auto-évaluation ou hétéro-évaluation à l'évaluation multiple des capacités cognitives 113
4.1.2 Evaluation multiple et réussite scolaire 121
4.2 Objectifs de la présente étude 122
4.3 Méthodologie 123
4.3.1 Participant.e.s 123
4.3.2 Mesures 124
4.4 Résultats 127
4.4.1 Etude des trois questionnaires 127
4.4.2 Procédure multi-informateur.trice.s 131
4.4.3 Liens entre l'évaluation multiple et la performance scolaire 133
4.5 Discussion 136

5. Etude 3 : Développement d'une échelle contextualisée des capacités cognitives auprès d'un échantillon d'adolescent.e.s français.e.s : étude exploratoire 143
5.1 Introduction 143
5.1.1 Limites de l'évaluation classique de l'intelligence 144
5.1.2 Evaluation contextualisée 146
5.2 Création d'un outil contextualisé d'évaluation des capacités cognitives pour des sujets adolescent.e.s 152
5.2.1 Développement de l'échelle 153
5.2.2 Etude de validation 157
5.3 Discussion 171
Conclusion et perspectives 180
Résumé 252
Résumé en anglais 252

Liste des tableaux

Tableau 1. Nombre de publications recensées 45
Tableau 2. Questionnaires d'auto-évaluation des capacités cognitives, présentation et qualités psychométriques. 58
Tableau 3. Questionnaires d'hétéro-évaluation des capacités cognitives, présentation et qualités psychométriques. 64
Tableau 4. Analyse du fonctionnement différentiel des 51 items du QAE selon le sexe 90
Tableau 5. Matrice des composantes après rotation Oblimin 92
Tableau 6. Matrice des corrélations entre les facteurs 94
Tableau 7. Différences de scores à chaque échelle selon le sexe (N=321) 95
Tableau 8. Corrélations entre les six facteurs du QAE et la moyenne générale au 3ème trimestre (N=87) 96
Tableau 9. Comparaison aux normes des scores obtenus par l'échantillon (N=33) au WISC-IV 97
Tableau 10. Corrélations entre les six facteurs du QAE et les résultats au WISC-IV (N=33) 98
Tableau 11. Comparaison des scores obtenus par l'échantillon (N=25) à la Vineland II aux normes 99
Tableau 12. Corrélations entre les six facteurs au QAE et les scores à la Vineland II (N=25) 100
Tableau 13. Description des facteurs du QAE 125
Tableau 14. Alpha pour les six facteurs du QHP 126
Tableau 15. Statistiques descriptives du QAE, QHP et QHE (N=139) 128
Tableau 16. Inter-corrélations pour les 15 facteurs des 3 questionnaires 129
Tableau 17. Comparaisons de moyennes entre le QAE et le QHP 130
Tableau 18. Analyse en composantes principales des facteurs Gc, Gf et Gs en utilisant la méthode de Kraemer et al. (2003) 132
Tableau 19. Corrélations entre les facteurs du QAE, QHP et QHE et la moyenne générale au troisième trimestre 133

Tableau 20. Régressions hiérarchiques 135
Tableau 21. Panorama des différentes méthodologies d'évaluation des capacités cognitives 148
Tableau 22. Répartition des sujets de notre échantillon selon la classe (N=61) .. 158
Tableau 23. Répartition des sujets selon le niveau de diplôme le plus haut des parents (N=57) 158
Tableau 24. Fonctionnement différentiel des items selon le sexe 160
Tableau 25. Difficulté, écart-type et indice de discrimination des items (N=61) 162
Tableau 26. Alpha de Cronbach des épreuves comportant plusieurs items 163
Tableau 27. Matrice des corrélations entre les épreuves 163
Tableau 28. Indices d'ajustement pour les deux modèles 165
Tableau 29. Corrélations entre les six épreuves et la moyenne générale au 3ème trimestre (N=21) 166
Tableau 30. Comparaison aux normes des scores obtenus par l'échantillon (N=59) au WISC-IV 166
Tableau 31. Corrélations entre les six épreuves et les résultats au WISC-IV (N=57) 167
Tableau 32. Comparaison des scores obtenus par l'échantillon (N=59) à la Vineland II aux normes 168
Tableau 33. Corrélations entre les six épreuves et les scores à la Vineland II (N=59) 168
Tableau 34. Différences de scores à chaque échelle selon le sexe (N=61) 170
Tableau 35. Différences de scores à chaque échelle selon le niveau de diplôme des parents (N=57) 171
Tableau 36. Proposition de modélisation d'une tâche d'arithmétique en réalité virtuelle 187

Liste des annexes

Annexe 1. Autorisation de diffusion 211
Annexe 2. Poster ICAP 212
Annexe 3. Questionnaire initial (QAE), version 52 items 213
Annexe 4. Questionnaire enseignant.e 216
Annexe 5. Cahier de passation de l'échelle d'évaluation contextualisée 217
Annexe 6. Matériel pour l'échelle d'évaluation contextualisée 232

Introduction

L'évaluation des capacités intellectuelles est un enjeu majeur en France comme dans le monde. Elle est centrale dans la pratique des psychologues, dans le cadre de bilans psychologiques. En effet, cette évaluation répond à plusieurs objectifs dont les principaux sont de recueillir des informations cliniques sur la personne, repérer et diagnostiquer certains troubles ou difficultés comme par exemple les troubles spécifiques des apprentissages, répondre à des exigences administratives (pour des orientations notamment ou pour la reconnaissance d'une situation de handicap) ou comprendre un fonctionnement cognitif pour apporter des aides ciblées (Grégoire, 2006). Pour les psychologues de l'Education Nationale amené.e.s à faire un bilan psychologique, la question initiale est souvent de savoir si le.la jeune peut suivre l'enseignement classique. Offrir aux praticien.ne.s des outils qui peuvent expliquer et prédire les résultats scolaires est donc primordial pour répondre à cette demande. La question sous-jacente est celle de la validité des résultats au test utilisé. Peut-on faire des inférences à partir des résultats obtenus ? De nombreuses recherches sont donc menées depuis plus d'un siècle sur cette thématique. Des revues y sont mêmes consacrées à l'instar d'*Intelligence*. Les tests d'intelligence et notamment ceux de Binet et Simon (1905) ont été créés pour répondre à cette question. De nombreuses études ont montré leur validité pour expliquer et prédire la réussite scolaire avec des corrélations qui sont généralement autour de .54 pour le français et .64 pour les mathématiques (Zimmerman & Woo-Sam, 1997). Ces prédictions demeurent néanmoins imparfaites. Comment les améliorer ?

Par ailleurs, un autre point central auquel sont confronté.e.s les praticien.ne.s, c'est l'observation, parfois, d'un décalage entre les résultats obtenus au test et ce qui peut être constaté dans la vie quotidienne. Ainsi, certains jeunes peuvent avoir des résultats faibles aux tests cognitifs et être tout à fait adaptés dans la vie de tous les jours et inversement. Pour certains auteurs (e.g Gottfredson, 2003), si des personnes réussissent bien aux tests classiques mais ne s'adaptent pas dans la vie de tous les jours, c'est que d'autres facteurs explicatifs interviennent comme la personnalité, la motivation, etc. Pour d'autres (e.g. Sternberg et al., 2000), l'explication vient du fait que les tests classiques n'évaluent qu'un type d'intelligence : l'intelligence académique. Or, il existerait également une intelligence pratique liée au contexte. Les processus sous-jacents seraient les mêmes mais des personnes qui appliquent correctement des processus dans un contexte peuvent ne pas le faire correctement dans un autre contexte. La question sous-jacente est à nouveau celle de la validité du test, plus particulièrement écologique, définie comme le degré auquel les résultats obtenus à un test peuvent être corrélés à ceux obtenus dans un environnement naturel (Tupper & Cicerone, 1990). Aussi serait-il nécessaire de développer des tests contextualisés pour se rapprocher du contexte de vie quotidienne, bien différent de celui du laboratoire. Cette forme d'évaluation alternative est déjà très présente en ressources humaines et en neuropsychologie, beaucoup moins en psychologie scolaire. Ainsi, les tests contextualisés ont-ils des preuves de validité écologique ?

L'objectif de cette thèse, sur articles, est de montrer les apports de l'évaluation contextualisée, notamment en termes de preuves de validité (liens avec la réussite

scolaire et/ou l'adaptation sociale) dans la mesure de l'intelligence chez des adolescent.e.s tout venants. Pour y répondre, elle est organisée sous forme de deux parties : une première théorique composée de deux articles, une deuxième constituée de trois contributions empiriques. Pour pallier à l'absence actuelle d'outils français et internationaux reposant sur le modèle théorique Cattell, Horn, Carroll (CHC), contextualisés et pour un public adolescent, nous avons dû créer des outils pour répondre à notre objectif. La thèse rend également compte du développement de ces derniers.

Le premier article fait l'état des lieux de l'évaluation contextualisée de l'intelligence, sa définition, sa mesure et les relations entretenues avec l'évaluation classique. Nous prenons le parti de puiser dans différents courants de la psychologie (ressources humaines, cognition située, neuropsychologie, etc.) et de faire l'état des lieux des recherches, y compris auprès d'adultes, bien que notre sujet se centre sur les adolescent.e.s. En effet, la recherche sur l'évaluation contextualisée auprès d'un public adolescent et tout venant est quasi inexistante.

Le deuxième article développe plus particulièrement les évaluations auto et hétéro rapportées de l'intelligence. Nous les définissons, puis nous précisons les méthodes utilisées pour les mesurer et enfin, nous tentons de voir si elles présentent un niveau de validité satisfaisant et si elles peuvent permettre, en complément de l'évaluation classique, d'augmenter le pourcentage de variance expliquée de la réussite scolaire.

Le troisième rend compte du développement d'un questionnaire d'auto-évaluation des capacités cognitives reposant sur six grands facteurs du modèle de Cattell-

Horn-Carroll (CHC) et de ses liens avec la performance scolaire en fin d'année et le comportement adaptatif.

L'article suivant se propose de voir si l'évaluation multiple des capacités cognitives peut permettre de prédire la réussite scolaire. Il présente ainsi trois questionnaires qui peuvent être utilisés en évaluation multiple avec des adolescent.e.s, parents et enseignant.e.s de la classe où est scolarisé.e le.la jeune.

Enfin, le dernier article rend compte du développement d'un test de performance contextualisé, reposant sur le modèle CHC, composé de sept épreuves. Nous montrons les premiers résultats des analyses psychométriques, en particulier les tentatives de preuves de validité.

Ce travail est donc exploratoire du fait du peu de recherches sur cette thématique, bien que n'étant pas nouvelle.

Première partie

Cadre théorique

1. L'évaluation contextualisée de l'intelligence

Introduction

Dans le domaine de l'évaluation de l'intelligence, les tests constitués de tâches décontextualisées ou ayant peu de rapport avec les contextes dans lesquels les sujets sont impliqués ont une position hégémonique (Huteau & Lautrey, 1999). Ainsi, les quatre tests les plus utilisés dans le monde (Oakland & HU, 1992), la *WAIS* (Wechsler, 2011), le *WISC* (Wechsler, 2016), les *Matrices de Raven* (Raven, 1998) et le *Stanford Binet* (Roid, 2003) sont des tests décontextualisés. Ils reposent sur des modèles, issus de la psychologie classique, centrés sur la mesure de capacités générales supposées fonctionner également avec tous les types de contenus (Bartholomew, 2004 ; Binet, 1911). Or, d'autres modèles insistent sur l'importance des répertoires de connaissances spécifiques et de la contextualisation des processus cognitifs (McLellan, 1996). Une forme d'évaluation alternative, constituée de tâches contextualisées, existe ainsi depuis le début du siècle dernier. Elle s'est principalement développée dans les champs de la neuropsychologie et des ressources humaines. La mise en perspective de ces deux formes d'évaluation active un débat ancien qui a marqué l'histoire de la philosophie puis de la psychologie, opposant, dans l'explication des conduites, les tenants d'une centration sur l'individu à ceux qui privilégient l'environnement. Ce débat semble aujourd'hui dépassé avec l'avènement de ces nouvelles formes

d'évaluation qui permettent d'étudier l'interaction entre le sujet et son environnement, comme le prônait déjà Wallon en 1942.

Le présent article a pour objectif de faire le point sur l'évaluation contextualisée de l'intelligence ; ce que ce concept recouvre (définitions et exemples) et quelles relations elle entretient avec les formes d'évaluation plus classiques.

1.1 Qu'entend-on par une évaluation contextualisée ?

Le contexte peut être défini comme « l'ensemble des circonstances dans lesquelles se produit un évènement où se situe une action » (Larousse, 2015). En parcourant la littérature scientifique, plusieurs expressions plus ou moins synonymes sont employées en référence à cette notion : évaluation contextualisée, écologique, située ou situationnelle.

L'approche de l'évaluation contextualisée n'est pas nouvelle. Dès 1948, aux États-Unis, l'Office of Strategic Services[1] (OSS) développe une méthode d'évaluation contextualisée ou situationnelle, dite organismique, qui veut reproduire les conditions « naturelles » dans lesquelles le sujet réalise des performances complexes. Pour recruter des militaires et prédire leur efficacité dans l'organisation, l'OSS propose des situations d'évaluation au plus proche de la vie réelle. Il oppose cette méthode à une méthode dite « élémentalistique » qui consiste à analyser différentes composantes d'un fonctionnement et à évaluer

[1] L'OSS était une agence de renseignement du gouvernement des Etats-Unis, créée en 1942 après leur entrée en guerre dans la Seconde Guerre mondiale pour collecter des informations. Pour mieux remplir ces fonctions, ils ont développé le premier centre d'évaluation psychologique aux Etats-Unis

séparément chacune d'entre elles. Selon l'OSS, cette dernière méthode, abstraite et non réaliste, ne parvient pas à reproduire les conditions dans lesquelles un sujet réalise ses performances quotidiennes. En effet, dans la vie réelle, le sujet organise ces différentes composantes. Cette orchestration n'est pas évaluée par la méthode élémentalistique qui ne se centre que sur les processus élémentaires relativement simples, partiels, et évalués les uns après les autres. Pour l'OSS, le critère essentiel d'un bon test est sa congruence avec la réalité. Des sujets peuvent en effet obtenir un haut score dans une tâche abstraite et être en difficulté sur le même type de tâche dans la vie quotidienne. Les performances mesurées lors de l'évaluation ne sont pas toujours représentatives du comportement usuel du sujet.

Dans un autre champ, en neuropsychologie, le terme actuellement utilisé est celui d'évaluation écologique. L'approche écologique est née avec Neisser (1987) qui propose de prendre en compte l'interaction entre le sujet et l'environnement dans lequel il vit. Selon lui, l'étude des processus d'acquisition des connaissances doit cesser de se faire uniquement en laboratoire. Messick (1988) propose également d'évaluer les sujets dans des conditions proches de véritables situations de travail afin d'obtenir de meilleures prédictions. Actuellement, l'approche écologique est entendue plus largement comme une conception s'intéressant à la vie quotidienne du sujet, c'est-à-dire au sujet en contexte. Il s'agit d'une évaluation réalisée au plus près des conditions de vie de la personne. Cette approche fait écho à celle de McClelland (1973) qui proposait, pour prédire la performance professionnelle, de privilégier l'évaluation de conduites complexes en situation plutôt que de faire passer des tests d'intelligence classiques.

Dans le champ des ressources humaines et du recrutement en particulier, on parle davantage de tests situationnels qui ont en commun de présenter aux candidats des situations relevant d'un emploi auxquelles ils doivent répondre (Oostrom, Born, Serlie & Van der Molen, 2011).

Bien que toutes ces approches contextualisées aient en commun de mettre en relation le sujet avec l'environnement dans lequel il vit, leur mise en application revêt des formes bien différentes que nous allons maintenant détailler.

1.2 Les formes d'évaluations contextualisées

L'évaluation contextualisée inclut différentes méthodes que l'on pourrait placer sur un continuum allant des tests papier-crayon dans lesquels le contexte est évoqué et représenté à une évaluation en situation dite naturelle, dans laquelle le sujet est directement plongé dans une situation de la vie réelle.

1.2.1 Les tests papier-crayon

Les tests papier-crayon ont en commun de faire appel au contexte sous une forme écrite. Le sujet doit dès lors se représenter mentalement le contexte évoqué. Dans cette catégorie d'outils, on retrouve principalement des tests d'auto ou d'hétéro-évaluation, des tests dits de jugement situationnel et des tests de performance.

1.2.1.1 Les auto et hétéro-évaluations

Les auto et hétéro-évaluations consistent à demander au sujet ou à son entourage d'évaluer sa capacité de réaliser certaines tâches cognitives en lien avec la vie quotidienne, comme se repérer sur un plan, suivre les conversations des autres, ou se souvenir où les choses sont rangées…

Nous avons recensé (Terriot, Grégoire & Loarer, 2017[2]) cinq questionnaires d'auto-évaluation en lien avec la vie quotidienne : le *Self-Report Measure of Cognitive Abilities* (SRMCA ; Jacobs & Roodenburg, 2014), composé de 19 items évaluant les intelligences cristallisée et fluide ainsi que le traitement visuel ; le *Spatial ability in everyday life questionnaire* (Quaiser-Pohl & Lehmann, 2002) qui mesure les capacités spatiales ; l'*Everyday Cognition questionnaire* (Ecog ; Rueda, Lau, Saito, Harvey, Risacher, Aisen, Petersen, Saykin & Farias, 2015) dans lequel six domaines sont investigués, la mémoire quotidienne, le langage, les capacités visuo-spatiales, la planification, l'organisation et l'attention divisée ; le *Multiple Abilities Self Report Questionnaire* (MASQ ; Seidenberg, Haltiner, Taylor, Hermann & Wyler, 1994) qui évalue le langage, le visuo-perceptif, la mémoire verbale, la mémoire visuelle et l'attention-concentration ; et enfin le *Multiple Intelligences Profiling Questionnaire* (MIPQ III ; Tirri & Nokelainen, 2008) reposant sur la théorie des intelligences multiples de Gardner (1983, 1999, 2006).

[2] Voir Partie 1, Chapitre 2

L'examen de ces différents questionnaires d'auto-évaluation met en évidence d'assez bonnes qualités psychométriques, lorsque ces données sont disponibles. Les alphas sont tous supérieurs à .70 pour les sous-échelles de ces différents questionnaires, sauf pour la sous-échelle d'intelligence spatiale du MIPQ III (α=.53). Les corrélations avec des tests classiques de mesure des capacités cognitives sont variables d'une étude à l'autre, mais le degré de liaison, lorsqu'il existe, est généralement faible ou modéré, excepté dans le domaine spatial pour lequel l'auto-évaluation est plus étroitement liée aux performances spatiales (r=.52, p<.001 entre le *Spatial ability in everyday life questionnaire* et le *Mental Rotations Test*, MRT).

Bien que présentant certaines qualités psychométriques, ces méthodes ne semblent pouvoir remplacer les tests de performance. Il est vraisemblable qu'elles mesurent un autre construit, tel que le concept de soi en lien avec des expériences passées.

L'hétéro-évaluation, quant à elle, consiste à demander à l'entourage du sujet d'évaluer la capacité de ce dernier de réaliser certaines tâches cognitives en lien avec la vie quotidienne. Les recherches ont apporté des preuves de leur validité dans différents domaines (Achenbach, McConaughy & Howell, 1987; Fabrega, Ulrich & Loeber, 1996 ; Hart, Lahey, Loeber & Hanson, 1994 ; Loeber, Green, Lahey & Stouthamer-Loeber, 1990 ; Loeber, Green, Lahey & Stouthamer-Loeber, 1991), mais très peu concernant le domaine du fonctionnement cognitif (Oliver, Dale, Saudino, Petrill, Pike & Plomin, 2002). A notre connaissance, seuls trois questionnaires d'hétéro-évaluation des capacités cognitives sont parus récemment

: le *Behavior Rating Inventory of Executive Function* (BRIEF ; Gioia, Isquith, Guy & Kenworthy, 2000), le *Parent Report of Children's Abilities* (PARCA 3 ; Oliver et al., 2002), qui évalue les capacités spatiales, la planification, l'organisation, le raisonnement, la mémoire, le raisonnement quantitatif, les nombres et les jeux symboliques ; et le questionnaire de Waschbusch, Daleiden et Drabman (2000), composé de six échelles en référence au modèle de Cattell, Horn, Carroll (CHC ; Newton & McGrew, 2010).

Ces questionnaires semblent présenter certaines qualités psychométriques correctes, telles que la cohérence interne (Alpha de Cronbach supérieurs à .70 pour l'ensemble des sous-échelles de chaque questionnaire à l'exception d'une échelle du PARCA 3), mais d'autres preuves de validité doivent encore être récoltées. Les corrélations avec des mesures objectives sont faibles ou modérées, ce qui semble indiquer que les hétéro-évaluations ne peuvent pas, là encore, remplacer les tests d'intelligence classiques.

Pour améliorer les corrélations avec les questionnaires auto et hétéro-rapportés, Paulhus, Lysy & Yik (1998) suggèrent de développer de nouveaux tests d'intelligence qui prendraient en compte les conceptions quotidiennes de l'intelligence.

1.2.1.2 Les tests de jugement situationnel

Les tests de jugement situationnel (TJS), aussi appelés tests d'intelligence pratique ou tests de connaissances contextuelles, sont des tests papier-crayon ou

informatisés qui mesurent le jugement dans des situations professionnelles. Un scénario, décrivant une situation que le candidat est susceptible de rencontrer professionnellement, est proposé et le répondant doit indiquer ce qu'il ferait ou ce qu'il faut faire pour faire face à la situation. La première utilisation de ces tests date des années 20. Durant la Deuxième Guerre mondiale, les psychologues de l'armée proposèrent également ces tests sous forme de scénarios avec un nombre de réponses alternatives. Les réponses faisaient appel au sens commun, à l'expérience et aux connaissances générales plus qu'au raisonnement logique. A partir des années 40, un certain nombre de tests furent développés pour mesurer le potentiel des managers, comme le *Practical Judgement Test* (Cardall, 1942), *How Supervise ?* (File & Remmers, 1948), le *Supervisory Practices Tests* (Bruce & Learner, 1958)… A partir de 1985, d'autres outils sont nés, comme le *Tacit Knowledge Inventory for Managers* (TKIM ; Wagner & Sternberg, 1991), *Tacit Knowledge Inventory for Military Leaders* (TKLM ; Hedlund, Horvath, Forsythe, Snook, Williams, Bullis, Dennis, Sternberg, 1998), *Tacit Knowledge in Sales* (TKIS ; Wagner, Sujan, Sujan, Rashotte & Sternberg, 1999), le *College Life Questionnaire* (CLQ ; Sternberg, Forsythe, Hedlund, Horvath, Wagner, Williams, Snook & Grigorenko, 2000), le *Common Sense Questionnaire* (CSQ ; Sternberg et al., 2000), testés auprès de différentes populations : managers, responsables militaires, commerciaux.les, enseignant.e.s, enfants scolarisé.e.s.

Il existe deux types de TJS (McDaniel, Hartman, Whetzel & Grubb, 2007) : ceux qui évaluent les connaissances (très proches des tests de connaissances tacites de Wagner et Sternberg, 1985) et ceux qui évaluent la tendance à produire certains

comportements. Dans cette dernière forme, il est demandé au sujet d'identifier comment il se comporterait dans une situation donnée. Les tests de connaissances générales ont des corrélations plus élevées avec les capacités cognitives (r[3]$=.35$ vs r=.19 pour les tests de comportement), et les tests de comportement avec les construits de personnalité (McDaniel *et al.,* 2007). Pour trois des cinq grandes dimensions de la personnalité se référant au modèle du Big Five, les corrélations entre les tests de jugement situationnel et les tests de personnalité sont supérieures pour les tests de comportement que pour les tests de connaissances générales : Agréabilité (.37 vs .19), Conscience (.34 vs. .24), et Stabilité émotionnelle (.35 vs .12). Par ailleurs, les corrélations entre les TJS et les performances dans le travail sont de l'ordre de .26, ce qui est équivalent aux corrélations entre les capacités cognitives et les performances au travail (.25) (McDaniel, Morgeson, Finnegan, Campion & Braverman, 2001). Soulignons toutefois que Schmidt et Hunter dans leur méta-analyse de 1998, qui reste une référence sur la question, trouvaient une corrélation de .51 entre l'aptitude cognitive générale et la performance professionnelle, ce qui est le double de ce qu'avancent McDaniel *et al.* (2001). Ces corrélations peuvent être augmentées lorsque les TJS sont utilisés en complément des capacités cognitives (.31), même si l'effet est peu important (McDaniel *et al.*, 2001).

En ce qui concerne les preuves de cohérence interne, dans leur méta-analyse portant sur trente-neuf TJS parus jusqu'en 2000, McDaniel et al. (2001) ont reporté des coefficients variant de .43 à .94.

[3] Moyenne estimée de la population

1.2.1.3 Les tests de performance

Enfin, plusieurs tests en format papier-crayon ont été développés en lien avec la vie quotidienne. C'est notamment le cas du Sternberg Triarchic Abilities test (STAT), non publié, mais utilisé dans plusieurs recherches (Grigorenko & Sternberg, 2001 ; Sternberg, Grigorenko, Ferrari & Clinkenbeard, 1999). Cet outil comprend douze subtests, quatre pour chaque facteur (analytique, créatif et pratique) combinant des contenus verbaux, quantitatifs et figuratifs. Par exemple, pour évaluer l'intelligence pratique, Sternberg propose des problèmes mathématiques quotidiens comme changer les quantités dans un récipient, faire des cookies, acheter des tickets pour un jeu (contenus quantitatifs); une description d'un problème typique rencontré dans la vie quotidienne pour lequel le sujet doit choisir la meilleure solution (contenus verbaux) ; le choix d'un itinéraire le plus approprié pour se déplacer d'un point à un autre en utilisant une carte (contenus figuratifs). La validité de ce test a été vivement critiquée (Brody, 2003 ; Gottfredson, 2003 ; Koke & Vernon, 2003). En fin de compte, les épreuves de ce test ne se distinguent pas radicalement des mesures classiques de l'intelligence dans la mesure où les contextes restent très simplifiés et assez artificiels puisqu'ils sont présentés sous un format papier-crayon (Grégoire, 2006).

Dans le domaine de la neuropsychologie, d'autres outils ont été élaborés dans une approche écologique, tels que le *Test of Everyday Attention* (TEA ; Robertson, Ward, Ridgeway & Nimmo-Smith, 1994), le *Rivermead Behavioral Memory Test*

3ème édition (RBMT 3 ; Wilson, Greenfield, Clare, Baddeley, Cockburn, Watson, Tate, Sopena, Nannery & Crawford, 2010), le *Behavioral Assessment of the Dysexecutive Syndrome* (BADS ; Wilson, Alderman, Burgess, Emslie & Evans, 1996), la *Batterie de Crook* (Crook, Salama & Gobert, 1986) dans laquelle le sujet doit, par exemple, apprendre 14 paires de noms/visages ; rappeler la localisation de vingt objets dans une maison schématisée de douze pièces, mémoriser un numéro de téléphone, ou un trajet effectué en voiture ; ou encore le *Test Lillois de Calcul II* (TLC2 ; Bout-Forestier & Depoorter, 2008) dans lequel le nombre est utilisé de manière écologique en demandant au sujet de retrouver une date précise sur un calendrier, de manipuler de l'argent ou une calculatrice…

Quelques études font le lien entre des tests de performance dans des domaines spécifiques (principalement, celui de la mémoire) et les aptitudes dans la vie quotidienne. Ainsi, Chaytor et Schmitter-Edgecombe (2003) dans leur revue de littérature ont relevé six études s'intéressant au lien entre la mémoire et les capacités dans la vie quotidienne (Goldstein, McCue, Rogers & Nussbaum, 1992 ; Higginson, Arnett & Voss, 2000 ; Johnson, 1994 ; Kaitiro, Koskinen & Kaipio, 1995 ; Makatura, Lam, Leahy, Castillo & Kalpakjian, 1999 ; Sunderland, Harris & Baddeley, 1983). Deux d'entre elles comparent les tests psychologiques traditionnels et les tests écologiques. L'étude de Makatura et al. (1999) montre que le test écologique, le *RBMT*, est le seul qui identifie les problèmes de mémoire quotidiens dans les quatre groupes de patients déterminés par un classement effectué par les cliniciens (précision de 60,5% dans la classification). L'étude de Higginson et al. (2000) aboutit aux mêmes conclusions : les tests

écologiques (*RBMT* et *TEA*) sont de meilleurs prédicteurs d'une déficience fonctionnelle (mesurée par un entretien) que les tests traditionnels de mémoire et d'attention chez des sujets ayant une sclérose en plaques (r^2= .10). Par ailleurs, 67% des tests écologiques corrèlent significativement avec le statut fonctionnel de la personne contre seulement 38% des tests traditionnels.

1.2.2 Simulations et mises en situation

Dans ce type d'évaluation, les sujets sont confrontés à des tâches de laboratoire qui sont similaires à ce qui est rencontré dans la vie quotidienne ou au travail en utilisant non plus un support papier-crayon, mais du matériel concret, un support vidéo ou une réalité virtuelle. Le niveau d'immersion dans la tâche varie donc en fonction de ce type de matériel. Ces méthodologies sont principalement utilisées dans le domaine du recrutement et de la neuropsychologie. Pour le recrutement, les simulations ont pour objectif d'évaluer directement les comportements au travail en mettant en scène des situations professionnelles réelles. Les réponses des sujets sont considérées représenter leur comportement réel en situation de travail. En neuropsychologie, l'objectif est de reproduire une situation de la vie quotidienne en créant des épreuves standardisées et objectives.

1.2.2.1 Utilisation d'un matériel concret

Dans le domaine des ressources humaines, les tâches de simulation sont utilisées depuis longtemps pour évaluer les personnes (Brouwers & van de Vijver, 2015). Pour plusieurs auteurs (Cooper, Robertson & Tinline, 2003 ; Salgado, Viswesvaran & Ones, 2001), il semble important de créer des épreuves basées sur des mises en situation proches du contexte de travail. Aujourd'hui, la tâche la plus connue est le test « in basket » ou corbeille à papier, utilisé par les centres d'évaluation pour prédire la performance au travail, notamment dans des postes managériaux. Le sujet doit traiter des notes, mails, papiers ou mémentos dans un délai contraint et prendre des décisions (Frederiksen, 1966; Frederiksen, Saunders & Wand, 1957). La performance est évaluée selon la manière dont les dossiers sont traités.

Pour les 31 outils recensés, la méta-analyse de Whetzel, Rotenberry et McDaniel (2014) met en évidence une fidélité inter-cotateurs moyenne de .76, une validité prédictive de la performance dans le travail de .42, comparable à d'autres tâches de simulation, mais qui reste, néanmoins, un peu plus faible que pour les capacités cognitives (.51 selon Schmidt & Hunter, 1998) et enfin une corrélation moyenne de .26 avec des tests de facteur g. Les construits mesurés par cette tâche semblent plus cognitifs que ceux mesurés dans les autres tâches, telles que les jeux de rôle qui renvoient davantage à des compétences relationnelles.

En neuropsychologie ont été créées différentes épreuves de simulation dites écologiques, telles que le *Behavioural Inattention Test* (BIT ; Wilson, Cockburn

& Halligan, 1987), les échelles de Zoccolotti et Judica (1991) ou des épreuves d'exécution de scripts. Le BIT comprend neuf épreuves qui simulent des activités de la vie quotidienne, comme décrire trois photographies représentant des scènes familières, composer un numéro de téléphone, lire un menu ou un article de journal, mettre à l'heure une pendule.

Dans les tâches d'exécution de scripts, il est demandé au sujet de réaliser, par exemple, des courses, des recettes de cuisine, un trajet, de rédiger une lettre…Elle est utilisée auprès d'un public varié : personnes avec un syndrome dysexécutif (Chevignard, Pradat-Diehl & Pillon, 1999), une dyspraxie développementale (Toussaint-Thorin, Marchal, Pradat-Diehl & Chevignard, 2011). Chevignard et ses collaborateurs (Chevignard, Pillon, Pradah-Diehl, Taillefer, Rousseau, Le Bras, 2000 ; Chevignard, Taillefer, Picq, Poncet, Noulhiane, Pradat-Diehl, 2008) ont notamment développé une tâche spécifique : la tâche de cuisine. Dans cette épreuve, il est demandé au sujet de réaliser deux recettes de difficulté croissante : une omelette pour deux personnes et un gâteau au chocolat. Les consignes et la disposition des ustensiles sont standardisées. La cotation est réalisée en fonction des erreurs (oublis, estimations, non prise en compte du contexte…). L'exécution de scripts permet d'analyser les fonctions exécutives (planification, organisation, inhibition…). Les auteurs ont fait passer cette épreuve à 45 patients présentant un syndrome dysexécutif suite à diverses pathologies (Chevignard et al., 2008). Ces sujets font plus d'erreurs (M=107.5, ET=74.3), que les patients du groupe contrôle (N=14, M=18.3, ET=7.8), et sont plus lents dans l'exécution de la tâche. La fidélité inter-cotateurs est élevée ($r = .85$; $p < .0001$). L'étude ne précise cependant pas si ce test écologique prédit davantage les répercussions du dysfonctionnement que les tests neuropsychologiques.

1.2.2.2 Utilisation d'un support vidéo

Le support vidéo est principalement utilisé dans le cadre des TJS déjà présentés précédemment. Les situations ne sont plus présentées en format papier-crayon, mais en format vidéo. Les réponses des sujets peuvent, quant à elles, être sous forme écrite ou filmée par une webcam (Oostrom et al., 2011). Christian, Edwards et Bradley (2010) ont noté quatre avantages quant à l'utilisation du format vidéo : 1) il permet de présenter des détails contextuels de l'environnement (indices non verbaux, ton de la voix, objets ambiants…) ; 2) il montre directement les comportements de travail réel ; 3) il peut améliorer la validité apparente du test, qui est liée à la performance au test ; 4) enfin, le format vidéo est adapté pour la mesure de compétences sociales. Cucina, Su, Busciglio, Harris Thomas, et Thompson Peyton (2015) ajoutent cinq autres avantages à l'utilisation de ce format à la fois en mode de présentation, mais également en mode de réponse : 1) il permet de réduire la surcharge cognitive, importante lors de la lecture et de la compréhension des scénarios présentés sous forme écrite ; 2) il augmente la validité de contenu ; 3) il offre la possibilité de mesurer les réponses émotionnelles des sujets ; 4) il permet de mesurer la performance réelle et non ce que ferait potentiellement la personne dans telle situation ; 5) enfin, en lien avec la théorie sociocognitive de Bandura (1977), les TJS sous format vidéo permettent de mettre en œuvre l'ensemble des étapes du modèle de l'apprentissage par modelage de Bandura (attention, rétention, reproduction et motivation) alors que

dans les TJS classiques sous format papier-crayon nécessitent uniquement l'attention et la rétention.

Dans la méta-analyse réalisée par Christian et al. (2010), seules quelques études utilisant le format vidéo ont été répertoriées. Les TJS ont une validité critériée moyenne de .36, plus importante que celle des tests papier-crayon qui est de .25, pour l'ensemble des construits mesurés (leadership, compétences inter-personnelles…).

1.2.2.3 Utilisation de la réalité virtuelle

La réalité virtuelle permet de proposer des environnements interactifs dans lesquels le sujet est acteur. Il peut, en effet, avoir une activité sensori-motrice et cognitive dans un monde simulant le monde réel (Klinger, Marie & Fuchs, 2006). Les stimuli peuvent être présentés de manière écologique en lien avec un contexte significatif et familier (une salle de classe, une maison, une rue…). Par ailleurs, cette technologie offre la possibilité de gérer le chronomètre, de contrôler les distracteurs et la complexité des tâches proposées (Zhang, Abreu, Masel, Scheibel, Christiansen, Huddleston, Ottenbacher, 2001). Elle peut améliorer la fiabilité des évaluations en diminuant la variabilité due aux différences entre examinateurs (même si la standardisation permet déjà cela) étant donné que toutes les réponses sont directement enregistrées. Elle permet également d'obtenir des preuves de validité écologique plus solides (Pugnetti, Mendozzi, Attree, Barbieri, Brooks, Cazzullo, Motta & Rose, 1998), cette dernière étant définie comme le

degré auquel les tests de performance correspondent à la performance dans la vie quotidienne (Tupper & Ciccerone, 1990) et offre la possibilité de transfert des compétences apprises de l'environnement virtuel au monde réel. Elle est utilisée avec des patients souffrant de diverses pathologies : troubles de l'attention (et notamment dans le dépistage Rizzo, Bowerly, Buckwalter, Schultheis, Matheis, Shahabi, Neumann, Kim & Sharifzadeh, 2002 ; Rizzo, Bowerly, Shahabi, Buckwalter, Klimchuk & Mitura, 2004), des fonctions exécutives (Klinger, Marie & Fuchs, 2006), la mémoire (Rose, Brooks & Rizzo, 2005), des capacités visuo-spatiales…Plusieurs environnements virtuels ont été créés comme une classe virtuelle (Rizzo et al., 2002 ; Rizzo et al., 2004), un bureau (Schultheis & Rizzo, 2002), un magasin de fruits (Lo Priore, Castelnuovo & Liccione, 2003), un appartement (Zalla, Plassiart, Pillon, Grafman & Sirigu, 2001), des supermarchés (Lee, Ku, Cho, Hahn, Kim, Lee, Kang, Kim, Yu, Wiederhold, Wiederhold, Kim, 2003 ; Klinger, Chemin, Lebreton, Marie, 2004), une cuisine (Zhang et al., 2001). Cette méthode est encore peu développée du fait de son coût important en termes de développements informatiques et de matériel nécessaire pour l'évaluation.

Quelques études se sont intéressées à la comparaison d'une épreuve neuropsychologique classique à une même épreuve en réalité virtuelle. Par exemple, Pugnetti et al. (1998) ont utilisé le *Wisconsin Card Sorting Test* (WSCT ; Grant & Berg, 1948) et un environnement virtuel représentant un bâtiment et reproduisant les mêmes consignes et conditions que dans le WSCT. Les résultats montrent que les deux tâches permettent de différencier les sujets, mais qu'elles semblent évaluer des processus différents, du fait de corrélations faibles entre les

deux, alors qu'elles sont basées sur un même principe. Les auteurs font l'hypothèse que le test de réalité virtuelle sonde davantage l'orchestration des processus cognitifs que chaque capacité cognitive prise isolément.

1.2.3 Evaluation en situation naturelle

L'évaluation en situation naturelle diffère principalement des autres méthodes décrites jusqu'à présent par le fait que le sujet n'est pas contraint de réaliser la tâche (comme en situation expérimentale) et qu'il est soumis à de véritables enjeux, à des problèmes qui le concernent véritablement, et ce dans son propre environnement.

Elle trouve son origine notamment dans le courant de la cognition située, associé généralement aux travaux de Suchman (1987). Elle met en avant le rôle fondamental joué par les facteurs contextuels dans l'action et prône une analyse de la cognition en situation naturelle et non en laboratoire. De nombreuses études ont montré que les compétences apprises à l'école n'étaient pas facilement transférables dans la vie quotidienne. Les approches de la cognition située (Hutchins, 1995) se démarquent des approches traditionnelles issues des sciences cognitives, qui conçoivent uniquement la cognition comme un processus de traitement de l'information. Au contraire, les approches de la cognition située voient la cognition comme le produit de l'interaction entre le sujet et son environnement et non comme dépendant uniquement du traitement individuel. Les

processus cognitifs sont indissociables de la situation dans laquelle ils sont en œuvre (Greeno, 1994).

De nombreuses recherches ont été réalisées dans le domaine des mathématiques. Ainsi, Lave (1988) part du constat qu'il est impossible d'analyser les apprentissages scolaires sans considérer leurs relations avec le monde auquel l'école est censée préparer. Dans cette perspective, cette chercheuse s'est intéressée à l'utilisation des mathématiques dans la vie quotidienne. Elle observe que la réussite dans de résolution de problèmes mathématiques varie selon les contextes. Lave a fait passer à des femmes américaines un test de connaissances arithmétiques générales, un questionnaire à choix multiple pour évaluer leurs compétences en résolution de problèmes mathématiques, ainsi que les mêmes problèmes à résoudre, mais dans le contexte d'un supermarché. Lave n'observe pas de corrélation significative entre les performances de calcul en contexte de supermarché et les scores au test de mathématiques et au questionnaire à choix multiple. Cette observation va à l'encontre de l'hypothèse d'un transfert des connaissances scolaires dans les contextes de la vie quotidienne. Les femmes au foyer américaines se révèlent en effet capables de résoudre correctement des problèmes mathématiques au supermarché, alors qu'elles sont incapables de résoudre les mêmes problèmes quand elles sont placées dans un contexte scolaire. Dans une autre étude menée avec Murtaugh et de la Roche (1984), Lave arrive aux mêmes conclusions. Des consommateurs peuvent calculer facilement lorsqu'ils effectuent des achats, mais ne sont pas capables de réussir avec des tests

papier-crayon. Scribner (1984) a réalisé une étude auprès d'ouvriers d'une usine de traitement de lait, toujours dans le domaine mathématique. Ces ouvriers ne semblent pas appliquer des algorithmes mathématiques appris à l'école, mais ils utilisent des stratégies complexes qu'ils ont eux-mêmes développées. Les chercheurs n'ont pas trouvé de corrélation significative entre les performances professionnelles de ces ouvriers et des mesures classiques comme des tests d'intelligence ou des tests de connaissances scolaires.

Aujourd'hui, les évaluations en situation naturelle sont utilisées notamment en neuropsychologie sous la forme d'observations directes du comportement et de performances dans le milieu de vie, observations qui peuvent être standardisées. Il n'existe pas, à notre connaissance de revue de littérature sur l'ensemble des outils existants et leurs qualités psychométriques. Nous pouvons, à titre d'exemple, citer le *Profil de Compétences Cognitives du Jeune* (P2CJP ; Pereira Da Costa & Scelles, 2012), outil d'évaluation cognitive des enfants et adolescent.e.s atteints de polyhandicap qui permet d'évaluer les capacités sensorielles, attentionnelles, spatiales, verbales et communicatives, socio-émotionnelles, d'apprentissage, la mémoire et le raisonnement. Cet outil peut être utilisé en observation indirecte ou directe à l'aide de 69 items. Le pourcentage moyen d'accord inter-juges est de 83,31% (avec des pourcentages variant de 67% pour les capacités attentionnelles à 94% pour les capacités d'apprentissage), ce qui démontre une fidélité inter-juges satisfaisante. La consistance interne pour l'ensemble des items est très satisfaisante (α =.96) et seule la sous-échelle des capacités sensorielles présente

un alpha inférieur à .70. Les auteurs ne précisent pas les corrélations avec d'autres outils d'évaluation cognitive.

L'observation directe présente toutefois certains inconvénients qui expliquent peut-être qu'elle ne soit pas davantage utilisée. En particulier, elle est très coûteuse en termes de temps. Elle nécessite, en effet, que l'observateur.trice se déplace dans le lieu de vie ou de travail du sujet, parfois à plusieurs reprises si la capacité qu'il souhaite observer ne s'est pas d'emblée manifestée.

Discussion

L'objectif de cet article était de faire le point sur l'évaluation contextualisée de l'intelligence, sa définition, sa mesure et les relations entretenues avec l'évaluation classique.

L'évaluation classique de l'intelligence a depuis longtemps montré son efficacité pour mesurer les capacités cognitives. Les tests d'intelligence cités dans l'introduction sont fiables, fidèles et valides. Ils répondent aux critères posés par Jensen (1992) pour mesurer objectivement une capacité, tels, par exemple, le fait que le construit mesuré doit impliquer des unités de comportement observables qui aboutissent à certains types de produits (réponse orale, écrite, motrice…). Ils prédisent la réussite scolaire avec des corrélations entre les résultats aux tests de performance et les notes aux examens scolaires variant de .40 à .70 selon les tests utilisés et les recherches (Brody, 1992 ; Jensen, 1980). Certaines études montrent également des corrélations de l'ordre de .50 à .60 entre le quotient intellectuel

(QI) et les performances professionnelles (Brody, 1992 ; Schmidt & Hunter, 1998). Leurs atouts ne sont donc plus à démontrer. Pour répondre aux objectifs fixés (mesurer objectivement une capacité), les tâches proposées dans ces tests classiques sont généralement assez abstraites (l'exemple prototypique serait les Matrices) afin de pouvoir isoler la capacité que l'on souhaite mesurer et pour ne pas être sensibles à l'expérience de la personne évaluée (Gottfredson, 2003). La consigne ne fait également pas appel à la vie quotidienne. Les auteurs de ces tests tentent de concevoir des tâches « pures », qui évaluent successivement des processus élémentaires. Le sujet doit résoudre chaque tâche une à une, dans un environnement fermé et structuré, à l'abri des sollicitations extérieures, et généralement dans un temps contrôlé.

A l'inverse, dans des évaluations contextualisées, l'objectif serait de se rapprocher des conditions de la vie quotidienne. Au quotidien, le sujet est confronté à des épreuves non structurées, simultanées et multiples, ce qui nécessite de coordonner plusieurs activités cognitives (McClelland, 1973 ; Sternberg, Conway, Keton & Bernstein, 1981) et ce dans un environnement ouvert avec de possibles stimulations interférentes. Les tâches rencontrées sont complexes, requérant l'orchestration des différentes capacités cognitives. Par ailleurs, dans la vie de tous les jours, un sujet peut utiliser des stratégies de compensation qu'il n'utilise pas dans les tests de performance, qui peuvent ainsi sous-évaluer ses capacités réelles ou à l'inverse, les utiliser dans la situation de test classique et oublier de le faire au quotidien. Enfin, le temps n'est pas toujours contraint.

Selon le niveau de contextualisation des épreuves d'évaluation de l'intelligence, l'objectif de se rapprocher le plus possible de la vie quotidienne est plus ou moins atteignable. Si l'on prend le cas des épreuves contextualisées sous format papier-crayon (exemple du STAT), les tâches ne sont pas complexes, réalisées avec des sollicitations extérieures et dans des environnements ouverts. Elles ne sont en lien avec la vie quotidienne que par le contenu de la tâche (calculer des quantités pour faire une recette de cuisine par exemple), voire le matériel utilisé.

Les différents outils d'évaluation contextualisée présentés montrent d'assez bonnes qualités psychométriques quand les études existent. Nous avons vu que dans certaines études, les tests d'évaluation contextualisée étaient de meilleurs prédicteurs d'une déficience dans la vie quotidienne que les tests classiques (Higginson et al., 2000 ; Makatura et al. 1999). Ils semblent également permettre une plus grande implication du sujet dans la passation, voire une plus grande motivation, point non négligeable dans le processus d'évaluation. Elle offre également, des preuves de validité écologique plus importantes par rapport à l'évaluation classique, ainsi que des preuves de validité incrémentielle concernant la réussite professionnelle notamment, mais également la réussite scolaire en augmentant la prédiction de ces dernières en complément des évaluations classiques

Ces deux approches sont à considérer comme complémentaires (Grigorenko, Meier, Lipka, Mohatt, Yanez & Sternberg, 2004 ; Sternberg, Grigorenko & Bundy 2001) afin d'avoir une évaluation plus complète de l'intelligence d'un sujet.

De nouvelles recherches sont néanmoins nécessaires, notamment pour comparer ces différentes méthodes entre elles. Par ailleurs, la plupart d'entre elles sont surtout utilisées en ressources humaines et en neuropsychologie, presque exclusivement auprès d'adultes. Il n'existe pas, à notre connaissance, de tels outils pour évaluer les capacités cognitives d'enfants et/ou d'adolescent.e.s tout venants. Développer un outil d'évaluation de l'intelligence contextualisé pour ce public semblerait une voie de recherche intéressante pour les avantages cités précédemment.

2. Les évaluations auto et hétéro-rapportées sont-elles des alternatives à la mesure psychométrique classique de l'intelligence ? Revue de 25 ans de recherche

Paru dans le Bulletin de Psychologie sous la référence [4]:

Terriot, K., Grégoire, J. & Loarer, E. (2017). Les évaluations auto et hétéro-rapportées sont-elles des alternatives à la mesure psychométrique classique de l'intelligence ? Revue de 25 ans de recherche. *Bulletin de psychologie*, 550(4), 275-289. doi:10.3917/bupsy.550.0275.

Introduction

Depuis Binet et Simon (1905, 1908) et le premier test d'évaluation de l'intelligence, cette dernière fait l'objet de très nombreuses recherches. L'évaluation intellectuelle repose traditionnellement sur l'utilisation de tests psychométriques, comme les tests différentiels d'aptitudes (DAT 5 ; Bennett, Seashore & Wesman, 2002) ; la batterie pour l'examen psychologique de l'enfant, deuxième édition (K-ABC-II ; Kaufman & Kaufman, 2008) ; les Progressive Matrices (PM38 ; Raven, 1998) ; l'échelle d'intelligence de Wechsler pour

[4] La revue a donné l'autorisation de diffusion (voir annexe n°1)

adultes, quatrième édition (WAIS-IV, Wechsler, 2011) ; l'échelle d'intelligence de Wechsler pour enfants et adolescent.e.s, quatrième édition (WISC-IV ; Wechsler, 2005) ; l'échelle d'intelligence de Wechsler pour la période pré-scolaire et primaire, quatrième édition (WPPSI-IV ; Wechsler, 2014) ; l'échelle non verbale de Wechsler (WNV ; Wechsler & Naglieri, 2009). Une enquête (Oakland & HU, 1992) montre ainsi que les cinq tests les plus utilisés dans le monde sont la WAIS, la WISC, les Matrices de Raven, le Stanford Binet et le DAT, qui sont tous des tests qui mesurent l'intelligence des sujets sur la base de leurs performances à des épreuves standardisées. Est-ce la seule manière d'évaluer l'intelligence ? Eysenck et Eysenck (1985) ont proposé, à côté de cette méthode classique, deux autres méthodologies de mesure de l'intelligence appelées génotypiques et auto ou hétéro-rapportées. L'intelligence génotypique ne peut être directement mesurée et doit l'être à travers des performances à des tâches élémentaires, censées évaluer des processus cognitifs au cœur de l'activité intellectuelle, comme le temps de réaction ou d'inspection, ce qui n'est pas sans rappeler les travaux de Wundt ou de Galton au XIXe siècle. L'intelligence auto-rapportée peut être mesurée par le sujet lui-même et être hétéro-rapportée, par des observateur.trice.s de celui-ci. Cette dernière méthodologie existe dans différents domaines pour évaluer des compétences sociales et adaptatives (Achenbach, 1991 ; Conners, 2001 ; Reynolds & Kamphaus, 2004), la personnalité (Bernaud, 2014), la qualité de vie (Balboni, Coscarelli, Giunti & Schalock, 2013), l'intelligence émotionnelle (Choi & Kluemper, 2012 ; Petrides & Furnham, 2001), l'intelligence sociale (Loarer, 2005), et a montré ses atouts, mais elle est très peu développée

pour l'évaluation des processus cognitifs, surtout en France. Les intelligences auto et hétéro-rapportées sont les objets du présent article. Nous définirons ce que sont les évaluations auto et hétéro-rapportées de l'intelligence, nous préciserons les méthodes utilisées pour les mesurer et tenterons de répondre aux questions suivantes : ces méthodes présentent-elles un niveau de validité satisfaisant ? Évaluent-elles le même construit sous-jacent que les tests de performance ? Dans quelles situations est-il intéressant de les utiliser ? À la place des tests classiques ou en complément de ceux-ci ? En particulier, sachant que les tests classiques de performance n'expliquent généralement que 25 à 30 % de la variance de la réussite scolaire (Mackintosh, 2004 ; McGrew & Knopick, 1993), peut-on augmenter le pourcentage de variance expliquée en utilisant les mesures auto et hétéro-rapportées en complément des tests classiques ?

Le travail de recension a été réalisé à partir de la base de données PsychINFO avec au départ les mots-clés suivants : « auto », ou « subjectivement », « hétéro », « autrui », « parental » associés à « évalué », « évaluation », « estimé », « perçu », « estimation » et « intelligence » ou « capacités » ainsi que dans certaines revues ciblées non référencées dans PsychINFO : *Educational Research and Reviews*, *Psychology Science Quarterly, Zeitschrift für Sozialpsychologie*. Le tableau 1 indique le nombre de publications recensées.

Mots clés associés	Nombre de publications
Auto-estimé.s ou auto-estimation ou auto-évaluation ou auto-évalué.s ou auto-perçu ou auto ou subjectivement ou auto-évaluation ou auto-estimation et intelligence	65
Auto-estimé.s ou auto-estimation ou auto-évaluation ou auto-évalué.s ou auto-perçu ou auto ou subjectivement ou auto-évaluation ou auto-estimation et capacité.s	46
Autrui.s ou parental ou hétéro et intelligence et capacité.s	52

Tableau 1. Nombre de publications recensées

Nous avons sélectionné les articles uniquement centrés sur les mesures hétéro et auto-rapportées de l'intelligence en lien avec nos questionnements et parus depuis 25 ans.

Le nombre d'articles recensés est relativement modeste, ce qui témoigne d'un intérêt limité pour les questions traitées au niveau mondial et plus particulièrement en France. On constate également un nombre très variable d'articles en fonction des différentes questions abordées. Par exemple, le nombre d'articles traitant de l'auto-évaluation est nettement plus élevé que le nombre d'articles s'intéressant à l'hétéro-évaluation, ce qui complique évidemment la comparaison de l'efficacité de ces deux formes d'évaluation de l'intelligence.

Nous aborderons en premier lieu les mesures auto-rapportées de l'intelligence, en distinguant l'auto-estimation de l'auto-évaluation, puis les mesures hétéro-rapportées qui comprennent les hétéro-estimations et les hétéro-évaluations. Lorsque les études existent et le permettent, nous indiquerons les facteurs qui influencent les mesures, les preuves de validité et de fidélité.

2.1 Les mesures auto-rapportées de l'intelligence

Ces dernières années s'est développé un intérêt pour les estimations subjectives de l'intelligence (Chamorro-Premuzic & Furnham, 2006 ; Furnham, 2001 ; Holling & Preckel, 2005 ; Visser, Ashton & Vernon, 2008). Dans ce champ des mesures auto-rapportées, de multiples définitions existent (Freund & Kasten, 2012) et il convient de distinguer l'auto-estimation (ou auto-perception) de l'auto-évaluation de l'intelligence. La première peut être définie comme la perception qu'a une personne de ses propres capacités, de son « intelligence » (Chamorro-Premuzic & Furnham, 2006). Elle est mesurée en demandant directement au sujet d'estimer son niveau d'intelligence, en utilisant un ou plusieurs items avec, à titre d'exemple, la consigne suivante : *«* Estimez votre intelligence verbale *»*. L'auto-évaluation, elle, permet d'évaluer la capacité perçue par le sujet lui-même de réaliser certaines tâches comme se repérer sur un plan, suivre les conversations des autres ou se souvenir où des choses sont rangées. Les items ne réfèrent pas explicitement à l'intelligence, mais font référence aux compétences du sujet dans diverses tâches cognitives. Ces mesures sont considérées comme indirectes et permettraient de contourner les éventuelles défenses du sujet activées par des questions directes comme dans l'auto-estimation (Paulhus, Lysy & Yik, 1998).

Les travaux sur les mesures auto-rapportées de l'intelligence soulignent des avantages à leur emploi pour favoriser le développement de la connaissance de soi, notamment dans le cadre d'un choix de carrière (Ackerman & Wolman, 2007). Un sujet qui sous-estime ses capacités peut, en effet, renoncer à des

opportunités professionnelles et, à l'inverse, un individu qui se surévalue peut viser des objectifs difficilement atteignables. L'écart entre la mesure subjective et les résultats aux tests de performance peut être utile pour le conseiller dans une démarche d'accompagnement, reflétant le degré de connaissance de soi de la personne et le réalisme au sujet de ses capacités, deux points essentiels dans un choix de carrière (Freund & Kasten, 2012).

La mesure auto-rapportée des capacités cognitives est également pratiquée en neuropsychologie. Elle permet de donner une idée de l'impact écologique du dysfonctionnement cognitif, d'évaluer la capacité d'introspection et de décrire des capacités cognitives au plus proches de la vie quotidienne (Banos, Lagory, Sawrie, Faught, Knowlton, Prasad, Kuzniecky & Martin, 2004).

Dans tous les domaines où ils sont utilisés, les questionnaires d'auto-estimation ou d'auto-évaluation sont généralement considérés comme ayant des avantages pratiques : peu onéreux, rapides à administrer, flexibles dans leur utilisation (Jacobs & Roodenburg, 2014). D'un autre côté, ils sont décrits comme ayant des avantages pour les personnes évaluées : ils diminuent l'anxiété (Furnham, 2001 ; Paulhus, Lysy & Yik, 1998 ; Simms, 2008), permettent d'investiguer des comportements difficilement observables (Tarrant & Cordell, 1997) et bénéficient d'une validité de surface généralement plus importante que les tests classiques et sont source de meilleure adhésion et coopération.

À côté de ces avantages, les mesures auto-rapportées présenteraient, néanmoins, plusieurs inconvénients. Le premier, le plus souvent mis en avant, est en lien avec la question de la subjectivité. Les personnes auraient des difficultés à produire des

jugements objectifs (Meyer, Finn, Eyde, Kay, Moreland, Dies, Eisman, Kubiszyn & Reed, 2001), seraient sujets à des distorsions ou des reconstructions (Hofstee, 1994), et auraient tendance à surestimer leurs compétences (Barr, 2007 ; Fuj, Hennessy & Mak, 1985 ; Geller, 1981 ; Warriner, McDougall & Claxton, 1984). Par ailleurs, ces difficultés seraient encore plus prégnantes chez des personnes ayant des troubles des fonctions exécutives (Spinella, Yang & Lester, 2004), une mémoire limitée ou de faibles connaissances (Warriner et al., 1984), un trouble de l'attention avec ou sans hyperactivité, une dépression ou des troubles psychiatriques, manquant de capacités méta-cognitives (Cyders & Coskunpinar, 2011 ; Ehrlinger & Dunning, 2003) ou encore ayant de faibles capacités cognitives (Kruger & Dunning, 1999).

2.1.1 L'auto-estimation

L'auto-estimation est la méthode d'évaluation qui a fait l'objet du plus grand nombre de recherches, comparativement à l'auto-évaluation définie plus haut. Nous nous intéressons, dans un premier temps, aux facteurs susceptibles d'influencer les auto-estimations.

2.1.1.1 Facteurs influençant les auto-estimations

L'auto-estimation, comme mesure dite subjective, peut présenter des variations selon différents facteurs que nous allons décrire.

Le premier, étudié à de maintes reprises, est le sexe. Les résultats semblent contradictoires selon les études. Pour certaines, l'auto-évaluation de l'intelligence générale varie selon le sexe (Bennett, 1996 ; Beloff, 1992 ; Byrd & Stacey, 1993 ; Furnham & Rawles, 1995 ; Hogan, 1978 ; Zhang & Gong, 2001) avec des scores auto-estimés plus élevés pour les garçons, alors que d'autres ne constateraient pas de différence à ce niveau, mais seulement selon les types d'intelligence. Les garçons se jugeraient supérieurs dans les domaines logique, mathématique et spatial et les filles dans les domaines de la musique et des relations interpersonnelles (Bennett, 2000 ; Furnham, Clark & Bailey, 1999 ; Furnham, Fong & Martin, 1999). Ces résultats se retrouveraient dans différents pays : Chine (Zhang & Gong, 2001), Allemagne (Rammstedt & Rammsayer, 2000), Écosse (Bennett, 2000).

En référence au modèle de Cattell, Horn, Carroll (CHC ; Newton & McGrew, 2010), Kaufman (2012) observe que les filles se considèrent moins fortes en intelligence cristallisée (Gc) et en intelligence fluide (Gf). Selon Furnham et Mansi (2014), on trouverait des différences au profit des garçons sur huit des dix compétences intellectuelles qu'ils ont évaluées : l'intelligence fluide, l'intelligence quantitative, la mémoire de travail, le traitement visuel, la vitesse de traitement, le traitement auditif, la vitesse de réaction et l'écriture/lecture.

Selon une méta-analyse de Szymanowicz et Furnham (2011), réalisée sur ce sujet, la plus grande taille d'effet apparaît dans les différences entre garçons et filles

pour les intelligences logique et mathématique (d[5]=.44), suivie de l'intelligence spatiale (d=.43), de l'intelligence générale (d=.37) et l'intelligence verbale (d=.07), avec des scores auto-estimés plus élevés pour les garçons dans tous les domaines, excepté l'intelligence verbale, pour laquelle les filles s'estiment plus fortes.

Un second facteur susceptible d'influencer les auto-estimations est la personnalité. L'auto-estimation de l'intelligence présenterait des corrélations positives avec l'extraversion et l'ouverture et des liens négatifs avec le névrosisme et l'agréabilité (Chamorro-Premuzic, Moutafi & Furnham, 2005 ; Furnham, Kidwai & Thomas, 2001 ; Furnham, Moutafi & Chamorro-Premuzic, 2005). Les personnes qui s'estiment un niveau de conscience élevé se notent, en moyenne, plus faiblement sur les tests d'intelligence et celles avec une ouverture auto-estimée élevée tendent à s'auto-estimer de manière plus importante (Chamorro-Premuzic, Furnham & Moutafi, 2004). Les tests de personnalité expliqueraient ainsi de 8 à 17% des auto-estimations (Furnham & Dissou, 2007 ; Furnham & Thomas, 2004). À titre d'exemple, dans l'étude de Chamorro-Premuzic et al. (2005), l'auto-estimation intellectuelle est prédite négativement par le névrosisme (β=-.23 ; t=-2.37 ; p=.02) et l'agréabilité (β=-.23 ; t=-2.68 ; p<.01), et positivement par l'ouverture (β=.26 ; t=2.83 ; p<.01).

5. Le d de Cohen permet de mesurer une taille d'effet, qui est la force de l'effet observé d'une variable sur une autre. Traditionnellement, un d autour de 0,2 est décrit comme un effet « faible », 0,5 « moyen » et 0,8 comme « fort ».

2.1.1.2 Corrélations avec les tests de performance

Dans cette section, nous nous intéressons en premier lieu aux études de validité convergente avec les tests de performance. Le but est de répondre à notre question initiale, qui est de savoir si les auto-estimations sont une alternative aux tests de performance ou un complément, si elles évaluent le même concept.

Mabe et West (1982) dans leur méta-analyse à partir de 55 études trouvent une corrélation moyenne entre l'auto-estimation et les tests de performance .29 (valeur de p non précisée). Mais ils ont réuni des études très diverses, qui mêlent l'auto-estimation, l'auto-évaluation et différentes compétences. Lorsque ces études sont uniquement centrées sur les capacités cognitives, la corrélation moyenne est de .34 (p-value non précisée).

Freund et Kasten (2012), dans une méta-analyse portant sur 41 études, observent, pour leur part, une corrélation moyenne de .32 ($p<.001$). Ces corrélations modérées semblent indiquer que l'auto-estimation n'est pas redondante avec l'évaluation des performances et ne peut donc remplacer cette dernière (Holling & Preckel, 2005).

Il est cependant possible d'améliorer les corrélations de l'auto-estimation avec les tests de performance en introduisant, par exemple, une comparaison sociale dans la consigne d'auto-estimation ou encore en créant des items portant spécifiquement sur les capacités évaluées par les tests de performance.

2.1.1.3 La comparaison sociale

Différentes consignes sont utilisées dans le cadre de la recherche sur les questionnaires d'auto-estimation. Pour rappel, il y est demandé aux sujets d'évaluer leur niveau d'intelligence. Holling et Preckel (2005) distinguent 1° l'estimation directe, dans laquelle les sujets donnent un chiffre pour estimer leur intelligence comparé à la distribution normale (Bennett, 1996 ; Furnham & Rawles, 1999 ; Hogan, 1978) ; 2° l'utilisation de rang percentile (Bailey & Mettetal, 1977 ; Bossong, 1982) ; 3° l'usage d'échelles visuelles analogiques, divisées en six segments (de -3 extrêmement faible à +3 extrêmement élevé, la moyenne étant 0) (Rammstedt & Rammsayer, 2000) ; 4° ou encore l'utilisation d'échelles de Likert de 5 à 9 points, allant de « pas du tout intelligent » à « très intelligent », sans comparaison sociale (Borkenau & Liebler, 1993 ; Fingerman & Perlmutter, 1994 ; Paulhus et al., 1998). Les deux premières méthodologies proposent une comparaison sociale du fait de la référence à la distribution normale dans le premier cas et du rang percentile dans le second. Les étiquettes des échelles visuelles analogiques activent, selon Holling et Preckel (2005), également une comparaison sociale. Ces trois méthodologies sont celles, selon Mabe et West (1982), qui permettent d'obtenir les corrélations les plus importantes avec les tests d'intelligence. Les résultats de la méta-analyse de Freund et Kasten (2012) aboutissent au même constat. Ces auteurs précisent, en outre, que, dans le cas d'une comparaison sociale, les corrélations sont encore

plus élevées lorsqu'un groupe de référence spécifique est introduit (par exemple, les élèves de la classe ou les étudiant.e.s de la promotion).

2.1.1.4 Le contenu des items

D'après l'étude de Freund et Kasten (2012), les corrélations entre l'auto-estimation et le test de performance seraient plus fortes quand on demande au sujet de s'auto-estimer directement en fonction du test utilisé, c'est-à-dire quand les items portent sur le contenu du test de performance utilisé. Ainsi, Furnham, Zhang et Chamorro-Premuzic (2005-2006) observent des corrélations de .29 ($p<.05$) entre l'auto-estimation sur la base des performances aux Matrices de Raven et les performances objectives à ce test ; des corrélations de .47 ($p<.01$) entre l'auto-estimation sur la base des performances au *Wonderlic Personnel Test* (Wonderlic, 1992) et les performances objectives à ce test et, enfin, des corrélations de .32 ($p<.05$) entre l'auto-estimation sur la base des performances aux *Baddeley Reasoning Test* (Baddeley, 1968) et les performances objectives à ce test. Jacobs et Roodenburg (2014) précisent que les corrélations excèdent généralement .30, quand les auto-estimations sont faites sur des capacités spécifiques, avec des mesures dans les domaines correspondants (Chamorro-Premuzic et al., 2004 ; Fingerman & Perlmutter, 1994 ; Furnham & Dissou, 2007 ; Visser, Ashton & Vernon, 2008).

2.1.1.5 Le nombre d'items

Un dernier facteur étudié, pouvant améliorer les corrélations avec les mesures objectives, est le nombre d'items. Dans la méta-analyse réalisée par Freund et Kasten (2012) sur 41 études et 154 tailles d'effet, 73,4% des auto-estimations sont faites avec un seul item (par exemple : « je suis aussi intelligent que les autres jeunes de mon âge »), 11% avec plusieurs items. Pour 15,6% d'entre elles, aucune information sur le nombre d'items n'a été donnée. Les résultats apparaissent assez contradictoires selon les auteurs. Pour certains, les corrélations avec les tests de performance semblent meilleures quand plusieurs items sont utilisés (Beyer, 1998) ; pour d'autres, aucune différence ne serait observée (Paulhus et al., 1998). Ces recherches montrent ainsi que la comparaison sociale avec un groupe spécifique, ainsi que des auto-estimations centrées sur les tests, conduisent à des corrélations plus fortes avec des tests psychométriques classiques. En revanche, les résultats divergent quant à l'utilisation d'un item unique ou d'items multiples.

2.1.1.6 Corrélations avec les résultats scolaires

Plusieurs études ont établi l'existence d'une relation entre l'auto-estimation de l'efficience cognitive et les résultats scolaires (Kornilova & Novikova, 2013). Dans la méta-analyse réalisée par Kornilov (2011), sur neuf études recensées, une corrélation moyenne positive et significative de .28 (valeur de p non précisée) est trouvée. Kornilova et Novikova (2013) observent une corrélation de .16 entre

l'auto-estimation et la réussite scolaire ($p<.01$). Bien que significatives, les corrélations entre auto-estimation et réussite scolaire restent faibles et moins élevées que celles entre les tests de performance et la réussite scolaire, qui sont généralement comprises entre .40 et .50 (Sternberg, Grigorenko & Bundy, 2001).

2.1.1.7 Validité prédictive

Très peu de chercheurs se sont intéressés à la validité prédictive des auto-estimations. Dans notre revue de littérature, nous avons relevé une seule recherche, celle de Kornilova, Kornilov et Chumakova (2009), auprès d'étudiant.e.s russes, centrée sur la validité prédictive de la réussite universitaire. En complément de l'âge, du sexe et des résultats aux tests de performance, l'auto-estimation ajouterait 3% dans la prédiction de la variance des résultats universitaires six mois après la passation des tests (Kornilova et al., 2009). Selon cette seule étude, l'auto-estimation permettrait d'augmenter la prédiction des résultats scolaires, même si cette augmentation reste très faible.

2.1.2 L'auto-évaluation

Les recherches sur l'auto-évaluation des capacités cognitives ont été beaucoup moins nombreuses que pour l'auto-estimation. Nous recensons seulement quelques questionnaires que nous allons détailler, en nous attachant, en

particulier, aux études psychométriques, qui permettent de répondre à nos questions initiales.

La plupart des questionnaires proposent plusieurs items. Un seul item est, en effet, insuffisant pour limiter la variance d'erreur et garantir la validité de contenu (Epstein, 1983 ; Hoyt, Warbasse & Chu, 2006). Ces échelles reposent, par ailleurs, sur des théories sous-jacentes différentes : le modèle CHC (Jacobs & Roodenburg, 2014), ou celui des intelligences multiples de Gardner (1983, 1999, 2006) (Tirri & Nokelainen, 2008), ou semblent athéoriques (Gignac, Stough & Loukomitis, 2004 ; Rueda, Lau, Saito, Harvey, Risacher, Aisen, Petersen, Saykin, & Farias 2015 ; Seidenberg, Haltiner, Taylor, Hermann & Wyler, 1994). Enfin, certaines ne mesurent qu'un seul domaine cognitif, comme les capacités spatiales (Quaiser-Pohl & Lehmann, 2002). Elles ont, cependant, en commun de proposer des énoncés, sur lesquels les sujets doivent se situer (échelles de Likert), de type « comprendre les conversations des autres » (intelligence cristallisée), « trouver une solution à un problème jamais expérimenté » (intelligence fluide), « lire une carte routière » (traitement visuel).

Un examen des différents questionnaires d'auto-évaluation, que nous avons recensés depuis 1990 (voir tableau 2), montre d'assez bonnes qualités psychométriques de ces derniers (cohérence interne notamment), lorsque les données sont disponibles.

Études	Test	Échantillon	Domaine(s) évalué(s)	Cohérence interne	Corrélations test-retest	Corrélations avec des mesures objectives
Gignac, et al. (2004)	Self-Report Intelligence Questionnaire (SRIQ) 9 items	N= 62 Etudiants	Les items se rapportent aux subtests des échelles de Wechsler (vocabulaire, similitudes…) et un item global	α =.73 (échelle globale)		r=.30 entre le score total à l'auto-évaluation et le facteur g (p<.05).
Jacobs et Roodenburg (2014)	Self-Report Measure of Cognitive Abilities (SRMCA) 19 items	N= 253 Adultes	Intelligence cristallisée (Gc), fluide (Gf) et traitement visuel (Gv)	α =.86 (Gf) α =.84 (Gc) α =.89 (Gv)		Une seule corrélation significative avec le Woodcock-Jonhson Tests of Cognitive Abilities-third edition (WJ III cog ; Woodcock, McGrew & Mather, 2001) de .37 (p<.001)
Quaiser-Pohl et Lehmann (2002)	Spatial ability in everyday life questionnaire 11 items	N= 183 Etudiants	Capacités spatiales	α =.72 Guttman's Split Half=0,80		r=.52 (p<.001) avec le Mental Rotations Test (MRT).
Rueda et al. (2015)	Everyday Cognition questionnaire (Ecog) 39 items	N= 1080 Adultes	Six domaines : mémoire quotidienne, langage, capacités visuo-spatiales, planification, organisation, attention divisée	α =.97 (échelle globale)	r=.82 (p<.001).	Pas de corrélations significatives avec des tests neuropsychologiques.

Seidenberg, et al. (1994)	Multiple Abilities Self Report Questionnaire (MASQ) 38 items	N= 118 Adultes	Cinq domaines cognitifs sont évalués : le langage, le visuo-perceptif, la mémoire verbale, la mémoire visuelle et l'attention-concentration	L'alpha du score global est de .92 et sont tous supérieurs à .70 pour les différentes échelles	La corrélation test-retest est de .71 pour le score global et varie de .55 (langage) à .74 (mémoire visuelle) pour les scores dans chaque domaine (valeurs de p non précisées)	Corrélations entre le MASQ et les résultats obtenus à la WAIS et la WMS (Wechsler, 2009) de -.29 ($p<.05$) pour le facteur verbal à .36 ($p<.01$) pour la mémoire verbale immédiate.
Tirri et Nokelainen (2008)	Multiple Intelligences Profiling Questionnaire (MIPQ III) 35 items	N= 410 Pré-adolescent.e.s (N= 183) Adultes (N= 227)	Évalue 9 dimensions : les intelligences verbale, logico-mathématique, musicale, spatiale, corporelle-kinesthésique, interpersonnelle, intra-personnelle, spirituelle et environnementale	Les alphas varient de .53 (intelligence spatiale) à .83 (intelligence musicale)		

Tableau 2. Questionnaires d'auto-évaluation des capacités cognitives, présentation et qualités psychométriques.

Les corrélations avec des tests classiques de mesure des capacités cognitives restent variables d'une étude à l'autre et généralement faibles ou modérées, ce qui confirmerait que ni l'auto-évaluation, ni l'auto-estimation ne peuvent remplacer les tests d'intelligence classiques. Nous pouvons, toutefois, noter que, dans le domaine spatial, l'auto-évaluation est davantage liée aux performances spatiales.

Deux études, ont cherché à comparer auto-évaluation et auto-estimation, afin de savoir laquelle présentait les meilleures corrélations avec un test de performance. Jacobs et Roodenburg (2014) ont proposé le *Self-Report Measure of Cognitive Abilities*[6] (SRMCA), ainsi qu'une auto-estimation reposant sur une comparaison sociale (distribution normale) et les *Woodcock-Jonhson Tests of Cognitive Abilities Third Edition*[7] (WJ III Cog ; Woodcock et al., 2001). Trois corrélations sont significatives entre les performances et l'auto-estimation : .52 ($p<.001$) entre Gc (intelligence cristallisée) du test et Gc de l'auto-estimation ; .28 ($p<.001$) pour Gf (intelligence fluide) et .24 ($p<.001$) pour Gv (traitement visuel), alors que, seules, deux corrélations sont significatives entre l'auto-évaluation et les performances (.37 ($p<.001$) pour Gc et .25 ($p<.001$) pour Gv). Ils concluent que l'auto-évaluation ne permet pas d'obtenir des estimations valides de l'intelligence.

Contrairement aux études réalisées sur l'auto-estimation, aucune étude ne l'a été sur les facteurs influençant l'auto-évaluation, ni sur les corrélations entre cette dernière et la réussite scolaire, ni sur les facteurs permettant d'améliorer les corrélations avec les performances, ni, enfin, sur la validité prédictive.

[6] Mesure auto-rapportée des capacités cognitives
[7] Tests de Woodcock-Johnson des capacités cognitives, 3e éd.

2.2 Les évaluations hétéro-rapportées

Comme pour le domaine de l'auto-évaluation, il est nécessaire de distinguer l'hétéro-estimation de l'hétéro-évaluation. La première consiste à demander à une personne dans l'entourage du sujet (parent, pair ou enseignant.e) d'estimer son intelligence ; la seconde consiste à demander à son entourage d'évaluer sa capacité de réaliser des tâches cognitives.

Les mesures hétéro-rapportées sont une technique très largement utilisée dans la perception des symptômes (Barkley, Knouse & Murphy, 2011), la description de pathologies chez l'enfant (De Los Reyes & Kazdin, 2005) ou encore de troubles des apprentissages (Cole, Maxwell & Martin, 1997). Les recherches ont apporté les preuves de la validité des hétéro-évaluations réalisées par les parents dans différents domaines (Achenbach, McConaughy & Howell, 1987 ; Fabrega, Ulrich & Loeber, 1996 ; Hart, Lahey, Loeber & Hanson, 1994 ; Loeber, Green, Lahey & Stouthamer-Loeber, 1990, 1991), mais très peu sur la performance cognitive (Oliver et al., 2002). En outre, certains auteurs (Rushton, Brainerd & Pressley, 1983) ont suggéré qu'agréger des scores issus de différents rapporteurs sur un même objet pouvait permettre d'améliorer la validité de la mesure, en capturant une plus grande part de variance vraie.

Les hétéro-évaluations menées par les parents présentent plusieurs avantages (Waschbusch, Daleiden & Drabman, 2000) : les parents sont les premiers présents auprès de l'enfant ou de l'adolescent ; ils l'observent dans de nombreuses

situations et de manière beaucoup plus fréquente qu'un.e observateur.trice extérieure.

2.2.1 L'hétéro-estimation

2.2.1.1 Facteurs influençant les hétéro-estimations

Très peu d'études ont été réalisées sur l'hétéro-estimation de l'intelligence et ont été principalement axées sur les facteurs influençant leur mesure. Deux d'entre eux ont été particulièrement étudiés : le sexe de l'enfant et le sexe des parents. Pour le premier facteur, le sexe de l'enfant, les études montrent que les parents donnent des scores plus élevés aux garçons par rapport aux filles, surtout dans le domaine logico-mathématique (Furnham, 2000). Quant au second facteur, le sexe du parent, les mères font des estimations plus faibles pour leur fille que pour leur fils dans le domaine spatial (d=.40) (Pérez, González & Beltrán, 2010).

2.2.1.2 Corrélations avec les tests de performance, les résultats scolaires et validité prédictive

Une seule étude (Kornilova, Kornilov & Chumakova, 2009) a porté sur les corrélations entre l'hétéro-estimation de l'intelligence, les tests de performance et les résultats scolaires, ainsi qu'à la validité prédictive. L'hétéro-estimation a été évaluée en demandant aux étudiant.e.s de classer chacun de leurs camarades de

promotion, par rapport à leur niveau d'intelligence. Un rang moyen pour chaque étudiant.e a été calculé. Les corrélations avec l'*Intelligenz-Structur-Test* (IST-70 ; Amthauer, 1973) sont toutes significatives : de -.22 ($p<.01$) pour l'intelligence spatiale à -.37 ($p<.001$) pour l'intelligence générale. Les corrélations avec les résultats universitaires sur trois trimestres sont de -.65 ($p<.01$). Enfin, l'hétéro-estimation ajoute 24 % dans la prédiction des résultats universitaires à la fin du semestre. Elle semble ainsi montrer de meilleures preuves de validité prédictive que l'auto-estimation, qui ajoutait seulement 3% dans la prédiction même si d'autres études seraient nécessaires pour le confirmer.

Aucune recherche n'a étudié les facteurs susceptibles d'améliorer les corrélations avec les tests de performance.

2.2.2 L'hétéro-évaluation

À notre connaissance, seuls trois questionnaires d'hétéro-évaluation des capacités cognitives sont parus depuis 25 ans : ceux élaborés par Gioia, Isquith, Guy et Kenworthy (2000), Oliver et al. (2002) et Waschbusch, Daleiden et Drabman (2000). Les principales données psychométriques de ces questionnaires sont présentées ci-dessous (voir tableau 3).

Étude	Test	Échantillon	Domaine(s) évalué(s)	Cohérence interne	Corrélations inter-cotateurs	Corrélations avec des mesures objectives
Gioia et al., 2000.	BRIEF 86 items	N= 951	Fonctions exécutives 8 échelles : inhibition, flexibilité, contrôle émotionnel, initiation, mémoire de travail, planification/organisation, organisation du matériel et contrôle.	Pour l'échelle globale : α =.86 (questionnaire parent) α =.92 (questionnaire enseignant.e)	La fidélité inter-cotateur est de .50 ; test-retest de .85 pour les parents et de .86 pour les enseignant.e.s (valeurs de p non précisées)	
Oliver et al. (2002)	Parent Report of Children's Abilities (PARCA 3) 22 items	N= 85	Capacités spatiales, planification et organisation, raisonnement, mémoire, raisonnement quantitatif, nombres et jeux symboliques	α =.67		r=.31 (p<.01) à .49 (p<.01) avec la McCarthy Scales of Children's Abilities (McCarthy, 1972)
Waschbusch et al., (2000)	25 items	N= 145	Six échelles en référence au modèle CHC : intelligences cristallisée (Gc), fluide (Gf), traitement visuel (Gv), auditif (Ga), mémoire (Gsm) et globale	α =.89 (échelle globale) α =.60 (Gf) α =.66 (Gc) α =.74 (Gv) α =.68 (Ga) α =.76	Vont de .24 (p<.05) à .83 (p<.001) pour les garçons Et de .09 (NS) à .83 (p<.001) pour les	Les corrélations avec les tests (WISC et le Woodcock-Jonhson Psycho-Educational Battery – Revised, WJR ;

			(Gsm) filles	Woodcock, 1989) vont de .08 (NS) pour la mémoire à .54 pour Gv (p<.001) chez les filles et de .26 (p<.05) pour Gv à .54 (p<.001) pour l'aptitude générale chez les garçons

Tableau 3. Questionnaires d'hétéro-évaluation des capacités cognitives, présentation et qualités psychométriques.

Les questionnaires d'hétéro-évaluation semblent présenter certaines qualités psychométriques satisfaisantes, telles que la cohérence interne, mais des preuves de validité demandent encore à être récoltées. Les corrélations avec des mesures de performances sont faibles ou modérées, ce qui semble indiquer que les hétéro-évaluations ne peuvent pas, là encore, remplacer les tests d'intelligence classiques. Ces trois recherches n'ont pas étudié les corrélations avec la réussite scolaire, ni les preuves de validité prédictive.

Discussion

Cette revue de littérature avait pour objectif d'évaluer si les mesures auto et hétéro-rapportées peuvent être utilisées à la place des tests d'intelligence ou pour les compléter, si elles évaluent le même construit sous-jacent et, enfin, le degré de validité de ces méthodes.

À partir des différentes études recensées, pour la très grande majorité réalisées en dehors de France, il apparaît que les mesures auto et hétéro-rapportées ne peuvent être substituées aux tests classiques d'intelligence, fondés sur la mesure de performances, les corrélations avec ces derniers étant de faibles à modérées. Elles mesureraient des construits distincts, essentiellement représentationnels. Les mesures auto-rapportées seraient davantage l'expression du concept de soi, défini comme une conception composite d'une personne sur elle-même, formée par ses expériences et l'interprétation de son environnement (Kornilova et al., 2009 ; Shavelson, Hubner & Stanton, 1976). Le concept de soi est influencé par les évaluations faites par les autres, les renforcements et les attributions. Shavelson et al. (1976) ont dégagé plusieurs caractéristiques du concept de soi : 1° il possède une dimension descriptive (« je suis malheureux ») et une dimension évaluative (« je suis bon en français ») ; 2° il est multidimensionnel : les visions que l'individu a de lui-même sont organisées en différentes catégories ; 3° il est organisé hiérarchiquement avec, au sommet de la hiérarchie, le concept de soi global, à la base, les perceptions relatives à un comportement particulier et, au milieu, les inférences émises dans un domaine plus large. Dans le champ de

l'intelligence, les questionnaires d'auto-estimation et d'auto-évaluation renverraient à un concept de soi cognitif (Ackerman, Bowen, Beier & Kanfer, 2001). Ce dernier serait influencé par les traits de personnalité, les stéréotypes de sexe et le niveau d'intelligence. En effet, concernant les traits de personnalité, nous avons vu que l'auto-estimation de l'intelligence présentait des corrélations positives avec l'extraversion et l'ouverture et des liens négatifs avec le névrosisme et l'agréabilité (Chamorro-Premuzic et al., 2005 ; Furnham et al., 2001 ; Furnham et al., 2005). Les personnes extraverties sont plus actives, confiantes et optimistes et l'ouverture est un trait lié à l'intelligence générale (Ackerman & Heggestad, 1997 ; Judge, Jackson, Shaw, Scott & Rich, 2007), ce qui peut expliquer des corrélations positives avec l'auto-estimation de l'intelligence. À l'inverse, le névrosisme implique une estime de soi plus faible, qui affecte la confiance en ses aptitudes et l'agréabilité, est généralement liée à la modestie (Costa & McCrae, 1992), pouvant ainsi altérer l'estimation que font les personnes de leur degré d'intelligence. Quant aux différences liées au sexe, systématiquement relevées dans les auto-estimations, elles seraient le reflet des stéréotypes associés au sexe (Beloff, 1992). Les mesures hétéro-rapportées évalueraient une représentation d'autrui pouvant être fondée sur des appréciations factuelles du point de vue de l'efficience et sur des comparaisons sociales, et pouvant être marquée par les représentations sociales associées et sujettes à divers biais de perception.

Ces mesures auto et hétéro-rapportées semblent cependant présenter certaines qualités psychométriques du point de vue de la validité et de la fidélité, mais

l'étude de celles-ci mériterait d'être approfondie. À titre d'exemple, très peu d'études indiquent les corrélations test-retest ou encore les corrélations item-test corrigées. D'un autre côté, les questionnaires pourraient être améliorés en se référant à des modèles validés scientifiquement, comme le modèle CHC, et en étant testés sur de plus larges échantillons. Les futures études devraient comparer l'auto-estimation (en utilisant la comparaison sociale), l'hétéro-estimation sur des domaines spécifiques évalués par les tests de performance ; l'auto-évaluation et l'hétéro-évaluation, en utilisant un modèle théorique solide, et étudier les corrélations avec des tests de performance, mais, également, avec des tests dits écologiques. Il semblerait, en effet, qu'il y ait des corrélations plus élevées entre les questionnaires auto et hétéro-rapportés avec les évaluations de type écologique, les deux se référant à des situations de la vie quotidienne, alors que les tests psychométriques classiques font référence à des situations nouvelles. À ce sujet, Paulhus, Lysy et Yik (1998) conseillent de développer de nouveaux tests d'intelligence, qui prennent en compte les conceptions quotidiennes de l'intelligence.

Nous retiendrons également que ces autres formes d'évaluation et, en particulier, l'hétéro-estimation, permettent, en complément des tests classiques d'intelligence, d'augmenter la part de variance prédite de la réussite scolaire, au même titre que d'autres variables conatives, comme la motivation, le sentiment d'efficacité personnelle, etc. D'un autre côté, elles sont particulièrement intéressantes à utiliser pour confronter les perceptions du sujet ou celles de son entourage aux données objectives fournies par le test (évaluation à 360°). Comme nous l'avions

souligné dans l'introduction, ces formes d'évaluation peuvent être particulièrement utiles dans le cadre des pratiques d'évaluation associées à l'orientation. Il semble alors important de repérer les personnes qui se sous-évaluent ou se sur-évaluent et de les assister dans le développement de la connaissance d'elles-mêmes, car le réalisme à propos de ses propres aptitudes et la clairvoyance du concept de soi sont deux facteurs importants pour une décision de carrière réussie (Freund, Kasten, 2012).

Synthèse du cadre théorique

Le premier chapitre se proposait de faire un état des lieux de l'évaluation contextualisée de l'intelligence, sa définition, sa mesure et les relations entretenues avec l'évaluation classique. Le second a développé plus particulièrement les évaluations auto et hétéro rapportées de l'intelligence. Nous avons fait appel à différents courants de la psychologie : cognitive, du travail et des organisations, neuropsychologie. Tous ces courants se sont posé la question de l'évaluation des capacités cognitives au plus proche des situations réelles, telle qu'est définie l'évaluation contextualisée. Nous avons repéré différents outils et fait le choix de les classer sur un continuum, selon le degré de contextualisation, allant des tests papier-crayon (auto et hétéro-évaluations, tests de jugement situationnel, tests de performance) aux situations dites naturelles en passant par les mises en situation (en utilisant du matériel concret, des supports vidéo ou la réalité virtuelle). A l'un des extrêmes de ce continuum, le contexte peut être totalement absent, alors qu'à l'autre extrémité, il est intégré totalement dans la procédure d'évaluation (la personne étant évaluée dans son contexte). Plus ce niveau de contextualisation augmente, plus les processus évalués deviennent complexes. Les outils contextualisés restent très peu nombreux et presque exclusivement conçus pour des adultes. Dans les tests papier-crayons, nous avons repéré les TJS utilisés en ressources humaines. Les TJS évaluant les connaissances présentent des corrélations très modestes avec les tests

d'intelligence (McDaniel et al., 2007) et les performances dans le travail (McDaniel et al., 2001). Nous avons également décrit le STAT de Sternberg, non publié et faisant l'objet de vives critiques (Brody, 2003 ; Gottfredson, 2003 ; Koke & Vernon, 2003). Enfin, quelques outils dans le domaine de la neuropsychologie ont été élaborés. Les seules études sont très ciblées en termes de population et de capacité cognitive évaluée et montrent chez des sujets ayant une sclérose en plaques que les tests contextualisés sont de meilleurs prédicteurs d'une déficience fonctionnelle (mesurée par un entretien) que les tests traditionnels de mémoire et d'attention (Higginson et al., 2000 ; Makatura Lam et al. ,1999). Enfin, dans les tests papier-crayon, nous avons également inclus les auto et hétéro-évaluations de l'intelligence qui ne sont pas des tests de performance comme les autres évoqués dans cette partie théorique. Ils sont basés sur une autre méthodologie (questionnaire). Les corrélations avec les tests classiques d'intelligence sont faibles mais elles pourraient augmenter en introduisant la comparaison sociale dans la consigne (Freund & Kasten, 2012 ; Mabe & West, 1982) et en proposant des items en lien avec les capacités évaluées par les tests de performance (Jacobs & Roodenburg, 2014). Les outils recensés présentent certaines qualités psychométriques du point de vue de la validité et de la fidélité, mais l'étude de celles-ci mériterait d'être approfondie. Ils permettent, en complément des tests classiques d'intelligence, d'augmenter la part de variance prédite de la réussite scolaire, au même titre que d'autres variables conatives, comme la motivation, le sentiment d'efficacité personnelle, etc.

Les tests de simulation comprennent les tests « in-basket », les TJS sous format vidéo et la réalité virtuelle. Les deux premiers types sont nombreux et ont fait l'objet de méta-analyses. Celle de Whetzel, Rotenberry et McDaniel (2014) montre que les tests « in-basket » présentent une corrélation faible avec des tests de facteur g et une validité prédictive modeste au regard de la performance dans le travail. Les TJS sous format vidéo ont quant à eux, selon Christian, Edwards & Bradley (2010), une validité critériée moyenne de .36, plus importante que celle des tests papier-crayon qui est de .25, pour l'ensemble des construits mesurés (leadership, compétences inter-personnelles…). Enfin, toujours dans les outils de simulation, seulement quelques outils de réalité virtuelle existent pour évaluer les capacités cognitives. Une étude (Pugnetti et al., 1998) montre que le test de réalité virtuelle corrèle peu avec le test papier-crayon même les deux ont les mêmes consignes et sont basés sur un principe identique.

Enfin, concernant les évaluations en situation naturelle, nous n'avons repéré qu'un seul test (le P2CJP ; Pereira Da Costa & Scelles, 2012) mais les auteurs ne précisent pas les corrélations avec d'autres outils d'évaluation cognitive.

A l'issue de cette première partie, nous pouvons faire quelques constats : l'approche contextualisée est principalement développée dans le champ des ressources humaines et tend à l'être dans celui de la neuropsychologie. Elle reste très marginale dans le domaine de la psychologie cognitive. Les tests sont donc de fait destinés à être utilisés avec un public adulte et/ou avec une pathologie (neuropsychologie). Les quelques études réalisées, difficilement généralisables,

ont toute en commun de montrer que les liens entre l'évaluation classique et l'évaluation contextualisée sont modestes voire faibles. En ce qui concerne la validité écologique et/prédictive, les études sont tellement peu nombreuses qu'il parait difficile de tirer des conclusions. Dans le domaine des ressources humaines, la prédiction de la performance au travail ne semble pas meilleure qu'avec des tests classiques d'intelligence. Par contre, les deux combinés augmentent cette prédiction. Les auto et hétéro-évaluations permettent, en complément des tests classiques d'intelligence, d'augmenter la part de variance prédite de la réussite scolaire. En neuropsychologie, les quelques rares études montrent que les tests écologiques prédisent mieux les problèmes en vie quotidienne que les tests classiques (Higginson et al., 2000 ; Makatura et al. 1999).

Suite à ces constats, nous souhaitons montrer l'apport de l'approche contextualisée à l'évaluation de l'intelligence chez les adolescent.e.s tout-venants, public qui n'a pas ou très peu été étudié, notamment en termes de preuves de validité, basées sur les relations avec la performance scolaire et le comportement adaptatif. La seconde partie se compose de trois articles rendant compte d'études empiriques : la première porte sur l'auto-évaluation des capacités cognitives; la seconde sur l'évaluation multiple des capacités cognitives et la dernière sur l'évaluation contextualisée avec un test de performance.

Seconde partie

Contributions empiriques

3. Etude 1 : Le développement d'un outil d'auto-évaluation des capacités cognitives pour adolescent.e.s : le QAE

Les résultats de cette étude ont été présentés sous forme de poster sous la référence (Annexe 2) : Terriot, K, Grégoire, J., Loarer, E. *The development of a self-assessment questionnaire of adolescent cognitive abilities*: QAE. 29th International Congress of Applied Psychology, 26-30 juin 2018, Montréal, Canada.

3.1 Introduction

Depuis Binet et Simon (1905, 1908), l'évaluation intellectuelle repose traditionnellement sur l'utilisation de tests psychométriques, comme les échelles de Wechsler qui restent les plus utilisées au monde. Une autre manière d'évaluer l'intelligence serait d'utiliser une mesure auto-rapportée (Eysenck et Eysenck, 1985). Cette méthodologie existe déjà pour évaluer des domaines tels que les compétences sociales et adaptatives (Achenbach, 1991 ; Conners, 2001), la personnalité (Bernaud, 2014), l'intelligence émotionnelle (Choi & Kluemper, 2012 ; Petrides & Furnham, 2001) et a montré ses atouts, mais elle est très peu développée pour l'évaluation des processus cognitifs, surtout en France. L'auto-

évaluation des capacités cognitives permet d'évaluer la capacité perçue par le sujet lui-même de réaliser certaines tâches comme se repérer sur un plan, suivre les conversations des autres ou se souvenir où des choses sont rangées. Les items ne réfèrent pas explicitement à l'intelligence, mais font référence aux compétences du sujet dans diverses tâches cognitives.

Les questionnaires d'auto-évaluation présentent plusieurs avantages : ils sont peu onéreux, rapides à administrer, flexibles dans leur utilisation (Jacobs & Roodenburg, 2014) ; ils diminuent l'anxiété chez la personne évaluée (Furnham, 2001 ; Paulhus, Lysy & Yik, 1998 ; Simms, 2008), permettent d'investiguer des comportements difficilement observables (Tarrant & Cordell, 1997) et sont source de meilleure adhésion et coopération des personnes évaluées. Un autre avantage est qu'ils proposent des items en lien avec la vie quotidienne. En effet, alors que la plupart des expert.e.s dans le domaine de l'intelligence s'accordent à dire que cette dernière a pour finalité l'adaptation à l'environnement (Chartier & Loarer, 2008), l'approche classique de mesure de l'intelligence se réfère à des tâches qui n'ont que peu de rapport avec les situations rencontrées dans la vie quotidienne et qui sont généralement abstraites et décontextualisées (Huteau & Lautrey, 1999). A titre d'exemple, les liens entre les échelles de performance et le comportement adaptatif dans la vie quotidienne tel que mesuré par la Vineland-II sont généralement quasi nuls (Sparrow, Cicchetti & Balla, 2015 ; Wechsler, 2016).

Malgré leurs nombreux avantages, les questionnaires d'auto-évaluation des capacités cognitives sont très peu nombreux (Terriot, Grégoire & Loarer, 2017).

Par ailleurs, alors que le modèle le plus valide actuellement pour rendre compte de l'intelligence est le modèle de Cattell-Horn-Carroll (CHC ; Schneider & McGrew, 2012), un seul questionnaire d'auto-évaluation de l'intelligence repose sur ce dernier (Jacobs & Roodenburg, 2014) et il ne prend en compte que trois dimensions du modèle. En outre, tous ces questionnaires sont destinés à un public adulte et aucune étude, à notre connaissance, n'a été menée chez des adolescent.e.s alors que la question de l'évaluation des capacités cognitives dans ce public est prégnante notamment dans le champ scolaire. L'objectif principal de cet article, après avoir fait un rapide point sur le modèle CHC et les questionnaires d'auto-évaluation des capacités cognitives, sera de valider un tel questionnaire sur une population d'adolescent.e.s français.e.s et notamment de voir les preuves de sa validité, en particulier convergente, en examinant ses relations avec la réussite scolaire, les résultats obtenus à une épreuve de performance (WISC-IV) et à une échelle d'évaluation du comportement adaptatif (Vineland II).

3.1.1 Le modèle Cattell-Horn-Carroll (CHC)

Le modèle de Cattell-Horn-Carroll (Schneider & McGrew, 2012) est devenu dans les années 90 le modèle de référence de l'organisation des aptitudes cognitives. Il permet de dépasser le débat qui a longtemps subsisté entre approche unidimensionnelle ou multidimensionnelle de l'intelligence. Il articule le modèle hiérarchique en trois strates de Carroll (93) et la théorie de Cattell-Horn. Cattell introduit en 1941, la distinction entre intelligence fluide (Gf) et intelligence

cristallisée (Gc). La première permet de résoudre de nouveaux problèmes, inhabituels pour le sujet alors que la seconde correspond à l'étendue et la profondeur des connaissances acquises ainsi qu'à l'habileté à utiliser ces connaissances. Puis, en collaboration avec Horn dans les années 60, Cattell poursuit ses travaux en introduisant de nouvelles aptitudes telles que le traitement visuel (Gv), le traitement auditif (Ga), la vitesse de traitement (Gs), la vitesse de réaction (CDS), la mémoire à long terme (Glr), à court terme (Gsm) et les connaissances quantitatives (Gq) (Horn & Noll, 2018). Carroll (1993), quant à lui, réalise des méta-analyses en collectant les résultats issus de 461 études publiées entre 1925 et 1987 et propose un modèle hiérarchique de l'intelligence, à trois strates : la strate III correspond au facteur g de Spearman ; la strate II à huit facteurs de groupe tels que l'intelligence fluide (Gf) ou l'intelligence cristallisée (Gc) ; la strate I à une trentaine de facteurs spécifiques.

La dernière mise à jour du modèle CHC (Schneider & McGrew, 2012) comprend : un facteur général (le facteur g) ; seize aptitudes globales de niveau II telles que l'intelligence fluide (Gf), cristallisée (Gc), les connaissances quantitatives[8] (Gq), la mémoire à court terme[9] (Gsm), le traitement visuel[10] (Gv), ou la vitesse de traitement[11] (Gs) ; et quatre-vingt-dix aptitudes primaires.

---

[8] Connaissances déclaratives et procédurales à la disposition de la personne. Ce facteur quantifie la capacité d'utiliser l'information quantitative et de manipuler des symboles numériques.

[9] Capacité à saisir une information, à la retenir à court terme et à l'utiliser au cours des prochaines secondes

[10] Habileté à générer, percevoir, analyser, emmagasiner, manipuler à transformer des stimuli visuels pour résoudre des problèmes

[11] Habileté à effectuer des tâches automatiques rapidement et efficacement, en maintenant notre attention et concentration.

Ce modèle permet de faire le lien entre les aptitudes cognitives et les performances académiques (Floyd, Evans & McGrew, 2003 ; Floyd, Keith, Taub & McGrew, 2007 ; Gottfredson & Saklofske, 2009 ; Mc Grew & Wendling, 2010 ; Taub, Floyd, Keith & McGrew, 2008). La réussite en mathématiques serait, par exemple, liée aux aptitudes en intelligence fluide, intelligence cristallisée, et connaissances quantitatives. Plusieurs outils ont été créés ou révisés pour tenir compte de ce modèle, comme tout récemment le WISC-V (Wechsler, 2016). Malgré la dominance actuelle du modèle CHC, très peu de questionnaires d'auto-évaluation des capacités cognitives reposent sur celui-ci.

3.1.2 L'auto-évaluation des capacités cognitives

Les recherches sur l'auto-évaluation des capacités cognitives sont très peu nombreuses, bien moins que celles sur l'auto-estimation qui peut être définie comme la perception qu'a une personne de ses propres capacités, de son « intelligence » (Chamorro-Premuzic & Furnham, 2006). L'auto-estimation est mesurée en demandant directement à la personne d'estimer son niveau d'intelligence, en utilisant un ou plusieurs items avec, à titre d'exemple, la consigne suivante : « Estimez votre intelligence verbale », alors que l'auto-évaluation est une mesure indirecte qui permet d'évaluer la capacité perçue par l'individu de réaliser certaines tâches.

Depuis 1990, seuls six questionnaires ont été recensés (Terriot, Grégoire & Loarer, 2017). Ils proposent tous des énoncés relatifs à des situations à propos

desquelles les personnes doivent se situer en utilisant une échelle de Likert. Les situations sont du type « comprendre les conversations des autres » (intelligence cristallisée), « trouver une solution à un problème jamais expérimenté » (intelligence fluide), « lire une carte routière » (traitement visuel). Les modèles théoriques sur lesquels ils reposent sont divers. Il peut s'agir du modèle CHC (Jacobs & Roodenburg, 2014) ou de celui des intelligences multiples de Gardner (Tirri & Nokelainen, 2008), mais il peut aussi s'agir de questionnaires athéoriques (Gignac, Stough & Loukomitis, 2004 ; Rueda *et al.,* 2015 ; Seidenberg, Haltiner, Taylor, Hermann & Wyler, 1994). Enfin, certains questionnaires ne mesurent qu'un seul domaine cognitif, comme les capacités spatiales (Quaiser-Pohl & Lehmann, 2002). Tous ces questionnaires ont été créés et validés pour les adultes. Ces questionnaires possèdent d'assez bonnes qualités psychométriques en ce qui concerne la cohérence interne : α = .73 pour l'échelle globale du Self-Report Intelligence Questionnaire (SRIQ ; Gignac et al., 2004), α = .86 pour Gf, α = .84 pour Gc et α = .89 pour Gv du Self-Report Measure of Cognitive Abilities (SRMCA ; Jacobs & Roodenburg, 2014), α = .72 pour le Spatial ability in everyday life questionnaire (Quaiser-Pohl & Lehmann, 2002), α = .97 pour l'échelle globale de l'Everyday Cognition questionnaire (Ecog ; Rueda et al., 2015), α= .92 pour le score global du Multiple Abilities Self Report Questionnaire (MASQ ; Seidenberg et al., 1994) et des alphas qui varient de .53 (intelligence spatiale) à .83 (intelligence musicale) pour le Multiple Intelligences Profiling Questionnaire (MIPQ III de Tirri & Nokelaine, 2008).

Par contre, les corrélations avec des tests classiques de mesure des capacités cognitives sont variables d'une étude à l'autre. Elles sont généralement faibles ou modérées, ce qui indiquerait que l'auto-évaluation ne peut remplacer les tests d'intelligence classiques. A titre d'exemple, pour le SRMCA, une seule corrélation positive et significative a été trouvée avec le Woodcock-Jonhson Tests of Cognitive Abilities – third Edition (WJ-III cog ; Woodcock, McGrew & Mather, 2001) de .37 ($p<.01$). Notons cependant que, dans le domaine spatial, l'auto-évaluation est davantage liée aux performances spatiales. Les mesures auto-rapportées mesureraient donc un autre construit que celui mesuré par les tests classiques de performance. Elles seraient sans doute l'expression du concept de soi, défini comme une conception composite d'une personne sur elle-même, formée par ses expériences et l'interprétation de son environnement (Shavelson, Hubner & Stanton, 1976). Dans le champ de l'intelligence, les questionnaires d'auto-évaluation renverraient à un concept de soi cognitif (Ackerman, Bowen, Beier & Kanfer, 2001).

Aucune étude n'a été faite sur les corrélations entre l'auto-évaluation et la réussite scolaire. Par contre, quelques recherches menées en utilisant un questionnaire d'auto-estimation, ont établi l'existence d'une relation entre l'auto-estimation de l'efficience cognitive et les résultats scolaires (Kornilova & Novikova, 2013). Ces auteurs observent une corrélation de 0,16 entre l'auto-estimation et la réussite scolaire ($p<.01$). Dans la méta-analyse réalisée par Kornilov (2011), une corrélation moyenne positive et significative de 0,28 est trouvée.

En ce qui concerne la validité prédictive des questionnaires d'auto-évaluation, à nouveau aucune étude n'a été réalisée, à notre connaissance, en ce qui concerne la réussite scolaire. Par contre, quelques chercheurs se sont intéressés à la validité prédictive des auto-estimations. Ainsi, Kornilova, Kornilov et Chumakova (2009), ont mené une étude auprès d'étudiant.e.s russes. En complément de l'âge, du sexe et des résultats aux tests de performance, l'auto-estimation ajouterait 3% dans la prédiction de la variance des résultats universitaires six mois après la passation des tests.

3.2 Objectifs

L'étude présente se donne comme objectif principal de valider sur une population d'adolescent.e.s français.e.s un questionnaire d'auto-évaluation des capacités cognitives reposant sur six grands facteurs du modèle CHC : intelligence cristallisée (Gc), fluide (Gf), Traitement visuo-spatial (Gv), mémoire à court terme (Gsm), vitesse de traitement (Gs) et connaissances quantitatives (Gq). Ce sont les principaux facteurs évalués dans les échelles les plus utilisées dans le monde telles que le WISC et le KABC.

Ainsi, nous souhaitons confirmer la structure en six dimensions spécifiques présentées plus haut, évaluer la fiabilité de ces facteurs à partir de l'examen de la consistance interne de chacune des dimensions puis examiner les relations entre ces facteurs. Les preuves de validité convergente seront évaluées à partir des corrélations des six facteurs avec d'autres critères : les résultats obtenus au WISC-

IV, la réussite scolaire en fin d'année et le comportement adaptatif. La validité différentielle selon le sexe sera enfin examinée à partir de l'étude des différences de scores entre filles et garçons sur les six facteurs mesurés.

3.3 Méthodologie

3.3.1 Développement du questionnaire

Une première version d'un questionnaire à destination des adolescent.e.s a été élaborée, puis évaluée auprès d'un groupe de sept expert.e.s en psychologie cognitive, et auprès d'un échantillon de cent sujets. Suite aux analyses psychométriques et en particulier l'analyse factorielle, il s'est avéré que cette première version n'était pas concluante, en particulier du fait des items inversés mal compris par les jeunes.

Une refonte a eu lieu en reprenant plusieurs items du SRMCA de Jacobs et Roodenburg (2014), en simplifiant les items, la consigne et le format de réponse. Ce second questionnaire de 52 items a été proposé aux mêmes expert.e.s en leur demandant d'identifier le facteur de niveau II du modèle CHC auquel était rattaché chaque item. Un indice d'accord entre les juges, le kappa de Fleiss (1971), a été calculé. Quinze items dont la concordance inter-juges était moyenne ou faible ont été retirés ou remplacés par d'autres items pour aboutir à une version finale mesurant six facteurs de niveau II du modèle CHC : l'intelligence

cristallisée (Gc) avec des items tels que « trouver le mot exact pour exprimer mes pensées » ou « comprendre tous les mots d'un texte de journal » ; l'intelligence fluide (Gf) (par exemple, « résoudre des problèmes logiques » ou « résoudre des énigmes »); le traitement visuo-spatial (Gv) (par exemple, « comprendre les graphiques » ou « organiser l'emplacement des meubles dans une pièce ») ; les connaissances quantitatives (Gq) (par exemple, « calculer des pourcentages » ou « faire mentalement des soustractions ») ; la mémoire à court terme (Gsm) (par exemple, « retenir des numéros de téléphone » ou « retenir des dates de rendez-vous ») ; la vitesse de traitement (Gs) (par exemple, « recopier rapidement un texte »).

Les items ont été formulés en lien avec la vie quotidienne (dont le milieu scolaire). Ils ont été présentés aléatoirement mais toujours de manière identique pour tous en un seul questionnaire.

La consigne donnée est la suivante : *« Indiquez, pour chaque phrase, si cette activité vous est beaucoup plus difficile, plus difficile, pareil, plus facile ou beaucoup plus facile que pour les autres jeunes de votre âge »*. Les personnes doivent répondre sur une échelle de Likert en cinq points allant de « Beaucoup plus difficile » à « Beaucoup plus facile ». Le principe de comparaison sociale a été introduit en demandant aux personnes, comme cela est induit dans la consigne, de se comparer aux autres jeunes du même âge. Cette méthode permet, selon Mabe et West (1982) et Freud et Kasten (2012), d'obtenir les corrélations les plus importantes avec les tests d'intelligence.

Cette version a été pré-testée auprès de jeunes pour s'assurer de la compréhension de la consigne, des items et du format de réponse.

3.3.2 Participant.e.s

Les participant.e.s ont été recrutés au sein d'établissements scolaires de onze départements (22, 52, 60, 65, 75, 77, 78, 86, 92, 93, 95). Pour ce faire, un courrier a été adressé aux inspections académiques, puis en cas d'accord de ces dernières, aux chef.fe.s d'établissement. Un formulaire de consentement a été envoyé aux familles. Suite à leur acceptation, le questionnaire a été administré, soit en passation collective, soit en passation individuelle.

Notre échantillon se compose de 324 jeunes âgés de 11 à 20 ans (M=13,85 ; ET=1,26), scolarisés de la 6ème à la première dont 72,9% en classe de 4ème. L'échantillon est équilibré au niveau du sexe (49,2% de garçons et 50,8% de filles).

3.3.3 Mesure

Tous les participant.e.s ont complété le questionnaire d'auto-évaluation des capacités cognitives (QAE) présenté en annexe n°3.

Les résultats scolaires en fin d'année scolaire ont également été recueillis (moyenne générale au 3ème trimestre ou au 2ème semestre selon la classe) pour 87 d'entre eux.elles.

Trente trois d'entre eux ont également passé le WISC-IV (Wechsler, 2005). Il s'agit d'une échelle dévaluation de l'intelligence pour enfants et adolescent.e.s de 6 à 16 ans 11 mois. Elle est composée de dix épreuves obligatoires qui permettent de calculer quatre indices : de compréhension verbale (ICV), raisonnement perceptif (IRP), mémoire de travail (IMT) et vitesse de traitement (IVT) ainsi que le quotient intellectuel total (QIT).

L'ICV se compose de trois épreuves : *Similitudes*, *Vocabulaire* et *Compréhension* ; l'IRP de *Cubes*, *Matrices* et *Identification de Concepts* ; l'IMT de *Mémoire des chiffres* et de *Séquence-lettres-chiffres* et enfin l'IVT de *Code* et *Symboles*. Bien que le WISC-IV n'ait pas été élaboré en s'appuyant sur le modèle CHC, plusieurs études (Keith, Fine, Taub, Reynolds & Kranzler, 2006 ; Lecerf, Rossier, Favez, Reverte & Coleaux, 2010 ; Lecerf, Golay, Reverte, Senn, Favez & Rossier, 2012) ont montré que l'interprétation des scores aux indices et subtests pouvait être effectuée en référence à ce modèle. Ainsi l'ICV correspond à l'intelligence cristallisée (Gc), l'IMT à la mémoire à court terme (Gsm), l'IVT à la vitesse de traitement (Gs) et l'IRP au traitement visuo-spatial (Gv) pour *Cubes* et au raisonnement fluide (Gf) pour *Matrices* et *Identification de concepts*. Les subtests sont notés de 1 à 19, ont une moyenne de 10 et un écart-type de 3 ; les indices et le QIT une moyenne de 100 et un écart-type de 15.

Pour vingt cinq sujets, un des deux parents a rempli la Vineland II (VABS-II ; Sparrow, Cicchetti & Balla, 2015) forme destinée aux parents. Il s'agit d'une échelle d'évaluation des comportements socio-adaptatifs. Elle permet de recueillir les compétences dans trois domaines : communication, socialisation et autonomie dans la vie quotidienne ainsi qu'un score global : le quotient socio-adaptatif total. Le comportement adaptatif se définit ici comme l'exécution des activités de la vie quotidienne nécessaires pour être autonome.

Il est demandé aux parents d'indiquer la note qui décrit le mieux la fréquence avec laquelle leur enfant manifeste l'activité décrite quand il le faut ou bien s'il.elle peut la réaliser : une note de 2 correspond au fait qu'il.elle manifeste habituellement l'activité sans aide ni rappel ; la note de 1 quand il.elle la manifeste parfois ou en partie sans aide ni rappel et la note de 0 quand il.elle ne la manifeste jamais sans aide ni rappel.

L'échelle de communication se compose de trois sous-domaines :

- Réceptif : la manière dont la personne évaluée écoute, comprend, suit les instructions ;
- Expressif : ce que dit le sujet, comment il utilise les mots et les phrases pour s'exprimer, transmet l'information ;
- Ecrit : compétences lexicales, orthographiques, utilisation et/ou investissement des supports écrits et multimédias pour communiquer.

L'échelle de vie quotidienne se compose de :

- Personnel : la manière dont la personne prend soin de soi : comment il s'habille, mange, prend soin de sa santé et de son corps ;

- Domestique : comment la personne s'occupe de son domicile, les tâches ménagères qu'il.elle effectue ;
- Communauté : la manière dont il.elle vit dans une communauté, comment il.elle gère son temps, son argent, le téléphone, l'ordinateur…

Et enfin, le domaine de la socialisation se compose de :

- Relations interpersonnelles, domaine qui renvoie au contact avec les autres ;
- Jeu et temps libre : comment le sujet joue et utilise ses temps libres (inscriptions dans des clubs sportifs, activités extra-scolaires) ;
- Adaptation : comment il.elle fait preuve de responsabilité et de sensibilité aux autres (rapports aux règles, respect des règlements).

Les échelles ont une moyenne de 100 et un écart-type de 15 ; les sous-échelles une moyenne de 15 et un écart-type de 3.

3.4 Résultats

Les données ont été analysées en utilisant SPSS 20.0 (IBM Corp. Released 2011), le logiciel R (R Core Team, 2014), en particulier les modules Psych (Revelle, 2015), Psychometric (Fletcher, 2010), et GPArotation (Bernaards & Jennrich, 2005).

Nous avons réalisé une analyse par item en appliquant la théorie classique des tests et les réponses à l'item (Microsoft Excel tools using the eirt add-in ; Valois,

Houssemand, Germain & Belkacem, 2011). Nous avons trouvé un item (le 20) avec de faibles caractéristiques psychométriques, en particulier une faible discrimination. Il a donc été retiré pour la suite des analyses. Nous avons également mené une analyse différentielle des items selon le sexe à l'aide du programme développé par Zumbo (1999), adapté aux échelles ordinales, en utilisant la régression logistique (tableau 4). En effet, comparée à d'autres approches, contrôler les fonctionnements différentiels d'items est une stratégie optimale dans les processus de développement de questionnaires (Wetzel, Hell & Päßler, 2012). Plusieurs standards pour le développement des tests évoquent les biais liés au sexe et la nécessité de les limiter pour s'assurer d'évaluations équitables entre hommes et femmes (ITC, 2003).

Zumbo (1999) propose de mesurer le fonctionnement différentiel d'items à partir d'un coefficient, le R^2 de Nagelkerke (1991). La taille d'effet correspond à la différence entre le groupe focal et le groupe de référence. Selon Jodoin et Gierl (2001), l'effet est négligeable si ΔR^2 est inférieur à 0,035, modéré entre 0,035 et 0,070 et important s'il est supérieur à 0,070.

Facteurs initiaux	Items	Δ Chi2	p-value	ΔR2	Effet
Gc	1	5.22	.07	.020	
	10	2.18	.34	.005	
	14	1.83	.40	.005	
	16	.06	.97	.000	
	23	4.55	.10	.013	
	26	5.46	.07	.016	
	35	8.73	.01	.023	Négligeable
	43	1.02	.60	.002	
	47	14.07	.00	.034	Négligeable
Gf	4	2.22	.33	.007	
	22	10.14	.01	.023	Négligeable
	27	.52	.77	.002	
	28	.55	.76	.002	
	39	.53	.77	.001	
	49	2.13	.34	.003	
Gv	2	14.71	.00	.048	Modéré
	5	.83	.66	.003	
	7	1.86	.39	.004	
	11	14.42	.00	.036	Modéré
	12	1.60	.45	.005	
	13	1.07	.59	.001	
	15	17.59	.00	.039	Modéré
	18	1.22	.54	.004	
	21	8.77	.01	.026	Négligeable
	24	5.14	.08	.015	
	25	23.78	.00	.066	Modéré
	29	2.69	.26	.007	
	32	30.7	.00	.073	Important
	34	1.45	.48	.003	
	38	8.85	.01	.028	Négligeable
	44	.49	.78	.002	
	48	3.74	.15	.010	
	51	35.08	.00	.080	Important
	52	17.29	.00	.033	Négligeable
Gsm	3	.61	.74	.003	
	8	2.60	.27	.006	
	17	6.97	.03	.022	Négligeable
	30	8.27	.02	.023	Négligeable
	31	6.60	.04	.019	Négligeable
	36	20.24	.00	.059	Modéré
	40	21.58	.00	.057	Modéré
	42	5.78	.06	.018	
	46	16.57	.00	.050	Modéré
Gs	6	2.67	.26	.008	
	19	7.89	.02	.023	Négligeable
	37	6.06	.05	.013	Négligeable
	50	5.96	.05	.014	Négligeable
Gq	9	12.64	.00	.035	Modéré
	33	11.68	.00	.026	Négligeable
	41	1.38	.50	.003	
	45	14.63	.00	.036	Modéré

Tableau 4. Analyse du fonctionnement différentiel des 51 items du QAE selon le sexe

Suite à l'analyse différentielle des items selon le sexe, vingt-cinq items ont été identifiés comme présentant un effet : treize avec un effet négligeable (items 17, 19, 21, 22, 30, 31, 33, 35, 37, 38, 47, 50 et 52), neuf avec un effet modéré (2, 9, 11, 15, 25, 36, 40, 45 et 46) et deux avec un effet important (32 et 51). Ces deux items ont donc été retirés pour conduire l'analyse factorielle.

Une analyse factorielle exploratoire, pour examiner la validité interne, a donc été conduite à partir de 49 items avec les packages Psych et GPArotation de R (méthode du minimum résiduel et rotation oblimin) adaptés pour les échelles ordinales. Le tableau 5 montre les résultats.

Items	Facteur 1 Gc	Facteur 2 Gf	Facteur 3 Gvc	Facteur 4 Gvo	Facteur 5 Gsm	Facteur 6 Gs	h^2
1	**.29**	.01	.17	.17	.03	.08	.24
10	**.58**	-.07	.09	.22	.08	.01	.40
14	**.58**	.11	.01	.13	-.12	.05	.42
16	**.67**	.12	-.12	.12	-.12	.06	.52
23	**.45**	.03	-.08	.12	.11	.05	.34
26	**.61**	.03	-.04	-.09	-.07	.13	.42
35	**.61**	-.09	.04	-.09	.02	.16	.44
43	**.60**	-.13	.07	.14	.16	.03	.48
47	**.43**	.14	.21	-.14	.28	-.04	.46
9	-.14	**.87**	-.05	-.04	-.03	.10	.70
22	.01	**.69**	.12	.10	-.11	-.02	.56
27	.03	**.40**	.19	.13	.01	.00	.28
28	-.05	**.43**	.21	.11	.05	-.08	.28
33	.19	**.68**	-.09	.05	-.01	-.04	.56
41	.03	**.70**	-.02	-.08	.13	.12	.57
45	-.03	**.70**	.03	.09	.17	-.03	.55
49	.16	**.49**	.29	.06	-.11	-.14	.47
2	-.09	-.13	**.49**	.02	.08	.21	.30
5	-.06	.10	**.54**	.23	-.06	-.02	.41
13	.06	.07	**.54**	.06	.01	.04	.36
21	-.05	.06	**.67**	-.01	-.13	.04	.45
24	-.01	-.13	**.28**	.24	.21	.17	.29
38	.03	-.19	**.37**	.11	.32	.04	.32
44	.00	.06	**.50**	.17	.03	.08	.36
11	.05	.01	.03	**.65**	.03	.12	.50
12	.23	.03	.02	**.46**	.08	-.09	32
15	.06	.13	.11	**.59**	-.05	.07	.46
18	-.07	.00	-.01	**.41**	.18	.12	.24
29	.13	-.03	.09	**.48**	.22	.01	.40
3	.24	.14	.06	..00	**.25**	.14	.29
17	.29	.05	.10	.06	**.40**	-.02	.30
30	-.09	.16	.11	.15	**.45**	.04	.26
31	-.09	.06	-.17	.17	**.66**	.01	.49
36	.04	.01	-.07	.04	**.53**	.26	.44
40	-.06	-.06	.15	-.07	**.59**	.00	.37
42	.34	.13	.22	-.10	**.38**	.02	.48
46	.20	.07	.03	-.09	**.45**	.07	.32
6	.03	.02	-.06	.06	-.04	**.88**	.78
37	.02	.02	.11	-.02	.05	**.80**	.71
50	-.01	-.06	.15	-.01	.16	**.37**	.22
Alpha	.83	.86	.75	.75	.75	.74	

Tableau 5. Matrice des composantes après rotation Oblimin

La matrice de structure conforte partiellement le modèle théorique : nous retrouvons bien les facteurs Gc, Gsm et Gs. Les facteurs Gf et Gq ne forment qu'un seul facteur et Gv se subdivise en deux facteurs : l'un que l'on pourrait nommer Gv Construction (Gvc) et l'autre Gv Orientation (Gvo). Nous retenons donc six facteurs : Gc (items 1, 10, 14, 16, 23, 26, 35, 43 et 47), Gsm (items 3, 17, 30, 31, 36, 40, 42 et 46), Gs (items 6, 37 et 50), Gf (items 9, 22, 27, 28, 33, 41, 45 et 49), Gv Construction (items 2, 5, 13, 21, 24, 38 et 44), Gv Orientation (items 11, 12, 15, 18 et 29). Neuf items saturant autant d'autres facteurs que celui d'appartenance (items 4, 7, 8, 19, 25, 34, 39, 48 et 52) ont été retirés du QAE qui comprend désormais 40 items.

Comme indiqué dans le tableau 5, les alpha calculés à partir des corrélations polychoriques vont de .74 à .86 pour les six facteurs, suggérant une bonne consistance interne pour toutes les échelles.

3.4.1 Les relations entre dimensions du QAE

Le Tableau 6 indique des corrélations entre les variables latentes issues de l'analyse factorielle exploratoire conduite

	Gc	Gf	Gvc	Gvo	Gsm	Gs
Gc	1.00					
Gf	.31	1.00				
Gvc	.19	.20	1.00			
Gvo	.30	.18	.11	1.00		
Gsm	.23	.04	.11	.24	1.00	
Gs	.22	.19	.28	.13	.21	1.00

** p < 0.01 (bilatéral)

Tableau 6. Matrice des corrélations entre les facteurs

3.4.2 Preuves de validité concourante et différentielle.

3.4.2.1 Validité différentielle selon le sexe

Un test *t* de Student a été effectué en vue d'examiner l'existence d'éventuelles différences selon le sexe dans les facteurs au QAE. Comme indiqué dans le tableau 7, on observe des différences significatives entre les garçons et les filles au niveau des facteurs Gf, Gvc et Gvo. C'est sur les facteurs Gf et Gvo qu'on retrouve la plus grande taille d'effet (respectivement $\eta^2 = 0.06$ et $\eta^2 = 0.07$ qui sont des tailles d'effet moyennes voire faibles) avec une différence en faveur des garçons.

	Garçons			Filles				
	N	M	ET	N	M	ET	Test *t*	η^2
Gc	158	100.62	14.48	163	99.40	15.51	.73	.00
Gf	158	103.65	14.06	163	96.46	15.08	4.41**	.06
Gvc	158	100.66	15.42	163	99.36	14.60	.78	.00
Gvo	158	104.01	14.17	163	96.12	14.80	4.88**	.07
Gsm	158	97.87	14.45	163	102.06	15.28	-2.53**	.02
Gt	158	99.41	15.44	163	100.57	14.58	-.70	.00

*. p < 0.05 (bilatéral)
**. p < 0.01 (bilatéral)
Note : les scores ont été normalisés et transformés selon la formule de Hull (moyenne de 100, écart-type de 15) pour permettre des comparaisons intra-individuelles

Tableau 7. Différences de scores à chaque échelle selon le sexe (N=321)

3.4.2.2 Validité concourante

3.4.2.2.1 Corrélations avec les résultats scolaires

Comme indiqué précédemment, nous avons recueilli les résultats scolaires au 3ème trimestre pour 87 sujets, soit de 2 à 6 mois après le recueil des données au QAE. Des corrélations Bravais-Pearson ont été calculées sur l'échantillon total (tableau 8) à partir de la version à 40 items du QAE en vue d'examiner les relations entre les deux. Nous avons également calculé des corrélations corrigées pour tenir compte de la restriction de l'étendue en utilisant la correction de Cohen et al. (1988) à l'aide du package Psychometric de R (Fletcher, 2010).

	Moyenne
Gc	.26* (.26*)
Gf	.45** (.43**)
Gvc	.16 (.16)
Gvo	-.07 (-.07)
Gsm	.10 (.09)
Gs	.02 (.02)

*. p < 0.05 (bilatéral)
**. p < 0.01 (bilatéral)
Note : la valeur entre parenthèse correspond à la corrélation corrigée

Tableau 8. Corrélations entre les six facteurs du QAE et la moyenne générale au 3ème trimestre (N=87)

Selon les normes établies par Cohen (1988), les corrélations positives et significatives entre les résultats au 3ème trimestre et Gf (*r* =.45) puis Gc (*r* =.26) sont de taille modérée. Les corrélations sont quasi nulles avec les autres facteurs.

Nous avons, en outre, effectué une analyse de régression multiple pour identifier le ou les facteurs les plus prédictifs de la moyenne générale obtenue au 3ème trimestre.

Seul le facteur Gf prédit la réussite scolaire au 3ème trimestre et explique 18,9% de la variance de la moyenne au 3ème trimestre.

3.4.2.2.2 Corrélations avec les résultats au WISC-IV

Nous avons vérifié que notre échantillon était représentatif en regardant si les moyennes obtenues aux indices et au QIT étaient conformes à la norme (voir

tableau 9). C'est le cas pour toutes les variables sauf l'ICV pour laquelle notre échantillon présente une moyenne et un écart-type plus élevé.

	Moyenne	Ecart-type	t de Student	p-value
ICV	108.33	22.34	2.14	.04
IRP	102.24	16.62	.78	.44
IMT	95.39	19.16	-1.38	.18
IVT	100.33	18.65	.10	.92
QIT	103.42	20.59	.96	.35

Tableau 9. Comparaison aux normes des scores obtenus par l'échantillon (N=33) au WISC-IV

Comme l'indique le tableau 10, seul Gf corrèle de manière positive et significative avec les subtests (*r* =.54 avec *Cubes*, *r* =.38 avec *Similitudes*, *r* =.36 avec *Vocabulaire* et *r* =.35 avec *Séquence-lettres chiffres*) et les indices au WISC-IV (*r* =.53 avec l'IRP, r =.37 avec l'IVT et *r* =.42 avec le QIT).

	Gc	Gf	Gvc	Gvo	Gsm	Gs
Cubes	.23 (.24)	.54** (.57**)	.33 (.34)	.11 (.09)	.16 (.12)	.25 (.21)
Similitudes	.05 (.05)	.38* (.41*)	.08 (.08)	.03 (.03)	.04 (.03)	-.17 (-.14)
Mémoire des chiffres	-.11 (-.11)	.21 (.23)	.11 (.12)	.18 (.15)	.08 (.06)	.07 (.06)
Identification de concepts	-.19 (-.19)	.12 (.13)	.17 (.18)	-.10 (-.08)	-.05 (-.04)	-.07 (-.06)
Code	.07 (.07)	.32 (.34)	.33 (.34)	.09 (.08)	.06 (.04)	.32 (.27)
Vocabulaire	.11 (.11)	.36* (.39*)	.00 (.00)	.08 (.07)	.03 (.02)	.01 (.01)
Séquence-lettres-chiffres	-.00 (-.00)	.35* (.37*)	.26 (.27)	.06 (.05)	.13 (.10)	-.06 (-.05)
Matrices	.12 (.12)	.25 (.27)	.26 (.27)	.01 (.01)	.17 (.13)	-.04 (-.03)
Compréhension	-.06 (-.06)	.22 (.24)	.07 (.07)	.05 (.04)	-.16 (-.11)	-.30 (-.25)
Symboles	-.02 (-.02)	.33 (.35)	.14 (.15)	.00 (.00)	.10 (.07)	-.16 (-.13)
ICV	.09 (.09)	.27 (.29)	-.01 (-.01)	.09 (.09)	-.05 (-.04)	-.04 (-.03)
IRP	-.01 (-.01)	.53** (.56**)	.28 (.29)	-.05 (-.04)	.22 (.16)	-.08 (-.07)
IMT	-.03 (-.03)	.21 (.23)	.17 (.18)	.11 (.09)	.05 (.04)	.06 (.05)
IVT	.03 (.03)	.37* (.40*)	.26 (.27)	.00 (.00)	.10 (.07)	.11 (.09)
QIT	.04 (.04)	.42* (.45*)	.20 (.21)	.07 (.06)	.09 (.07)	-.01 (-.01)

*. $p < 0.05$ (bilatéral)
**. $p < 0.01$ (bilatéral)
La valeur entre parenthèse correspond à la corrélation corrigée

Tableau 10. Corrélations entre les six facteurs du QAE et les résultats au WISC-IV (N=33)

3.4.2.2.3 Corrélations avec la Vineland II

Nous avons vérifié que notre échantillon était conforme en ce qui concerne les résultats obtenus à la Vineland II (tableau 11). Nos participant.e.s obtiennent des résultats conformes dans toutes les dimensions sauf pour les trois dernières : adaptation, socialisation et la note composite dans lesquelles ils obtiennent une moyenne supérieure.

	Moyenne	Ecart-type	t de Student	p-value
Réceptif	14.71	2.44	-.59	.56
Expressif	15.40	3.91	.51	.61
Ecrit	14.72	3.70	-.38	.71
Communication	97.92	23.76	-.43	.67
Personnel	15.87	4.63	.92	.36
Domestique	15.64	2.29	1.40	.17
Communauté	15.71	5.32	.65	.52
Vie quotidienne	108.39	24.47	1.64	.11
Relations	16.08	4.35	1.22	.23
Jeu	15.52	3.19	.78	.44
Adaptation	16.61	2.62	2.94	.01
Socialisation	107.61	15.04	2.43	.02
Note composite	108.76	17.51	2.29	.03

Tableau 11. Comparaison des scores obtenus par l'échantillon (N=25) à la Vineland II aux normes

Comme indiqué dans le tableau 12, seul le facteur Gc du QAE n'entretient aucune relation avec les épreuves et les échelles de la Vineland II. Gf est lié positivement et significativement avec les sous-domaines Personnel (r=.46), Communauté (r=.47) et le domaine Vie quotidienne (r=.46) ; Gvc avec les sous-domaines Réceptif (r=.46), Ecrit (r=.42), Personnel (r=.46), Domestique (r=.38), Communauté (r=.49) et les domaines de la Communication (r=.42) et de la Vie Quotidienne (r=.53) ; le facteur Gvo avec Ecrit (r=.44), Communauté (r=.65) et Relations (r=.41), et le domaine de la Vie quotidienne (r=.56). Le facteur Gsm n'entretient qu'une seule relation positive et significative avec Communauté (r=.46). Enfin, c'est le facteur Gs du QAE qui entretient les relations les plus fortes avec la Vineland-II et le seul facteur pour lequel existe une corrélation positive, significative et relativement forte avec la note composite (r=.69) et le

domaine de la Socialisation (r=.60). On retrouve des liens positifs avec tous les sous-domaines et domaines de la Vineland-II exceptés Expressif et Adaptation.

	Gc	Gf	Gv	Go	Gsm	Gs
Réceptif	.23 (.31)	.38 (.40)	.46* (.50*)	.27 (.23)	.14 (.11)	.42* (.37)
Expressif	.24 (.33)	.32 (.34)	.35 (.39)	.29 (.25)	.35 (.28)	.28 (.25)
Ecrit	.36 (.48*)	.38 (.40)	.42* (.46*)	.44* (.38)	.22 (.17)	.51* (.46*)
Communication	.30 (.40)	.38 (.40)	.42* (.46*)	.37 (.32)	.27 (.21)	.44* (.39)
Personnel	.37 (.49*)	.46* (.49*)	.46* (.50*)	.34 (.29)	.32 (.25)	.46* (.41*)
Domestique	.13 (.18)	.20 (.21)	.38* (.42*)	.36 (.31)	.18 (.14)	.50* (.45*)
Communauté	.23 (.31)	.47* (.50*)	.49* (.53*)	.65** (.59**)	.46* (.37)	.65** (.60**)
Vie quotidienne	.24 (.33)	.46* (.49*)	.53* (.58*)	.56* (.50*)	.36 (.28)	.70* (.65**)
Relations	.32 (.43*)	.28 (.30)	.30 (.33)	.41* (.36)	.21 (.16)	.68** (.63**)
Jeu	.09 (.13)	-.17 (-.18)	.10 (.11)	.18 (.15)	-.07 (-.05)	.54** (.49**)
Adaptation	.25 (.34)	.02 (.02)	.07 (.08)	-.12 (-.10)	-.01 (.01)	.32 (.28)
Socialisation	.17 (.24)	-.13 (-.14)	.06 (.07)	.04 (.03)	-.06 (.05)	.60** (.55**)
Note composite	.07 (.10)	.15 (.16)	.32 (.36)	.24 (.20)	.12 (.09)	.69** (.64**)

*. p < 0.05 (bilatéral)
**. p < 0.01 (bilatéral)

Tableau 12. Corrélations entre les six facteurs au QAE et les scores à la Vineland II (N=25)

3.5 Discussion

L'objectif principal de cette étude était de valider, sur un échantillon d'adolescent.e.s français.e.s, un questionnaire d'auto-évaluation des capacités cognitives basé sur le modèle CHC. Nous avons vérifié la structure factorielle de l'outil, la fiabilité des six dimensions et les relations entre les facteurs. Puis, nous avons examiné la validité convergente ainsi que les différences de scores entre filles et garçons pour les facteurs du QAE.

La structure factorielle

Concernant la validité interne du QAE, les résultats issus de l'analyse factorielle exploratoire avec rotation Oblimin témoignent d'une structure factorielle en six facteurs. Les résultats confortent partiellement le modèle théorique initial : nous retrouvons bien les facteurs d'intelligence cristallisée (Gc), de mémoire à court terme (Gsm) et de vitesse de traitement (Gs). Les facteurs d'intelligence fluide (Gf) et de connaissances quantitatives (Gq) ne forment qu'un seul facteur et le traitement visuel (Gv) se subdivise en deux facteurs : l'un que l'on pourrait nommer Gv Construction et l'autre Orientation.

Les adolescent.e.s sont ainsi capables de produire des réponses différenciées selon les domaines. Le fait que Gf et Gq ne forment qu'un seul facteur peut être lié au principe que dans le modèle CHC, le raisonnement quantitatif est une aptitude primaire de Gf. Le raisonnement quantitatif est défini comme la capacité de raisonner avec des nombres, des relations mathématiques et des opérations. Les tests mesurant le raisonnement quantitatif n'exigent pas de connaissances avancées en mathématiques (Schneider & McGrew, 2012). Le facteur de connaissances quantitatives (Gq), quant à lui, concerne la profondeur et la grandeur des connaissances mathématiques (connaissance des symboles, des opérations, des procédures). Les items se référant à cette compétence dans le QAE sont « Calculer mentalement » (item 9), « Calculer des pourcentages » (item 33), « Faire mentalement des soustractions » (item 41), et « Calculer mentalement le prix total pour l'achat de plusieurs produits » (item 45). Bien que l'idée de

raisonnement ne soit pas directement évoquée dans les items, le fait que les mots « calculer » ou « mental » aient été utilisés a pu induire cela.

En ce qui concerne la distinction entre Gvc et Gvo, cela irait dans le sens d'autres études (Hergarty, Montello, Richardson, Ishikawa & Lovelace, 2006) qui montrent une distinction entre les compétences spatiales réduites (visualisation, rotation mentale, labyrinthes, puzzles) et les compétences spatiales larges qui incluent l'environnement (tâches d'orientation dans l'espace, de navigation). Ces dernières capacités sont plus complexes et incluent l'environnement. Les deux entretiendraient néanmoins un lien positif et significatif ($r=.30$), comme on le retrouve dans notre étude ($r=.40$, $p<.01$). Schneider et McGrew (2012) précisent que beaucoup d'aspects de Gv n'ont pas encore été explorés et que le modèle CHC actuel ne tient pas compte des capacités d'orientation dans l'espace. Ce champ a, en effet, été très peu exploré. Ils précisent que la réalité virtuelle pourrait apporter des réponses à ce sujet.

La consistance interne des échelles

Les alpha de Cronbach de la version du QAE composée de 40 items vont de .74 à .86 pour les six facteurs, suggérant une bonne consistance interne pour toutes les échelles.

Les relations entre les facteurs du QAE

Tous les facteurs entretiennent entre eux des liens bien que modestes. Ceci est conforme au modèle CHC, dans lequel les facteurs de second ordre sont reliés entre eux.

La corrélation la plus forte est observée entre Gc et Gf, comme indiqué plus haut.

Validité convergente

Liens avec la réussite scolaire

Pour quatre-vingt-sept sujets de notre étude, les résultats scolaires en fin d'année scolaire ont été recueillis (moyenne générale au 3ème trimestre ou au 2ème semestre selon la classe). Nous observons des corrélations positives, significatives modérées entre les résultats au 3ème trimestre (soit quelques mois après avoir passé le QAE) et Gf (r =.45) puis Gc (r =.26). Les corrélations sont quasi nulles avec les autres facteurs. Par ailleurs, le facteur Gf prédit la réussite scolaire au 3ème trimestre et explique 18,9% de la variance de la moyenne au 3ème trimestre. Il n'est pas étonnant de trouver ces résultats, Gf étant composé pour moitié d'items en lien avec les mathématiques (items 9, 33, 41 et 45) et Gc d'items en lien avec des compétences en français. Comme Grégoire (à paraître) le précise, les corrélations entre l'intelligence cristallisée (Gc) et l'intelligence fluide (Gf), et les performances scolaires sont élevées. Ces capacités jouent un rôle prépondérant dans l'efficacité du comportement intelligent.

Il n'existe pas d'étude comparative ayant examiné les relations entre questionnaire d'auto-évaluation des capacités cognitives et la réussite scolaire. Mais en

s'appuyant sur la méta-analyse réalisée par Kornilov (2011), dans laquelle une corrélation moyenne positive et significative de 0,28 était trouvée, les résultats que nous obtenons semblent donc bien prometteurs, puisque la corrélation entre Gf et la moyenne au troisième trimestre est bien supérieure à la corrélation moyenne habituellement observée et celle de Gc est identique.

La corrélation trouvée entre Gf et la moyenne au troisième trimestre se rapproche de celles habituellement trouvée entre les tests de performance et la réussite scolaire, qui sont généralement comprises entre 0,40 et 0,50 (Sternberg, Grigorenko & Bundy, 2001).

Liens avec les tests de performance

Trente-trois adolescent.e.s de notre échantillon ont également passé le WISC-IV (Wechsler, 2005) en plus du QAE. Seul Gf corrèle de manière positive et significative avec les subtests (r =.54 avec Cubes, r =.38 avec Similitudes, r =.36 avec Vocabulaire et r =.35 avec Séquence-lettres chiffres) et les indices au WISC-IV (r =.53 avec l'IRP, r =.37 avec l'IVT et r =.42 avec le QIT). Ceci est assez conforme aux recherches précédentes même si les corrélations sont un peu plus élevées. En effet, comme indiqué dans la première partie, pour le SRMCA, une seule corrélation positive et significative a été trouvée avec le Woodcock-Jonhson Tests of Cognitive Abilities – third Edition (WJ III cog ; Woodcock, McGrew & Mather ; 2001) de .37 ($p<.01$).

De même, Mabe et West (1982) ont mené une méta-analyse à partir de 55 études et ont trouvé une corrélation moyenne entre l'auto-estimation/auto-évaluation et

les tests de performance centrés sur les capacités cognitives de 0.34 (valeur de p non précisée). Freund et Kasten (2012), dans leur méta-analyse ont observé une corrélation moyenne de 0.32 ($p<.001$).

Nous nous attendions à ce que les autres facteurs du QAE corrèlent avec les indices du WISC-IV : Gc avec l'ICV, Gsm avec l'IMT, Gvc et Gvo avec l'IRP et Gs avec l'IVT. Il semble donc que ces autres facteurs évaluent un tout autre construit que celui mesuré par ces indices dans le WISC-IV. Il est également possible que les adolescent.e.s ne parviennent pas à s'auto-évaluer dans ces domaines.

Concernant Gf, nous pourrions faire l'hypothèse qu'il est très saturé en facteur g car il corrèle fortement avec les subtests les plus saturés également en facteur g à part Compréhension et Matrices (Grégoire, 2006) et avec le QIT du WISC-IV.

Liens avec le comportement adaptatif

Pour vingt-cinq sujets de notre échantillon, un des deux parents a rempli la Vineland II (VABS-II ; Sparrow, Cicchetti & Balla, 2015) forme destinée aux parents. Il s'agit d'une échelle d'évaluation des comportements socio-adaptatifs. Elle permet de recueillir les compétences dans trois domaines : communication, socialisation et autonomie dans la vie quotidienne ainsi qu'un score global, le quotient socio-adaptatif total. L'étude des corrélations montre que le facteur Gc du QAE n'entretient aucune relation avec les épreuves et les échelles de la Vineland II. Gf est lié positivement et significativement avec le domaine Vie quotidienne ($r=.46$) ; Gvc avec les domaines de la Communication ($r=.42$) et de la

Vie Quotidienne (r=.53) ; Gvo avec le domaine de la Vie quotidienne (r=.56). Enfin, c'est le facteur Gs du QAE qui entretient les relations les plus fortes avec la Vineland-II et le seul facteur pour lequel existe une corrélation positive, significative et relativement forte avec la note composite (r=.69), le domaine de la Socialisation (r=.60) et de la Vie quotidienne (r=.70).

Les études antérieures montraient systématiquement des liens entre les échelles de performance et le comportement adaptatif quasi nuls (Sparrow, Cicchetti & Balla, 2015 ; Wechsler, 2016). Or, dans notre échantillon, les corrélations existent, notamment entre le QIT du WISC-IV et le score composite de la Vineland-II ($r=.45$, $p<.05$). Cela pourrait être dû à la spécificité de notre échantillon, recruté sur volontariat. Les parents ayant rempli la Vineland-II ont peut-être une connaissance de leur enfant plus développée que les parents ayant participé aux recherches antérieures, par ailleurs toutes effectuées aux Etats-Unis.

Le QAE semble donc entretenir des liens avec le comportement adaptatif, défini comme l'exécution des activités de la vie quotidienne nécessaires pour être autonome, en particulier le facteur Gs, lié à la rapidité pour exécuter des tâches.

Validité différentielle selon le sexe

Notre étude met en exergue des différences significatives entre les garçons et les filles au niveau des facteurs Gf, Gvc et Gvo. C'est sur les facteurs Gvo et Gf qu'on retrouve la plus grande taille d'effet (η^2 = .07 et .06, effets de taille moyenne) avec une différence en faveur des garçons. Ceci est conforme aux autres études menées dans ce cadre : les garçons se jugeraient supérieurs dans les

domaines logique, mathématique et spatial (Bennett, 2000 ; Furnham, Clark & Bailey, 1999 ; Furnham, Fong & Martin, 1999 ; Kaufman, 2012). Selon une méta-analyse de Szymanowicz et Furnham (2011), réalisée sur ce sujet, la plus grande taille d'effet apparaît dans les différences entre garçons et filles pour les intelligences logique et mathématique (d=0,44), suivie de l'intelligence spatiale (d=0,43), de l'intelligence générale (d=0,37) et l'intelligence verbale (d=0,07), avec des scores auto-estimés plus élevés pour les garçons dans tous les domaines, excepté l'intelligence verbale, pour laquelle les filles s'estiment plus fortes.

Comment expliquer ces différences entre filles et garçons ? Deux courants théoriques s'opposent (Beloff, 1992) : le premier social qui attribue la différence aux stéréotypes de sexe, aux rôles de sexe, aux concepts de soi genrés ; le second biologique qui se focalise sur des facteurs biologiques comme les gênes, les hormones, le développement neuronal. Les différences de sexe observées au niveau des scores d'auto-évaluation entre filles et garçons seraient a priori liées aux facteurs sociaux : aux rôles sociaux par exemple, les hommes étant souvent assignés à certains rôles comme l'ingénierie, les mathématiques et, aux stéréotypes, les filles étant souvent qualifiées moins performantes en mathématiques, logique, spatial…D'ailleurs, au niveau des performances observées au WISC, Grégoire (2006) montre que les différences de performance entre filles et garçons ont diminué entre WISC-III et WISC-IV grâce à l'uniformisation progressive de l'éducation des filles et des garçons depuis les années 60.

Bien que nous retrouvions des différences entre filles et garçons dans certains facteurs, elles ont été réduites par l'analyse différentielle des items, qui mériterait d'être développée davantage pour les questionnaires d'auto-évaluation comme cela est prôné par différentes commissions internationales.

Limites de cette étude

Cette étude présente un certain nombre de limites : en premier lieu, pour compléter la validation du questionnaire QAE, une étude test-retest serait nécessaire pour étudier la fidélité temporelle ; deuxièmement, une étude complémentaire sur un autre échantillon permettrait de mener une analyse factorielle confirmatoire pour valider le modèle théorique en six facteurs.

Il conviendrait, en outre, de revoir les items du facteur Gsm qui ne présentent aucun lien avec le WISC-IV, ou la réussite scolaire. Gc pourrait également être amélioré car contrairement à ce qui était attendu, ce facteur ne corrèle significativement ni avec l'ICV du WISC-IV, ni avec le domaine de la Communication de la Vineland-II.

Pour l'étude de la validité convergente, les échantillons sont restreints et les résultats observés mériteraient d'être confirmés sur une population plus importante.

Enfin, pour pouvoir montrer l'apport du QAE en terme de validité prédictive incrémentielle en ce qui concerne la réussite scolaire, une étude complémentaire avec les résultats au WISC, les résultats scolaires et les résultats au QAE pourrait être intéressante.

Ainsi, bien que présentant quelques résultats intéressants, les items du questionnaire pourraient encore être améliorés et des études complémentaires de validité sont nécessaires, notamment pour confirmer sa structure.

Implications pratiques

Utiliser un questionnaire d'évaluation des capacités cognitives présente de nombreux avantages pour les praticien.ne.s comme expliqué en introduction : ils sont peu onéreux, rapides à administrer, flexibles dans leur utilisation (Jacobs & Roodenburg, 2014) ; ils diminuent l'anxiété chez la personne évaluée (Furnham, 2001 ; Paulhus, Lysy & Yik, 1998 ; Simms, 2008), permettent d'investiguer des comportements difficilement observables (Tarrant & Cordell, 1997), sont source de meilleure adhésion et coopération et enfin ils proposent des items en lien avec la vie quotidienne, ce qui permet éventuellement de leur octroyer une certaine validité écologique. L'approche multidimensionnelle du questionnaire permet également d'affiner la compréhension du.de la jeune de son fonctionnement cognitif en différenciant les domaines. Le QAE, une fois amélioré, pourrait être intéressant à utiliser pour confronter les perceptions de l'individu aux données objectives fournies par le test. Cette forme d'évaluation pourrait à terme être utile dans le cadre des pratiques d'évaluation associées à l'orientation. En effet, le.la praticien.ne pourrait repérer les personnes qui se sous-évaluent ou se surévaluent et les aider dans le développement de la connaissance d'elles-mêmes.

4. Etude 2 : Evaluation multiple ou multi-source : apports pour la prédiction de la réussite scolaire ?

4.1 Introduction

L'évaluation de l'intelligence repose, depuis Binet et Simon (1905), surtout sur des tests de performance. Les épreuves sont conçues pour mesurer les aptitudes basées sur la réalisation d'un certain nombre de situations-problèmes. Depuis de nombreuses années, dans d'autres champs de la psychologie, comme la personnalité (Bernaud, 2014), la motivation (Forner, 2010), l'estime de soi (Harter, 1988/2012), l'adaptation (Achenbach, 1991 ; Conners, 2001 ; Reynolds & Kamphaus, 2004), on demande directement à la personne évaluée d'exprimer la perception de son comportement dans diverses situations ou bien à des observateur.trice.s externes (famille, amis, collègues, professionnel.le.s). Il s'agit d'une évaluation dite subjective. Dans le domaine des capacités cognitives, bien que ces dernières années ait émergé un intérêt pour les estimations subjectives de l'intelligence (Chamorro-Premuzic & Furnham, 2006 ; Furnham, 2001 ; Holling & Preckel, 2005 ; Visser, Ashton & Vernon, 2008) et que quelques outils aient été développés, ils ne sont pas utilisés dans la pratique. Comment expliquer que cette

méthodologie, pourtant répandue dans l'évaluation de facteurs conatifs ne soit pas également utilisée pour l'évaluation des capacités cognitives ? Une des premières limites qu'on leur reproche est de présenter un certain nombre de biais dont, pour l'auto-évaluation, une tendance à surestimer ses capacités (désirabilité sociale). Une manière de les limiter serait de proposer une évaluation multiple ou multi-sources dans laquelle plusieurs observateur.trice.s sont sollicités en plus de la personne évaluée. Une autre limite aux évaluations auto ou hétéro-rapportées est liée aux preuves de validité, notamment en ce qui concerne la réussite scolaire. Les tests de performance et notamment ceux de Binet et Simon (1905) ont été créés pour repérer les enfants qui ne pouvaient suivre l'enseignement scolaire classique. De nombreuses études ont montré leur validité pour expliquer et prédire la réussite scolaire avec des corrélations qui sont généralement autour de .54 pour le français et .64 pour les mathématiques (Zimmerman & Woo-Sam, 1997).). Il en est de même avec les variables conatives telles que la personnalité (et notamment les dimensions ouverture et conscience du Big Five), l'estime de soi, la motivation qui sont liées à la réussite scolaire également. Or, dans le champ des évaluations de l'intelligence subjectives, très peu d'études ont été menées sur leurs preuves de validité à ce niveau. L'évaluation multiple des capacités cognitives pourrait-elle permettre de prédire la réussite scolaire ?

Offrir des outils valides aux praticien.ne.s pour repérer des élèves en difficulté scolaire est un enjeu important. Par ailleurs, les évaluations subjectives présentent quelques avantages comme un coût moins élevé, une rapidité et une facilité d'utilisation. Elles permettent aussi de travailler sur la connaissance de soi.

4.1.1 De l'auto-évaluation ou hétéro-évaluation à l'évaluation multiple des capacités cognitives

4.1.1.1 Auto-évaluation

L'auto-évaluation permet d'évaluer la capacité perçue par la personne de réaliser certaines tâches comme retenir des numéros de téléphone, se repérer dans une ville qu'il.elle ne connaît pas. Les items ne réfèrent pas explicitement à l'intelligence, mais font référence aux capacités de la personne dans diverses tâches cognitives.

Les quelques questionnaires que nous avons recensés (Terriot, Grégoire & Loarer, 2017) présentent d'assez bonnes qualités psychométriques, lorsque ces données sont disponibles. Les alphas sont tous supérieurs à .70 pour les sous-échelles de ces différents questionnaires. Les corrélations avec des tests classiques de mesure des capacités cognitives sont variables d'une étude à l'autre, mais le degré de liaison, lorsqu'il existe, est généralement faible ou modéré, indiquant que ces méthodes mesurent un autre construit que celui des tests de performance. Il est vraisemblable qu'elles mesurent le concept de soi, défini comme la perception de soi influencée par les évaluations faites par les autres, les renforcements et les attributions (Bong & Clark, 1999; Bong & Skaalvik, 2003; Shavelson, Hubner & Stanton, 1976). Shavelson et al. (1976) ont dégagé plusieurs caractéristiques du concept de soi : 1° il possède une dimension descriptive (« je suis heureux ») et une dimension évaluative (« je suis bon en mathématiques ») ; 2° il est

multidimensionnel : les perceptions qu'a la personne d'elle-même sont organisées en différentes catégories ; 3° il est organisé hiérarchiquement avec, au sommet le concept de soi global, à la base, les perceptions relatives à des domaines spécifiques et, au milieu, les inférences émises dans un domaine plus large. Dans le champ de l'intelligence, les questionnaires d'auto-évaluation renverraient à un concept de soi cognitif (Ackerman, Bowen, Beier & Kanfer, 2001).

4.1.1.2 L'hétéro-évaluation

L'hétéro-évaluation consiste à demander à l'entourage du sujet (proche ou éloigné) d'évaluer la capacité de ce dernier de réaliser certaines tâches cognitives. Les recherches ont apporté des preuves de leur validité dans différents domaines (Achenbach, McConaughy & Howell, 1987; Fabrega, Ulrich & Loeber, 1996 ; Hart, Lahey, Loeber & Hanson, 1994 ; Loeber, Green, Lahey & Stouthamer-Loeber, 1990 ; Loeber, Green, Lahey & Stouthamer-Loeber, 1991), mais très peu concernant le domaine de l'intelligence (Oliver et al., 2002). A notre connaissance (Terriot, Grégoire & Loarer, 2017), seuls trois questionnaires d'hétéro-évaluation des capacités cognitives sont parus récemment : le *Behavior Rating Inventory of Executive Function* (BRIEF ; Gioia, Isquith, Guy & Kenworthy, 2000) qui évalue les fonctions exécutives, le *Parent Report of Children's Abilities* (PARCA 3 ; Oliver et al., 2002), qui évalue les capacités spatiales, la planification, l'organisation, le raisonnement, la mémoire, le raisonnement quantitatif, les nombres et les jeux symboliques ; et le questionnaire de Waschbusch, Daleiden et

Drabman (2000), composé de six échelles en référence au modèle de Cattell, Horn, Carroll (CHC ; Schneider & McGrew, 2012). Ces questionnaires possèdent certaines qualités psychométriques, telles que la cohérence interne, mais d'autres preuves de validité doivent encore être récoltées. Les corrélations avec des mesures objectives sont faibles ou modérées, ce qui semble indiquer que les hétéro-évaluations ne mesurent pas le même construit que les tests d'intelligence classiques.

4.1.1.3 Avantages des mesures auto et hétéro-rapportées

Dans tous les domaines où ils sont utilisés, les questionnaires d'autoévaluation sont généralement considérés comme ayant des avantages pratiques : peu onéreux, rapides à administrer, flexibles dans leur utilisation (Jacobs & Roodenburg, 2014). Du côté des personnes évaluées, ils favorisent le développement de la connaissance de soi, notamment dans le cadre d'un choix de carrière (Ackerman & Wolman, 2007), diminuent l'anxiété (Furnham, 2001 ; Paulhus, Lysy & Yik, 1998 ; Simms, 2008), permettent d'investiguer des comportements difficilement observables (Tarrant & Cordell, 1997) et bénéficient d'une validité de surface généralement plus importante que les tests classiques et sont source de meilleure adhésion et coopération des personnes évaluées. Quant aux hétéro-évaluations, le fait de solliciter les parents présente plusieurs avantages (Waschbusch, Daleiden & Drabman, 2000) : ils sont les premiers présents auprès du.de la jeune; ils l'observent fréquemment et dans de nombreuses situations.

4.1.1.3 Limites des mesures auto et hétéro-rapportées

Malgré ces avantages, quelques inconvénients sont relevés, notamment en lien avec la question de la subjectivité pour les mesures auto-rapportées. En premier lieu, pour conserver le plus possible une bonne image d'elles-mêmes et d'après la théorie de l'optimisme comparatif (Weinstein, 1980), les personnes ont tendance à s'auto-évaluer plus favorablement qu'elles n'évalueraient les autres. De plus, les individus présentant une forte confiance en eux auraient tendance à se surestimer comparativement aux individus ayant une faible confiance en eux (Harris & Schaubroeck, 1988). Ils auraient donc des difficultés à produire des jugements objectifs (Meyer et al., 2001), seraient sujets à des distorsions ou des reconstructions (Hofstee, 1994), et auraient tendance à surestimer leurs compétences (Barr, 2007 ; Fuj, Hennessy & Mak, 1985 ; Geller, 1981 ; Warriner et al., 1984).

Pour limiter certains de ces biais, Rushton, Brainerd et Pressley (1983), par exemple, ont suggéré qu'agréger des scores issus de rapporteurs multiples sur un même objet pouvait permettre de capturer une plus grande partie de la variance vraie et donc d'améliorer la généralisation des mesures.

4.1.1.4 Vers l'évaluation multiple

On parle d'évaluation multiple lorsqu'une évaluation est réalisée par un ensemble de personnes qui peuvent inclure le sujet lui-même.

Dans le domaine des ressources humaines et du management en particulier, le terme consacré est celui d'évaluation à 360 degrés (Levy-Leboyer, 2000), qui comprend une évaluation par le sujet concerné d'une part et, d'autre part par des observateur.trice.s (supérieur.e.s hiérarchiques, subordonné.e.s, collègues, ami.e.s, familles, client.e.s…). Lorsque le sujet est lui-même impliqué, le 360 degrés correspond à une évaluation dans laquelle on recueille donc l'auto-évaluation et les hétéro-évaluations de ses observateur.trice.s. L'objectif est de mettre à jour les écarts et les disparités existant entre la perception que le sujet a de lui-même et celles que ses observateur.trice.s ont de lui.

Outre le domaine du management, l'évaluation multiple existe également dans le champ du comportement adaptatif (Achenbach, 1991), commence à se développer dans celui des intérêts (Vrignaud & Cuvillier, 2006), des fonctions exécutives (Gioia, Isquith, Guy & Kenworthy, 2000), mais très peu, au niveau des capacités cognitives.

Aucune information seule ne peut donner une image fiable et infaillible de la personne évaluée. L'utilisation de plusieurs sources d'information est donc utile pour croiser les informations (Dubowitz & Sawyer, 1994 ; Stanger & Lewis, 1993 ; Starr, Dubowitz, Harrington & Feigelman, 1998). Comparer les évaluations faites par plusieurs personnes permet également de voir si le comportement du sujet est différent selon le contexte (ex : maison versus école). La complexité d'une personne, qui est en développement (adolescence) nécessite la complexité de la démarche et l'évaluation multi-source permet d'approcher cette complexité. Par ailleurs, elle est perçue comme efficace, car les informations qu'elle fournit

proviennent de sources multiples et de personnes dont l'opinion est importante (Hazucha, Hezlett & Shneider, 1993). Cette information acquiert donc une plus grande crédibilité et augmente ainsi la motivation de la personne pour changer son comportement (Latham & Wexley, 1981).

La difficulté principale de l'évaluation multiple tient au fait que, généralement, des décalages sont observés entre les évaluations.

4.1.1.5 Les écarts entre les évaluateur.trice.s

4.1.1.5.1 Entre l'auto-évaluation et les hétéro-évaluations

Quel que soit le construit étudié, plusieurs études ont relevé un décalage entre les évaluations, avec des taux d'accords moyens notamment entre l'auto-évaluation et les hétéro-évaluations (Achenbach, 1991; Achenbach, McConaughy & Howell, 1987 ; Biederman, Mick & Faraone, 1998 ; Campbell & Lee, 1988 ; Shrauger & Schoeneman, 1979). Dans la méta-analyse réalisée par Achenbach, McConaughy et Howell (1987) dans le domaine de la psychopathologie, une corrélation moyenne de 0,22 a été trouvée entre l'évaluation réalisée par les jeunes et celle menée par les autres informateur.trice.s. L'auto-évaluation ne prédit pas l'évaluation faite par les observateur.trice.s (Mabe & West, 1982). Ceci serait expliqué en partie par la théorie de l'attribution (Jones & Nisbett, 1972; Lord, Vader & Alliger, 1986) : les personnes évaluées auraient tendance à attribuer les bonnes performances à leur comportement et les mauvaises à l'environnement, alors que les observateur.trice.s attribueraient plus volontiers les bonnes

performances a l'environnement et les mauvaises à la personne évaluée. D'autres auteurs, comme Furman et Jacob (1997), évoquent des différences réelles dans les perspectives de chacun des évaluateur.trice.s ; chaque mesure évaluerait un construit différent même si il s'agit de la même dimension.

4.1.1.5.2 Entre les hétéro-évaluations faites par les parents et celles faites par les enseignant.e.s

La littérature rapporte également des décalages systématiques entre les évaluations rapportées par les parents lors de l'évaluation du comportement de leur enfant et celles provenant des enseignant.e.s. Achenbach, McConaughy et Howell (1987) dans leur méta-analyse évoquent une corrélation de 0,28 entre différents types d'informateur.trice.s (par exemple parent / enseignant.e). Dans le champ du comportement adaptatif, notamment, les parents tendent à rapporter systématiquement plus de difficultés que ne le fait l'enseignant.e (Achenbach, McConaughy & Howell, 1987 ; Grietens, Onghena, Prinzie, Gadeyne, Van Assche, Ghesquiere & Hellinckx, 2004 ; Van der Ende & Verhulst, 2005 ; Winsler & Wallace, 2002). Ainsi, les corrélations sont modérées à élevées entre les deux parents, mais faibles à modérées entre les parents et les enseignant.e.s (Achenbach et al., 1987 ; De Los Reyes & Kazdin, 2005).

4.1.1.5.3 Comment évaluer l'accord entre évaluateur.trice.s ?

James, Demaree et Wolf (1993) décrivent deux aspects de l'accord entre évaluateur.trice.s :

- l'accord peut être compris comme l'étendue de la similarité sur le classement de la personne évaluée. Il sera étudié par une approche corrélationnelle et si la corrélation est forte, cela reflétera un modèle commun (Kenny, 1991) ;
- il peut sinon être entendu comme la différence au niveau moyen des évaluations de chacun. Cela sous-entend une approche par ANOVA.

Kraemer, Measelle, Ablow, Essex, Boyce et Kupfer (2003) proposent une autre méthodologie pour prendre en compte les désaccords entre informateur.trice.s. Pour ces derniers, les désaccords ne sont pas considérés comme des erreurs de mesure à corriger, mais comme des renseignements précieux sur l'expression différente des capacités du.de la jeune selon le contexte. Les décalages observés entre les informateur.trice.s sont la conséquence du fait même de combiner plusieurs sources d'information. Il n'existe pas de mesure idéale, chaque informateur.trice. détenant une partie de la

« vérité ». L'addition des sources permet de tendre vers ce score vrai. Ainsi, au lieu d'étudier les différences de moyenne, ils proposent une méthode basée sur l'analyse factorielle. Les auteurs définissent trois dimensions et modélisent leur proposition par cette formule : I= T+C+P+E. I correspond à l'information donnée par une source, T est la dimension « Trait » qui représente la variable mesurée, C est la dimension « Contexte » et correspond aux facteurs influençant la manière

dont les sujets expriment la dimension trait, P est la dimension « Perspective » et renvoie aux caractéristiques des sources d'information qui peuvent influer l'évaluation de la dimension et enfin E correspond à l'erreur de mesure. Ils proposent de mener une analyse factorielle en composantes principales, car les données émanant des différentes sources d'information sont une combinaison de variables orthogonales latentes (T, C et P).

4.1.2 Evaluation multiple et réussite scolaire

Plusieurs études ont établi l'existence d'une relation entre les mesures auto-rapportées de l'efficience cognitive et les résultats scolaires (Kornilova & Novikova, 2013). Dans la méta-analyse réalisée par Kornilov (2011), sur neuf études recensées, une corrélation moyenne positive et significative de 0,28 (valeur de p non précisée) est trouvée. Kornilova et Novikova (2013) observent une corrélation de 0,16 entre la mesure auto-rapportée et la réussite scolaire ($p<.01$). Bien que significatives, les corrélations entre auto-évaluations et réussite scolaire restent faibles et moins élevées que celles entre les tests de performance et la réussite scolaire, qui sont autour de .54 pour le français et .64 pour les mathématiques (Zimmerman & Woo-Sam, 1997).

Très peu de chercheurs se sont intéressés à la validité prédictive des auto-évaluations. Dans notre revue de littérature (Terriot, Grégoire & Loarer, 2017), nous avons relevé une seule recherche, celle de Kornilova, Kornilov et Chumakova (2009), auprès d'étudiant.e.s russes, centrée sur la validité prédictive

de la réussite universitaire. En complément de l'âge, du sexe et des résultats aux tests de performance, l'auto-estimation ajouterait 3% dans la prédiction de la variance des résultats universitaires six mois après la passation des tests (Kornilova, Kornilov & Chumakova, 2009). Selon cette seule étude, l'auto-évaluation permettrait d'augmenter la prédiction des résultats scolaires, même si cette augmentation reste très faible.

Concernant les hétéro-évaluations, une seule étude (Kornilova, Kornilov & Chumakova, 2009) a porté sur les corrélations entre l'hétéro-estimation de l'intelligence et les résultats scolaires, ainsi qu'à la validité prédictive. Les corrélations avec les résultats universitaires sur trois trimestres sont de -0,65 ($p<.01$). Enfin, l'hétéro-évaluation ajoute 24 % dans la prédiction des résultats universitaires à la fin du semestre. Elle semble ainsi montrer de meilleures preuves de validité prédictive que l'auto-évaluation, qui ajoutait seulement 3% dans la prédiction même si d'autres études seraient nécessaires pour le confirmer.

4.2 Objectifs de la présente étude

Cette étude se pose pour objectifs de développer une évaluation multiple des capacités cognitives en prenant trois sources d'informations : le sujet lui-même, un de ses parents et son enseignant.e ; d'observer les écarts éventuels entre les résultats aux trois questionnaires et enfin de voir dans quelle mesure elle permet de prédire la réussite scolaire en fin d'année, et l'apport de chaque questionnaire à cette prédiction.

4.3 Méthodologie

4.3.1 Participant.e.s

Les participant.e.s ont été recrutés au sein d'établissements scolaires de plusieurs départements. Pour ce faire, un courrier a été adressé aux chef.fe.s d'établissement, puis un formulaire de consentement a été adressé aux familles. Suite à leur acceptation, trois questionnaires ont été administrés : un auprès des élèves en passation collective, un second par un des deux parents à la maison et un troisième par le.la professeur.e principal.e de la classe.

Un échantillon initial de 198 sujets a été constitué. Onze personnes aux valeurs extrêmes au questionnaire rempli par un des deux parents, ainsi que deux personnes aux valeurs extrêmes au questionnaire rempli par l'enseignant.e ont été retirées en utilisant à la fois les scores Z et la distance de Mahalanobis (Tabachnick & Fidell, 2013).

Enfin nous avons retiré tous les sujets pour lesquels il manquait des données à l'un des trois questionnaires. Notre échantillon final se compose de 139 sujets, âgés de 12 à 18 ans (M=3,66 ; ET=1,07), avec 54% de garçons et 46% de filles, scolarisés en quatrième (90%), seconde professionnelle (4,2%), et CAP (5,8%).

Ce sont principalement les mères qui ont rempli le questionnaire à destination des parents (78,5%). Le niveau de qualification le plus élevé dans la famille a été également recueilli (pour 88 sujets) : 19,3% possèdent le niveau Bac+4/Bac+5,

15,9% le niveau Bac+3, 13,6% le niveau Bac+2, 21,6% le niveau Baccalauréat, 18,2% le niveau CAP/BEP et 11,4% sont sans diplôme.

Concernant le questionnaire à destination des enseignant.e.s, 62,2% de femmes et 37,8% d'hommes ont répondu.

4.3.2 Mesures

Pour un même élève, trois questionnaires ont été complétés : un par le.la jeune, un par l'un des deux parents et un par le.la professeur.e principal.e.

Les jeunes ont rempli le questionnaire d'auto-évaluation des capacités cognitives (QAE). Il se compose de 40 items portant sur des activités en lien avec la vie quotidienne et le champ scolaire, pour lesquelles le sujet doit se positionner en indiquant si l'activité lui est beaucoup plus difficile, plus difficile, pareil, plus facile ou beaucoup plus facile que pour les autres jeunes de son âge. Le questionnaire repose sur le modèle de Cattell-Horn-Carroll (CHC ; Schneider & McGrew, 2012) et permet d'évaluer six facteurs :

Nom du facteur	Description	Exemple d'item	Alpha Validation	Alpha Echantillon
Intelligence cristallisée (Gc)	Mesure l'étendue et la profondeur des connaissances acquises ainsi que l'habileté à utiliser ces connaissances	« Trouver le mot exact pour exprimer mes pensées »,	.83	.82
Intelligence fluide (Gf)	Mesure la capacité de résoudre de nouveaux problèmes, inhabituels pour le sujet et le raisonnement quantitatif	« Résoudre des problèmes logiques »	.86	.86
Traitement visuel Construction (Gvc)	Mesure l'habileté à générer, percevoir, analyser, emmagasiner, manipuler, transformer des stimuli visuels pour résoudre des problèmes (compétences réduites)	«Assembler les pièces d'un puzzle »	.75	.71
Traitement visuel Orientation (Gvo)	Mesure les compétences spatiales larges qui incluent l'environnement	« M'orienter dans une ville »	.75	.67
Mémoire à court terme (Gsm)	Capacité à saisir une information, à la retenir à court terme et à l'utiliser au cours des prochaines secondes	« Retenir les numéros de téléphone »	.75	.76
Vitesse de traitement (Gs)	Habileté à effectuer des tâches automatiques rapidement et efficacement, en maintenant notre attention et concentration	« Recopier rapidement un texte »	.74	.71

Tableau 13. Description des facteurs du QAE

Comme indiqué dans le tableau 13, les alpha calculés sur les corrélations polychoriques dans l'étude de validation vont de .74 à .86 pour les six facteurs, suggérant une bonne consistance interne pour toutes les échelles. Dans notre échantillon, l'alpha pour Gvo est plus réduit et à la limite du seuil acceptable.

Dans l'étude de validation, les corrélations inter-facteurs vont de .04 (entre Gf et Gsm) à .31 (entre Gc et Gf). Des études de validité convergente ont été effectuées avec la performance scolaire : les facteurs Gf et Gc corrèlent avec la moyenne générale recueillie au troisième trimestre, avec des corrélations respectives de .45

et .26. Avec le WISC-IV (Wechsler, 2005), le facteur Gf corrèle de manière positive et significative (*r* =.42 avec le QIT). Et enfin avec la Vineland II (VABS-II ; Sparrow, Cicchetti & Balla, 2015), échelle d'évaluation des comportements socio-adaptatifs. le QAE entretient des liens positifs et significatifs en particulier avec Gs (r=.69 avec la note composite),

Un des deux parents a rempli le questionnaire d'hétéro-évaluation pour parents (QHP) qui se compose des mêmes items, et présentés dans le même ordre que ceux du QAE. Seule la consigne change puisqu'on leur demande d'indiquer pour chaque phrase, si l'activité est, pour leur enfant, beaucoup plus difficile, plus difficile, pareille, plus facile ou beaucoup plus facile que pour les autres jeunes de son âge.

Comme indiqué dans le tableau 14, les alpha du QHP vont de .67 à .90 pour les six facteurs, suggérant une bonne consistance interne pour toutes les échelles, sauf pour le traitement visuel Orientation (Gvo) à la limite du seuil acceptable.

	Gc	Gf	Gvc	Gvo	Gsm	Gs
QHP	.89	.90	.73	.67	.82	.70

Tableau 14. Alpha pour les six facteurs du QHP

Enfin, un troisième questionnaire a été développé : le questionnaire d'hétéro-évaluation pour les enseignant.e.s (QHE) en ne conservant que les items en lien avec le domaine scolaire.

Il se compose de 17 items. Suite à une analyse factorielle, les items se répartissent en trois facteurs seulement : intelligence cristallisée (Gc), intelligence fluide (Gf) et vitesse de traitement (Gs). La consigne est identique à celle du QAE/QHP mais en désignant l'élève. Les alpha de chaque facteur sont très élevés : .94 pour Gc, .92 pour Gf et .87 pour Gs

Les résultats scolaires en fin d'année scolaire, soit de 3 à 6 mois plus tard, ont également été recueillis (moyenne générale au 3ème trimestre ou au 2ème semestre selon la classe) pour 109 sujets en excluant la note du.de la professeur.e ayant rempli le questionnaire.

4.4 Résultats

Les données ont été analysées en utilisant SPSS 20.0 (IBM Corp. Released 2011) et AMOS 20.0 (Arbuckle, 1983–2011).

4.4.1 Etude des trois questionnaires

4.4.1.1 Statistiques descriptives

Le tableau 15 présente les statistiques descriptives pour les facteurs des trois questionnaires utilisés dans l'étude. Les valeurs des coefficients d'asymétrie et d'aplatissement sont comprises entre 1 et -1 indiquant que les variables sont proches de la normalité (Carricano, Poujol & Bertrandias, 2010). Les facteurs

manquent parfois de sensibilité puisque toutes les valeurs extrêmes ne sont pas utilisées, sauf pour le facteur Gs quel que soit le questionnaire.

	Min	Max	M	ET	Asymétrie	Aplatissement
QAE						
Gc	15.00	43.00	28.76	5.55	.23	-.50
Gf	10.00	38.00	23.85	6.40	.07	-.66
Gvc	13.00	35.00	24.26	4.28	-.03	-.19
Gvo	6.00	25.00	15.99	3.44	-.06	.16
Gsm	13.00	40.00	27.45	5.10	-.02	-.01
Gs	4.00	15.00	10.59	2.57	-.06	-.68
QHP						
Gc	17.00	42.00	28.89	5.70	.39	-.38
Gf	9.00	38.00	23.80	6.14	.09	-.32
Gvc	15.00	35.00	23.28	3.73	.53	.46
Gvo	6.00	24.00	15.01	2.85	-.14	1.34
Gsm	17.00	39.00	27.79	4.65	.14	-.40
Gs	3.00	15.00	9.48	2.42	-.07	-.12
QHE						
Gc	11.00	40.00	25.56	6.87	.16	-.52
Gf	5.00	25.00	14.76	4.33	.22	.14
Gs	4.00	20.00	13.03	3.31	.18	.00

Tableau 15. Statistiques descriptives du QAE, QHP et QHE (N=139)

4.4.1.2 Corrélations

Le tableau 16 présente les corrélations entre les facteurs des trois questionnaires utilisés dans cette étude. Chaque facteur d'un questionnaire tend à davantage corréler avec les autres facteurs du même questionnaire et/ou le même facteur des deux autres questionnaires. A titre d'exemple, Gc du QAE entretient davantage de liens positifs et significatifs avec les autres facteurs du QAE (les corrélations vont de .25 avec Gvc à .54 avec Gsm) et Gc du QHP (r=.50) ou Gc du QHE (r=.39)

qu'avec les autres facteurs du QHP ou du QHE. Seul Gs du QHE ne suit pas cette tendance car il tend à davantage corréler avec Gc et Gf du QAE et du QHP qu'avec Gs de ces deux questionnaires.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
QAE															
(1) Gc	1.00														
(2) Gf	.45**	1.00													
(3) Gvc	.25**	.36**	1.00												
(4) Gvo	.40**	.42**	.46**	1.00											
(5) Gsm	.54**	.32*	.35**	.42**	1.00										
(6) Gs	.45**	.28**	.18	.23**	.33**	1.00									
QHP															
(7) Gc	.50**	.25**	.05	.15	.23**	.11	1.00								
(8) Gf	.20*	.63**	.20*	.27**	.13	.01	.48**	1.00							
(9) Gv	-.04	.27**	.42**	.31**	.05	.02	.32**	.61**	1.00						
(10) Go	.19*	.20**	.10	.45**	.16	.13	.39**	.40**	.43**	1.00					
(11) Gsm	.26**	.17**	-.02	.13	.29**	.17*	.69**	.48**	.38**	.43**	1.00				
(12) Gs	.20*	.12	-.01	0.09	.21*	.40**	.45**	.27**	.35**	.29**	.65**	1.00			
QHE															
(13) Gc	.39**	.36**	.09	.12	.15	.11	.44**	.41**	.18*	.18*	.29**	.15	1.00		
(14) Gf	.31**	.50**	.08	.14	.09	.05	.33**	.50**	.21*	.15	.16	.00	.83**	1.00	
(15) Gs	.36**	.34**	.17*	.22**	.21*	.13	.36**	.41**	.21*	.17	.27**	.11	.79**	.74**	1.00

* $p < 0.05$ (bilatéral)
** $p < 0.01$ (bilatéral)

Tableau 16. Inter-corrélations pour les 15 facteurs des 3 questionnaires

4.4.1.3 Coefficients intra-classes et comparaisons de moyennes

Afin de préciser les liens entre les questionnaires et surtout voir les éventuels décalages entre les réponses des différentes sources (élève et parent), nous avons calculé le coefficient intra-classe (CIC) et avons procédé à des comparaisons de moyennes entre les facteurs, en indiquant la taille d'effet à chaque fois.

Concernant les comparaisons entre élèves ou parents et enseignant.e.s, nous n'avons pas pu effectuer ces analyses puisque les facteurs ne comportent pas les mêmes items.

	QAE		QHP					
	M	ET	M	ET	CIC	Différence moyenne	t	η^2[12]
Gc	28.76	5.5	28.89	5.70	.50**	-.12	-.26	.00
Gf	23.85	6.40	23.80	6.14	.63**	.06	.12	.00
Gvc	24.26	4.28	23.28	3.73	.42**	.98	2.67**	.05
Gvo	15.99	3.44	15.01	2.85	.45**	.98	3.47**	.08
Gsm	27.45	5.10	27.79	4.65	.29**	-.34	-.70	.00
Gs	10.59	2.57	9.48	2.42	.40**	1.11	4.78**	.14

** p-value<.01

Tableau 17. Comparaisons de moyennes entre le QAE et le QHP

Les coefficients intra-classes sont modérés mais positifs et significatifs, plus forts pour les facteurs Gf (=.63) puis Gc (=.50). Les différences de moyennes sont non significatives pour Gc, Gf et Gsm, ce qui indique qu'il n'existe pas de différence entre l'évaluation par le.la jeune et celle de son parent. Par contre, nous notons une différence significative (p-value<.01) de taille moyenne pour les facteurs Gvc (t=2.67), Gvo (t=3.47) et un effet de grande taille pour Gs (t=4.78). Les jeunes s'auto-évaluent significativement plus favorablement que leur parent pour ces trois facteurs.

[12] L'éta-carré est la proportion de variance de la variable testée expliquée par la variable groupe et permet de tester la taille d'effet. Cohen (1988) parle d'un effet de petite taille quand l'éta-carré est autour de .01, de taille moyenne autour de .06 et de grande taille autour de .14

4.4.2 Procédure multi-informateur.trice.s

Afin de combiner les données provenant des différentes sources d'information (élève, parent, enseignant.e), la procédure proposée par Kraemer et al. (2003), explicitée plus haut, a été appliquée.

L'évaluation réalisée par les parents, les enseignant.e.s et les élèves a ainsi été introduite dans une analyse factorielle en composantes principales. Des analyses séparées ont été réalisées pour chaque facteur commun : Gc, Gf et Gs.

Trois facteurs ont été extraits : le premier représente donc le facteur évalué (Trait), le deuxième, la différence entre l'évaluation des parents et celle des professionnel.le.s (Perspective) et le troisième, la différence entre le comportement général dans les situations de la vie quotidienne et les comportements spécifiques à l'école (Contexte). Ainsi, le trait représente le comportement de l'adolescent.e tel que tous les informateur.trice.s ont contribué à l'évaluer mais débarrassé des éléments liés au contexte et à la perspective de chacun des informateur.trice.s. Ces aspects de contexte et de perspective sont représentés par les deux autres facteurs résultant de l'analyse.

Le facteur trait est le premier facteur issu de l'analyse, celui expliquant la plus grande part de variance et sur lequel les différents informateur.trice.s saturent fortement. Les deuxième et troisième facteurs correspondent au contraste entre les perspectives et les contextes.

L'analyse factorielle combinant les informations provenant des parents, des enseignant.e.s et des élèves fournit une solution factorielle intéressante indiquant

l'adéquation de la procédure proposée par Kraemer et al. (2003) dans le cadre de l'évaluation des capacités cognitives. Le facteur trait constitue en effet le facteur expliquant la plus grande part de variance, quel que soit le domaine CHC évalué, soit 62,76% pour Gc, 69,64% pour Gf et 48,68% pour Gs. Les saturations factorielles sur le trait sont ainsi élevées (> 0,40) pour toutes les sources d'information mais particulièrement pour Gf puis Gc. Le second facteur « Perspective » contraste les réponses des enfants et parents à celles des enseignant.e.s (famille versus professionnel.le.s) et le troisième « Contexte » oppose les évaluations des élèves avec celles des parents, les réponses des enseignant.e.s étant intermédiaires.

	Trait	Perspective	Contexte
Gc			
Elève	.80	-.44	.42
Parent	.82	-.17	-.54
Enseignant.e	.76	.64	.15
Variance attribuée au facteur	62.76%	20.74%	16.50%
Gf			
Elève	.86	-.29	.43
Parent	.86	-.28	-.43
Enseignant.e	.78	.62	.07
Variance attribuée au facteur	69.64%	18.17%	12.19%
Gs			
Elève	.81	-.22	-.55
Parent	.80	-.25	.55
Enseignant	.41	.91	.02
Variance attribuée au facteur	48.68%	31.30%	20.03%

Tableau 18. Analyse en composantes principales des facteurs Gc, Gf et Gs en utilisant la méthode de Kraemer et al. (2003)

4.4.3 Liens entre l'évaluation multiple et la performance scolaire

4.4.3.1 Corrélations

Les corrélations entre les facteurs des trois questionnaires et la moyenne générale obtenue en fin d'année scolaire, soit de 3 à 6 mois après la passation des questionnaires figurent dans le tableau 19.

	QAE						QHP						QHE		
	Gc	Gf	Gvc	Gvo	Gsm	Gs	Gc	Gf	Gvc	Gvo	Gsm	Gs	Gc	Gf	Gs
T3	.33**	.38**	.11	.15	,28**	,14	,42**	,45**	,19*	,04	,41**	,24*	,77**	,67**	,59**

**. La corrélation est significative au niveau 0.01 (bilatéral).
*. La corrélation est significative au niveau 0.05 (bilatéral).

Tableau 19. Corrélations entre les facteurs du QAE, QHP et QHE et la moyenne générale au troisième trimestre

Les corrélations sont nulles avec Gvc, Gvo et Gs du QAE, Gvo du QHP, positives et significatives mais faibles avec Gsm du QAE (r=.28), Gvc du QHP (r=.19), modérées faibles avec Gc (r=.33) et Gf (r=.38) du QAE, Gc (r=.42), Gf (r=.45), Gsm (r=.41) et Gs (r=.24) du QHP et modérées fortes voire fortes avec les trois facteurs du QHE (.59 pour Gs à .77 pour Gc).

4.4.3.2 Analyse de régression hiérarchique

À partir d'une analyse de régression multiple, nous avons cherché à évaluer la contribution conjointe des trois questionnaires dans la prédiction de la moyenne

générale en fin d'année et leur validité prédictive incrémentielle. Ces résultats sont présentés dans le tableau 20.

Comme nous n'avons pas observé de différence significative au niveau des résultats scolaires entre les filles et les garçons, nous n'avons pas introduit la variable sexe comme variable contrôle. L'effet conjoint des six facteurs du QAE (modèle 1) explique 21% de la variance de la moyenne générale au troisième trimestre. Les six facteurs du QHP introduits dans un second temps (modèle 2) ajoutent 14% de plus. Enfin, les trois facteurs du QHE rentrés en dernier (modèle 3) ajoutent 34%.

Au final, les trois questionnaires permettent de prédire 69% de la moyenne générale obtenue par les jeunes au troisième trimestre.

	β	t	Résumé
Modèle 1			R^2=.21
Gc QAE	.23	1.89	F(6,102)=4.46**
Gf QAE	.32	3.17**	ER=2.25
Gvc QAE	-.05	-.40	
Gvo QAE	-.07	-1.25	
Gsm QAE	.11	1.63	
Gs QAE	-.11	-1.30	
Modèle 2			R^2=.35
Gc QAE	.12	.80	F(12,96)=4.31**
Gf QAE	.11	.71	ER=2.10
Gvc QAE	-.07	-.57	ΔR^2=.14
Gvo QAE	.04	.31	
Gsm QAE	.09	.79	
Gs QAE	-.08	-.62	
Gc QHP	.17	1.08	
Gf QHP	.29	1.74	
Gvc QHP	.01	-.06	
Gvo QHP	-.28	-2.59*	
Gsm QHP	.16	1.04	
Gs QHP	.04	.29	
Modèle 3			R^2=.69
Gc QAE	-.07	-.69	F(15,93)=13.47**
Gf QAE	.09	.79	ER=1.49
Gvc QAE	.03	.37	ΔR^2=.34
Gvo QAE	.05	.59	
Gsm QAE	.14	1.69	
Gs QAE	-.11	-1.30	
Gc QHP	.03	.29	
Gf QHP	.04	.34	
Gv QHP	-.01	-.09	
Go QHP	-.25	-3.26**	
Gsm QHP	.11	1.01	
Gs QHP	.11	1.27	
Gc QHE	.75	5.71**	
Gf QHE	.08	.55	
Gs QHE	-.13	-1.20	

* p<.05.

** p<.01.

Tableau 20. Régressions hiérarchiques

4.5 Discussion

Cette étude avait pour objectif de développer une évaluation subjective multiple des capacités cognitives en prenant trois sources d'informations : le sujet lui-même, un de ses parents et son enseignant.e ; d'observer les écarts éventuels entre les résultats aux trois questionnaires et enfin de voir dans quelle mesure l'évaluation multi-sources permet de prédire la performance scolaire en fin d'année, avec l'apport de chaque questionnaire à cette prédiction.

Pour cela, nous avons fait passer le même questionnaire à des jeunes et l'un de leurs parents ainsi qu'un questionnaire abrégé avec les items ne portant que sur le champ scolaire aux enseignant.e.s. de la classe où ils sont scolarisés. Le questionnaire à destination des adolescent.e.s (le QAE), celui pour les parents (le QHP) comprennent 40 items. Celui pour les enseignant.e.s n'en comporte que 17. Les trois proposent des items en lien avec la vie quotidienne (dont le champ scolaire) et reposent sur le modèle CHC. Le QAE et le QHP permettent d'évaluer six facteurs : l'intelligence cristallisée (Gc), fluide (Gf), le traitement visuo-spatial réduit (Gvc), les compétences visuo-spatiales élargies(Gvo), la mémoire à court terme (Gsm) et la vitesse de traitement (Gs). Le QHE permet d'évaluer seulement trois facteurs : Gc, Gf et Gs. Les alpha de chaque questionnaire sont bons.

Les résultats scolaires en fin d'année scolaire, soit de 3 à 6 mois plus tard, ont également été recueillis (moyenne générale au 3ème trimestre ou au 2ème semestre selon la classe) pour 109 sujets.

En ce qui concerne les écarts entre évaluateur.trice.s, nous avons utilisé trois indicateurs : 1) les coefficients intra-classes ; 2) les différences de moyennes pour chaque facteur avec la taille d'effet associée ; 3) la méthodologie de Kraemer et al. (2003) qui permet, par l'utilisation d'une analyse factorielle de montrer, pour chaque facteur, à la fois ce qui est commun aux trois évaluateur.trice.s et ce qui les différencie.

Entre les jeunes et leurs parents, les coefficients intra-classes sont modérés mais positifs et significatifs, plus forts pour les facteurs Gf (=.63) puis Gc (=.50). Dans la méta-analyse réalisée par Achenbach, McConaughy et Howell (1987) mais dans le domaine de la psychopathologie, une corrélation moyenne de 0.22 a été trouvée entre l'auto-évaluation et l'hétéro-évaluation. Nous n'avons pas trouvé d'étude équivalente dans le domaine des évaluations subjectives de l'intelligence. Ainsi, bien qu'un décalage soit observé dans les évaluations entre les jeunes et leurs parents, puisque les corrélations restent modérées, ce décalage semble moins important dans le domaine de l'évaluation subjective des capacités cognitives.

Avec le second indicateur utilisé, à savoir les différences de moyenne pour chaque facteur entre les scores des élèves et ceux des parents, nous observons des différences de moyennes non significatives pour Gc, Gf et Gsm, ce qui indique qu'il n'existe pas de différence entre l'évaluation par le.la jeune et celle de son parent. Par contre, nous notons une différence significative (p-value<.01) de taille moyenne pour les facteurs Gvc (t=2.67), Gvo (t=3.47) et de grande taille pour Gs (t=4.78). Les jeunes s'auto-évaluent significativement plus favorablement que leur parent pour ces trois facteurs.

Ainsi, il existe des différences au niveau de l'accord entre élève et parent selon le domaine investigué : un accord plus élevé pour Gc et Gf. Les jeunes ont peut-être davantage de connaissances d'eux-mêmes en ce qui concerne les domaines de l'intelligence cristallisée et fluide, ayant des feed-backs réguliers dans ces domaines, notamment à l'école.

Ces analyses n'ont pas pu être faites entre enseignant.e.s et élèves/parents du fait que les facteurs ne sont pas composés des mêmes items.

Pour résumer, l'accord entre élève et parents semble plus fort pour Gf puis Gc. C'est ce qui semble également se dégager des résultats issus de la procédure proposée par Kraemer et al. (2003). En effet, c'est pour Gf que le facteur Trait explique la plus grande part de variance (72,70%) puis Gc (63,71%) et enfin Gs (seulement 54,70%).

Le fait que les accords moyens entre évaluateur.trice.s soient plus élevés entre Gf et Gc peut être lié à une meilleure connaissance et donc évaluation de ces facteurs. Les items qui composent ces facteurs sont en partie en lien avec le champ scolaire, pour lequel le.la jeune et ses parents ont des retours réguliers, par le biais des évaluations.

Dans un second temps, cette étude avait pour objectif de montrer l'apport de l'évaluation multiple à la prédiction de la performance scolaire 3 à 6 mois après.

Les corrélations entre chaque facteur de chaque questionnaire et la moyenne générale en fin d'année scolaire ont été étudiées. Les corrélations sont nulles avec Gv, Go et Gs du QAE, Go du QHP, positives et significatives mais faibles avec

Gsm du QAE (r=.28), Gvc du QHP (r=.19), modérées faibles avec Gc (r=.33), Gf (r=.38) du QAE, Gc (r=.42), Gf (r=.45), Gsm (r=.41), Gs (r=.24) du QHP et modérées fortes voire fortes avec les trois facteurs du QHE (.59 pour Gs à .77 pour Gc).

Ces résultats peuvent être analysés du point de vue du répondant et du point de vue des facteurs. En ce qui concerne la source, l'enseignant.e semble plus « objectif » dans son évaluation que les parents, puis que les élèves par rapport au critère externe choisi qui est la performance scolaire. Ce résultat n'est pas étonnant. Il.elle a une très bonne connaissance, a priori, des capacités de ses élèves dans le champ scolaire. Il.elle peut également facilement établir des comparaisons entre élèves dans la classe et ainsi produire des classements. Les élèves ont, semble-t-il, plus de difficultés à évaluer de manière juste leurs capacités. Du point de vue des facteurs, les deux les plus en lien avec la performance scolaire, quel que soit le questionnaire sont Gc et Gf. On peut analyser cela du point de vue des items qui composent ces deux facteurs : Gf est composé pour moitié d'items en lien avec le calcul (items 9, 33, 41 et 45) et Gc d'items en lien avec des compétences nécessaires en français. Il est plus étonnant de constater que les autres facteurs, et notamment Gs n'entretiennent aucun lien avec la moyenne au 3ème trimestre alors que ce sont également des capacités utiles pour la réussite scolaire.

Puis, à partir d'une analyse de régression multiple, nous avons cherché à évaluer la contribution conjointe des trois questionnaires dans la prédiction de la performance scolaire en fin d'année. L'effet conjoint des six facteurs du QAE

explique 21% de la variance de la moyenne générale au troisième trimestre. Les six facteurs du QHP introduits dans un second temps ajoutent 14% de plus. Enfin, les trois facteurs du QHE rentrés en dernier ajoutent 34%.

Au final, les trois questionnaires permettent de prédire 69% de la moyenne générale obtenue par les jeunes au troisième trimestre.

Comparativement aux quelques études antérieures (et non équivalentes en termes de population, d'outils, de variables contrôles), ces résultats sont très encourageants. En effet, Kornilova, Kornilov et Chumakova (2009) indiquaient que l'auto-estimation ajoutait 3% dans la prédiction de la variance des résultats universitaires six mois après la passation des tests en complément de l'âge, du sexe et des résultats aux tests de performance, et l'hétéro-évaluation 24 %.

Limites

Notre étude présente un certain nombre de limites : la faible taille de l'échantillon en premier lieu, notamment pour les analyses de régression. Une étude ultérieure serait donc nécessaire avec un échantillon beaucoup plus important.

Par ailleurs, il serait intéressant d'intégrer une mesure de performance cognitive dans le modèle afin de voir l'apport de chaque questionnaire par rapport aux tests d'intelligence. Enfin, la validité prédictive concernant la performance scolaire pourrait être étudiée sur plus long terme (un à deux ans après la passation des questionnaires).

Intérêts pratiques

Comme indiqué en introduction, offrir des outils valides aux praticien.ne.s pour repérer des élèves en difficulté scolaire est un enjeu important. L'évaluation multiple telle que présentée dans cette étude permet de prédire la moyenne générale de 3 à 6 mois après la passation puisque les trois questionnaires utilisés conjointement expliquent 69% des résultats scolaires en fin d'année. Utiliser des évaluations subjectives présentent de nombreux avantages pour les praticien.ne.s : ils sont peu onéreux, rapides à administrer, flexibles dans leur utilisation (Jacobs & Roodenburg, 2014) ; ils diminuent l'anxiété chez la personne évaluée (Furnham, 2001 ; Paulhus, Lysy & Yik, 1998 ; Simms, 2008), sont source de meilleure adhésion et coopération Cette approche multidimensionnelle peut être intéressante à utiliser pour confronter les perceptions du sujet évalué aux perceptions qu'on de lui les autres informateur.trice.s (parent, enseignant.e).

5. Etude 3 : Développement d'une échelle contextualisée des capacités cognitives auprès d'un échantillon d'adolescent.e.s français.e.s : étude exploratoire

5.1 Introduction

Les tests d'intelligence sont aujourd'hui utilisés dans de nombreux domaines : Education, Travail et Santé ; ils sont, en effet, de bons prédicteurs de la réussite scolaire (Mackintosh, 2004) et professionnelle (Schmidt & Hunter, 1998). Ils permettent, par ailleurs et en particulier dans le champ de l'Education, le diagnostic de certains troubles, le recueil d'informations cliniques et la satisfaction d'exigences administratives (Grégoire, 2004). Les tests les plus utilisés dans le monde (Oakland & Hu, 1992) reposent sur des modèles centrés sur la mesure de capacités générales supposées fonctionner également avec tous les types de contenus (Binet & Simon, 1905). Ainsi, ces tests sont constitués de tâches décontextualisées (Huteau & Lautrey, 1999, Chartier & Loarer, 2008) qui ne prennent pas en compte les aptitudes pratiques de la vie quotidienne (Neisser, 1976 cité par Mackintosh, 2004). Or, d'autres modèles insistent sur la contextualisation des processus cognitifs (McLellan, 1996). Quelques outils ont

été développés, principalement dans les champs des ressources humaines et de la neuropsychologie (Chapitre 1). Il n'existe cependant pas, actuellement en France, d'outil d'évaluation contextualisée des capacités cognitives pour un public d'adolescent.e.s. L'objectif principal de cet article, après avoir fait un point sur l'évaluation contextualisée, sera de rendre compte, de manière exploratoire, du développement d'un tel test sur une population d'adolescent.e.s français.e.s.

5.1.1 Limites de l'évaluation classique de l'intelligence

L'évaluation classique de l'intelligence a depuis longtemps montré son efficacité pour prédire la réussite scolaire avec des corrélations entre les résultats aux tests de performance et les notes aux examens scolaires variant de .40 à .70 selon les tests utilisés et les recherches (Brody, 1992 ; Jensen, 1980) ainsi que pour prédire la réussite professionnelle avec des corrélations de l'ordre de .50 à .60 entre le quotient intellectuel (QI) et les performances professionnelles (Brody, 1992 ; Heinstein & Murray, 1994 , Jencks, 1972 ; Schmidt & Hunter, 1998). Dans les tests classiques, les processus évalués sont généralement élémentaires, simples, partiels et mesurés les uns après les autres. La personne doit résoudre consécutivement chaque tâche dans un environnement fermé, à l'abri des sollicitations extérieures. Bien que ces tests classiques s'avèrent être les outils les plus efficaces pour prédire la réussite scolaire et professionnelle qui sont des composantes importantes de la vie quotidienne, ces prédictions restent très imparfaites pour le comportement adaptatif au sens large. Ce dernier se définit

comme la façon dont le sujet fait face aux exigences de la vie courante et à sa capacité à se conformer aux normes escomptées à son âge dans son milieu culturel (American Psychiatric Association, 2013). On constate, par exemple, que certains sujets, malgré des performances « normales » aux tests psychométriques présentent des dysfonctionnements adaptatifs dans la vie de tous les jours (Chan & Chen, 2005, Shallice & Burgess, 1991). Selon l'Office of Strategic Services[13] (OSS, 1948), les tests classiques ne rendent pas suffisamment compte de l'orchestration des processus, nécessaire dans la vie quotidienne du fait de la complexité des tâches rencontrées. Au quotidien, le sujet est confronté à des épreuves non structurées, simultanées et multiples, ce qui nécessite de coordonner plusieurs activités cognitives (McClelland, 1973 ; Sternberg, Conway, Kehon & Bernstein, 1981) et ce, dans un environnement ouvert avec de possibles stimulations interférentes. Par ailleurs, dans la vie de tous les jours, un sujet peut utiliser des stratégies de compensation qu'il n'utilise pas dans les tests de performance, qui peuvent ainsi sous-évaluer ses capacités réelles, ou à l'inverse, les utiliser dans la situation de test classique et oublier de le faire au quotidien. Ainsi, les tests classiques sont très utiles pour repérer d'éventuelles composantes non-fonctionnelles qui pourraient impacter l'adaptation dans la vie quotidienne, mais ils pourraient être complétés par une évaluation contextualisée rendant davantage compte de l'organisation des processus et de leur mise en œuvre dans des situations complexes et changeantes.

[13] L'OSS était une agence de renseignement du gouvernement des Etats-Unis, créée en 1942 après leur entrée dans la Seconde Guerre mondiale pour collecter des informations. Pour mieux remplir ces fonctions, ils ont développé le premier centre d'évaluation psychologique aux Etats-Unis.

5.1.2 Evaluation contextualisée

L'évaluation contextualisée vise à placer le sujet dans les conditions les plus proches possibles des situations de la vie quotidienne. Cette forme d'évaluation alternative existe depuis le début du siècle dernier et s'est principalement développée dans les champs de la neuropsychologie et des ressources humaines. Elle inclut différents formats et méthodologies : des tests papier-crayon (auto et hétéro-évaluations, tests de jugement situationnel, tests de performance) dans lesquels le contexte est évoqué, des simulations et mises en situation (en utilisant du matériel concret, des supports vidéo ou la réalité virtuelle) qui mettent la personne évaluée face à des problèmes qui la concernent véritablement dans son propre environnement et, enfin ; des situations dites naturelles, pour lesquelles la personne est directement immergée dans le contexte. Une revue de littérature sur ces différents outils d'évaluation contextualisée a été menée (Chapitre 1). Nous présentons ici (tableau 21) un panorama de ces derniers au regard de la méthode classique d'évaluation des capacités cognitives. Ces outils d'évaluation contextualisée sont classés par ordre croissant de niveau de contextualisation. Le contexte peut en effet être totalement absent, seulement évoqué dans la consigne ou dans la tâche demandée, être simulé grâce à des supports vidéo ou de la réalité virtuelle ou intégré totalement dans la procédure d'évaluation (la personne étant évaluée dans son contexte).

		Tests papier-crayon			Simulations et mises en situation			Situation naturelle
		Tests de performance classiques	Tests de jugement situationnel (TJS)	Test contextualisé	Matériel à manipuler	Vidéo	Réalité virtuelle	
Exemple d'outil prototypique		WISC IV (Wechsler, 2005)	TKIM (Wagner & Sternberg, 1991)	STAT (non publié mais utilisé dans plusieurs recherches ; Grigorenko & Sternberg, 2001 ; Sternberg, Grigorenko, Ferrari & Clinkenbeard, 1999	Test « in basket »	TJS avec un support vidéo	Classe virtuelle (Rizzo et al., 2002 ; Rizzo et al., 2004)	Profil de Compétences Cognitives du Jeune (P2CJP ; Pereira Da Costa & Scelles, 2012)
Référence à la vie quotidienne (dont contexte scolaire/professionnel)	Consigne / tâche	Répondre à des questions, manipuler des cubes, restituer des suites de chiffres et de nombres prononcés oralement…sans lien avec la vie quotidienne	Indiquer ce que la personne évaluée ferait ou ce qu'il faudrait faire pour faire face à la situation. Le contexte professionnel est évoqué par écrit	Résoudre différents problèmes. La vie quotidienne est évoquée (ex : changer les quantités dans un récipient, faire des cookies, acheter des tickets pour un jeu (contenus quantitatifs); une description d'un problème typique rencontré dans la vie quotidienne pour lequel le sujet doit choisir la meilleure solution (contenus verbaux) ; le choix d'un itinéraire le plus approprié pour se déplacer d'un point à un autre en utilisant une carte (contenus figuratifs)	Traiter des notes, mails, papiers ou mémentos dans un délai contraint et prendre des décisions (Frederiksen, 1966; Frederiksen, Saunders & Wand, 1957).	Indiquer ce que la personne évaluée ferait ou ce qu'il faudrait faire pour faire face à la situation. Le contexte professionnel est évoqué par la vidéo.	Résoudre des tâches cognitives (mémoriser des objets, un itinéraire…)	Le sujet n'est pas contraint de réaliser une tâche (comme en situation expérimentale). Il est soumis à de véritables enjeux, à des problèmes qui le concernent véritablement, et ce dans son propre environnement.
	Matériel	Epreuves avec supports verbaux,	Ecrit : scenarii qui décrivent une	Subtests combinant des contenus verbaux,	Concret, manipulable (notes,	Vidéos (mode de présentation et	Matériel informatique	Une grille d'observation

	visuels (abstraits, figuratifs) sans lien avec la vie quotidienne	situation professionnelle concrète	quantitatifs et figuratifs	mémos, mails…)	mode de réponse)	(casque de réalité virtuelle, ordinateur ou téléphone…)	
Lieu de passation	Laboratoire	Laboratoire	Laboratoire	Laboratoire	Laboratoire	Laboratoire	Milieu de vie
Immersion sujet	Très faible : peu de lien avec la vie quotidienne et très faible ressemblance avec la réalité	Limitée : le sujet doit se représenter mentalement le contexte	Limitée : le sujet doit se représenter mentalement le contexte	Importante : le sujet est plongé dans un environnement proche de l'environnement réel	Importante : la vidéo permet de présenter des détails contextuels de l'environnement (indices non verbaux, ton de la voix, objets ambiants et de montrer directement les comportements de travail réel)	Importante : le sujet est plongé dans un environnement qui simule l'environnement réel (au niveau des tâches à réaliser, des modalités sensorielles, des réponses motrices attendues, de l'interaction homme/machine, du graphisme…)	Quasi-totale : la personne évaluée est immergée dans son contexte de vie mais la situation reste artificielle de par la présence d'un.e observateur. trice

Tableau 21. Panorama des différentes méthodologies d'évaluation des capacités cognitives

Les outils contextualisés restent très peu nombreux et principalement développés en ressources humaines et en neuropsychologie, presque exclusivement pour des adultes. Nous ferons un rapide point sur les preuves de validité de ces derniers quand les informations sont disponibles.

Validité des tests papiers-crayons. En ce qui concerne les épreuves papier-crayon, nous distinguons les tests de jugement situationnel (TJS) et les tests contextualisés. Il existe deux types de TJS (McDaniel, Hartman, Whetzel & Grubb, 2007) : ceux qui évaluent les connaissances et ceux qui évaluent la tendance à produire certains comportements. Les tests de connaissances générales présentent des corrélations plus élevées avec les capacités cognitives (r[14]$=.35$) que les tests de comportement ($r=.19$), ceux-ci corrélant plus fortement avec les construits de personnalité (McDaniel *et al.,* 2007). Les corrélations entre les TJS et les performances dans le travail sont de l'ordre de .26 (McDaniel, Morgeson, Finnegan, Campion & Braverman, 2001). Ces corrélations peuvent être augmentées (.31) lorsque les TJS sont utilisés en complément de l'évaluation des capacités cognitives (McDaniel *et al.*, 2001).

Concernant les tests papier-crayon contextualisés, le Sternberg Triarchic Abilities test (STAT), non publié, a été vivement critiqué (Brody, 2003 ; Gottfredson, 2003 ; Koke & Vernon, 2003) du fait que les tâches proposées ne se distinguent pas radicalement des mesures classiques de l'intelligence. En effet, les contextes restent très simplifiés et assez artificiels (Grégoire, 2006). Outre le STAT, quelques outils dans le domaine de la neuropsychologie ont été élaborés dans une approche écologique, tels que le *Test of Everyday Attention* (TEA ; Robertson, Ward, Ridgeway & Nimmo-Smith, 1994), le *Rivermead Behavioral Memory Test* 3ème édition (RBMT 3 ; Wilson et al., 2010). Quelques études font le lien entre des tests de performance dans des domaines spécifiques (principalement, celui de la mémoire) et les

[14] Corrélation moyenne estimée

aptitudes dans la vie quotidienne. Ainsi, Makatura Lam, Leahy, Castillo et Kalpakjian (1999) et Higginson, Arnett et Voss (2000) montrent que les tests écologiques sont de meilleurs prédicteurs d'une déficience fonctionnelle (mesurée par un entretien) que les tests traditionnels de mémoire et d'attention chez des sujets ayant une sclérose en plaques (gain de r^2= .10). Par ailleurs, selon Higginson et al. (2000), 67% des tests écologiques corrèlent significativement avec le statut fonctionnel de la personne (défini comme ce que peut faire la personne dans la vie de tous les jours) contre seulement 38% des tests traditionnels.

Validité des tests de simulation. Parmi les outils de simulation, la méta-analyse de Whetzel, Rotenberry et McDaniel (2014), relative aux tests « in-basket », met en évidence que ces tests présentent une corrélation moyenne de .26 avec des tests de facteur g et une validité prédictive de .42 au regard de la performance dans le travail, comparable à d'autres tâches de simulation, mais qui reste, néanmoins, inférieure à celle relevée pour les capacités cognitives (.51 selon Schmidt & Hunter, 1998). Les TJS sous format vidéo ont, selon la méta-analyse réalisée par Christian, Edwards & Bradley (2010), une validité critériée moyenne de .36, plus importante que celle des tests papier-crayon qui est de .25, pour l'ensemble des construits mesurés (leadership, compétences inter-personnelles…). Enfin, toujours dans les outils de simulation, quelques études se sont intéressées à la comparaison d'une épreuve neuropsychologique classique à une même épreuve en réalité virtuelle. Par exemple, Pugnetti et al. (1998) ont utilisé le *Wisconsin Card Sorting Test* (WSCT ; Grant & Berg, 1948) et un environnement virtuel représentant un bâtiment et reproduisant les mêmes consignes et conditions que dans le WSCT. Les auteurs constatent que les deux épreuves corrèlent faiblement et semblent évaluer des processus différents, alors même qu'elles sont basées sur un principe identique.

Enfin, concernant les évaluations en situation naturelle, il n'existe pas, à notre connaissance de revue de littérature recensant les outils existants et leurs qualités psychométriques. Un seul

test a été étudié : le *Profil de Compétences Cognitives du Jeune* (P2CJP ; Pereira Da Costa & Scelles, 2012), outil d'évaluation cognitive des enfants et adolescent.e.s atteints de polyhandicap mais les auteurs ne précisent pas ses corrélations avec d'autres outils d'évaluation cognitive.

Par ailleurs, l'évaluation contextualisée présenterait aux yeux des sujets, une validité faciale plus importante, générant de leur part une attitude plus favorable voire une plus grande motivation, élément favorable au processus d'évaluation. Sachant que la communication des résultats obtenus à un test en direction de la personne évaluée et sa famille (pour un mineur) est la finalité du processus d'évaluation, et que cette dernière se doit d'être, selon le code de déontologie des psychologues (2002), les recommandations internationales sur l'utilisation des tests (2003) ou la conférence de Consensus sur le bilan psychologique (Voyazopoulous, Vannetzel & Eynard, 2011) compréhensible et concrète en fournissant des informations qui soient utilisables pour la personne évaluée (Castro, 2011), l'évaluation contextualisée semble plus propice à répondre à ces objectifs. Plusieurs études (Ownby Wallbrown & Brown, 1982 ; Salvagno & Teglasi, 1987) ont, par ailleurs, montré que les personnes préfèrent que les propositions du.de la psychologue ayant réalisé le bilan psychologique soient concrètes et illustrées d'exemples extraits de la vie quotidienne. Pour le.la psychologue, les liens entre la vie quotidienne et les résultats obtenus à une épreuve contextualisée d'intelligence semblent plus évidents à faire, qu'avec les résultats obtenus à un test classique d'intelligence du fait du support utilisé. Sa validité écologique, définie comme le degré auquel les résultats obtenus dans des conditions expérimentales contrôlées peuvent être corrélés à ceux obtenus dans un environnement naturel (Tupper & Cicerone, 1990), est supérieure à celle de l'évaluation classique.

En résumé, les différents outils d'évaluation contextualisée étudiés montrent d'assez bonnes qualités psychométriques et présentent certains avantages en termes de validité faciale et écologique. Généralement, les corrélations avec les mesures classiques de l'intelligence restent faibles, indiquant que les deux mesures n'évaluent pas le même construit, y compris quand la consigne et les conditions sont identiques (Pugnetti et al., 1998). Nous expliquons ces écarts par le fait que les tests contextualisés amènent une complexité plus grande à la situation qui oblige à la gestion plus importante de variables situationnelles et, en même temps, laisse une place plus importante à des stratégies individuelles. Les deux approches ne sont donc pas à considérer comme opposées, mais plutôt complémentaires (Grigorenko et al., 2004 ; Sternberg et al., 2001) afin d'avoir une évaluation plus complète de l'intelligence d'un sujet.

L'intérêt de disposer d'épreuves cognitives contextualisées est donc confirmé. A ce jour, la plupart de celles qui existent ont été développées pour les ressources humaines et la neuropsychologie, et presque exclusivement pour des sujets adultes. Il n'existe pas, à notre connaissance, de tels outils pour évaluer les capacités cognitives d'enfants et/ou d'adolescent.e.s tout venants. Notre travail présenté dans cet article vise à combler ce manque.

5.2 Création d'un outil contextualisé d'évaluation des capacités cognitives pour des sujets adolescent.e.s

Cette étude a pour objectif de développer une échelle d'évaluation des capacités cognitives, reposant sur le modèle Cattell-Horn Carroll (CHC ; Schneider & McGrew, 2012), contextualisée, à destination d'adolescent.e.s français.es (12 à 18 ans) et de montrer les premiers résultats des analyses psychométriques.

5.2.1 Développement de l'échelle[15]

Nous avons choisi de développer un nouvel outil en reprenant les mêmes capacités cognitives que celles évaluées dans le WISC-IV : intelligence cristallisée, fluide, traitement visuo-spatial, mémoire de travail et vitesse de traitement. Nous souhaitions, au départ, utiliser les mêmes épreuves en les adaptant par l'introduction d'un contexte. Etant donné la nature des items dans le WISC-IV, cela n'était pas envisageable. A titre d'exemple, l'épreuve des *Matrices* est une tâche abstraite ayant pour supports des figures géométriques formant une suite qu'il faut compléter, sans lien avec la vie quotidienne.

A l'instar du WISC-IV, nous avons créé des épreuves papier-crayon (mais avec des supports audio pour deux subtests) et une passation individuelle.

Contextualisation. Nous avons fait le choix d'introduire le contexte de deux façons : dans le choix des tâches (en relation avec la vie quotidienne) et au niveau des consignes (consignes générales de chaque épreuve, voire pour certaines d'entre elles, consignes, de chaque item). Cette façon de contextualiser les épreuves, à savoir en faisant référence à des faits de la vie quotidienne, est commode à mettre en œuvre mais ne sera cependant pas naturelle : le contexte est seulement évoqué et le sujet doit se le représenter mentalement, ou simulé (support audio pour deux épreuves). Par ailleurs, le lien avec la vie quotidienne n'est pas équivalent pour chaque épreuve. Pour certaines d'entre elles, il est explicite, et pour d'autres moins implicite.

Modèle théorique. Le modèle théorique sous-jacent est le modèle CHC (Schneider & McGrew, 2012), qui fait actuellement consensus dans la compréhension de l'organisation des aptitudes cognitives. Nous avons choisi de ne prendre que les capacités de niveau 2 du modèle évaluées dans le WISC-IV, à savoir l'intelligence cristallisée (Gc), le raisonnement fluide

[15] Le cahier de passation est présenté en annexe n°5 et le matériel en annexe n°6

(Gf), le traitement visuel (Gv), la mémoire à court terme (Gsm) et la vitesse de traitement (Gs).

Description du matériel. La version finale comprend sept épreuves, qui évaluent les capacités cognitives des jeunes de 12 à 18 ans à partir de tâches contextualisées.

La consigne générale proposée est la suivante : « Nous allons faire ensemble plusieurs exercices en lien avec la vie quotidienne. Ces exercices évaluent les capacités améliorables dans différents domaines : verbal, non verbal, la mémoire, la vitesse et le raisonnement. Certaines choses te paraitront faciles, d'autres plus difficiles. As-tu des questions avant que l'on commence ? »

La passation globale dure environ 1h00. Le.la psychologue se place à côté du sujet durant la passation. Chaque épreuve comprend un ou plusieurs exemples pour que le.la jeune puisse se familiariser avec la tâche.

Les sept épreuves retenues sont les suivantes :

Différences : le sujet doit trouver les différences entre un tableau et sa copie. Cette épreuve se compose de cinq items. La consigne proposée est la suivante « Ton ami.e adore reproduire des tableaux. Il/elle te montre le tableau original et ce qu'elle a peint et te demande ton avis. Il/elle l'a déjà montré à une autre personne qui a relevé à chaque fois des erreurs. Il/elle te demande de lui dire si tu en vois également. Indique-lui pour qu'il.elle puisse les corriger ». Le nombre d'erreurs à relever diffère d'un item à l'autre (3 erreurs pour le premier item et 8 pour le dernier). Vingt-huit erreurs sont à retrouver au total (chacune est cotée 1 point). La passation de chaque item est chronométrée et limitée dans le temps (de 1 minute trente seconde à 3 minutes trente secondes selon les items). Le score global peut donc varier de 0 à 28. Le temps global de passation de cette épreuve est de 15 minutes environ. Cette épreuve a été conçue pour évaluer le traitement visuel (Gv) dans le modèle CHC.

Numéros de téléphone : le sujet doit répéter un numéro de téléphone prononcé oralement suite à la consigne « Tu vas entendre un numéro de téléphone que tu vas devoir mémoriser pour me le redonner ». L'épreuve comprend huit items. Chaque item est noté sur 5 points (un point par élément de deux chiffres correctement restitué et dans l'ordre, c'est-à-dire bien placés dans la série de cinq). Le score total est donc de 40 points. Cette épreuve évalue la mémoire à court terme (Gsm).

Chaussettes : à partir d'une photographie sur laquelle sont disposées une quarantaine de chaussettes, le sujet doit relier par un trait les chaussettes constituant une paire. La consigne proposée est : « On te demande chez toi de trier rapidement les chaussettes et de former les paires. Les voici (montrer la feuille). Peux-tu indiquer celles qui vont ensemble en les reliant grâce au stylo le plus vite possible ? Certaines restent seules. N'oublie pas qu'il faut faire le plus vite possible ». Dix-sept paires de chaussettes sont à relier en 120 secondes au maximum. Chaque paire trouvée est cotée un point. Les erreurs sont décomptées du score global. Le score maximum est donc de 17 points. Cette épreuve a été pensée pour évaluer la vitesse de traitement (Gs) et le traitement visuel (Gv).

Connaissances générales : tout en écoutant une conversation entre jeunes (fichier audio), le sujet doit réagir dès qu'il entend une information erronée (exemple : « Ankara est la capitale de la Grèce ») et la corriger s'il le peut en proposant la réponse qu'il estime correcte. La consigne est la suivante « Tu vas entendre plusieurs extraits de conversations entre jeunes. Comme tu pourrais le faire dans la vie de tous les jours, réagis à chaque fois que ce que tu entends te parait inexact, faux, non approprié. Tu dois à la fois dire que la personne se trompe et la corriger si possible. Il ne faut pas avoir peur de le dire. J'arrêterai la conversation». A la fin de l'épreuve, les erreurs non identifiées sont reprises avec le sujet afin qu'il propose une correction. En effet, certains jeunes hésitent à se manifester spontanément (manque de confiance en soi, timidité, inhibition…). Cette épreuve comporte 22 items regroupant diverses

thématiques telles que l'histoire, la géographie, la littérature ou encore la vie quotidienne. Chaque item est coté sur trois points (un point par erreur repérée et 2 points pour une bonne réponse) avec un total maximal de 66 points. Elle dure environ 15 minutes et évalue l'intelligence cristallisée (Gc) du modèle CHC.

Raisonnement : Plusieurs problèmes utilisant un support visuo-spatial ou numérique sont présentés au sujet qui doit les résoudre suivant la consigne suivante « Tu vas voir, observer plusieurs problèmes que tu dois résoudre. Tu as le droit de dessiner et ou d'écrire si tu le souhaites ». Un crayon ainsi que des feuilles de papier sont à sa disposition pour accompagner sa réflexion. Cette épreuve comporte 20 items, chaque bonne réponse étant notée 1 point (soit un total maximum de 20 points). Elle dure environ 10 minutes. Elle permet d'évaluer l'intelligence fluide (Gf) dont le raisonnement quantitatif (Rq) et les connaissances quantitatives (Gq) du fait de calculs algébriques (additions, soustractions, multiplications, racines carrées…).

Lexique : tout en écoutant une conversation entre jeunes (support audio), le sujet doit repérer et corriger des erreurs de lexique ou bien proposer un mot lorsque celui-ci manque selon la consigne suivante « Je vais te lire une histoire racontée par une personne ayant des difficultés pour s'exprimer : elle utilise parfois des mots qui sont faux, ou à l'opposé de ce qu'il faut dire ou parfois pas assez précis ; ou encore, la personne ne trouve pas le mot du tout. Comme tu pourrais le faire dans la vie de tous les jours, réagis à chaque fois que ce que tu entends te parait inexact, faux, non-approprié ou pas tout à fait adapté ou aide-la à trouver le bon mot. Tu dois à la fois dire que la personne se trompe et la corriger si possible. N'hésite pas à m'interrompre pendant que je lis. Je peux aussi répéter si tu n'as pas entendu ou compris ». L'épreuve comporte 24 items notés sur trois points sur le même modèle que Connaissances générales. Les erreurs non décelées à la première écoute sont reprises afin que le sujet propose

des corrections. La passation est de 15-20 minutes environ. Cette tâche évalue l'intelligence cristallisée (Gc).

Majuscules : le.la jeune doit repérer et corriger le plus vite possible les erreurs de majuscules au sein d'une rédaction de français. La consigne donnée est la suivante « Un voisin qui est en 3ème te demande de relire sa rédaction en français. Il l'a tapée sur ordinateur. Il a oublié certaines fois de mettre une majuscule en début de phrase et d'autres fois, il a mis une majuscule alors que cela n'était pas nécessaire. Il est très pressé de rentrer et n'a que deux minutes. Corrigez-le le plus vite possible. Vous n'avez pas besoin de corriger le reste, uniquement les majuscules ». 36 erreurs sont à déceler (un point par erreur trouvée). Cette épreuve est chronométrée et limitée dans le temps (120 secondes). Elle évalue la vitesse de traitement (Gs).

5.2.2 Etude de validation

5.2.2.1 Echantillon

Les participant.e.s ont été recrutés au sein de divers établissements scolaires français. Pour ce faire, un courrier a été adressé aux chef.fe.s d'établissement, puis un formulaire de consentement a été signé par les familles.

L'échantillon se compose de 61 jeunes âgés en moyenne de 14 ans (ET=1.54), avec 26 filles (42,6%) et 35 garçons (57,4%), scolarisés de la 6ème à la 1ère selon la répartition indiquée dans le tableau 22.

Classe	Effectif	Pourcentage
6ème	8	13,1%
5ème	17	27,9%
4ème	14	23,0%
3ème	6	9,8%
2nde G&T	7	11,5%
1ère	7	11,5%
1ère année CAP	1	1,6%
2nde professionnelle	1	1,6%

Tableau 22. Répartition des sujets de notre échantillon selon la classe (N=61)

Ces jeunes ont été recrutés dans différentes régions de France, avec une majorité en Ile de France (84,6%).

L'information relative au niveau de diplôme des parents a été recueillie (pour 57 sujets, 4 données étaient manquantes) et le plus haut du couple a été retenu. Cinquante pour cent des jeunes ont au moins un des deux parents ayant un diplôme supérieur à bac+2 (cf. tableau 23).

Niveau	Effectif	Pourcentage
Aucun diplôme	4	6,6%
Certificat d'études primaires	1	1,6%
BEPC, Brevet des collèges	2	3,3%
CAP/BEP	7	11,5%
Baccalauréat	3	4,9%
Bac+2	9	14,8%
Supérieur à bac+2	31	50,8%

Tableau 23. Répartition des sujets selon le niveau de diplôme le plus haut des parents (N=57)

5.2.1.2 Procédure

La procédure comportait 4 étapes : 1/ un premier entretien avec le.la jeune et sa famille pour expliquer la démarche et recueillir leur consentement, 2/ un second rendez-vous pour passer le WISC-IV, 3/ un troisième rendez-vous pour la passation de l'échelle d'évaluation contextualisée, puis 4/ un dernier entretien de communication des résultats. Le questionnaire

de la Vineland II a été confié aux représentant.e.s légaux.les de l'élève et remis au.à la psychologue au troisième rendez-vous. Certains parents ont demandé un compte-rendu écrit qui leur a été fourni. Un retour qualitatif des résultats de l'échelle d'évaluation contextualisée a été réalisé sur la motivation du.de la jeune, sa persévérance dans la tâche, les stratégies utilisées, … à partir des observations receuillies par le.la psychologue. La démarche complète était ainsi couteuse en temps pour le.la jeune et sa famille mais également pour l'expérimentateur.trice (quatre rendez-vous de 1h à 2h, cotation, analyse et interprétation, préparation de la communication orale, voire rédaction d'une synthèse). Deux jeunes ont abandonné avant la fin de la démarche.

5.2.1.3 Mesures

Outre les réponses obtenues à l'échelle d'évaluation contextualisée, nous avons recueilli les résultats scolaires, les résultats au WISC-IV et à la Vineland-II.

Résultats scolaires. Les résultats scolaires en fin d'année scolaire, soit de 3 à 6 mois plus tard, ont également été recueillis (moyenne générale au 3ème trimestre ou au 2ème semestre selon la classe) pour 21 sujets.

WISC-IV. Cinquante-huit jeunes ont passé le WISC-IV (Wechsler, 2005).

Vineland-II. Enfin, pour 59 sujets, un des deux parents a rempli la Vineland II (VABS-II ; Sparrow, Cicchetti & Balla, 2015) forme destinée aux parents.

5.2.1.4 Analyse de données

Les données ont été analysées en utilisant SPSS 20.0 (IBM Corp. Released 2011) et jMetrik (Meyer, 2014).

5.2.1.4.1 Fonctionnement différentiel des items selon le sexe

Dans les processus de développement de questionnaires, contrôler les fonctionnements différentiels d'items est une stratégie optimale par rapport à d'autres méthodes (Wetzel, Hell & Päßler, 2012). L'ITC (2003) évoque la nécessité de limiter en particulier les biais liés au sexe dans le cadre du développement des tests. Pour analyser le fonctionnement différentiel selon le sexe, nous avons utilisé la méthode de Cochran-Mantel-Haenszel (CMH ; Cochran, 1954 ; Mantel, 1963) implantée dans jMetrik. Elle apparait la plus adaptée pour des échantillons réduits. Nous ne reportons ici que les items pour lesquels nous avons trouvé un important fonctionnement différentiel selon le sexe. Selon Zwick et Erickan (1989), le fonctionnement différentiel est considéré comme important quand l'Odds ratio commun est inférieur à .53 ou supérieur à 1.89 pour les items binaires, et quand la taille d'effet est égale ou supérieure à .10 pour les items polytomiques.

Epreuves	Type d'item	X^2	Taille d'effet
Connaissances	Quel est le numéro des pompiers ?	3.73*	.41
Raisonnement	Suite logique (cartes à jouer)	8.72**	8.75
Lexique	Trouver le mot courage	2.27	.31

*. p < 0.05 (bilatéral)
**. p < 0.01 (bilatéral)
Note : l'indicateur de taille d'effet diffère selon le type d'item (binaire ou polytomique)

Tableau 24. Fonctionnement différentiel des items selon le sexe

Ces trois items sont donc retirés pour la suite des analyses.

5.2.1.4.2 Analyse d'items

Les items dont l'indice de discrimination était inférieur à .20 et l'indice de difficulté inférieur à .15 ou supérieur à .85 (MacAlpine, 2002) ont été retirés. La difficulté des items restants, les écarts-types et les corrélations item-test sont reportés dans le tableau 25.

Echelle	Items	p	ET	r_{itc}
Différences	1	.75	.89	.41
	2	.84	1.10	.64
	3	.69	1.15	.70
	4	.68	1.45	.56
	5	.76	1.31	.63
Numéros de téléphone	1	.76	1.09	.69
	1bis	.72	1.34	.62
	2	.66	1.52	.45
	2bis	.78	.97	.41
	3	.69	1.28	.45
	3bis	.60	1.44	.58
	4	.36	1.63	.28
	4bis	.32	1.56	.37
Connaissances	1	.51	1.20	.54
	2	.42	1.16	.41
	3	.74	1.14	.62
	4	.60	1.29	.56
	5	.44	1.32	.69
	6	.23	1.14	.56
	7	.16	.81	.44
	8	.69	1.14	.41
	9	.63	1.32	.51
	10	.58	1.36	.35
	11	.78	1.02	.50
	12	.71	1.17	.44
	13	.28	1.17	.40
	14	.49	1.31	.54
	15	.18	.98	.29
	16	.61	1.28	.46
	17	.75	1.02	.30
	18	.79	.99	.35
	19	.43	1.18	.39
	20	.53	1.20	.37
	21	.64	1.33	.42
	22	.53	1.33	.42

Raisonnement	1	.73	.45	.33
	2	.73	.45	.33
	3	.83	.38	.22
	4	.82	.39	.58
	5	.55	.50	.43
	6	.70	.46	.29
	7	.82	.39	.48
	8	.30	.46	.31
	9	.78	.42	.44
	10	.62	.49	.31
	11	.38	.49	.26
	12	.77	.43	.33
	13	.68	.47	.44
	14	.23	.43	.34
	15	.75	.44	.66
	16	.55	.50	.35
	17	.58	.50	.34
	18	.78	.42	.49
	19	.28	.45	.31
	20	.47	.50	.29
Lexique	1	.59	1.28	.54
	2	.76	1.10	.38
	3	.74	1.23	.32
	4	.48	1.36	.51
	5	.77	1.05	.47
	6	.51	1.32	.54
	7	.79	1.07	.41
	8	.68	1.24	.54
	9	.51	1.18	.50
	10	.31	1.21	.37
	11	.52	1.22	.65
	12	.50	1.25	.47
	13	.21	.92	.39
	14	.67	1.17	.63
	15	.63	1.27	.71
	16	.60	1.32	.39
	17	.76	1.07	.64
	18	.78	1.14	.57
	19	.45	1.18	.43
	20	.38	1.09	.47
	21	.62	1.26	.47
	22	.43	1.16	.51
	23	.15	.72	.43
	24	.23	.92	.38

Seules les épreuves comportant plusieurs items sont reportées

Tableau 25. Difficulté, écart-type et indice de discrimination des items (N=61)

La première épreuve comporte des items dont l'indice de difficulté est élevé. Pour les autres épreuves, l'indice de difficulté varie davantage d'un item à l'autre. Dans la version finale de l'épreuve, les items seront proposés par ordre croissant de difficulté.

5.2.1.4.3 Fiabilité

Les alpha de Cronbach sont acceptables pour les cinq épreuves composées de plusieurs items comme indiqué dans le tableau 26.

	Différences	Numéros de téléphone	Connaissances	Raisonnement	Lexique
Alpha	.80	.80	.87	.80	.90

Tableau 26. Alpha de Cronbach des épreuves comportant plusieurs items

5.2.1.4.4 Tentatives de validation de l'échelle

Analyse des inter-corrélations

	1	2	3	4	5	6	7
1. Différences	1.00						
2. Numéros de téléphone	.27*	1.00					
3. Connaissances	.46**	.44**	1.00				
4. Raisonnement	.35**	.31*	.55**	1.00			
5. Lexique	.49**	.28*	.73**	.55**	1.00		
6. Chaussettes	.41**	-.03	.41**	.36**	.52**	1.00	
7. Majuscules	.33**	.32*	.52**	.57**	.51**	.34**	1.00

**. p < 0.01 (bilatéral)

Tableau 27. Matrice des corrélations entre les épreuves

Les corrélations entre les épreuves ont été calculées. Elles indiquent des corrélations positives et significatives au seuil .01, de grande taille (Cohen, 1988) entre les épreuves de Connaissances et Lexique (*r* =.73), et Raisonnement (*r* =.55), entre Raisonnement et Lexique (*r* =.55), entre Lexique et Chaussettes (*r* =.52) et Majuscules (*r* =.51) ; de taille moyenne pour toutes les autres corrélations hormis entre Chaussettes et Numéros de téléphones, épreuves entre lesquelles on observe une corrélation nulle.

Analyse factorielle confirmatoire

Deux modèles théoriques ont été testés :

- le modèle 1 avec un facteur unique, les 7 épreuves étant sous-tendues par un facteur général et ;
- le modèle 2 avec deux facteurs : un facteur verbal (épreuves Lexique, Connaissances, Numéros de téléphone) et un facteur visuo-spatial (avec les épreuves Différences, Raisonnement, Chaussettes et Majuscules).

Suivant les recommandations de nombreux auteurs (Hooper, Coughlan & Mullen, 2008 ; Hu & Bentler, 1999 ; Kline, 1998), plusieurs indices d'ajustement sont examinés pour déterminer l'adéquation du modèle : le Chi2, le Chi2 rapporté aux degrés de liberté (Chi2/ddl), l'indice comparatif d'adéquation (CFI), l'index de Tucker-Lewis (TLI), l'erreur quadratique moyenne de l'approximation (RMSEA) et l'indice de la racine du carré moyen d'erreur (SRMR). La valeur du Chi2 ne doit pas être significative mais elle est très sensible à la taille de l'échantillon. Ainsi, les auteurs suggèrent d'utiliser plutôt le rapport Chi2/ddl qui doit être inférieur à 3 (Hooper et al.,2008 ; Kline, 1998). Pour une bonne adéquation des données au modèle, les valeurs du TLI et CFI doivent être supérieures à .95 mais .90 est également considéré comme un seuil acceptable. La valeur du RMSEA doit être inférieure à .06 ou une

limite supérieure stricte de .07 (Steiger, 2007) et celle du SRMR à .08. Les résultats sont présentés dans le Tableau 28.

Modèle	χ^2	Ddl*	p	χ^2 / ddl	CFI	TLI	RMSEA	SRMR
1. Modèle à 1 facteur	18.35	14	.19	1.31	.97	.95	.07	.07
2. Modèle à 2 facteurs	16.93	13	.20	1.30	.97	.95	.07	.07

*ddl : degrés de libertés. N = 59

Tableau 28. Indices d'ajustement pour les deux modèles

Comme nous pouvons le constater sur le tableau précédent, tous les indicateurs d'ajustement atteignent les seuils recommandés. Les deux modèles s'ajustent correctement aux données. La différence entre les deux modèles n'est pas significative ($Chi^2[1]=1.42$, p=.23). Ces résultats sont à prendre avec beaucoup de précaution du fait de l'effectif très restreint.

5.2.1.4.5 Tentatives de validation basées sur les relations avec d'autres variables

Corrélations avec les résultats scolaires

Comme indiqué précédemment, nous avons recueilli les résultats scolaires au 3ème trimestre pour 21 sujets, soit de 2 à 6 mois après le recueil des données à l'évaluation contextualisée. Des corrélations Bravais-Pearson ont été calculées en vue d'examiner les relations entre les deux.

	Moyenne générale
Différences	.39
Numéros de téléphone	.19
Connaissances	.42*
Raisonnement	.35
Lexique	.01
Chaussettes	-.08
Majuscules	.21

*. p < 0.05 (bilatéral)

Tableau 29. Corrélations entre les six épreuves et la moyenne générale au 3ème trimestre (N=21)

Seule l'épreuve de Connaissances corrèle de manière positive et significative avec la moyenne générale. La corrélation (*r* =.42) est de taille modérée, selon les normes établies par Cohen (1988) avec seulement 16% de variance commune entre les deux. Les corrélations sont faibles ou quasi nulles avec les autres épreuves.

Corrélations avec les résultats au WISC-IV

Nous avons vérifié que notre échantillon était représentatif en regardant si les moyennes obtenues aux indices et au QIT étaient conformes à la norme (voir tableau 30). C'est le cas pour toutes les variables sauf l'ICV pour laquelle notre échantillon présente une moyenne et un écart-type plus élevés.

	Moyenne	Ecart-type	t de Student	p-value
ICV	107.22	23.67	2.34	.02
IRP	100.14	17.62	.06	.95
IMT	96.24	15.45	-1.85	.07
IVT	97.24	17.93	-1.18	.24
QIT	102.00	19.02	.80	.43

Tableau 30. Comparaison aux normes des scores obtenus par l'échantillon (N=59) au WISC-IV

	Différences	Numéros de téléphone	Connaissances	Raisonnement	Lexique	Chaussettes	Majuscules
Cubes	.24	.29*	.44**	.63**	.45**	.36**	.67**
Similitudes	.28*	.10	.61**	.59**	.68**	.30*	.47**
MCH	.28*	.45**	.48**	.44**	.28*	.27*	.44**
IDC	.39**	.04	.50**	.49**	.51**	.34*	.44**
Code	.45**	.32	.38**	.50**	.51**	.40**	.70**
Vocabulaire	.34**	.45**	.53**	.41**	.69**	.22	.54**
SLC	.33*	.46**	.55**	.59**	.51**	.35**	.49**
Matrices	.18	.23	.47**	.52**	.49**	.43**	.44**
Compréhension	.36**	.19	.55**	.39**	.57**	.28*	.46**
Symboles	.35**	.29*	.53**	.56**	.48**	.54**	.63**
ICV	.22	.14	.58**	.46**	.60**	.14	.41**
IRP	.29*	.20	.54**	.67**	.55**	.39**	.58**
IMT	.27*	.46**	.52**	.50**	.30*	.11	.40**
IVT	.36**	.27*	.35*	.52**	.47**	.41**	.62**
QIT	.35**	.32*	.64**	.69**	.62**	.30*	.63**

Tableau 31. Corrélations entre les six épreuves et les résultats au WISC-IV (N=57)

Comme l'indique le tableau 31, toutes les épreuves corrèlent positivement et significativement avec le QIT du WISC-IV, fortement pour les épreuves de Connaissances (r=.64), Raisonnement (r=.69), Lexique (r=.62) et Majuscules (r=.63) et modérément pour les épreuves Différences (r=.35), Numéros de téléphone (r=.32) et Chaussettes (r=.30). Elles corrèlent également toutes positivement et significativement avec les indices sauf Différences avec l'ICV (r=.22), Numéros de téléphone avec l'ICV (r=.14) et l'IRP (r=.20) et Chaussettes avec l'ICV (r=.14) et l'IMT (r=.11) ; de même qu'avec les subtests du WISC IV sauf Différences avec Cubes (r=.18) et Matrices (r=.22), Chaussettes avec Vocabulaire (r=.22) et Numéros de téléphone qui ne corrèle qu'avec six épreuves sur les dix et principalement avec celles évaluant la mémoire à court terme (MCH et SLC).

Corrélations avec la Vineland II

Nous avons vérifié que notre échantillon était représentatif en regardant si les moyennes obtenues aux sous-échelles et échelles de la Vineland II étaient conformes à la norme (tableau 32). En effet, la Vineland II est un outil utilisé principalement auprès d'une population avec déficit cognitif ou trouble du spectre autistique. Or, notre échantillon est composé de jeunes tout venant. Nos participant.e.s obtiennent des résultats conformes dans toutes les dimensions sauf pour l'épreuve Réceptif et la sous-dimension Communication.

	Moyenne	Ecart-type	t de Student	p-value
Réceptif	13.34	3.57	-3.53	.00
Expressif	14.17	3.90	-1.64	.11
Ecrit	14.22	3.50	-1.71	.09
Communication	90.43	23.44	-3.11	.00
Personnel	15.59	4.01	1.12	.27
Domestique	15.12	3.07	.30	.77
Communauté	15.31	4.36	.54	.59
Vie quotidienne	105.02	22.02	1.72	.09
Relations	15.62	4.19	1.13	.26
Jeu	14.07	4.26	-1.65	.11
Adaptation	15.49	2.78	1.34	.19
Socialisation	100.14	20.08	.05	.96
Note composite	99.44	21.37	-.20	.85

Tableau 32. Comparaison des scores obtenus par l'échantillon (N=59) à la Vineland II aux normes

	Différences	Numéros	Connaissances	Raisonnement	Lexique	Chaussettes	Majuscules
Réceptif	.12	.16	.37**	.50**	.35**	.24	.42**
Expressif	.14	.17	.36**	.41**	.28*	.07	.36**
Ecrit	.45**	.22	.62**	.50**	.71**	.25	.64**
Communication	.17	.21	.46**	.54**	.46**	.09	.46**
Personnel	.07	.00	.16	.22	.29*	.17	.15
Domestique	.18	-.01	.22	.13	.38**	.19	.12
Communauté	.09	.05	.20	.16	.21	.14	.25
Vie quotidienne	-.02	-.03	.15	.15	.25	.13	.10
Relations	.16	.02	.27*	.08	.30*	.19	.11
Jeu	.05	-.03	.19	.24	.26*	.06	.40**
Adaptation	.28*	.01	.28*	.18	.39**	.07	.35**
Socialisation	-.10	-.10	.03	.05	.13	-.15	.17
Note composite	-.04	-.04	.18	.29*	.29*	-.01	.27*

Tableau 33. Corrélations entre les six épreuves et les scores à la Vineland II (N=59)

Comme indiqué dans le tableau 33, trois épreuves de l'évaluation contextualisée corrèlent positivement avec la note composite globale : Raisonnement (r=.29), Lexique (r=.29) et Majuscules (r=.27). Toutes les épreuves corrèlent avec au moins une épreuve de la Vineland II sauf Chaussettes et Numéros de téléphone. C'est l'épreuve Lexique qui présente le plus de liens avec l'échelle d'évaluation du comportement socio-adaptatif (en nombre de corrélations et en force), suivie de Majuscules et Connaissances. C'est avec la sous-dimension Ecrit de la Vineland II que les liens sont les plus nombreux et les plus forts, et le sous-domaine Communication.

5.2.1.4.5 Différences de scores aux épreuves selon le sexe

Un *t* de student a été effectué, en vue d'examiner l'existence d'éventuelles différences dans les scores aux différentes épreuves entre garçons et filles. Comme indiqué dans le tableau 34, on observe des différences significatives entre les garçons et les filles pour toutes les épreuves sauf celles de Numéros de téléphone, Connaissances et Raisonnement. C'est sur les subtests Différences, Chaussettes et Majuscules qu'on retrouve la plus grande taille d'effet (respectivement $\eta^2 = 0.16$, $\eta^2 = 0.13$ et $\eta^2 = 0.12$, ce qui est un effet important). Les différences sont toujours en faveur des filles qui obtiennent en moyenne des résultats supérieurs à ceux des garçons hormis pour l'épreuve de Numéros de téléphone.

	Garçons			Filles				
	N	M	ET	N	M	ET	Test *t*	η^2
Différences	35	9.01	3.12	26	11.33	2.27	-3.35**	.16
Numéros tél.	35	10.34	2.85	26	9.54	3.19	1.03	.02
Connaissances	34	9.74	2.95	26	10.34	3.09	-.76	.01
Raisonnement	34	9.82	3.18	26	10.24	2.79	-.53	.01
Lexique	35	9. 33	2.98	26	10.90	2.83	-2.07*	.07
Chaussettes	35	9.08	2.95	26	11.25	2.64	-2.97**	.13
Majuscules	35	9.09	2.64	26	11.22	3.07	-2.90**	.12

*. p < 0.05 (bilatéral)
**. p < 0.01 (bilatéral)
Note : les scores ont été normalisés et transformés selon la formule de Hull (moyenne de 10, écart-type de 3) pour permettre des comparaisons intra-individuelles

Tableau 34. Différences de scores à chaque échelle selon le sexe (N=61)

5.2.1.4.6 Différences selon le niveau de diplôme des parents

Pour déterminer d'éventuelles différences de scores selon le niveau de diplôme des parents, nous avons réalisé un *t* de student. Afin d'avoir suffisamment de sujets dans chaque catégorie, nous avons décidé de regrouper les sept niveaux initiaux et de constituer deux groupes : le premier comprend les personnes sans diplôme jusqu'à CAP/BEP ; le second les personnes avec au minimum un baccalauréat. Comme indiqué dans le tableau 35, on observe des différences significatives entre les deux groupes pour trois épreuves : Connaissances, Raisonnement et Majuscules. C'est sur le subtest Majuscules qu'on retrouve la plus grande taille d'effet ($\eta^2 = 0.12$, ce qui est un effet moyen). Les différences sont toujours en faveur des enfants dont un des deux parents possède au minimum le baccalauréat, qui obtiennent en moyenne des résultats supérieurs.

Nous avons également calculé ces différences de résultats selon le niveau de diplôme le plus élevé des parents aux dix épreuves du WISC-IV. On constate des différences significatives pour 6 épreuves sur les dix (Similitudes, Mémoire des Chiffres, Identification

de concepts, Code, Vocabulaire et Compréhension), toujours en faveur des enfants dont les parents ont un diplôme égal ou supérieur au baccalauréat. Par ailleurs, nous observons trois effets de grande taille pour Mémoire des Chiffres (η^2 =.14), Vocabulaire (η^2 = 0.15) et Compréhension (η^2 = 0.27), ainsi que trois effets de taille moyenne : Similitudes (η^2 = 0.09), IDC (η^2 = 0.10) et Code (η^2 = 0.08).

	Inférieur au Bac			Supérieur au bac				
	N	M	ET	N	M	ET	Test *t*	η^2
Différences	17	9.85	3.93	40	10.20	2.56	-.40	.00
Numéros tél.	17	9.82	2.04	40	9.96	3.31	-.16	.00
Connaissances	17	8.60	2.28	40	10.52	2.93	-2.35*	.09
Raisonnement	17	8.72	3.58	40	10.48	2.69	-2.03*	.07
Lexique	17	9.69	3.17	40	10.30	2.97	-.70	.01
Chaussettes	17	9.56	3.02	40	10.50	2.83	-1.12	.02
Majuscules	17	8.35	2.43	40	10.48	2.81	-2.72**	.12

*. p < 0.05 (bilatéral)

**. p < 0.01 (bilatéral)

Note : les scores ont été normalisés et transformés selon la formule de Hull (moyenne de 10, écart-type de 3) pour permettre des comparaisons intra-individuelles

Tableau 35. Différences de scores à chaque échelle selon le niveau de diplôme des parents (N=57)

5.3 Discussion

Cette étude exploratoire avait pour objectifs de développer une échelle d'évaluation des capacités cognitives, reposant sur le modèle CHC (Schneider & McGrew, 2012), contextualisée, à destination d'adolescent.e.s français.e.s (12 à 18 ans) et de montrer les premiers résultats des analyses psychométriques. Pour cela, une analyse classique des items a été menée ainsi qu'une analyse différentielle des items selon le sexe. Nous avons évalué la

fiabilité des épreuves à partir de l'examen de leur consistance interne, examiné les relations entre les épreuves et mené une analyse factorielle confirmatoire. Les preuves de validité convergente ont été étudiées à partir des corrélations des épreuves avec la réussite scolaire en fin d'année, les résultats obtenus au WISC-IV, et le comportement adaptatif, mesuré par la Vineland-II. Nous avons étudié, enfin, les différences de scores selon le sexe et le niveau socio-économique.

Suite à l'analyse du fonctionnement différentiel des items selon le sexe et l'analyse classique, plusieurs items ont été supprimés. Les épreuves de Différences et Chaussettes restent un peu faciles par rapport aux autres épreuves et nécessiteraient d'être revues en ajoutant des items plus difficiles. Les alpha de Cronbach sont bons (variant de .80 Pour Différences, Numéros de téléphone et Raisonnement à .90 pour Lexique) pour les cinq épreuves composées de plusieurs items.

Concernant les corrélations entre les épreuves, nous nous attendions à trouver les plus forts liens entre les épreuves censées évaluer la même capacité cognitive, en lien avec le modèle CHC. C'est le cas pour Connaissances et Lexique, épreuves évaluant l'intelligence cristallisée avec une corrélation de .73. Par contre, le lien entre Chaussettes et Majuscules, censées évaluer la vitesse de traitement (Gs), bien que positif et significatif au seuil de .01 est modéré (r=.34), de même entre Différences et Chaussettes qui évaluent le traitement visuel (Gv). La corrélation, bien que positive et significative est moyenne (r=.41). Par ailleurs, nous constatons les corrélations les plus fortes entre Connaissances, Lexique, Raisonnement et Majuscules. Nous posons l'hypothèse que ces épreuves mesurent davantage le facteur g. Une analyse factorielle confirmatoire a été réalisée, en testant deux modèles théoriques : un modèle avec un facteur unique (le facteur g) et un modèle à deux facteurs (verbal et visuo-spatial). Les deux modèles s'ajustent correctement aux données mais il faut rester très prudent par rapport à ces résultats du fait de la taille de l'échantillon. Cela doit être confirmé avec un

échantillon beaucoup plus important. Concernant le modèle en un facteur unique (facteur g), les épreuves qui saturent le plus ce facteur par ordre décroissant sont Lexique (.83), Connaissances (.83), Raisonnement (.69) et Majuscules (.64), ce qui rejoint l'hypothèse énoncée plus haut.

Concernant les preuves de validité concourante et notamment les liens entre l'échelle d'évaluation contextualisée et la réussite scolaire définie à partir de la moyenne scolaire générale obtenue par le sujet, seule l'épreuve de Connaissances corrèle de manière positive et significative avec la moyenne générale. La Corrélation (r =.42) est de taille modérée. Les corrélations sont faibles ou quasi nulles avec les autres épreuves. L'épreuve de Connaissances, est de par ce qui est demandé à la personne évaluée et la nature des items, très en lien avec les apprentissages scolaires même si certains items reposent davantage sur la vie quotidienne. Ce sont des connaissances générales, acquises notamment à l'école. Par ailleurs, Connaissances évalue Gc et il est admis depuis plusieurs décennies que l'intelligence cristallisée est le meilleur prédicteur des performances scolaires (Grégoire, 2004).

Il est plus étonnant que les autres épreuves corrèlent peu avec la moyenne générale, en particulier l'épreuve de Lexique, autre mesure de Gc.

Ceci est peut-être lié au fait que le critère externe retenu, à savoir la moyenne générale du.de la jeune, est une mesure moins objective qu'un test de connaissances standardisé. L'évaluation faite par les enseignant.e.s est influencée par d'autres critères comme le comportement, l'apparence physique, le niveau de la classe (Grégoire, 2004). On pourrait donc attendre des corrélations plus importantes entre les épreuves de l'échelle d'évaluation contextualisée et des tests scolaires standardisés qu'avec la moyenne générale. De tels tests n'existent hélas pas en France.

Afin de vérifier que l'échelle d'évaluation contextualisée mesure bien les capacités cognitives, les corrélations avec le WISC-IV, échelle d'évaluation de l'intelligence, reconnue et validée, ont été étudiées. Toutes les épreuves corrèlent positivement et significativement avec le QIT du WISC-IV. Les quatre épreuves qui corrèlent le plus avec le QIT du WISC-IV sont aussi celles qui sont le plus saturées par le facteur g. Au niveau des indices, nous nous attendions à trouver un lien plus fort, significatif et positif entre les épreuves de l'évaluation contextualisée et les indices mesurant la même capacité. C'est le cas pour toutes les épreuves : Raisonnement corrèle davantage avec l'IRP (r=.67), Connaissances et Lexique avec l'ICV (r=.58 et .60), Numéros de téléphone avec l'IMT (r=.46), Majuscules avec l'IVT (=.62). Différences et Chaussettes sont plus complexes : elles corrèlent avec l'IRP (respectivement r=.29 et r=.39) et l'IVT (r=.36 et r=.41). Elles sont toutes deux chronométrées, ont un support visuel et sollicitent le geste graphique comme en IVT, en plus d'évaluer le traitement visuel.

Au niveau des subtests, les épreuves censées évaluer les mêmes capacités corrèlent davantage : Connaissances avec Similitudes (r=.61), Lexique avec Vocabulaire (r=.69), Majuscules avec Code (r=.70), Chaussettes avec Symboles (=.54). L'épreuve de Raisonnement entretient le lien le plus fort avec Cubes (r=.63) puis SLC (r=.59), alors qu'on se serait attendu à une corrélation plus élevée avec Matrices, épreuve de raisonnement fluide (elle n'est que de .52). Le point commun entre Raisonnement et Cubes est le fait que ces deux épreuves évaluent le raisonnement sur les relations spatiales. Quant à la corrélation positive et significative avec SLC, les liens entre raisonnement fluide et mémoire de travail sont depuis longtemps établis (Grégoire, 2004). Enfin, l'épreuve de Différences corrèle davantage avec Code (r=.45) puis Identification de concepts (r=.39). Elle partage avec Code le support visuel et la vitesse de traitement (c'est une épreuve chronométrée), avec Identification de Concepts (IDC), le support visuel et figuratif (objets concrets). IDC est une épreuve complexe qui évalue à la fois Gf, Gv et Gc.

Nous avons, enfin, vérifié les liens entre l'échelle d'évaluation contextualisée et le comportement adaptatif tel que mesuré par la Vineland II. Trois épreuves de l'évaluation contextualisée corrèlent positivement avec la note composite globale : Raisonnement (r=.29), Lexique (r=.29) et Majuscules (r=.27), les épreuves étant saturées, a priori, le plus par le facteur g. Ces corrélations, bien que positives et significatives restent faibles. Elles sont néanmoins plus fortes que celles entre les épreuves du WISC et la note composite globale (Sparrow, Cicchetti et Balla, 2015 ; Wechsler, 2016). Ces deux tests évalueraient des construits différents. Toutes les épreuves de l'échelle d'évaluation contextualisée corrèlent avec au moins une épreuve de la Vineland II sauf Chaussettes et Numéros de téléphone. C'est l'épreuve Lexique qui présente le plus de liens avec l'échelle d'évaluation du comportement socio-adaptatif (en nombre de corrélations et en force), suivie de Majuscules et Connaissances. Enfin, c'est avec la sous-dimension Ecrit de la Vineland II que les liens sont les plus nombreux et les plus forts, et le sous-domaine Communication. Ce domaine impliquerait des capacités cognitives plus complexes que les autres domaines (Wechsler, 2016).

Enfin, nous avons trouvé des différences significatives entre les garçons et les filles pour toutes les épreuves sauf celles de Numéros de téléphone, Connaissances et Raisonnement. C'est sur les subtests Différences, Chaussettes et Majuscules qu'on retrouve la plus grande taille d'effet (respectivement $\eta^2 = 0.16$, $\eta^2 = 0.13$ et $\eta^2 = 0.12$) même si elle reste modeste. Ces trois épreuves sont chronométrées et évaluent notamment la vitesse de traitement. Ces résultats sont concordants avec d'autres études, qui montrent qu'aux épreuves de vitesse, les filles obtiennent des résultats en moyenne supérieurs à ceux des garçons (Grégoire, 2004 ; Pezzuti &Orsini, 2016). Selon Grégoire (2004), ces épreuves sont particulièrement sensibles aux problèmes d'attention et de concentration et les difficultés attentionnelles seraient plus fréquentes chez les garçons que chez les filles. « On peut penser que les épreuves qui exigent

un degré élevé de concentration tendent à favoriser les filles par rapport aux garçons » (p.107).

Nous avons, également, observé les différences de scores selon le niveau de diplôme des parents sur trois épreuves : Connaissances, Raisonnement et Majuscules. C'est sur le subtest Majuscules qu'on retrouve la plus grande taille d'effet ($\eta^2 = 0.12$, ce qui est un effet moyen). Les différences sont toujours en faveur des enfants dont un des deux parents possède au minimum le baccalauréat.

Ces différences semblent néanmoins moins fortes qu'au WISC-IV où sont observées des différences significatives pour six épreuves sur les dix (Similitudes, Mémoire des Chiffres, Identification de concepts, Code, Vocabulaire et Compréhension), toujours en faveur des enfants dont les parents ont un diplôme égal ou supérieur au baccalauréat. Par ailleurs, les tailles d'effet sont moyennes pour trois épreuves (Similitudes, IDC et Code) et grande pour les trois autres (Mémoire des Chiffres, Vocabulaire et Compréhension). Il semblerait qu'à l'échelle d'évaluation contextualisée, les différences selon le niveau de diplôme des parents et donc le milieu socio-culturel soient moindres.

Ainsi, l'échelle d'évaluation contextualisée semble présenter des preuves de sensibilité, fidélité et validité acceptables qu'il conviendrait néanmoins de confirmer par de nouvelles études. En effet, de nombreuses limites sont à noter, que nous détaillerons ci-après.

Limites de cette étude

Cette étude présente un certain nombre de limites : la taille de l'échantillon est restreinte. Une nouvelle étude sur un échantillon plus important, représentatif de la population française et stratifié est nécessaire. Cela nous permettrait également de pouvoir étudier les preuves de validité incrémentielle concernant la réussite scolaire par exemple. Deux épreuves ont un indice de difficulté un peu élevé : Différences et Numéros de téléphone. Des items de niveau

difficile pourraient être ajoutés. Des études complémentaires concernant la fidélité, notamment inter-juges et temporelle, doivent être menées. Par ailleurs, notre échelle propose des contextes qui restent artificiels puisque le contexte est seulement évoqué par la consigne et/ou les items que le sujet doit se représenter mentalement, ce qui peut expliquer les corrélations assez élevées avec les épreuves du WISC-IV. Dans de futures études, il serait nécessaire de développer ces mêmes épreuves en renforçant le caractère naturel de la contextualisation. Cela pourrait se faire de différentes façons : par l'utilisation d'un matériel concret, par le recours à des vidéos ou à la réalité virtuelle ou encore par de l'évaluation en situation naturelle.

Implications pratiques

Comme nous l'avons indiqué en introduction de cette étude, l'évaluation contextualisée présente plusieurs intérêts et avantages : elle suscite, de par la nature des tâches proposées en lien avec la vie quotidienne, une attitude plus favorable de la part de la personne évaluée, voire une plus grande motivation ; les liens sont plus évidents à faire entre la vie quotidienne et les résultats au test, et de ce fait plus exploitables pour la communication des résultats. Le format papier-crayon est plus économique et pratique même si le contexte est simplement évoqué à travers la consigne, le matériel ou les items.

Enfin, d'un point de vue qualitatif, les différents expérimentateurs ayant participé à l'étude ont noté une différence dans la relation entre le.la psychologue et l'adolescent.e, entre le WISC et l'échelle d'évaluation contextualisée. Cette dernière, de par la nature de certaines épreuves, rompt la dysmétrie entre psychologue/jeune. A Lexique et Connaissances, on leur demande, en effet, de stopper le.la psychologue dès que ce qu'ils entendent leur parait faux, inexact ou inapproprié, ce qui renverse le positionnement classique où c'est le.la psychologue

qui pose des questions, relance. Par ailleurs, les jeunes ont fait, d'emblée, pour certaines épreuves le lien avec ce qu'ils faisaient dans la vie quotidienne.

Conclusion et perspectives

L’objectif de cette thèse était d'étudier les apports de l'évaluation contextualisée, notamment en termes de validité, à la prédiction de la réussite scolaire et du comportement adaptatif d'adolescent.e.s scolarisé.e.s. Cette prédiction est en effet centrale pour les psychologues intervenant auprès de ce public.

Pour répondre à cet objectif, nous avons conçu plusieurs outils car aucun, à notre connaissance, répondait à nos attentes : évaluer cinq facteurs du modèle CHC (Gc, Gf, Gv, Gsm et Gs), être à destination des adolescente.s et en lien avec la vie quotidienne. Ces outils ont été créés selon deux méthodologies différentes : une évaluation de la performance (échelle d’évaluation contextualisée) et une évaluation subjective (questionnaires d’auto et d’hétéro-évaluations). Pour les mesures auto et hétéro-rapportées, nous avons intégré dans la consigne une incitation à la comparaison sociale afin d’améliorer les corrélations avec les tests de performance classiques.

Les études que nous avons conduites montrent que ces différents outils présentent certaines qualités psychométriques acceptables, notamment en termes de preuves de fidélité. Concernant les preuves de validité, nous pouvons faire la synthèse selon trois points : les corrélations avec les tests de performance (WISC-IV dans notre étude) ; les liens avec la performance scolaire (mesurée ici par la moyenne générale) et les liens avec le comportement adaptatif (mesurée par la Vineland II).

- Liens avec les tests de performance

Nous avons choisi le WISC-IV car c’est actuellement l’outil le plus utilisé par les psychologues pour évaluer les capacités cognitives des adolescent.e.s, qu’il est validé scientifiquement et que ses indices peuvent être analysés sous l’angle du modèle CHC.

Concernant le QAE (questionnaire d'auto-évaluation), seul le facteur Gf corrèle de manière positive et significative avec les subtests et les indices au WISC-IV. Nous avons trouvé une corrélation de .42 avec le QIT (quotient intellectuel total) du WISC-IV, ce qui est un peu supérieur à ce qui a pu être trouvé dans les différentes études menées (Freund & Kasten, 2012; Mabe et West, 1982). Notons toutefois que les autres facteurs ne corrèlent pas avec les indices du WISC-IV. Il semble que ces autres facteurs évaluent un autre construit que celui mesuré par ces indices dans le WISC-IV. Il est également possible que les adolescent.e.s ne parviennent pas à s'auto-évaluer dans ces domaines. Ces résultats vont dans le sens des recherches précédentes : des liens modestes entre évaluation subjective et objective. Ces deux formes d'évaluations porteraient sur des construits distincts, ce qui soutient l'idée que leur utilisation conjointe pourrait être pertinente.

Concernant l'échelle d'évaluation contextualisée, il apparait que toutes les épreuves corrèlent positivement et significativement avec le QIT du WISC-IV (avec des corrélations allant de .62 à .69 pour quatre épreuves), les quatre indices et les subtests. Rappelons que nous avons conçu les épreuves de l'échelle d'évaluation contextualisée en lien avec celles du WISC-IV. Nous pouvons en déduire que l'échelle d'évaluation contextualisée semble bien mesurer les capacités cognitives.

- Liens avec la performance scolaire

Evaluations subjectives

Nous avons observé des corrélations positives, modérées mais significatives entre les résultats au 3ème trimestre (soit quelques mois après avoir passé le QAE) et Gf (r =.45) puis Gc (r =.26) du QAE. Les corrélations sont quasi nulles avec les autres facteurs. Par ailleurs, le

facteur Gf est le seul qui prédit la performance scolaire au 3ème trimestre. Il explique 18,9% de la variance de la moyenne au 3ème trimestre. Ce résultat est difficilement comparable à ceux des recherches antérieures puisqu'il n'existe pas d'étude comparative ayant examiné les relations entre questionnaire d'auto-évaluation des capacités cognitives et la réussite scolaire. La seule étude robuste sur le sujet est la méta-analyse réalisée par Kornilov (2011), dans laquelle une corrélation moyenne positive et significative de 0,28 était trouvée. Et lorsque nous combinons ce questionnaire aux questionnaires d'hétéro-évaluations remplis par les parents et l'enseignant.e, les trois questionnaires permettent de prédire 69% de la moyenne générale obtenue par les jeunes au troisième trimestre. Quel que soit le questionnaire, les deux facteurs les plus en lien avec la performance scolaire sont Gc et Gf.

Evaluation contextualisée

Concernant l'évaluation contextualisée, seule l'épreuve de Connaissances (qui évalue Gc) corrèle de manière positive et significative avec la moyenne générale (r=.42). Les corrélations sont faibles ou quasi nulles avec les autres épreuves. Ce résultat n'atteint pas les mêmes valeurs que ce qui est habituellement retrouvé dans la recherche, à savoir des corrélations qui sont généralement autour de .54 pour le français et .64 pour les mathématiques (Zimmerman & Woo-Sam, 1997). Les épreuves qui composent cette échelle ne sont donc que très peu ou pas liées à la performance scolaire alors que, par ailleurs, elles sont corrélées au WISC-IV.

- Liens avec l'adaptation sociale

Enfin, nous avons cherché les liens des évaluations subjective et objective de l'intelligence avec l'adaptation sociale telle que mesurée par la Vineland II (VABS-II ; Sparrow, Cicchetti

& Balla, 2015). Tous les facteurs du QAE, hormis Gc, entretiennent des liens avec les sous-échelles de la Vineland-II. C'est le facteur Gs (vitesse de traitement) qui présente les relations les plus fortes avec les dimensions de l'adaptation sociale, en particulier le domaine de la Socialisation (r=.60), de la Vie quotidienne (r=.70) et avec la note composite globale (r=.69).

Pour l'évaluation contextualisée, les liens sont beaucoup plus modestes puisque seules trois épreuves de l'évaluation contextualisée corrèlent positivement avec la note composite globale : Raisonnement (r=.29), Lexique (r=.29) et Majuscules (r=.27) même si ces corrélations sont faibles.

Pour résumer les apports de l'évaluation contextualisée, il semble nécessaire de distinguer les évaluations objectives des évaluations subjectives. Les évaluations subjectives, formalisées dans ce travail à travers les questionnaires QAE, QHP et QHE, montrent des liens probants avec la performance scolaire (en particulier Gf et Gc) et l'adaptation sociale (en particulier Gs). Elles semblent donc pertinentes à utiliser en complément de l'évaluation classique puisqu'elles mesurent des construits différents et que les deux prédisent la réussite scolaire. Nous n'avons malheureusement pas pu mener d'étude en utilisant les deux de manière conjointe afin d'étudier le poids de chacune dans la prédiction de la performance scolaire. Il s'agit là d'une piste de recherche future.

Concernant l'épreuve de performance, l'échelle d'évaluation contextualisée, les apports semblent très modestes concernant les preuves de validité, et notamment les liens avec la performance scolaire et l'adaptation sociale. Bien que cette échelle ait des liens significatifs avec le WISC-IV, qu'elle semble donc bien mesurer les capacités cognitives ; les liens sont faibles avec la performance scolaire (seule une épreuve corrèle de manière significative) et avec l'adaptation sociale telle que mesurée par la Vineland-II. Cette épreuve semble ainsi ne

pas se distinguer réellement de l'épreuve classique, certainement du fait que les contextes proposés restent artificiels. Deux aspects sont néanmoins à souligner. Nous avons pu voir qu'à l'échelle d'évaluation contextualisée, les différences selon le niveau de diplôme des parents et donc le milieu socio-culturel étaient moindres qu'au WISC-IV. Nous avons également eu des retours qualitatifs positifs sur la passation de cette échelle qui apparait « moins scolaire » pour les sujets. Cet aspect ne faisait pas partie de nos hypothèses initiales et n'a donc pas été évalué.

Des limites concernant nos trois études sont à souligner. Nous en dégageons principalement trois. 1/ Comme indiqué dans le développement, les outils créés, bien que présentant certaines qualités psychométriques, peuvent encore être largement améliorés ; 2/ Ces études ont été menées sur des échantillons réduits, en particulier celle sur l'échelle d'évaluation contextualisée ; 3/ Nous avons pris comme critère de la performance scolaire la moyenne générale. Cet indicateur n'est pas le plus pertinent. Idéalement, des mesures objectives des performances seraient plus adéquates même si les outils sont manquants actuellement.

Les perspectives de recherche sont donc très nombreuses. Concernant les évaluations de performance en particulier, nous avons vu que les tests contextualisés proposaient des tâches artificielles. La technologie proposée par la réalité virtuelle permettrait de dépasser cet inconvénient. Par ailleurs, nous avons indiqué, dans la première étude, que Schneider et McGrew (2012) précisaient que beaucoup d'aspects dans le modèle CHC n'avaient pas encore été explorés, car certaines capacités sont difficilement mesurables avec des épreuves classiques sous format papier-crayon. Ils précisent que la réalité virtuelle pourrait apporter des réponses à ce sujet.

Nous proposons dans les lignes qui suivent d'évoquer la réalité virtuelle, outil de simulation, comme voie de recherche future pour l'évaluation des capacités cognitives. Les avancées techniques permettent désormais de s'équiper pour un coût relativement modeste (tablette/ téléphone, casque de réalité virtuelle et application).

La réalité virtuelle est un « domaine scientifique et technique permettant à l'individu (ou à plusieurs) d'interagir en temps réel avec des entités 3D au moyen d'interfaces comportementales, dans un monde artificiel dans lequel il est plus ou moins immergé » (Klinger, 2006, p.13). Elle permet de simuler les activités de la vie quotidienne en contrôlant les informations. Elle répond donc aux inconvénients de l'évaluation en situation naturelle (temps d'observation très long, variables non contrôlables) tout en satisfaisant aux critères de l'évaluation classique (standardisation, contrôle des variables). Par ailleurs, elle s'adapte aux capacités du sujet afin d'éviter les échecs et de favoriser son apprentissage (Klinger, 2006).

Dans la réalité virtuelle, il est indispensable de pouvoir combiner immersion et interaction. L'immersion est définie comme l'état du sujet lorsque l'un ou plusieurs de ses sens n'enregistre que ce qui est issu de l'ordinateur ; l'interaction comme le fait de ne percevoir aucun décalage temporel entre l'action du sujet sur l'environnement virtuel et la réponse de ce dernier (Klinger, 2006). La notion de réalisme est également importante, c'est-à-dire le degré de ressemblance avec la réalité. Ce réalisme se décline au niveau du graphisme, de l'agencement des lieux, des tâches à réaliser (écologiques), de l'interaction avec l'environnement virtuel (naturel ou non) et peut être caractérisé par la notion de fidélité : fidélité perceptive (ressemblance avec le réel perceptif), physique (respect des lois physiques) et enfin psychologique (ressemblance avec les actions et les processus psychologiques qui seraient mis en œuvre dans la réalité).

Au vu de tous ces éléments, nous avons réfléchi au développement d'une tâche à réaliser en réalité virtuelle, qui permettrait d'évaluer les connaissances quantitatives et le raisonnement quantitatif du modèle CHC. Nous présentons cette tâche dans le tableau n° 36.

Nature de la tâche	Résoudre des problèmes arithmétiques, faire des calculs en utilisant diverses opérations (soustractions, additions, divisions, multiplications)
Contexte	Divers contextes : magasins de type boulangerie, prêt-à-porter ; cuisine
Activités du sujet	Actif : procéder à des calculs mentaux
Capacités évaluées dans le modèle CHC	• Connaissances quantitatives (Gq), en particulier les connaissances mathématiques (KM) • Raisonnement fluide (Gf), en particulier le raisonnement quantitatif (Rq)
Mode de présentation	Visuel et verbal
Mode de réponse	Oral (réponse orale du sujet)
Mesures	Réponse du sujet Temps de réponse
Degré d'immersion	Assez élevé du fait de l'utilisation des différents sens (vue, ouïe, odorat éventuellement, toucher)
Degré d'interaction	Assez élevé : le.la psychologue aurait un rôle actif dans la passation en rendant par exemple la monnaie (dépose des pièces dans la main du sujet)
Notion de réalisme	Elevée par l'utilisation de la vidéo à 360 degrés. Les fidélités perceptive, physique et psychologique seraient élevées.
Matériel utilisé	Casque de réalité virtuelle et tablette

Tableau 36. Proposition de modélisation d'une tâche d'arithmétique en réalité virtuelle

L'objectif serait de comparer les résultats d'un sujet à différentes tâches de niveau de contextualisation croissant afin d'évaluer l'impact du contexte sur les résultats :

- Une tâche d'arithmétique classique ;
- Une tâche d'arithmétique contextualisée en papier-crayon ;
- La tâche d'arithmétique en réalité virtuelle telle que présentée ci-dessus ;

mais également d'étudier les preuves de leur validité écologique (en évaluent les difficultés de calcul dans la vie quotidienne par questionnaire et/ou entretien).

Le travail mené dans le cadre de cette thèse nous apparait offrir des perspectives prometteuses et invite donc à la poursuite des recherches dans le domaine de l'évaluation contextualisée.

Bibliographie

ACHENBACH, T.M. *Integrative guide for the 1991 CBCL/4-18, YSR and TRF profiles*. Burlington : Department of Psychiatry, University of Vermont, 1991, p. 211.

ACKERMAN, P.L., BOWEN, K.R., BEIER, M.E., KANFER, R. Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology*, 2001, 93(4), 797-825.

ACKERMAN, P.L., HEGGESTAD, E.D. Intelligence, personality and interests: Evidence for overlapping traits. *Psychological Bulletin*, 1997, 121(2), 219-245.

ACHENBACH, T.M., MCCONAUGHY, S.H., HOWELL, C. Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, 1987, 101, 213-232.

ACKERMAN, P.L., WOLMAN, S.D. Determinants and Validity of Self-Estimates of Abilities and Self-Concept Measures. *Journal of Experimental Psychology*, 2007, 13(2), 57-78.

AMERICAN PSYCHIATRIC ASSOCIATION. *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition*. Washington: DC: Author, 2013, p.991.

AMTHAUER, R. *I-S-T 70. Intelligenz-Struktur-Test*. Göttingen: Hogrefe, 1973.

ARBUCKLE, J. L. *Amos 20 user's guide*. [en ligne], Chicago, IL: SPSS, 1983–2011, [référence du 24 octobre 2018].
ftp://public.dhe.ibm.com/software/analytics/spss/documentation/amos/20.0/en/Manuals/IBM_SPSS_Amos_User_Guide.pdf

BADDELEY, A. A 3 min reasoning test based on grammatical transformation. *Psychonomic Science*, 1968, 10, 341-342.

BAILEY, R.C., METTETAL, G. Perceived intelligence in married partners. *Social Behavior and Personality*, 1977, 1, 137-141.

BALBONI, G., COSCARELLI, A., GIUNTI, G., SCHALOCK, R.L. The assessment of the quality of life of adults with intellectual disability: The use of self-report and report of others assessment strategies. *Research in Developmental Disabilities*, 2013, 34(11), 4248-4254.

BANOS, J.H., LAGORY, J., SAWRIE, S., FAUGHT, E., KNOWLTON, R., PRASAD, A., KUZNIECKY, R., MARTIN, R.C. Self-report of cognitive abilities in temporal lobe epilepsy: Cognitive, psychosocial, and emotional factors. *Epilepsy and Behavior*, 2004, 5, 575-579.

BARKLEY, R.A., KNOUSE, L.E., MURPHY, K.R. Correspondence and disparity in the self- and other ratings of current and childhood ADHD symptoms and impairment in adults with ADHD. *Psychological Assessment*, 2011, 23(2), 437-446.

BARR, S. Factors influencing environmental attitudes and behaviors: A UK case study of household waste management. *Environment and Behavior*, 2007, 39, 435-473.

BARTHOLOMEW, D.J. *Measuring Intelligence: Facts and Fallacies 1st Edition*. Cambridge: Cambridge University Press, 2004, p. 186.

BELOFF, H. Mother father and me: Our IQ. *The Psychologist*, 1992, 5, 309-311.

BENNETT, G.K., SEASHORE, H.G., WESMAN, A.G. *DAT 5 Tests différentiels d'aptitudes*. Paris : ECPA, 2002.

BENNETT, M. Men's and women's self-estimates of intelligence. *Journal of Social Psychology*, 1996, 136, 411-412.

BENNETT, M. Gender differences in the self-estimates of ability. *Australian Journal of Psychology*, 2000, 52, 23-28.

BERNAARDS, C.A., JENNRICH, R.I. Gradient Projection Algorithms and Software for Arbitrary Rotation Criteria in Factor Analysis. *Educational and Psychological Measurement*, 2005, 65, 676–696.

BERNAUD, J.L. *Les méthodes d'évaluation de la personnalité*. Paris : Dunod, 2014, p. 128

BEYER, S. Gender differences in self-perception and negative recall bias. *Sex Roles*, 1998, 38, 103-133.

BIEDERMAN, J., MICK, E., FARAONE, S.V. Biased maternal reporting of child psychopathology? *Journal of the American Academy of Child & Adolescent Psychiatry*, 1998, 37, 10–12.

BINET, A., SIMON, T. Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Année psychologique*, 1905, 11, 191-244.

BINET, A., SIMON, T. Le développement de l'intelligence chez l'enfant. *L'Année psychologique*, 1908, 14, 191-244.

BONG, M., CLARK, R. E. Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist*, 1999, 34(3), 139–153.

BONG, M., SKAALVIK, E. M. Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 2003, 15, 1–40.

BORKENAU, P., LIEBLER, A. Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings, and measured intelligence. *Journal of Personality and Social Psychology*, 1993, 65(3), 546-553.

BOSSONG, B. Kognitive Dissonanz, Attribution und Leistung. *Zeitschrift für Sozialpsychologie*, 1982, 13, 194-208.

BOUT-FORESTIER, N., DEPOORTER, H. *TLC 2 Test lillois de Calcul 2*. Ballan-Miré : Unithèque, 2008.

BRODY, N. *Intelligence (2nd edition)*. New York: Academic, 1992, p. 395.

BRODY, N. Construct validation of the Sternberg Triarchic Abilities Test Comment and reanalysis. *Intelligence*, 2003, 31, 319-329.

BROUWERS, S.A., VAN DE VIJVER, J.R. Contextualizing intelligence in assessment: The next step. *Human Resource Management Review*, 2015, 25, 38–46.

BRUCE, M.M., LEARNER, D.B. A supervisory practices test. *Personnel Psychology*, 1958, 11, 207-216.

BYRD, M., STACEY, B. Bias in IQ perception. *The Psychologist*, 1993, 6, 16.

CAMPBELL, D. J., LEE, C. Self-Appraisal in Performance Evaluation: Development Versus Evaluation. *Academy of Management Review*, 1988, 13(2), 302-314.

CARDALL, A.J. *Preliminary manual for the Test of Practical Judgment*. Chicago : Science Research, 1942.

CARRICANO, M., POUJOL, F., BERTRANDIAS, L. Analyse de données avec SPSS®, 2e éd, Pearson Education France, 2010, p. 288.

CASTRO, D. *Pratique de l'examen psychologique en clinique adulte – 2e ed*. Paris : Dunod, 2011, p. 528.

CHAMORRO-PREMUZIC, T., FURNHAM, A. Personality and self-assessed intelligence: Can gender and personality distort self-assessed intelligence? *Educational Research and Reviews*, 2006, 1(7), 227-233.

CHAMORRO-PREMUZIC, T., FURNHAM, A., MOUTAFI, J. Self-estimated personality predicts psychometric intelligence better than well-established personality traits. *Journal of Research in Personality*, 2004, 38, 505-513.

CHAMORRO-PREMUZIC, T., MOUTAFI, J., FURNHAM, A. The relationship between personality traits, subjectively-assessed and fluid intelligence. *Personality and Individual Differences*, 2005, 38, 1517-1528.

CHAN, R.C.K., CHEN, E.Y.H. Assessment of Executive Function for Schizophrenia in Honk Kong. *Hong Kong Journal of Psychiatry*, 2005, 15, 23-28.

CHARTIER, P., LOARER, E. *Évaluer l'intelligence logique*. Paris : Dunod, 2008, p. 416.

CHAYTOR, N., SCHMITTER-EDGECOMBE, M. The ecological validity of neuropsychological tests: a review of the literature on everyday cognitive skills. *Neuropsychogy Revue*, 2003, 13(4), 181-97.

CHEVIGNARD, M., PILLON, B., PRADAH-DIEHL, P., TAILLEFER, C., ROUSSEAU, S., LE BRAS, C. An ecological approach to planning dysfunction: script execution. *Cortex*, 2000, 36, 649–669.

CHEVIGNARD, M.P., PRADAT-DIEHL, P., PILLON, B. Exécution de scripts et syndrome dysexécutif: une approche écologique. *Annales de Réadaptation et de Médecine Physique*, 1999, 42(7), 368-369.

CHEVIGNARD, M., TAILLEFER, C., PICQ, C., PONCET, F., NOULHIANE, M., PRADAT-DIEHL, P. Ecological assessment of the dysexecutive syndrome using execution of a cooking task. *Neurological Rehabilitation*, 2008, 18(4), 461-485.

CHOI, S., KLUEMPER, D.H. The relative utility of differing measures of emotional intelligence: Others-rated EI as prédicteur of social functioning. *European Review of Applied Psychology*, 2012, 62(3), 121-127.

CHRISTIAN, M., EDWARDS, B. D., BRADLEY, J. C. Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validity. *Personnel Psychology*, 2010, 63, 83–117.

COCHRAN, W. G. Some methods for strengthening the common χ 2 tests. *Biometrics*, 1954, 10, 417–451.

Code de déontologie des Psychologues. [en ligne], 1996, version actualisée de 2002, [référence du 24 octobre 2018]. http://www.codededeontologiedespsychologues.fr/LE-CODE.html

COHEN, J. *Statistical Power Analysis for the Behavioral Sciences, 2nd Edition*. Hillsdale, N.J.: Lawrence Erlbaum, 1988, p. 400.

COLE, D.A., MAXWELL, S.E., MARTIN, J.M. Reflected self-appraisals: Strength and structure of the relation of teacher, peer, and parent ratings to children's self-perceived competencies. *Journal of Educational Psychology*, 1997, 89, 55-70.

CONNERS, C.K. *CRS-R: Conners' Rating Scales-revised : Technical Manual*. New York : Multi-Health Systems, Inc, 2001, p. 214.

COOPER, D., ROBERTSON, I.T, TINLINE, G. *Recruitment and Selection: A Framework for Success*. London : Cengage Learning EMEA, 2003, p. 256.

COSTA, P.T., MCCRAE, R.R. *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manuel*. Odessa, Florida : Psychological Assessment Resources, 1992, p. 101.

CROOK, T. H., SALAMA, M., GOBERT, J. A computerized test battery for detecting and assessing memory disorders. In : BES, A., CAHN, J., HOYER, S., MARC-VERGENES, J.P.

& WISNIEWSKI, H.M. (éd.)., *Senile Dementias: Early Detection*. London-Paris : John Libbey Eurotext, 1986., p. 654.

CUCINA, J.M., SU, C., BUSCIGLIO, H.H., HARRIS THOMAS, P., THOMPSON PEYTON, S. Video-based Testing: A high-fidelity job simulation that demonstrates reliability, validity, and utility. *International Journal of Selection and Assessment*, 2015, 23(3), 197-209.

CYDERS, M.A., COSKUNPINAR, A. Measurement of constructs using self-report and behavioral lab tasks: is there overlap in nomothetic span and construct representation for impulsivity? *Clinical Psychology Review*, 2011, 31(6), 965-982.

DE LOS REYES, A, KAZDIN, A.E. Informant discrepancies in the assessment of childhood psychopathology: a critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, 2005, 131, 483–509.

DUBOWITZ, H., SAWYER, R. J. School behavior of children in kinship care. *Child Abuse and Neglect*, 1994, 18, 899-911.

EHRLINGER, J., DUNNING, D. How Chronic Self-Views Influence (and Potentially Mislead) Estimates of Performance. *Journal of Personality and Social Psychology*, 2003, 84(1), 5-17.

EPSTEIN, S. Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality*, 1983, 51(3), 360-392.

EYSENCK, H.J., EYSENCK, M.W. *Personality and individual differences: A natural science approach*. New York: Plenum Press, 1985, p. 424.

FABREGA, H., ULRICH, R.F., LOEBER, R. Adolescent psychopathology as a function of informant and risk status. *Journal of Nervous and Mental Disease*, 1996, 184, 27–34.

FILE, Q.W., REMMERS, H.H. *How Supervise? manual 1948 revision*. New York : Psychological Corporation, 1948.

FINGERMAN, K.L., PERLMUTTER, M. Self rating of past, present, and future cognitive performance across adulthood. *International Journal of Aging and Human Development*, 1994, 38, 303-322.

FLEISS, J. L. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 1971, 76(5), 378–382.

FLETCHER, T.D. *Psychometric: Applied Psychometric Theory. R package version 2.2*. [en ligne], 2010, [référence du 25 juillet 2017]. http://CRAN.R-project.org/package=psychometric

FLOYD, R. G., EVANS, J. J., MCGREW, K. S. Relations between measures of Cattell – Horn-Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools*, 2003, 40(2), 155-171.

FLOYD, R. G., KEITH, T. Z., TAUB, G. E., MCGREW, K. S. Cattell –Horn –Carroll cognitive abilities and their effects on reading decoding skills: g has indirect effects, more specific abilities have direct effects. *School Psychology Quarterly*, 2007, 22(2), 200-233.

FORNER, Y. *QMF-6, questionnaire de motivation à la réussite en situation de formation*. Paris : Eurotests Editions, 2010.

FREDERIKSEN, N. Validation of a simulation technique. *Organizational Behavior and Human Performance*, 1966, 1, 87–109.

FREDERIKSEN, N., SAUNDERS, D. R., WAND, B. The in-basket test. *Psychological Monographs: General and Applied*, 1957, 71(9), 1-28.

FREUND, P.A., KASTEN, N. How smart do you think you are? A metaanalysis on the validity of self-estimated cognitive ability. *Psychological Bulletin*, 2012, 138, 296-321.

FUJ, E.T., HENNESSY, M., MAK, J. An evaluation of the validity and reliability of survey response data on household electricity conservation. *Evaluation Review*, 1985, 9, 93-104.

FURMAN, J. M., JACOB, R. G. Psychiatric dizziness. *Neurology*, 1997, 48(5), 1161-1166.

FURNHAM, A. Parents' estimates of their own and their children's multiple intelligences. *British Journal of Developmental Psychology*, 2000, 18, 583-594.

FURNHAM, A. Self-estimates of intelligence: Culture and gender difference in self and other estimates of both general (g) and multiple intelligences. *Personality and Individual Differences*, 2001, 31, 1381-1405.

FURNHAM, A., CLARK, K & BAILEY, K. Sex differences in estimates of multiple intelligences. *European Journal of Personality*, 1999, 13(4), 247-259.

FURNHAM, A., DISSOU, G. The Relationship Between Self-Estimated and Test-Derived Scores of Personality and Intelligence. *Journal of Individual Differences*, 2007, 28(1), 3-44.

FURNHAM, A., FONG, G. & MARTIN, N. Sex and cross-cultural differences in the estimated multifaceted intelligence quotient score for self, parents and siblings. *Personality and Individual Differences*, 1999, 26, 1025-1034.

FURNHAM, A., KIDWAI, A., THOMAS, C. Personality, psychometric intelligence and self-estimated intelligence. *Journal of Social Behaviour and Personality*, 2001, 16, 97-114.

FURNHAM, A., MANSI, A. The self-assessment of the Cattell–Horn–Carroll broad stratum abilities. *Learning and Individual Differences*, 2014, 32, 233-237.

FURNHAM, A., MOUTAFI, J., CHAMORRO-PREMUZIC, T. Personality and intelligence: Gender, the "Big Five", self-estimated and psychometric intelligence. *International Journal of Selection and Assessment*, 2005, 13, 11-24.

FURNHAM, A., RAWLES, R. Correlations between self-estimated and psychometrically measured IQ. *The Journal of Social Psychology*, 1999, 139, 405-410.

FURNHAM, A., RAWLES, R. Sex differences in the estimation of intelligence. *Journal of Social Behavior and Personality*, 1995, 10, 741-745.

FURNHAM, A., THOMAS, C. Parents' gender and personality and estimates of their own and their children's intelligence. *Personality and Individual Differences*, 2004, 37, 887-903.

FURNHAM, A., ZHANG, J., CHAMORRO-PREMUZIC, T. The relationship between psychometric and self-estimated intelligence, creativity, personality and academic achievement. *Imagination, cognition and personality*, 2005-2006, 25(2), 119-145.

GARDNER, H. *Frames of mind*. New York : Basic Books, 1983, p. 440.

GARDNER, H. *Intelligence reframed: multiple intelligences for the 21st century*. New York : Basic Books, 1999, p.304.

GARDNER, H. *Development and education of the mind. The selected works of Howard Gardner*. London : Routledge, 2006, p. 280.

GELLER, E.S. Evaluating energy conservation programs: Is verbal report enough? *Journal of Consumer Research*, 1981, 8, 331-335.

GIGNAC, G.E., STOUGH, C., LOUKOMITIS, S. Openness, intelligence, and self-report intelligence. *Intelligence*, 2004, 32, 133-143.

GIOIA, G.A., ISQUITH, P.K., GUY, S.C., KENWORTHY, L. Behavior Rating Inventory of Executive Function. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, 2000, 6(3), 235-238, trad. fr. BRIEF: inventaire d'évaluation comportementale des fonctions exécutives, Paris : Hogrefe.

GOLDSTEIN, G., MCCUE, M., ROGERS, J., NUSSBAUM, P. D. Diagnostic differences in memory test based predictions of functional capacity in the elderly. *Neuropsychology Rehabilitation*, 1992, 2, 307–317.

GOTTFREDSON, L., SAKLOFSKE, D.H. Intelligence: Foundations and Issues in Assessment. *Canadian Psychology*, 2009, 50(3), 183–195.

GOTTFREDSON, L.S. Dissecting practical intelligence theory: Its claims and evidence. *Intelligence*, 2003, 31, 343 – 397.

GRANT, D.A., BERG, E.A. A behavioural analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *Journal of Experimental Psychology*, 1948, 38, 404–411.

GREENO, J.G. Gibson's affordances. *Psychological Review*, 1994, 101(2), 336-342.

GREGOIRE, J. *L'examen clinique de l'intelligence de l'adulte*. Sprimont : Mardaga, 2004, p. 263.

GREGOIRE, J. *L'examen Clinique de l'intelligence de l'enfant. Fondements et pratique du WISC-IV*. Sprimont : Mardaga, 2006, p. 302.

GREGOIRE, J. *L'examen clinique de l'intelligence de l'enfant. Fondements et pratique du WISC-V*. Sprimont : Mardaga, à paraître.

GRIETENS, H., ONGHENA, P., PRINZIE, P., GADEYNE, E., VAN ASSCHE, V., GHESQUIERE, P., HELLINCKX, W. Comparison of Mothers', Fathers', and Teachers' Reports on Problem Behavior in 5- to 6-Year-Old Children. *Journal of Psychopathology and Behavioral Assessment*, 2004. 26(2), 137-146.

GRIGORENKO, E.L., MEIER, E., LIPKA, J., MOHATT, G., YANEZ, E., STERNBERG, R.J. Academic and Practical Intelligence: A Case Study of the Yup'ik in Alaska. *Learning & Individual Differences*, 2004, 14(4), 183-207.

GRIGORENKO, E. L., STERNBERG, R. J. Analytical, creative, and practical intelligence as predictors of self-reported adaptive functioning: A case study in Russia. *Intelligence*, 2001, 28, 1-17.

HARRIS, M. M., SCHAUBROECK, J. A meta-analysis of self-supervisor, self-peer and peer-supervisor ratings. *Personnel Psychology*, 1988, 41(1), 43-62.

HART, E.L., LAHEY, B.B., LOEBER, R., HANSON, K.S. Criterion validity of informants in the diagnosis of disruptive behavior disorders in children: A preliminary study. *Journal of Consulting and Clinical Psychology*, 1994, 62, 410-414.

HARTER, S. *Manual for the Self-Perception Profile for Adolescents*. Denver: University of Denver, 1988/2012, p.50.

HAZUCHA, J. F., HEZLETT, S. A., SHNEIDER, R. J. The impact of 360-degree feedback on management skills development. *Human Ressource Management*, 1993, 32, 325-351.

HEDLUND, J., HORVATH, J. A., FORSYTHE, G. B., SNOOK, S., WILLIAMS, W. M., BULLIS, R. C., DENNIS, M., STERNBERG, R. J. *Tacit Knowledge in Military Leadership: Evidence of Construct Validity (Tech. Rep. 1080)*. Alexandria, VA: Army Research Institute, 1998, p. 116.

HEGARTY, M., MONTELLO, D.R., RICHARDSON, A.E., ISHIKAWA, T., LOVELACE, K. Spatial Abilities at Different Scales: Individual differences inaptitude-test performance and spatial-layout learning. *Intelligence*, 2006, 34, 151– 176.

HIGGINSON, C. I., ARNETT, P. A., VOSS, W. D. The ecological validity of clinical tests of memory and attention in multiple sclerosis. *Arch. Clin. Neuropsychol*, 2000, 15, 185–204.

HOFSTEE, W.K.B. Who should own the definition of personality? *European Journal of Personality*, 1994, 8(3), 149-162.

HOGAN, W.H. IQ self-estimates of males and females. *Journal of Social Psychology*, 1978, 106, 137-138.

HOLLING, H., PRECKEL, F. Self-estimates of intelligence-methodological approaches and gender differences. *Personality and Individual Differences*, 2005, 38, p. 503-517.

HOOPER, D., COUGHLAN, J., MULLEN, M. Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 2008. 6(1), 53–60.

HORN, J. L., NOLL, J. Human cognitive capabilities: Gf-Gc theory. In FLANAGAN, D. P., GENSHAFT, HARRISON, P.L. (éd.), *Contemporary intellectual assessment*. New York: Guilford, 2018, p. 926.

HOYT, W.T., WARBASSE, R.E., CHU, E.Y. Construct validation in counseling psychology research. *The Counseling Psychologist*, 2006, 34(6), 769-805.

HU, L., BENTLER, P. M. Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 1999, 6, 1–55.

HUTCHINS, E. How a cockpit remembers its speed. *Cognitive Science*, 1995, 19, 265-288.

HUTEAU, M. & LAUTREY, J. *Evaluer l'intelligence – Psychométrie cognitive*. Paris : PUF, 1999, p.310.

IBM Corp. Released. *IBM SPSS Statistics for Windows, Version 20.0*. Armonk, NY: IBM Corp, 2011.

INTERNATIONAL TEST COMMISSION (ITC). *International Guidelines for Test Use* [en ligne], 2003, [référence du 26 octobre 2018]. www.intestcom.org.

JACOBS, K., ROODENBURG, J. The development and validation of the Self-Report Measure of Cognitive Abilities: A multitrait–multimethod study. *Intelligence*, 2014, 42, 5–21.

JAMES, L. R., DEMAREE, R. G., WOLF, G. An assessment of withingroup interrater agreement. *Journal of Applied Psychology*, 1993, 78(2), 306-309.

JENSEN, A.R. Bias in mental testing. New York: Free Press, University of California, Berkeley, CA, 1980, p. 786.

JENSEN, A.R. Psychometric g and achievement. In: GIFFORD, B.R., O'CONNOR, M.C. (éd.), *Changing assessments: Alternative views of aptitude, achievement, and instruction*. Boston, MA: Kluwer Academic Publishers, 1992, p.337.

JODOIN, M.G., GIERL, M.J. Evaluating type I error and power rates using an effect size measure with logistic regression procedure for DIF detection. *Applied measurement in education*, 2001, 14, 329-349.

JOHNSON, J. L. Episodic memory deficits in Alzheimer's disease: A behaviorally anchored scale. *Arch. Clin. Neuropsychol*, 1994. 9, 337–346.

JONES, E. E., NISBETT, R. E. The actor and the observer: Divergent perceptions of the causes of behavior. In JONES, E.E., KANOUSE, D., KELLY, H.H., NISBETT, R.E., VALINS, S., WEINER, B. (éd.), *Attribution: Perceiving the causes of behavior*. Morristown, NJ: General Learning Press, 1972, p. 186.

JUDGE, T.A., JACKSON, C.L., SHAW, J.C., SCOTT, B.A., RICH, B.L., Self-efficacy and work-related performance: The integral role of individual differences. *Journal of Applied Psychology*, 2007, 92(1), 107-127.

KAITARO, T., KOSKINEN, S., KAIPIO, M. Neuropsychological problems in everyday life: A 5-year follow-up study of young severely closed-head-injured patients. *Brain Inj*, 1995, 9, 713–727.

KAUFMAN, A.S., KAUFMAN, N.L. *KABC-II Batterie pour l'examen psychologique de l'enfant, 2e éd*. Paris: ECPA, 2008.

KAUFMAN, J.C. Self estimates of general, crystallized, and fluid intelligences in an ethnically diverse population. *Learning and Individual Differences*, 2012, 22, 118-122.

KEITH, T.Z., FINE, J.G., TAUB, G.E., REYNOLDS, M.R., KRANZLER, J.H. Higher order, multisample, confirmatory factor analysis of the Wechsler intelligence scale for children-fourth edition: what does it measure? *School Psychology*, 2006, 35, 108–127.

KENNY, D. A. A general model of consensus and accuracy in interpersonal perception, *Psychological Review*, 1991, 98(2), 155-163.

KLINE, R. B. Principles and practice of structural equation modeling. New York: The Guilford Press, 1998, p.354.

KLINGER, E. *Apports de la réalité virtuelle à la prise en charge de troubles cognitifs et comportementaux*. Thèse, Télécom ParisTech, 07 février 2006, p. 229.

KLINGER, E., CHEMIN, I., LEBRETON, S., MARIE, R.M. A Virtual Supermarket to Assess Cognitive Planning. *CyberPsychology & Behavior*, 2004, 7(3), 292-293.

KLINGER, E., MARIE, R.M., FUCHS, P. Réalité virtuelle et sciences cognitives, applications en psychiatrie et neuropsychologie. *Cognito*, 2006, 3(2), 1-31.

KOKE, L.C., VERNON, P.A. The Sternberg Triarchic Abilities Test (STAT) as a measure of academic achievement and general intelligence. *Personality and Individual Differences*, 2003, 35(8), 1803-1807.

KORNILOV, S. Self-estimated intelligence and academic achievement: A mini meta-analysis. *Psychology*, 2011, 3, 56-66.

KORNILOVA, T., KORNILOV, S, CHUMAKOVA, M. Subjective evaluations of intelligence and academic self-concept predict academic achievement: Evidence from a selective student population. *Learning and Individual Differences*, 2009, 19, 596-608.

KORNILOVA, T.V., NOVIKOVA, M.A. Self-assessed intelligence, psychometric intelligence, personality, and academic achievement: Two structural models. In GOWDA, M., KHANDERIA, A. (éd.), *Educational Achievement: Teaching Strategies, Psychological Factors and Economic Impact* (pp.197-212). New York: Nova Science, 2013, p. 251.

KRAEMER, H.C., MEASELLE, J.R., ABLOW, J.C., ESSEX, M.J., BOYCE, W.T., KUPFER, D.J. A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *Am J Psychiatry*, 2003, 160(9), 1566–1577.

KRUGER, J., DUNNING, D. Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 1999, 77, 1121-1134.

LATHAM, G. P., WEXLEY, K. N. *Increasing productivity through performance appraisal.* Reading. MA: Addison-Wesley, 1981, p.262.

LAVE, J. *Cognition in practice*. Cambridge: Cambridge University Press, 1988, p. 229.

LAVE, J., MURTAUGH, M., DE LA ROCHE, O. The dialectic of arithmetic in grocery shopping. In ROGOFF, B., LAVE, J. (éd.), *Everyday Cognition: Its development in social context*. Cambridge, Mass: Harvard University Press, 1984, p. 352.

LECERF, T., GOLAY, P., REVERTE, I., SENN, D., FAVEZ, N., ROSSIER, J. Scores composites CHC pour le WISC-IV : normes francophones. Pratiques Psychologiques, 2012, 18(1), 37-50.

LECERF, T., ROSSIER, J., FAVEZ, N., REVERTE, I., COLEAUX, L. The four- vs. alternative six-factor structure of the French. WISC-IV: comparison using Confirmatory Factor Analyses. *Swiss Journal of Psychology,* 2010, 69, 221–232.

LEE, J.H., KU, J., CHO, W., HAHN, W.Y., KIM, I.Y., LEE, S.M., KANG, Y., KIM, D.Y., Yu, T., WIEDERHOLD, B.K., WIEDERHOLD, M.D., KIM, S.I. A virtual reality system for the assessment and rehabilitation of the activities of daily living. *CyberPsychology & Behavior*, 2003, 6(4), 383-388.

LEVY-LEBOYER, C. Le 360° Outil de développement personnel. Paris: Editions de l'organisation, 2000, p.152.

LO PRIORE C., CASTELNUOVO G., LICCIONE D. Experience with V-STORE: considerations on presence in virtual environments for effective neuropsychological rehabilitation of executive functions. *CyberPsychology & Behavior*, 2003, 6(3), 281-287.

LOARER, E. L'intelligence sociale et l'intelligence émotionnelle. In : LAUTREY, J., RICHARD, J. F. *L'intelligence. Traité des sciences cognitives*. Paris : Hermès, 2005, p. 303.

LOEBER, R., GREEN, S.M., LAHEY, B.B., STOUTHAMER-LOEBER, M. Optimal informants on childhood disruptive behaviors. *Development and Psychopathology*, 1990, 1, 317-337.

LOEBER, R., GREEN, S.M., LAHEY, B.B., STOUTHAMER-LOEBER, M. Differences and similarities between children, mothers, and teachers as informants on disruptive child behavior. *Journal of Abnormal Child Psychology*, 1991, 19, 75-95.

LOHBECK,A. Self-concept and self-determination theory: math self-concept, motivation, and grades in elementary school children. *Early Child Development and Care*, 2018, 188 (8),1031-1044.

LORD, R. G., DE VADER, C. L., ALLIGER, G. M. A meta-analysis of the relation between personality traits and leadership perceptions: An application of validity generalization procedures. *Journal of Applied Psychology*, 1986, 71(3), 402-410.

MABE, P.A., WEST, S.G. Validity of Self-Evaluation of Ability: A Review and Meta-Analysis. *Journal of Applied Psychology*, 1982, 67(3), 280-296.

MACALPINE, M. *A summary of methods of item analysis*. University of Luton : CAA Centre Bluepaper 2, 2002, p. 33.

MACKINTOSH, N. J. *QI & intelligence humaine.* (A. Brossard, Trad., P. Chartier, Éd.). Bruxelles : De Boeck, 2004, p. 490.

MAKATURA, T. J., LAM, C. S., LEAHY, B. J., CASTILLO, M. T. & KALPAKJIAN, C. Z. Standardized memory tests and the appraisal of everyday memory. *Brain Inj*, 1999, 13, 355–367.

MANTEL, N. Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association*, 1963, 58, 690–700.

MCCARTHY, D. *McCarthy Scales of Children's Abilities*. New York: The Psychological Corporation, 1972, p. 209.

MCCLELLAND, D. C. Testing for competence rather than for intelligence. *American Psychologist*, 1973, 28, 1-14.

MCDANIEL, M.A., HARTMAN, N.S., WHETZEL, D.L., GRUBB, W.L. Situational Judgment Tests, response instructions, and validity: a meta-analysis. *Personnel Psychology*, 2007, 60, 63–91.

MCDANIEL, M.A., MORGESON, F.P., FINNEGAN, E.B., CAMPION, M.A., BRAVERMAN, E. P. Use of situational judgment tests to predict job performance: a clarification of the literature. *Journal of Applied Psychology*, 2001, 86(4), 730-40.

MCGREW, K.S., KNOPIK, S.N. The Relationship between the WJ-RGf-Gc Cognitive Clusters and Writing Achievement across the Life-Span. *School Psychology Review*, 1993, 22, 687-695.

MCGREW, K.S., WENDLING, B.J. Cattell–Horn–Carroll cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools*, 2010, 47, 651 - 675.

MCLELLAN, H. (éd.). *Situated Learning Perspectives*. Englewood Cliffs, NJ: Educational Technology Publications, 1996, p. 311.

MESSICK, S. The once and future issues of validity: Assessing the meaning and consequences of measurement. In : WAINER, H., BRAUN, H.I. (éd.). *Test validity*. Hillsdale, NJ : Lawrence Erlbaum Associates, 1988, p. 288.

MEYER, G.J., FINN, S.E., EYDE, L.D., KAY, G.G., MORELAND, K.L., DIES, R.R., EISMAN, E.J., KUBISZYN, T.W., REED, G.M. Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, 2001, 56, 128-165.

MEYER, J.P. *Applied Measurement with jMetrik*. New York, NY: Routledge, 2014, p. 171.

NAGELKERKE, N. J. D. A note on a general definition of the coefficient of determination. *Biometrika*, 1991, 691-692.

NEISSER, U. *Concepts and conceptual development: Ecological and intellectual factors in categorization*. New York, NY US : Cambridge University Press, 1987, p. 317.

NEWTON, J.H., Mcgrew, K.S. Introduction to the special issue: Current research in Cattell-Horn-Carroll based assessment. *Psychology in the Schools*, 2010, 47(7), 621-634.

NUNNALY, J.C., BERNSTEIN, I. *Psychometric theory.* New York : McGraw-Hill, 1994, p. 640.

OAKLAND, T., HU, S. The top 10 tests used with children and youth worldwide. *Bulletin of the International Test Commission*, 1992, 19, 99-120.

OLIVER, B., DALE, P., SAUDINO, K., PETRILL, S., PIKE, A., PLOMIN, R. The Validity of a Parent-based Assessment of Cognitive Abilities in Three-year Olds. *Early Child Development and Care*, 2002, 172(4), 337-348.

OOSTROM, J. K., BORN, M. P., SERLIE, A. W., VAN DER MOLEN, H. T. A multimedia situational test with a constructed-response format: Its relationship with personality, cognitive ability, job experience, and academic performance. *Journal of Personnel Psychology*, 2011, 10, 78–88.

OWNBY, R. L., WALLBROWN, F. H., BROWN, D. Y. Special education teachers' perceptions of reports written by school psychologists. *Perceptual and Motor Skills*, 1982, 55, 955–961.

PAULHUS, D., LYSY, D., YIK, M. Self-Report Measures of Intelligence: Are they useful as proxy IQ Tests ? *Journal of Personality*, 1998, 66(4), 525-554.

PEREIRA DA COSTA, M., SCELLES, R. Un outil d'évaluation des compétences cognitives des jeunes polyhandicapés : le P2CJP. *Alter - Revue européenne de recherche sur le handicap*, 2012, 6, 110-123.

PEREZ, L.F., GONZÁLEZ, C., BELTRAN, J.A. Parental estimates of their own and their relatives' intelligence. *Learning and Individual Differences*, 2010, 20, 669-676.

PETRIDES, K.V., FURNHAM, A. Trait Emotional Intelligence: Psychometric Investigation with Reference to Established Trait Taxonomies. *European Journal of Personality*, 2001, 15, 425-448.

PEZZUTI, L., ORSINI, A. Are there sex differences in the Wechsler Intelligence Scale for Children — Forth Edition? *Learning and Individual Differences*, 2016, 45, 307-312.

PUGNETTI, L., MENDOZZI, L., ATTREE, E.A., BARBIERI, E., BROOKS, B.M., CAZZULLO, C.L., MOTTA, A., ROSE, F.D. Probing memory and executive functions with virtual reality : Past and present studies. *CyberPsychology and Behavior*, 1998, 2, 151-161.

QUAISER-POHL, C., LEHMANN, W. Girls' spatial abilities: Charting the contributions of experiences and attitudes in different academic groups. *British Journal of Educational Psychology*, 2002, 72, 245-260.

R CORE TEAM. *R: A language and environment for statistical computing*. Vienna, Austria : R Foundation for Statistical Computing, 2014. URL http://www.R-project.org/.

RAMMSTEDT, B., RAMMSAYER, T.H. Sex differences in self-estimates of different aspects of intelligence. *Personality and Individual Differences*, 2000, 29, 869-880.

RAVEN, J.C. *Progressive matrices de Raven*. Paris : ECPA, 1998.

REVELLE, W. psych: *Procedures for Personality and Psychological Research*. Evanston, Illinois, USA : Northwestern University, 2015. http://CRAN.R-project.org/package=psych Version = 1.5.8.

REYNOLDS, C.R., KAMPHAUS, R.W. *BASC-2: Behavior assessment system for children, second edition manual*. Circle Pines, MN : American Guidance Service, 2004.

RIZZO, A.A., BOWERLY, T., BUCKWALTER, J.G., SCHULTHEIS, M.T., MATHEIS, R., SHAHABI, C., NEUMANN, U., KIM, L., SHARIFZADEH, M. Virtual Environments for the Assessment of Attention and Memory Processes: The Virtual Classroom and Office. In: SHARKEY, P, LANYI, CS, STANDEN, P (éd.). *Proc. of the 4 th International Conference on Disability*, *Virtual Reality and Associated Techniques,* Veszprém, Hungary, 18-20 septembre 2002, Reading UK: University of Reading, 2002, 4.

RIZZO, A.A., BOWERLY, T., SHAHABI, C., BUCKWALTER, J.G., KLIMCHUK, D., MITURA, R. Diagnosing Attention Disorders in a Virtual Classroom. *IEEE Computer*, 2004, 37(6), 87-89.

ROBERTSON, I.H., WARD, T., RIDGEWAY, V., NIMMO-SMITH, I. *The Test of Everyday Attention*. Bury St. Edmunds : Thames Valley Test Company, 1994.

ROID, G. H. *Stanford-Binet Intelligence Scales, Fifth Edition (SB:V).* Itasca, IL : Riverside Publishing, 2003.

ROSE, F.D., BROOKS, B.M., RIZZO, A.A. Virtual reality in brain damage rehabilitation: review. *Cyberpsychol Behav*, 2005, 8(3), 241-271.

RUEDA, A.D., LAU, K.M., SAITO, N., HARVEY, D., RISACHER, S.L., AISEN, P.S., PETERSEN, R.C., SAYKIN, A.J., FARIAS, S.T. Self-rated and informant-rated everyday function in comparison to objective markers of Alzheimer's disease. *Alzheimer's & Dementia*, 2015, 11(9), 1080-1089.

RUSHTON, J. P., BRAINERD, C. J., PRESSLEY, M. Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin*, 1983, 94, 18-38.

SALGADO, J.F., VISWESVARAN, C., ONES, D.S. Predictors used for personal selection: an overview of construct, methods and techniques. In : ANDERSON, N., ONES, D.S., SINANGIL, H.K., VISWESVARAN, C. (éd.). *Handbook of Industrial, Work and Organizational Psychology, 1*. London: Sage, 2001, p. 480.

SALVAGNO, M., TEGLASI, H. Teacher perceptions of different types of information in psychological reports. *Journal of School Psychology*, 1987, 25, 415-424.

SCHMIDT, F.L., HUNTER, J.E. The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological Bulletin*, 1998, 124(2), 262-274.

SCHNEIDER, W. J., MCGREW, K. S. The Cattell-Horn-Carroll Model of Intelligence. In FLANAGAN, D. P., HARRISON, P. L. (éd.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed.). New York, NY: Guilford Press, 2012, p. 926.

SCHULTHEIS, M.T., RIZZO, A. A. The virtual office: Assessing and re-training vocationally relevant cognitive skills. In : The 10th Annual Medicine Meets Virtual Reality Conference, Newport Beach, CA, 26 January, 2002.

SCRIBNER, S., Studying working intelligence. In ROGOFF, B., LAVE, J. (éd.). *Everyday cognition: Its development in social context*. Cambridge, Mass: Harvard University Press, 1984, p. 352.

SEIDENBERG, M., HALTINER, A., TAYLOR, M.A., HERMANN, B.B., WYLER, A. Development and validation of a multiple ability self-report questionnaire. *Journal of Clinical and Experimental Neuropsychology*, 1994, 16(1), 93-104.

SHALLICE, T., BURGESS, P.W. Deficits in strategy application following frontal lobe damage in man. *Brain*, 1991, 114, 727–741.

SHAVELSON, R.J., HUBNER, J.J., STANTON, G.C. Self-concept: Validation of Construct Interpretations. *Review of Educational Research*, 1976, 46, 407-441.

SHRAUGER, J. S., SCHOENEMAN, T. J. Symbolic interactionist view of self-concept: Through the looking glass darkly. *Psychological Bulletin*, 1979, 86(3), 549-573.

SIMMS, L.J. Classical and modern methods of psychological scale construction. *Social and Personality Psychology Compass*, 2008, 2(1), 414-433.

SPARROW, S.S., CICCHETTI, D.V., BALLA, D.A. *Vineland-II, Echelle d'évaluation du comportement socio-adaptatif de Vineland, Adaptation française*, Paris : ECPA, 2015, p. 263.

SPINELLA, M., YANG, B., LESTER, D. Prefrontal system dysfunction and credit card debt. *International Journal Neuroscience*, 2004, 114, 1323-1332.

STANGER, C., LEWIS, M. Agreement Among Parents, Teachers, and Children on Internalizing and Externalizing Behavior Problems. *Journal of Clinical Child Psychology*, 1993, 22(1), 107-116.

STARR, R., DUBOWITZ, H., HARRINGTON, D., FEIGELMAN, S. Behavior problems of teens in kinship care: Cross-informant reports. In: HEGAR, R.L., SCANNAPIECO, M. (éd.).

Kinship foster care: Policy practice and research. New York, NY: Oxford University Press; 1998. p. 272.

STERNBERG, R., CONWAY, B., KETON, J., BERNSTEIN, M. People's conceptions of intelligence. *Journal of Personality and Social Psychology*, 1981, 41, 37-55.

STERNBERG, R. J., FORSYTHE, G. B., HEDLUND, J., HORVATH, J. A., WAGNER, R. K., WILLIAMS, W. M., SNOOK, S., GRIGORENKO, E. L. *Practical intelligence in everyday life*. New York: Cambridge University Press, 2000, p. 304.

STERNBERG, R.J., GRIGORENKO, E., BUNDY, D.A. The Predictive Value of IQ. Merrill-Palmer Quarterly, 2001, 47(1), 1-41.

STERNBERG, R.J., GRIGORENKO, E.L., FERRARI, M., CLINKENBEARD, P. A triarchic analysis of an aptitude-treatment interaction. *European Journal of Psychological Assessment*, 1999, 15(1), 1-11.

SUCHMAN, L. *Plans and situated action*. Cambridge, NJ: Cambridge University Press, 1987, p. 220.

SUNDERLAND, A., HARRIS, J. E., BADDELEY, A. D. Do laboratory tests predict everyday memory? A neuropsychological study. *J. Verbal Learn. Verbal Behav,* 1983, 22, 341–357.

SZYMANOWICZ, A., FURNHAM, A. Gender Differences in self-estimates of general, mathematical, spatial and verbal intelligence: Four Meta analyses. *Learning and Individual Differences*, 2011, 21, 493-504.

TABACHNICK, B. G., FIDELL, L. S. Using Multivariate Statistics. Boston, MA: Pearson, 2013, p. 1072.

TARRANT, M.A., CORDELL, H.K. The effect of respondent characteristics on environmental attitude-behavior correspondence. *Environment and Behavior*, 1997, 29(5), 618-637.

TAUB, G. E., FLOYD, R. G., KEITH, T. Z., MCGREW, K. S. Effects of general and broad cognitive abilities on mathematics achievement. *School Psychology Quarterly*, 2008, 23(2), 187-198.

TERRIOT, K., GREGOIRE, J., LOARER, E. L'auto et l'hétéro-évaluation de l'intelligence sont-elles des alternatives à la mesure psychométrique classique de l'intelligence? Revue de 25 ans de recherche. *Bulletin de Psychologie*, 2017, 550(4), 275-289.

TERRIOT, K., VIGNOLI, E., LALLEMAND, N., BOURCIER, B. Validation du Self-Perception Profile for Adolescents de Harter auprès d'un échantillon d'adolescents français. *Pratiques Psychologiques*, 2017, 23, 359-376.

TIRRI, K., NOKELAINEN, P. Identification of multiple intelligences with the Multiple Intelligence Profiling Questionnaire III. *Psychology Science Quarterly*, 2008, 50(2), 206-221.

TOUSSAINT-THORIN, M., MARCHAL, F., PRADAT-DIEHL, P., CHEVIGNARD, M. Évaluation des fonctions exécutives des enfants présentant une dyspraxie développementale : comparaison d'une approche neuropsychologique classique et d'une approche plus écologique. *Annals of Physical and Rehabilitation Medicine*, 2011, 54(S1), e210.

TUPPER, D.E., CICERONE, K.D. The Neuropsychology of Everyday Life: Assessment and Basic Competencies. Boston: Kluwer Academic Publishers, 1990, p.348.

VALOIS, P., HOUSSEMAND, C., GERMAIN, S. & BELKACEM, A. An open source tool to verify the psychometric properties of an evaluation instrument. *Procedia Social and Behavioral Sciences*, 2011, 15, 552–556.

VAN DEN ENDE J., VERHULST F.C. Informant, gender and age differences in ratings of adolescent problem behavior. *European Child and Adolescent Psychiatry*, 2005, 14, 117–126.

VISSER, B., ASHTON, M., VERNON, P. What Makes You Think You're so Smart? Measured Abilities, Personality, and Sex Differences in Relation to Self-Estimates of Multiple Intelligences. *Journal of Individual Differences*, 2008, 29(1), 35-44.

VOYAZOPOULOUS, R., VANNETZEL, L., EYNARD, L-A. *L'examen psychologique de l'enfant et l'utilisation des mesures. Conférence de consensus*. Paris : Dunod, 2011, p. 472.

VRIGNAUD, P., CUVILLIER, B. *Hexa3D, outil d'évaluation des intérêts professionnels*. Paris : ECPA, 2006.

WAGNER, R. K., SUJAN, H., SUJAN, M., RASHOTTE, C. A., STERNBERG, R. J. Tacit knowledge in sales. In : STERNBERG, R. J., HORVATH J. A. (éds.). *Tacit knowledge in professional practice*. Mahwah, NJ: Lawrence Erlbaum Associates, 1999, p. 272.

WAGNER, R.K., STERNBERG, R.J. Practical intelligence in real-world pursuits: The role of tacit knowledge. *Journal of Personality and Social Psychology*, 1985, 49(2), 436-458.

WAGNER, R.K. & STERNBERG, R.J. *Tacit Knowledge Inventory for Managers: User manual*. San Antonio, TX : Psychological Corporation, 1991.

WALLON H. De l'acte à la pensée, Paris : Flammarion, 1942, p. 248.

WARRINER, G.K., MCDOUGALL, G.H.G., CLAXTON, J.D. Any Data or None at All? Living with Inaccuracies in Self-Reports of Residential Energy Consumption. *Environment and Behavior*, 1984, 16(4), 503-526.

WASCHBUSCH, D.A., DALEIDEN, E., DRABMAN, R.S. Are Parents Accurate Reporters of Their Child's Cognitive Abilities? *Journal of Psychopathology and Behavioral Assessment*, 2000, 22(1), 61-77.

WECHSLER, D. *Manuel d'interprétation et de cotation du WISC-IV*. Paris : ECPA, 2005.

WECHSLER, D. *Wechsler Memory Scale – Fourth Edition (WMS-IV).* San Antonio: Pearson, 2009.

WECHSLER, D. *WAIS-IV : Echelle d'Intelligence de Wechsler pour Adultes – Quatrième édition*. ECPA, Paris, 2011.

WECHSLER, D. *WPSSI-IV : Échelle d'Intelligence de Wechsler pour la période préscolaire et primaire – Quatrième édition*. Paris : ECPA, 2014.

WECHSLER, D. *Manuel d'interprétation et de cotation du WISC-V*. Paris : ECPA, 2016.

WECHSLER, D., NAGLIERI, J.A. *WNV : échelle non verbale d'intelligence.* Paris : ECPA, 2009.

WEINSTEIN, N. D. Unrealistic optimism about future life events. *Journal of personality and social psychology*, 1980, 39(5), 806.

WETZEL, E., HELL, B., PÄßLER, K. Comparison of different test construction strategies in the development of a gender fair interest inventory using verbs. *Journal of Career Assessment*, 2012, 20(1), 86-102.

WHETZEL, D.L., ROTENBERRY, P.F., MCDANIEL, M.A. In-basket Validity: A systematic review. *International Journal of Selection and Assessment*, 2014, 22(1), 62-79.

WILSON, B.A., ALDERMAN, N., BURGESS, P.W., EMSLIE, H., EVANS, J.J. *Behavioral Assessment of the Dysexecutive Syndrome*. Bury St. Edmunds : Thames Valley Test Company, 1996.

WILSON, B.A., GREENFIELD, E., CLARE, L., BADDELEY, A., COCKBURN, J., WATSON, P., TATE, R., SOPENA, S., NANNERY, R., CRAWFORD, J. *Rivermead Behavioural Memory Test – Troisième Edition. Paris: ECPA*, 2010.

WILSON,.B., COCKBURN, J., HALLIGAN, P. Development of a behavioral test of visuospatial neglect. *Arch Phys Med Rehabil*, 1987, 68(2), 98-102.

WINSLER, A., WALLACE, G.L. Behavior problems and social skills in preschool children: Parent-teacher agreement and relations with classroom observations. *Early Education & Development*, 2002, 13, 41-58.

WONDERLIC, E.F. *Wonderlic Personnel Test*. Libertyville: WPTI, 1992.

WOODCOCK, R.W. *Woodcock-Johnson Psycho-Educational Battery—Revised*. Allen, TX: DLM Teaching Resources, 1989.

WOODCOCK, R.W., MCGREW, K.S., MATHER, N. *Woodcock-Johnson III Tests of Cognitive Abilities*. Itasca, IL: Riverside, 2001.

ZALLA, T., PLASSIART, C., PILLON, B., GRAFMAN, J., SIRIGU, A. Action planning in a virtual context after prefrontal cortex damage. *Neuropsychologia*, 2001, 39(8), 759-770.

ZHANG, Y., GONG, Y. Self-estimated intelligence and its related factors. *Chinese Journal of Clinical Psychology*, 2001, 9, 193-195.

ZHANG, L., ABREU, B.C., SEALE, G.S., MASEL, B., CHRISTIANSEN, C.H., OTTENBACHER K.J. A virtual reality environment for evaluation of a daily living skill in brain injury rehabilitation: reliability and validity. *Archives of Physical Medicine and Rehabilitation*, 2003, 84(8), 1118-1124.

ZIMMERMAN, I. L., WOO-SAM, J. M. Review of the criterion-related validity of the WISC-III: the first five years. *Perceptual and Motor Skills*, 1997, 85(2), 531-546.

ZOCCOLOTTI, P., JUDICA, A. Functional evaluation of hemineglect by means of a semistructured scale: personal extrapersonal differentiation. *Neuropsychological Rehabilitation*, 1991, 1, 33-44.

ZUMBO, B. D. *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense, 1999, p. 57.

ZWICK, R., ERCIKAN, K. Analysis of differential item functioning in the NAEP history assessment. *Journal of Educational Measurement*, 1989, 26, 55–66.

Annexes

Annexe 1. Autorisation de diffusion

bulletin de psychologie

17 rue de la Sorbonne, 75230 PARIS Cedex 05, *en Sorbonne, esc. J, 3e étage*

Stéphane Laurens
Directeur de publication

24 octobre 2018

Katia Terriot
Inetop CNAM,
Paris, France.

Courriel de K. Terriot du 24 octobre 2018

Chère Collègue,

Nous vous donnons bien volontiers l'autorisation de publier, dans votre thèse, l'article écrit avec Grégoire J. et Loarer E. « Les évaluations auto et hétéro-rapportées sont-elles des alternatives à la mesure psychométrique classique de l'intelligence ? Revue de 25 ans de recherche », paru dans le *Bulletin de psychologie*, *70*, 4, n° 550, 2017, p. 275-289, sous réserve que vous ayez obtenu l'accord de chacun de vos co-auteurs.

Vous aurez l'amabilité de veiller à ce que soient mentionnées, dans votre publication, les formules d'usage (en notant les références de l'article publié initialement).

Veuillez agréer, chère Collègue, l'expression de nos sentiments les plus cordiaux.

S. Laurens

bulletin@bulletindepsychologie.net

Annexe 2. Poster ICAP

JUNE 26-30 29th International Congress of Applied Psychology ICAP 2018 MONTREAL

le cnam

Montréal, 26-30 juin 2018

29th International Congress of Applied Psychology

The development of a self-assessment questionnaire of adolescent cognitive abilities: QAE

Katia Terriot, Jacques Grégoire & Even Loarer

Centre de recherche sur le travail et le développement (CRTD), EA 4132 - Equipe « Psychologie de l'orientation », CNAM, Paris, France

katia.terriot@lecnam.net, jacques.gregoire@uclouvain.be, even.loarer@lecnam.net

Introduction

Since Binet and Simon (1908), intellectual evaluation has traditionally relied on the use of psychometric tests, such as the Wechsler scales which are still the most used in the world. Another way of assessing intelligence would be to use a self-reported measure (Eysenck & Eysenck, 1985). The self-evaluation of cognitive abilities makes it possible to evaluate the capacity perceived by the subject himself to perform certain tasks. Items do not refer explicitly to intelligence, but refer to the subject's skills in various cognitive tasks. Despite numerous advantages (inexpensive, quick to administer, flexible in their use, reduce the anxiety of the person evaluated, are a source of better cooperation), the questionnaires for self-evaluation of cognitive abilities are very few (Terriot, Grégoire & Loarer, 2017).

Moreover, while the most current model for reporting intelligence is the Cattell-Horn-Carroll model (CHC; Schneider & McGrew, 2012), a single intelligence self-assessment questionnaire is on the latter (Jacobs & Roodenburg, 2014) and it takes into account only three dimensions of the model. In addition, all these questionnaires are intended for an adult audience and no study, to our knowledge, has been conducted in adolescents while the question of the evaluation of cognitive abilities in this audience is significant especially in the field school.

Aim of the study

The main objective of this study is to **validate a cognitive capacity self-assessment questionnaire based on the Cattel-Horn-Carroll model** (CHC; Schneider & McGrew, 2012) with French adolescents.

Method

Participants

321 young people (49.2% boys and 50.8% girls) between the ages of 11 and 20 (M = 13.85, SD =1.26) completed the questionnaire.

Mesures

The QAE questionnaire includes **52 questions measuring six CHC model II level factors**: crystallized intelligence (Gc); fluid intelligence (Gf); visuospatial processing (Gv); Quantitative knowledge (Gq) short-term memory (Gsm); the treatment speed (Gs).

The **academic performance** (school results) at the end of the school year were also collected (general average in the 3rd quarter or the 2nd semester according to the class) for 87 of them.

Thirty-three of them also passed the **WISC-IV** (Wechsler, 2005), scale assessment of intelligence for children and teens from 6 to 16;11 years old.

For twenty five subjects, one of the parents completed the **Vineland II** (VABS-II; Sparrow, Cicchetti & Balla, 2015). It is a scale of evaluation of socio-adaptive behaviors.

Results

3 items (with low psychometric characteristics or witch have a significant effect following the differential analysis of item by sex) were removed to conduct the factor analysis.

The structure matrix partially reinforces the theoretical model: we find the Gc, Gsm and Gs factors well. The factors Gf and Gq form only one factor and Gv is subdivided into two factors: one that could be called Gv Construction (Gvc) and the other Gv Orientation (Gvo). We thus retain six factors: Gc, Gsm, Gs , Gf, Gv Construction, Gv Orientation. Nine items saturating as many other factors as belonging were removed from the QAE which now includes 40 items.

As shown in Table 1, **Cronbach's alpha ranged from .74 to .86 for the six factors**, suggesting good internal consistency across all scales.

Table 1 : Cronbach's alpha

	Gc	Gf	Gvc	Gvo	Gsm	Gs
Cronbach' alpha	.83	.86	.75	.75	.75	.74

Relationships between dimensions

All factors are correlated with each other. Correlations range from .04 (between Gf and Gsm) to .31 (between Gc and Gf).

Evidence of concurrent and differential validity.

Differential validity according to gender

There are significant differences between boys and girls in Gf, Gvc and Gvo factors. It is on the Gf and Gvo factors that we find the largest effect size (respectively $\eta 2 = 0.06$ and $\eta 2 = 0.07$) with a difference in favor of boys.

Concurrent validity

Correlations with school results

We observe positive, moderate significant correlations between the results in the 3rd quarter (a few months after having passed the QAE) and Gf (r = .45) and then Gc (r = .26). The correlations are almost zero with the other factors. Moreover, the factor Gf predicts academic success and explains 18.9% of the variance of the school results.

Correlations with WISC-IV

Only Gf correlates positively and significantly with the subtests (r = .54 with Cubes, r = .38 with Similarities, r = .36 with Vocabulary and r = .35 with Sequence-letter digits) and the indices with WISC-IV (r = .53 with IRP, r = .37 with IVT and r = .42 with QIT).

Correlations with Vineland-II

Gs factor has the strongest relationship with Vineland-II and the only factor for which there is a positive, significant, and relatively strong correlation with the composite score (r = .69), the domain of Socialization (r = .60) and Daily Life (r = .70).

Discussion

The structure matrix partially reinforces the initial theoretical model. Teenagers are able to produce differentiated answers according to the domains. With regard to the **distinction between Gvc and Gvo**, this would be in line with other studies (Hergarty, Montello, Richardson, Ishikawa & Lovelace, 2006) that show a **distinction between reduced spatial skills (visualization, mental rotation, labyrinths, puzzles) and broad spatial skills that include the environment (orientation tasks in space, navigation).** Schneider and McGrew (2012) state that many aspects of Gv have not been explored yet and that the current CHC model does not take into account spatial orientation capabilities.

Analyzes show good psychometric qualities and promising results for proof of predictive validity. Thus, although presenting some interesting results, the items of the questionnaire could be further improved and further studies of validity are necessary, in particular to confirm its structure.

Références

Binet, A., Simon, T. (1908). Le développement de l'intelligence chez l'enfant. *L'Année psychologique*, 14, 191-244.

Eysenck, H.J., & Eysenck, M.W. (1985). *Personality and individual differences: A natural science approach*. New York: Plenum.

Jacobs, K., Roodenburg, J. (2014). The development and validation of the Self-Report Measure of Cognitive Abilities: A multitrait–multimethod study. *Intelligence*, 42, 5-21.

Hegarty, M., Montello, D.R., Richardson, A.E., Ishikawa, T., Lovelace, K. (2006). Spatial Abilities at Different Scales:Individual differences inaptitude-test performance and spatial-layout learning. Intelligence, 34, 151– 176.

Schneider, W. J., & McGrew, K. (2012). The Cattell-Horn-Carroll model of intelligence. In, D. Flanagan & P. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (3rd ed.) (p. 99-144). New York: Guilford.

Sparrow, S.S., Cicchetti, D.V., Balla, D.A. (2015). *Vineland-II, Echelle d'évaluation du comportement socio-adaptatif de Vineland, Adaptation française*. Paris : ECPA.

Terriot, K., Grégoire, J., Loarer, E. (2017). L'auto et l'hétéro-évaluation de l'intelligence sont-elles des alternatives à la mesure psychométrique classique de l'intelligence? Revue de 25 ans de recherche. *Bulletin de Psychologie*, 4, 275-289.

Wechsler, D. (2005). *Manuel d'interprétation et de cotation du WISC-IV*. Paris : ECPA.

Annexe 3. Questionnaire initial (QAE), version 52 items

Identifiant :	Date de passation :
Date de naissance :	Sexe :
Age :	Classe :

Indiquez, pour chaque phrase, si cette activité vous est beaucoup plus difficile, plus difficile, pareil, plus facile ou beaucoup plus facile que pour les autres jeunes de votre âge (entourez le chiffre correspondant à votre réponse).

n° item	Phrases	Beaucoup plus difficile	Plus difficile	Pareil	Plus facile	Beaucoup plus facile
1	Comprendre les informations à la télévision	1	2	3	4	5
2	Assembler les pièces d'un puzzle	1	2	3	4	5
3	Retenir les éléments importants d'une conversation	1	2	3	4	5
4	Imaginer de nouvelles utilisations d'un objet familier	1	2	3	4	5
5	Construire un objet à partir d'un plan	1	2	3	4	5
6	Recopier rapidement un texte	1	2	3	4	5
7	Comprendre les figures ou schémas visuels qui accompagnent les textes	1	2	3	4	5
8	Retenir des choses importantes entendues en classe	1	2	3	4	5
9	Calculer mentalement	1	2	3	4	5
10	Trouver le mot exact pour exprimer ses pensées	1	2	3	4	5
11	M'orienter dans une ville	1	2	3	4	5
12	Utiliser les transports en commun dans une ville que je ne connais pas	1	2	3	4	5
13	Comprendre les schémas contenus dans les notices d'objets à assembler	1	2	3	4	5
14	Comprendre tous les mots d'un texte de journal	1	2	3	4	5
15	Trouver mon chemin dans la forêt à l'aide d'une carte	1	2	3	4	5
16	Connaître des mots que les autres ne connaissent pas	1	2	3	4	5
17	Me rappeler de conversations intéressantes que j'ai entendues	1	2	3	4	5

1

n° item	Phrases	Beaucoup plus difficile	Plus difficile	Pareil	Plus facile	Beaucoup plus facile
18	Trouver un ami dans une foule	1	2	3	4	5
19	Faire rapidement mon travail scolaire	1	2	3	4	5
20	Jouer à des jeux vidéo	1	2	3	4	5
21	Construire une maquette en lego	1	2	3	4	5
22	Résoudre des problèmes logiques	1	2	3	4	5
23	Utiliser le mot correct pour désigner un objet	1	2	3	4	5
24	Estimer si des bagages pourront rentrer dans un coffre de voiture	1	2	3	4	5
25	Estimer la grandeur d'une pièce (en m2)	1	2	3	4	5
26	Utiliser les mots appropriés dans les rédactions	1	2	3	4	5
27	Imaginer un grand nombre de solutions possibles à un problème	1	2	3	4	5
28	Résoudre des casses têtes	1	2	3	4	5
29	Trouver mon chemin dans un bâtiment inconnu	1	2	3	4	5
30	Retenir des numéros de téléphone	1	2	3	4	5
31	Retenir le nom de personnes que je viens de rencontrer	1	2	3	4	5
32	Trouver un endroit précis sur une carte	1	2	3	4	5
33	Calculer des pourcentages	1	2	3	4	5
34	Comprendre des graphiques	1	2	3	4	5
35	Rédiger un texte	1	2	3	4	5
36	Retenir des dates de rendez-vous	1	2	3	4	5
37	Recopier rapidement ce qui est écrit au tableau	1	2	3	4	5
38	Organiser l'emplacement des meubles dans une pièce	1	2	3	4	5

n° item	Phrases	Beaucoup plus difficile	Plus difficile	Pareil	Plus facile	Beaucoup plus facile
39	Utiliser une information, apprise précédemment, dans un nouveau contexte	1	2	3	4	5
40	Retenir une liste de courses	1	2	3	4	5
41	Faire mentalement des soustractions	1	2	3	4	5
42	Retenir des consignes données oralement	1	2	3	4	5
43	Exprimer avec précision mes idées	1	2	3	4	5
44	Dessiner un plan	1	2	3	4	5
45	Calculer mentalement le prix total pour l'achat de plusieurs produits	1	2	3	4	5
46	Retenir une poésie	1	2	3	4	5
47	Comprendre des consignes données oralement	1	2	3	4	5
48	Trouver des objets ou des documents précis dans une pièce en désordre	1	2	3	4	5
49	Résoudre des énigmes	1	2	3	4	5
50	Ranger rapidement mes affaires	1	2	3	4	5
51	Estimer la distance entre deux endroits (en mètres ou en kilomètres)	1	2	3	4	5
52	Démonter, puis remonter un objet ou un appareil	1	2	3	4	5

Annexe 4. Questionnaire enseignant.e

Identifiant élève :		Date de passation :
Matière enseignée :		
Sexe :		
Sexe de l'élève :	Date de naissance élève :	Classe :

Indiquez, pour chaque phrase, si l'activité est, pour cet élève, beaucoup plus difficile, plus difficile, pareille, plus facile ou beaucoup plus facile que pour les autres élèves de son âge (entourez le chiffre correspondant à votre réponse).

n° item	Phrases	Beaucoup plus difficile	Plus difficile	Pareil	Plus facile	Beaucoup plus facile
1	Recopier rapidement un texte	1	2	3	4	5
2	Calculer mentalement	1	2	3	4	5
3	Trouver le mot exact pour exprimer ses pensées	1	2	3	4	5
4	Comprendre tous les mots d'un texte	1	2	3	4	5
5	Connaître des mots que les autres ne connaissent pas	1	2	3	4	5
6	Faire rapidement son travail scolaire	1	2	3	4	5
7	Résoudre des problèmes logiques	1	2	3	4	5
8	Utiliser le mot correct pour désigner un objet	1	2	3	4	5
9	Utiliser les mots appropriés dans les rédactions	1	2	3	4	5
10	Imaginer un grand nombre de solutions possibles à un problème	1	2	3	4	5
11	Calculer des pourcentages	1	2	3	4	5
12	Rédiger un texte	1	2	3	4	5
13	Recopier rapidement ce qui est écrit au tableau	1	2	3	4	5
14	Faire mentalement des soustractions	1	2	3	4	5
15	Exprimer avec précision ses idées	1	2	3	4	5
16	Comprendre des consignes données oralement	1	2	3	4	5
17	Ranger rapidement ses affaires	1	2	3	4	5

Annexe 5. Cahier de passation de l'échelle d'évaluation contextualisée

Pseudonyme ou identifiant :
Date de passation :
Date de naissance :
Age :
Sexe : ☐ fille ☐ garçon
Classe :
☐ Droitier ☐ Gaucher
Année(s) de retard ☐ Aucune ☐ 1 an ☐ 2 ans ☐ 3 ans ou plus
Année(s) d'avance ☐ Aucune ☐ 1 an ☐ 2 ans ☐ 3 ans ou plus
Profession du père ou représentant légal :
Profession de la mère ou représentante légale :

Diplôme le plus élevé obtenu par le père (cochez) :

aucun diplôme	
certificat d'études primaires	
BEPC, brevet des collèges	
CAP ou BEP	
Tout baccalauréat ou Brevet professionnel	
diplôme de niveau bac + 2	
diplôme de niveau supérieur à bac + 2	

** Source INSEE 2008*

Profession de la mère ou représentante légale :

Diplôme le plus élevé obtenu par la mère (cochez) :

aucun diplôme	
certificat d'études primaires	
BEPC, brevet des collèges	
CAP ou BEP	
Tout baccalauréat ou Brevet professionnel	
diplôme de niveau bac + 2	
diplôme de niveau supérieur à bac + 2	

** Source INSEE 2008*

« Diagnostic » ou « difficultés connues » (dyslexie, haut potentiel….) :
Département :
Lunettes : ☐ oui ☐ non

Durée de passation :

1

Résultats :

	Note brute	Intelligence cristallisée	Intelligence fluide	Traitement visuel	Mémoire à court terme	Vitesse de traitement
1. Différences	…/28					
2. Numéros de téléphone	…/40					
3. Chaussettes	…/17					
4. Connaissances générales	…/66					
5. Raisonnement	…/20					
6. Lexique	…/72					
7. Majuscules	…/36					

Consigne générale

Nous allons faire ensemble plusieurs exercices en lien avec la vie quotidienne. Ces exercices évaluent des capacités améliorables dans différents domaines : verbal, non verbal, la mémoire, la vitesse, le raisonnement. Cela dure environ 1h00. Certaines choses te/vous paraitront faciles, d'autres plus difficiles. As-tu/avez-vous des questions avant qu'on commence ?

Bien expliciter lors des exemples. Ne pas poursuivre si les exemples ne sont pas bien compris

Epreuve 1 : Différences

Matériel : tableaux (6 tableaux originaux + 6 tableaux réalisés)

Consigne « ton/votre amie adore copier/reproduire des tableaux. Elle te/vous montre le tableau original et ce qu'elle a fait et te/vous demande ton/votre avis. Elle l'a déjà montré à une autre personne qui a relevé à chaque fois des erreurs. Elle vous demande de lui dire si vous en voyez également. Indique-lui/Indiquez-lui pour qu'elle puisse les corriger ».

Temps limité pour chaque tableau et chronométrer. S'il/elle dépasse le temps, passer au tableau suivant.

	Temps en secondes	Solution	Réponses données (entourer les réponses)	Score
Exemple		3 erreurs à trouver		/3
Item 1 Temps limité à 1min30		3 erreurs à trouver		.../3
Item 2 Temps limité à 2min30		5 erreurs à trouver		.../5

3

Item 3 Temps limité à 3 min		6 erreurs à trouver 		.../6
Item 4 Temps limité à 3 min		6 erreurs à trouver 		.../6
Item 5 Temps limité à 3min30		8 erreurs à trouver 		.../8
Total	1 point par bonne réponse		/28	

Commentaires/notes

_____
_____
_____
_____
_____
_____
_____

Epreuve 2 : Mémoriser un numéro de téléphone (10 éléments)
Matériel : aucun

Consigne : Vous allez entendre un numéro de téléphone que vous allez devoir mémoriser pour me le redonner

Exemple :
Etes-vous prêt(e) ?
Essai 1 : 01-11-22-33-44 (dire « onze » et non « un un »)
Essai 2 : 01-11-22-33-55

Cotation :
1 point par série correctement placée ; 0,5 point pour une série (deux chiffres) exacte mais incorrectement placée et 0 point pour toute série fausse. 5 points au total

Items

Je vais vous dire un numéro de téléphone que vous allez devoir mémoriser pour me le redonner. Etes-vous prêt(e) ?

		Réponse donnée	Cotation			Total
			Série (2 chiffres) correcte et bien placée (1 point)	Série (2 chiffres) exacte mais incorrectement placée (0,5 points)	Série fausse (0 point)	
Item 1	01-11-22-33-48		/5	..../2,5		/5
	01-11-22-33-84		/5	..../2,5		/5
Item 2	01-11-22-52-97		/5	..../2,5		/5
	01-11-22-14-08		/5	..../2,5		/5
Item 3	01-11-17-28-54		/5	..../2,5		/5
	01-11-53-74-28		/5	..../2,5		/5
Item 4	01-36-28-47-59		/5	..../2,5		/5
	01-47-25-98-63		/5	..../2,5		/5
Total			.../40	..../20		..../40

Commentaires/notes

_____

_____

_____

_____

_____

Epreuve 3 : chaussettes

Matériel : feuille A4 exemple chaussettes + feuille A4 chaussettes

Consigne : « On vous demande chez vous de trier rapidement les chaussettes et de former les paires. Les voici. Pouvez-vous m'indiquez celles qui vont ensemble en les reliant grâce au stylo le plus vite possible ? Certaines restent seules. Il faut faire le plus vite possible »

Exemple

Item

Cotation : un point par paire et temps mis – 1 point par erreur fausse
Temps limité à 120 secondes

Nombre de paires repérées	Cotation	Erreur	Total (total sur 17 – erreurs)	Temps réalisé
	…/17 (paires) -	….		

Correction

6

Commentaires/notes

_____
_____
_____
_____
_____
_____

Epreuve 4 : connaissances générales
Matériel : fichier audio (connaissances) + fichier écrit

Consigne « Vous allez entendre plusieurs extraits de conversations entre jeunes. Comme vous pourriez le faire dans la vie de tous les jours, réagissez à chaque fois que ce que vous entendez vous parait inexact, faux, non approprié. Vous devez à la fois dire que la personne se trompe et la corriger si possible. Il ne faut pas avoir peur de le dire. J'arrêterai la conversation. La scène se passe début octobre»

Arrêter la lecture du fichier dès que le sujet réagit. Il faut ensuite faire un retour en arrière sur le fichier (si la lecture du fichier MP3 est trop difficile : lire à haute voix). Il est possible de répéter ce qu'ont dit les jeunes si le sujet dit ne pas avoir bien entendu ou compris.
Et quand le jeune a repéré une erreur, posez la question correspondant à la réponse (ex : quelle est la capitale de la Grèce ? Qui a écrit le tour du monde en 80 jours ? Qui a peint la Joconde ?...). Notez la réponse du sujet, NSP quand il ne sait pas.

Une fois la lecture complète terminée, revenir sur les erreurs non repérées en disant « à un moment il/elle dit cela, qu'en penses-tu ? » Noter les réponses

Cotation : 1 point pour chaque erreur repérée ; 2 points pour une bonne réponse donnée

Exemples :

Items	Erreur repérée (cocher)	Réponse attendue
Ours polaires qui vivent en France		Pôle Nord
Le corbeau et le renard de Jean Rivière		Jean de La Fontaine
Egypte Amérique		Afrique

Items

N°	Items	Erreur repérée	Réponse attendue	Réponse donnée	Cotation	Total
1	Ankara, capitale de la Grèce	…/1	Athènes		…/2	…/3
2	Le tour du monde en 80 jours de Victor Hugo	…/1	Jules Verne		…/2	…/3

3	la Joconde de Raphaël	…/1	Léonard de Vinci		…/2	…/3
4	Lire en Italie	…/1	Euro		…/2	…/3
5	1889 : prise de la Bastille	…/1	1789		…/2	…/3
6	125 : bataille d'Alésia	…/1	entre 50 et 60 avant JC, 52 pour être exact		…/2	…/3
7	Copernic : découverte de la loi de la relativité	…/1	La terre tourne autour du soleil		…/2	…/3
8	Le Président de la République est élu 7 ans	…/1	5 ans		…/2	…/3
9	Equipe qui a gagné la coupe de monde de football en 1998 : le Brésil	…/1	La France		…/2	…/3
10	Dans quel sport marque-t-on des essais : au football américain	…/1	Rugby		…/2	…/3
11	au tour de France, quelle est la couleur du maillot du vainqueur ? rose	…/1	jaune		…/2	…/3
12	si on avait plus de jours que les 364 jours de l'année	…/1	365 ou 366		…/2	…/3
13	Sur 50 semaines de l'année	…/1	52		…/2	…/3
14	le 01er mai ; pourquoi c'est un jour férié déjà ? ça doit être une fête religieuse ou un armistice également	…/1	Fête du travail		…/2	…/3
15	on aura moins de vacances quand on travaillera. Si ça ne change pas d'ici là, on n'aura plus que 6 semaines de congés payés	…/1	5 semaines		…/2	…/3
16	Il faut que l'eau bout, donc qu'elle atteigne 200 degrés	…/1	100 degrés		…/2	…/3
17	J'utilise un baromètre pour	…/1	thermomètre		…/2	…/3

	mesurer la température ?					
18	j'ai un rendez-vous médical avant de partir pour mes yeux. ah tu vas voir le dermatologue ?	…/1	Ophtalmologiste / ophtalmologue		…/2	…/3
19	Carte vitale pour payer le médecin	…/1	carte d'assurance maladie permettant de justifier les droits du titulaire de la carte		…/2	…/3
20	Le timbre sert juste à savoir d'où on l'a envoyée	…/1	Sert à payer la lettre		…/2	…/3
21	Travailler à 15 ans	…/1	16 ans		…/2	…/3
22	Titanic avec Tom Cruise	…/1	Léonardo Di Caprio		…/2	…/3
	Total	…/22			…/44	…/66

Une fois la lecture complète terminée, revenir sur les items non repérés en disant « à un moment il/elle dit cela, qu'en penses-tu ? » Noter les réponses. Coter 2 (bonne réponse) ou 0 (réponse fausse)

Items (n°)	Réponse donnée	Cotation
		…/2
		…/2
		…/2
		…/2
		…/2
		…/2
		…/2
		…/2
		…/2
		…/2
		…/2
		…/2
Total		

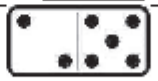
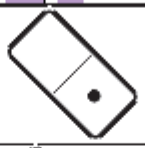
Total des deux : /66
Additionnez les deux totaux et reporter dans le tableau p.2

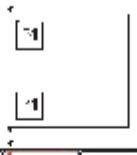
Commentaires/notes

_____
_____
_____

Epreuve 5 : raisonnement
Matériel : planches « raisonnement» + feuille blanche et crayons ou feutres
Consigne générale: « vous allez voir, observer plusieurs problèmes que vous devrez résoudre. Vous avez le droit de dessiner si vous le souhaitez »

Cotation : 1 point par bonne réponse

Items	Consigne	Réponse attendue	Réponse donnée	Cotation
Exemple 1	Un carreleur pose du carrelage dans un immeuble. Quel carrelage doit-il choisir pour poursuivre ?			X
Exemple 2	Vous faites une partie de dominos avec un ami. La règle du jeu est la suivante : il pose quelques dominos et vous devez deviner comment compléter.			X
Exemple 3	Vous faites une partie de jeu de cartes avec une amie. La règle du jeu est la suivante : elle pose quelques cartes et vous devez deviner comment compléter.			X
Exemple 4	Vous avez un nouveau téléphone. Vous rentrez un code avec une certaine logique. Quel est le chiffre qui suit ?	4		X
1. Mosaïque				…/1
2. Domino 2	Idem			…/1
3. Domino 3	Idem			…/1
4. Domino 4	Idem			…/1
5. Frise de décoration	vous aidez un voisin à terminer la frise de décoration dans la chambre. Comment continuez-vous ?			…/1

6. Deux robots	Vous observez le fonctionnement de deux robots de nettoyage de sol qui se déplacent sur le sol de cette manière. Où se situeront-ils ?			…/1
7.	Idem			…/1
8. Code	Un ami habite dans un immeuble qui possède un code d'entrée, qui change chaque mois. Voici les 3 premiers codes de l'année qui suivent la même logique. Quel sera le prochain code ?	E15		…/1
9. Jeu de cartes	Vous jouez à un nouveau jeu de cartes avec un ami. Il pose trois ou quatre cartes de suite selon une certaine logique et vous devez deviner la carte suivante	5 de trèfle		…/1
10. Jeu de cartes	Idem	4 de coeur		…/1
11. Code de casier	Votre voisin possède un casier au lycée. Il change de code tous les mois selon une règle précise. Il peut choisir des chiffres de 0 à 9. Voici les codes des trois premiers mois. Quel sera le code du mois d'avril ?	4/8/2/6 Ou 5/0/5/0		…/1
12. Classement	M. Albert est assistant de direction dans une entreprise. Il numérote ses factures selon une certaine logique. Le 03 décembre 2014, il établit 3 factures qu'il numérote ainsi (montrer feuille stimuli), le 05 décembre, il numérote la facture ainsi. Le 07 décembre, il fait 4 factures. Comment vat-il numéroter la 3ème facture ?	3/07-12-2014		…/1
13. Alarme	Votre voisin vient d'installer une alarme chez lui. Il doit créer un code de 5 chiffres (composés de 1 ou 2 chiffres). Il compose les 4 premiers selon une certaine logique. Quel sera le 5ème ?	22		…/1
14. Euromilli	Votre ami joue à l'euromillion selon une certaine logique. Que	48		…/1

ons	va-t-il jouer pour le chiffre numéro 5 ?			
15. Code coffre fort	Un directeur de banque change de code tous les jours pour des raisons de sécurité, selon une certaine logique. Le code est composé de lettres et de chiffres. Voici les codes de lundi à jeudi. Quel sera celui de vendredi ?	C/24		…/1
16. Joueur de loto	Un ami joue au loto chaque semaine de manière logique comme suit 7 13 20 28 37 Quel est d'après vous le numéro complémentaire ?	47		…/1
17. Télécom mande	la télécommande de votre téléviseur est déréglée. Quand on appuie sur un : on tombe sur la chaîne 7 ; quand on appuie sur 2 : on tombe sur la chaine 8… comme noté sur le tableau. Vous appuyer sur le 23, sur quelle chaine tombez-vous ?	29		…/1
18. Code parental	Votre famille a installé un code parental sur la télévision pour que vous ne puissiez pas regarder des chaînes interdites aux moins de 18 ans. Ils le changent chaque mois selon une certaine logique. Vous essayer de trouver le code de ce mois-ci. C'est un code à quatre chiffres	1205		…/1
19. Code journal intime	Une voisine possède un journal intime avec un cadenas à code de 3 chiffres. Elle le change chaque jour, selon une certaine logique, pour éviter qu'un membre de sa famille ne trouve le code et lise son journal. Voici les codes de lundi à jeudi. Quel sera celui de vendredi ?	191		…/1
20. Codes Internet	Une amie rentre des codes composés de chiffres comme mots de passe sur Internet. Elle suit une certaine logique. Voici	22		…/1

	les codes des 5 premiers. Quel est le code du dernier ?			
Total				…/20

Commentaires/notes

_____
_____
_____
_____
_____
_____

Epreuve 6 : lexique

Matériel : fichier papier Epreuve 6 lexique

Consigne Je vais vous lire une histoire racontée par une personne ayant des difficultés pour s'exprimer : elle utilise parfois des mots qui sont faux, ou à l'opposé de ce qu'il faut dire ou parfois pas assez précis ; ou encore, la personne ne trouve pas le mot du tout.

Comme vous pourriez le faire dans la vie de tous les jours, réagissez à chaque fois que ce que vous entendez vous parait inexact, faux, non approprié ou pas tout à fait adapté ou aidez-la à trouver le bon mot. Vous devez à la fois dire que la personne se trompe et la corriger si possible. N'hésitez pas à m'interrompre pendant que je lis. Je peux aussi répéter si vous n'avez pas entendu ou compris. On va déjà commencer par un exemple. Dans l'exemple, la personne parle de ce qu'elle a fait hier. »

Arrêts et répétitions possibles

Une fois la lecture complète terminée, revenir sur les items non repérés en disant « à un moment il/elle dit cela, qu'en penses-tu ? » Noter les réponses

Cotation : pour les erreurs de langage, cotez 1 point pour chaque erreur repérée +2 points pour une correction adéquate (accepter des mots adéquats même si ce ne sont pas ceux demandés. Les noter).

Pour les mots à trouver, cotez 3 points pour un bon mot ou 0

Items

Items	Erreur repérée (cochez d'une croix)	Cotation	Réponse donnée (si différente de la bonne réponse)	Cotation	Total
Exemple : fatigué		X		X	X
Exemple : contrôle		X		X	X
Exemple : main		X		X	X
Exemple : fleurs		X		X	X
1. comédie	/	/		…/2	…/3
2. habitude		…/1		…/2	…/3
3. nuage		…/1		…/2	…/3
4. fidélité	/	/		/	…/3
5. Chemin		…/1		…/2	…/3
6. Capital/épargne	/	/		/	…/3
7. Dispute	/	/		/	…/3

8. Flacons		…/1		…/2	…/3
9. Exceptionnel		…/1		…/2	…/3
10. Aventure	/	/		/	…/3
11. Aride		…/1		…/2	…/3
12. A la baguette		…/1		…/2	…/3
13. Distractions		…/1		…/2	…/3
14. Adversaire		…/1		…/2	…/3
15. Intention		…/1		…/2	…/3
16. Barrage		…/1		…/2	…/3
17. Célèbre		…/1		…/2	…/3
18. Fiable		…/1		…/2	…/3
19. Flûte		…/1		…/2	…/3
20. Crissement		…/1		…/2	…/3
21. Crinière		…/1		…/2	…/3
22. Conviction		…/1		…/2	…/3
23. Accalmie		…/1		…/2	…/3
24. Déchéance		…/1		…/2	…/3
Total		…/24		…/48	…/72

Une fois la lecture complète terminée, revenir sur les items non repérés en disant « à un moment il/elle dit cela, qu'en penses-tu ? » Noter les réponses. Coter 2 points si c'est une bonne réponse, sinon 0

Items (n°)	Réponse attendue	Réponse donnée	Cotation
			…/2
			…/2
			…/2
			…/2
			…/2
			…/2
Total			

Total : …. ./72

Additionnez les deux totaux et reporter dans le tableau p.2

<u>Commentaires/notes</u>

_____

_____

_____

_____

_____

_____

Epreuve 7 : majuscules

Matériel : texte des majuscules + feuille de correction

Consigne : « Un voisin qui est en 3ème vous demande de relire sa rédaction en français. Il l'a tapée sur ordinateur. Il a oublié certaines fois de mettre une majuscule en début de phrase et d'autres fois il a mis une majuscule alors que cela n'était pas nécessaire. Il est très pressé de rentrer et n'a que 2 minutes. Corrigez-le le plus vite possible. Vous n'avez pas besoin de corriger le reste, uniquement les majuscules.

Exemple : 6 corrections à apporter

Item : même consigne. Chronométrer. **120 s maximum**

Cotation : nombre de majuscules corrigées (sur 36) - erreurs et temps mis.

Nombre de majuscules corrigées	Erreurs	Total (nombre majuscules correctes – erreurs)	Temps réalisé
…/36			

Commentaires/notes

15

Annexe 6. Matériel pour l'échelle d'évaluation contextualisée

Epreuve 1 : différences

Item exemple

Modèle

Copie

Item 1

Modèle

Copie

Item 2

Modèle

Copie

Item 3

Modèle

Copie

Item 4

Modèle

Copie

Item 5

Modèle

Copie

Epreuve 3 : chaussettes

Exemple

Item

Epreuve 4 : connaissances générales

Item d'exemple
A : « salut B. Comment tu vas ? Tu as fait quoi ce week-end ? »
B : « j'ai regardé des reportages sur les animaux. Ils parlaient des **ours polaires qui vivent en France.** C'était superbe à regarder et j'ai appris pleins de choses. »
A : « moi je me suis bien reposée. J'ai essayé de nouvelles recettes de gâteaux. J'ai relu **Le corbeau et le renard de Jean Rivière** et le Médecin Malgré lui de Molière. Et j'ai aussi regardé une émission sur **l'Egypte en Amérique** ».

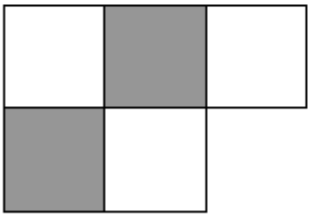
Item
A : « c'est bientôt les vacances. Vous partez ? »
B : « moi, je pars avec mes parents en Grèce, dans la capitale.
A : "Ah, tu vas à **Ankara** ? Et toi C?
C : « moi je reste ici. Je vais en profiter pour lire Le tour du monde en 80 jours, de **Victor Hugo** ou Harry Potter de J.K. Rowling ou des BD genre Tintin, ou encore aller au musée pour voir la Joconde de **Raphaël** ou les Nymphéas de Monet. Et toi A?
A : « moi je vais voir le Kilimandjaro. Je suis super heureux(se). C'était mon rêve. »
C : « vous avez beaucoup de chance de pouvoir voyager. Moi, je ne suis partie que deux fois : une pour visiter Big Ben et l'autre fois visiter le Colisée il y a un an »
B : « c'est déjà bien d'être allée à Londres et Rome. Par contre, l'inconvénient, c'est que tu as dû faire du change pour payer avec **la Lire** pour l'Italie et la livre Sterling en Angleterre. Et puis tu vas pouvoir réviser l'histoire et l'éducation civique. Je vous rappelle qu'on a un contrôle à la rentrée. Je ne suis pas très à l'aise. Quel est le programme déjà ?
G : « les inventeurs, les grandes dates, les grands hommes de l'histoire et l'éducation Civique. D'ailleurs, on pourrait s'entrainer un peu : partons sur les grandes dates : **1889** : prise de la bastille ; **125** : bataille d'Alésia ; 1515 : bataille de Marignan, 1914 : début de la première guerre mondiale
A : « bon il faut que je révise vraiment car je ne sais pas et je ne me rappelle de rien. Ah si il me semble qu'on doit aussi parler de Christophe Colomb. Et il y a aussi Copernic »
C : « celui qui a découvert la **loi de la relativité**. Et il y a aussi tout le cours en Education civique mais ça me parait plus facile »
A : « Alors, combien de temps est élu le Président de la République ? »
G : « bah **7 ans**.
C : « il va falloir qu'on s'y mette car on va bientôt voter. Il faudrait peut-être savoir de quoi on parle, non ? Je préférerais avoir un contrôle en sport. J'aurais 20 sans problème. Si je vous dis par exemple qui a gagné la coupe du monde de football en 1998 ?
B : « **Le Brésil** »
C : « dans quel sport marque-t-on des essais ?
A : « **en football américain** »
C : « Et combien de joueurs compte une équipe de basket-ball ?
A : « ah ça je sais 5 »
C : « et au tour de France, quelle est la couleur du maillot du vainqueur ? »
B : « **rose**. Et pourquoi on n'a jamais de contrôle en musique. On aurait tous des bonnes notes ».
A « cela dépend des questions qu'on te poserait. »

B : « si on avait plus de jours que les **364 jours** de l'année, on en saurait plus. Il n'y a pas assez d'heures dans une journée. 24h, c'est trop court. On peut pas en profiter. Ca passe trop vite. »
A : « heureusement qu'on a les jours fériés, les week-end et les vacances pour se reposer. Sur **50 semaines** de l'année, on a environ 16 semaines de vacances et 6/7 jours fériés comme le 14 juillet (prise de la Bastille), le 11 novembre (armistice), le 08 mai (armistice), le 01er mai ; pourquoi c'est un jour férié déjà ? »
C : « ça **doit être une fête religieuse ou un armistice également** »
B : « on aura moins de vacances quand on travaillera. Si ça ne change pas d'ici là, on n'aura plus que **6 semaines de congés payés** »
C : « Comme je pars pas et que mon père travaille, je vais devoir rester seul pendant les deux semaines de vacances et je vais devoir me faire à manger. Je l'ai jamais fait. »
A : « tu sais comment faire des pâtes, non ? Il faut chauffer de l'eau et qu'elle bout, donc qu'elle atteigne **200 degrés** »
C : « comment je sais si ça bout ? J'utilise **un baromètre** pour mesurer la température ?
A : « non tu vas voir des petites bulles »
A : « moi j'ai un rendez-vous médical avant de partir pour mes yeux »
C « ah tu vas voir le **dermatologue** ? Il ne faudra pas que tu oublies la carte vitale. Ca m'est arrivé la dernière fois et c'est embêtant. »
A : « la quoi ? »
C : « **la carte pour payer le médecin** »
A : « ah je savais pas »
C : « j'espère que je vais pas tomber malade. A chaque fois, c'est la même chose, je tombe malade pendants les vacances alors que je suis en pleine forme pendant les périodes de cours »
C : « et vous m'écrirez ? une petite carte postale ? et n'oublie pas de mettre un timbre cette fois-ci ! »
A : « Et alors, c'est pas grave, la lettre arrive quand même. **Le timbre sert juste à savoir d'où on l'a envoyée** »
B : « moi, je veux pas user du papier»
C : «Tu es un écolo en fait toi. Ah oui et je vais essayer aussi de me chercher un petit boulot déclaré pour les vacances pour me faire un peu d'argent. **A 15 ans**, on m'a dit que c'était possible. En fait, j'ai un programme bien chargé quand j'y pense !!! »
B : « Tu vas pouvoir aussi te faire des séances de cinéma en regardant des films cultes »
C : « comme quoi ? »
B : « bah par exemple Titanic avec **Tom Cruise** ; Grease avec Travolta,
C : « ce sont des vieux films…Ca me tente pas vraiment. Je préfère des films d'action »
A : « bon allez faut que je vous laisse. Je vais aller préparer ma valise. »
B : « ouais moi aussi. Allez bonnes vacances.
C : « bonnes vacances et m'oubliez pas ! »

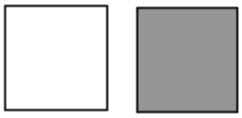
Epreuve 5 : raisonnement

Items exemple

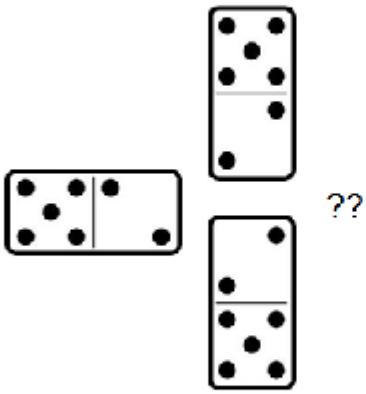
Exemple 1



Quel carrelage doit-il choisir pour poursuivre ?

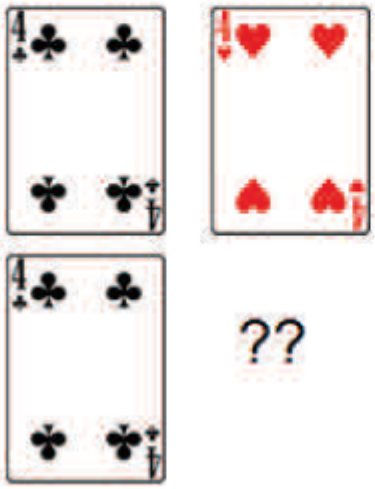


Exemple 2



1

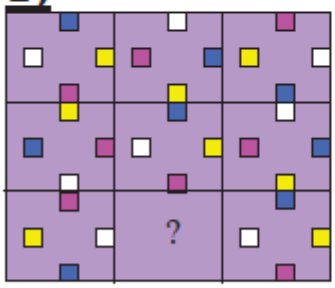
Exemple 3



Exemple 4

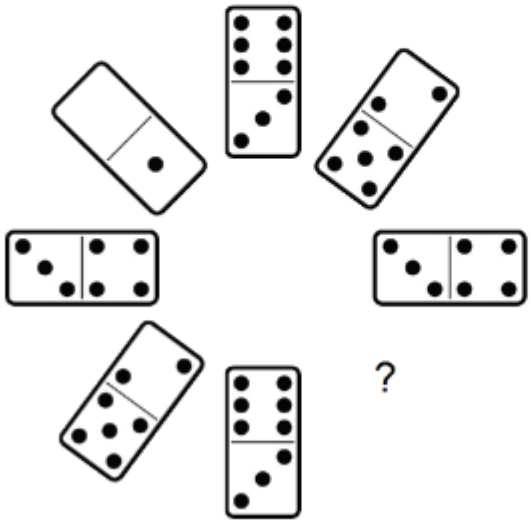
1/2/3/ ?

1)

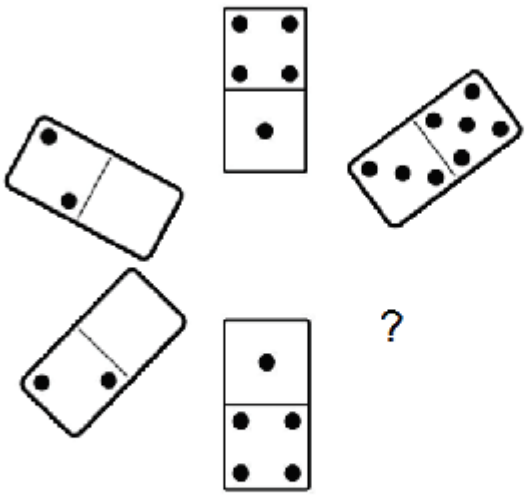


2

2)

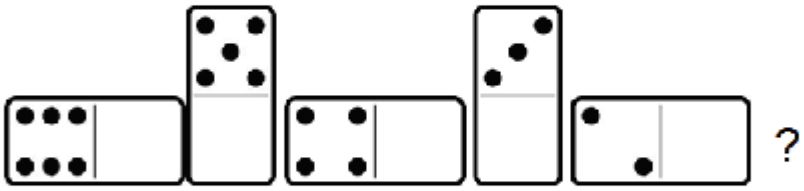


3)



3

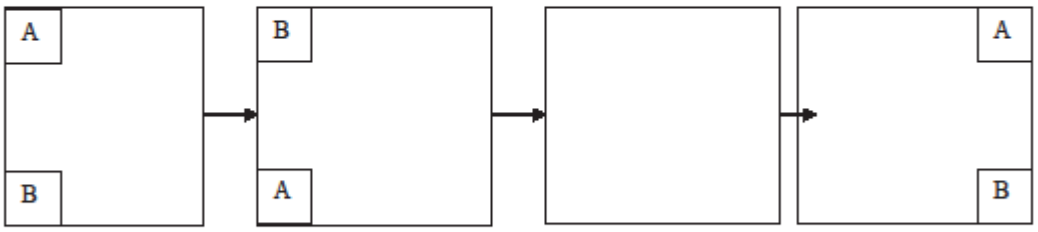
4)



5) Frise de décoration

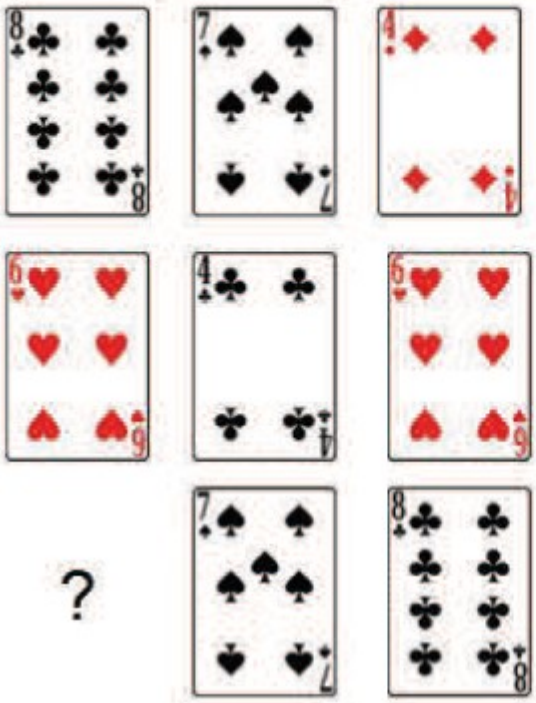
---

6) Deux robots



4

7)



8) Code d'entrée immeuble

A1 B3 C6 D10 ??

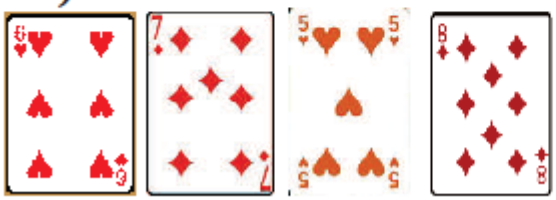
5

9) Jeu de cartes

a)

?

10)



?

---

11) Code casier

Janvier : 1/2/3/4
Février : 2/4/6/8
Mars : 3/6/9/2
Avril : ???

---

12) Classement

Le 03 décembre 2014 :
1/03-12-2014
2/03-12-2014
3/03-12-2014

Le 05 décembre 2014
1/05-12-2014

Le 07 décembre 2014
???

---

13) Alarme

8/10/13/17/ ?

---

14) Euromillions

2/2/4/12/ ?

---

15) Code coffre fort

Lundi	A/26
Mardi	Z/1
Mercredi	B/25
Jeudi	Y/2
Vendredi	??

---

16) Joueur de loto

7 13 20 28 37 ??

---

17) Télécommande

tableau

1	7
2	8
3	9
23	??

8

18) Code parental

1250
1241
1232
1223
1214
??

---

19) Code de journal intime

Lundi	011
Mardi	023
Mercredi	047
Jeudi	095
Vendredi	??

---

20)

102-100-050-048-024- ??

9

Epreuve 6 : lexique

Lire DOUCEMENT

Exemples :

Hier, je me suis couché tard pour réviser et levé très tôt. Je suis **reposé.**
J'espère que le…….euh….comment on dit……le……**(contrôle/devoir)** de maths ne va pas être trop dur. Je n'aurai pas dû me coucher aussi tard mais j'ai regardé un film avant de me mettre à travailler : l'histoire d'une fille très généreuse qui avait la **faim** sur le cœur. Elle était fleuriste et distribuait chaque matin des **arbres** aux personnes qu'elle croisait.

Item :

J'ai vu un film hier : **une……..euh comment on dit…. ? une (1)**_____ romantique. Je n'ai pas (2) **l'attitude** de regarder ce genre de film. En gros c'était une histoire d'amour, l'histoire de Yohann, un jeune de 19 ans, très grand, plutôt à faire craquer les filles, sportif, et assez beau. Il est sorti avec une fille pendant deux ans dont il était totalement fou (et elle également). Il a vécu une grande histoire d'amour. Il était sur un petit (3) **soleil**, n'avait aucun doute sur son histoire. Il vivait dans le bonheur total. Il n'avait qu'elle. Il ne l'a jamais trompée. Il faisait preuve d'une grande (4)**……. euh comment on dit, une**_____. Il s'imaginait faire sa vie avec elle. Il faisait beaucoup de sport, 1h ou 2h par jour en travaillant surtout le haut du corps. Le matin, il faisait (5) **l'autoroute** avec elle pour aller au lycée. Il en profitait pour lui faire des compliments. Elle adorait ça. Il lui achetait tout ce qu'elle voulait : bijoux, parfums, ….et tout l'argent mis de côté, son (6)_____ **euh son** _____diminuait rapidement.
Puis, un jour, il y a eu une grosse (7)……euh…..euh une_____ entre les deux. Il a mis son poing dans le mur, a jeté ses (8) **flocons** de parfum par terre. Elle lui avait en effet, annoncé qu'elle n'était plus amoureuse de lui. Elle ne lui trouvait rien de (9) **banal**. Elle le trouvait comme monsieur tout le monde. Elle avait besoin de connaître autre chose, de tester de nouvelles choses, de quelque chose qui la fasse rêver, pour résumer (10) **d'……euh…..euh d'**_____**.**
Il essaya alors de l'impressionner. Il passa le permis de conduire et acheta une voiture. Il l'emmena en week-end dans le désert (11) **acide** en Tunisie. C'était son rêve. Elle le menait à la (12) **planche.** Tout ce qu'elle voulait, il lui offrait. Un jour, elle rencontra un autre type qui la fit craquer. Il avait ce petit plus. Il lui faisait vivre des expériences nouvelles. Il lui proposait des (13) **attractions** incomparables. Tous les jours, elle faisait une nouvelle activité. C'était un (14) **allié** de taille pour Yohann. Il osait lui dire qu'il l'aime, ce que n'avait jamais fait Yohann.
Yohann regarda sa bien-aimée partir petit à petit. Il laissa faire, impuissant. Il n'avait pas (15) l'**attention** de se battre.
Puis cinq ou six années passèrent. Il s'est lancé dans une carrière prestigieuse d'ingénieur. Il construit des (16) **ponts** qui permettent de stocker l'eau dans un étang ou un lac. Tout le monde fait appel à lui. Il commence à se faire un nom dans le monde professionnel. Il devient (17) **anonyme**. Dans le travail, on peut lui faire confiance. Il est très (18) **viable.** Il enchaîne les histoires d'amour sans lendemain.
Pendant ce temps, elle est restée avec le même homme. Elle travaille dans la mode. Et un jour, ils se croisent dans la rue et commencent à discuter comme s'ils ne s'étaient jamais quittés.

Yohann est toujours sous son charme. Il l'a dans la peau. Il va alors se décider à prendre les choses en main et tenter de la reconquérir en se disant qu'il n'aurait jamais dû laisser tomber la première fois.
Il l'invite le soir même en lui disant qu'il lui prépare une surprise. Elle se laisse tenter, ayant toujours le goût de l'aventure. Ils se donnent rendez-vous devant un café. Il lui bande les yeux, lui donne un (19) **demi** de champagne, la fait monter dans une superbe voiture louée pour l'occasion et démarre sous un (20) **croisement** de pneus. Il l'emmène à l'aéroport et ils décollent pour l'Afrique, lieu de leur dernier voyage commun, survolent le désert qu'elle apprécie tant, la savane où les lions déploient leur superbe (21) **chevelure**. Il a le (22) **doute** absolu que c'est la femme de sa vie. Mais elle ne ressent rien. Elle ne vibre toujours pas, malgré tous ses efforts.
Ils repartent chacun de leur côté. Au début, il est plutôt serein. Mais cette (23) **exacerbation** est de courte durée. C'est le calme avant la tempête. La fin du film montre son (24) **échéance.** Cela n'est pas très gai.

2

Epreuve 7 : majuscules

Exemple

C'est le printemps. les oiseaux chantonnent gaiement dès le Petit matin, la pluie laisse place au soleil Radieux qui illumine la ville. Les gens ont une nouvelle expression, qu'on ne leur reconnaissait plus depuis quelques Temps : un sourire se dessine sur leur visage. ils ont l'air de renaitre comme s'ils se réveillaient après une longue Hibernation.

Item

Margot

Elle s'appelait Margot. Elle avait un doux regard empli de Compassion, un petit nez fin, une bouche bien dessinée, des joues roses, de beaux cheveux gris ondulés attachés en chignon. elle n'était pas très grande ; je la dépassais depuis mes 12 ans. lorsque j'avais un chagrin, je courrai chez elle. Elle me préparait alors un Bon chocolat chaud et mon dessert préféré : des muffins au Chocolat. Tous mes soucis disparaissaient alors en un instant. je pouvais lui parler de tout, de l'école, des Copains, de mes chagrins d'amour. Mais aussi de l'actualité. nous avions de vrais débats sur Tous les sujets : politique, historique, sociétal. Elle avait une culture générale infinie. J'aimais l'entendre me raconter des anecdotes sur sa vie, sa riche vie. elle avait, en effet, voyagé à travers le monde, rencontré de Multiples personnes, visité de très nombreux endroits. Tous les jours, Elle me racontait une nouvelle histoire. je buvais ses paroles, accoudé à la table de la cuisine. Nous avions ce rituel presque tous les jours. Elle habitait juste à quelques minutes de Chez moi. Je la connaissais depuis toujours et je lui vouais une admiration sans borne. nous étions comme les deux doigts de la main, inséparables. Elle vivait seule depuis plusieurs années et ma présence quasi Quotidienne lui faisait du bien. elle me disait que j'étais son petit rayon de soleil qui illuminait sa journée. Dès que je frappais à sa porte, elle m'ouvrait avec un large sourire. Nous jouions régulièrement aux cartes en en particulier au rami, son jeu préféré. nous écoutions en même temps de la musique classique…sur sa vieille chaine hifi. Elle me faisait découvrir des Univers inconnus pour moi. Parfois, je venais avec mon MP3 pour Lui faire aussi toucher du doigt mon univers musical très pop. elle m'apprenait également la cuisine et le jardinage. Elle possédait un immense jardin fruitier, et un potager. Je l'aidais à ramasser les pommes, prunes, framboises, groseilles, salades, carottes, pommes de terre, Tâche qui devenait

1

Yohann est toujours sous son charme. Il l'a dans la peau. Il va alors se décider à prendre les choses en main et tenter de la reconquérir en se disant qu'il n'aurait jamais dû laisser tomber la première fois.
Il l'invite le soir même en lui disant qu'il lui prépare une surprise. Elle se laisse tenter, ayant toujours le goût de l'aventure. Ils se donnent rendez-vous devant un café. Il lui bande les yeux, lui donne un (19) **demi** de champagne, la fait monter dans une superbe voiture louée pour l'occasion et démarre sous un (20) **croisement** de pneus. Il l'emmène à l'aéroport et ils décollent pour l'Afrique, lieu de leur dernier voyage commun, survolent le désert qu'elle apprécie tant, la savane où les lions déploient leur superbe (21) **chevelure**. Il a le (22) **doute** absolu que c'est la femme de sa vie. Mais elle ne ressent rien. Elle ne vibre toujours pas, malgré tous ses efforts.
Ils repartent chacun de leur côté. Au début, il est plutôt serein. Mais cette (23) **exacerbation** est de courte durée. C'est le calme avant la tempête. La fin du film montre son (24) **échéance.** Cela n'est pas très gai.

2

Katia TERRIOT

Quels apports de l'approche contextualisée à l'évaluation de l'intelligence chez les adolescent.e.s ?

Résumé

L'objectif principal de la thèse est de montrer les apports de l'évaluation contextualisée à l'évaluation de l'intelligence chez les adolescents.e.s, à travers cinq articles. Le premier fait l'état des lieux de l'évaluation contextualisée de l'intelligence. Le second développe les évaluations auto et hétéro-rapportées de l'intelligence. Les trois suivants rendent compte des apports d'outils d'évaluation contextualisée notamment en termes de preuve de validité. Seul le facteur de raisonnement fluide du questionnaire d'auto-évaluation prédit les résultats scolaires. Les liens avec le comportement adaptatif existent, notamment avec le facteur de vitesse de traitement du questionnaire (r=.69). Les questionnaires d'hétéro-évaluations entretiennent des liens modérés à forts avec les résultats scolaires. Enfin, pour le test de performance composé de sept subtests, seule une épreuve (intelligence cristallisée) corrèle positivement et significativement avec la moyenne scolaire générale (r=.42); et trois avec le comportement adaptatif (corrélations néanmoins faibles). Les résultats et implications de ces résultats sont discutés dans la conclusion et une ouverture sur la réalité virtuelle est proposée.

Mots clés : évaluation 360°, auto-évaluation, hétéro-évaluation, Cattell-Horn-Carroll, CHC, comportement adaptatif, test cognitif, WISC

Résumé en anglais

The main objective of the thesis is to show the contributions of the contextualized evaluation to the evaluation of the intelligence of teenagers, through five articles. The first takes stock of the contextualized evaluation of intelligence. The second develops self and hetero-reported assessments of intelligence. The following three report the contributions of contextualized evaluation tests, particularly in terms of proof of validity. Only the fluid reasoning factor of the self-report questionnaire predicts school results. The links with adaptive behavior exist, notably with the processing speed factor of the questionnaire (r = .69). Hetero-assessment questionnaires have moderate to strong links to school performance. Finally, for the performance test composed of seven subtests, only one subtest (crystallized intelligence) correlates positively and significantly with school results (r = .42); and three with adaptive behavior (yet weak correlations). The results and implications of these results are discussed in the conclusion and an opening on virtual reality is proposed.

Key words: 360° degree feedback, self-assessment, hetero-assessment, Cattell-Horn-Carroll, CHC, adaptive behavior, cognitive test, WISC