



**HAL**  
open science

## Améliorer l'exactitude de l'auto-évaluation : quels dispositifs pour quels apprenants ?

Adeline Maillard

### ► To cite this version:

Adeline Maillard. Améliorer l'exactitude de l'auto-évaluation : quels dispositifs pour quels apprenants ?. Philosophie. Université Toulouse le Mirail - Toulouse II, 2015. Français. NNT : 2015TOU20108 . tel-01355806

**HAL Id: tel-01355806**

**<https://theses.hal.science/tel-01355806>**

Submitted on 24 Aug 2016

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

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



# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse - Jean Jaurès

---

**Présentée et soutenue par :**

**Adeline Maillard**

**le** Vendredi 27 novembre 2015

**Titre :**

Améliorer l'exactitude de l'auto-évaluation : quels dispositifs pour quels apprenants ?

---

**École doctorale et discipline ou spécialité :**

ED CLESCO : Psychologie

**Unité de recherche :**

Laboratoire Cognition, Langues, Langage, Ergonomie

**Directeur/trice(s) de Thèse :**

Nathalie Huet, Professeure, CLLE-LTC, Université de Toulouse

Caroline Dupeyrat, Maître de Conférences, CLLE-LTC, Université de Toulouse

**Jury :**

Thérèse Bouffard, Professeure, Université du Québec à Montréal

Marie Izaute, Professeure, Université Blaise Pascal, Clermont-Ferrand

Jean-François Rouet, Directeur de Recherche CNRS, Université de Poitiers



---

## RÉSUMÉ

---

L'auto-évaluation est un processus métacognitif central dans le cadre d'un apprentissage en autonomie. Des décennies de recherche ont montré, toutefois, que les apprenants n'auto-évaluaient pas leurs performances futures avec exactitude, conduisant les chercheurs à essayer de mieux comprendre le processus d'auto-évaluation afin d'en améliorer l'exactitude. La majeure partie des travaux effectués sur le sujet ont, toutefois, été réalisés dans des situations peu écologiques et impliquent des tâches d'apprentissage simples (e.g., l'apprentissage associatif) ou classiques (e.g., compréhension de textes linéaires).

Cette thèse présente trois études qui se centrent autour de l'exactitude de l'auto-évaluation dans des contextes plus écologiques. Ces études testent l'efficacité de divers outils d'aide à l'auto-évaluation sur une tâche d'apprentissage intégrée aux enseignements des participants. Elles considèrent également des caractéristiques propres à la tâche ou à l'apprenant. Ainsi, la première étude combinait plusieurs outils d'aide à l'auto-évaluation (incitations à la prise de conscience des stratégies utilisées et de leur pertinence, et auto-explications) afin d'améliorer l'exactitude des auto-évaluations sur des connaissances déclaratives et procédurales. Une deuxième étude explorait l'effet des cartes conceptuelles en tant qu'outil d'aide à l'auto-évaluation. L'efficacité des cartes conceptuelles sur l'exactitude de l'auto-évaluation et l'apprentissage a été testée selon le mode de leur utilisation (construction vs. consultation), le moment de leur présentation (simultanément à l'apprentissage vs. en différé) et le niveau d'expertise des apprenants (experts vs. novices). Enfin, la troisième étude évaluait si un entraînement relativement long à l'utilisation des cartes conceptuelles permettait d'en faire un outil efficace pour améliorer l'exactitude de l'auto-évaluation et l'apprentissage.

Les résultats obtenus lors de ces trois études n'ont pas confirmé nos hypothèses relatives à l'efficacité des outils d'aide à l'auto-évaluation sur l'exactitude de l'auto-évaluation. Les résultats montrent, cependant, que les participants auto-évaluaient avec plus d'exactitude des connaissances déclaratives que procédurales (Étude 1). Les résultats ont également répliqué, sur une tâche plus écologique, l'effet positif du niveau des connaissances antérieures dans le domaine d'apprentissage sur l'exactitude de l'auto-évaluation, généralement démontré dans la littérature sur des tâches de laboratoire (Étude 2). Finalement, les résultats ont révélé que la durée de l'entraînement à la réalisation de cartes conceptuelles n'avait pas d'effet sur l'efficacité des cartes conceptuelles (Étude 3).

La discussion de ces résultats est organisée autour de quatre points principaux. Une première partie revient sur la question de l'adaptation des outils d'aide à l'auto-évaluation au niveau d'expertise des apprenants. Une seconde partie traite de l'adéquation des outils d'aide à l'auto-évaluation avec le type de connaissances évaluées. Une troisième hypothèse se centre sur la nature de l'exactitude de l'auto-évaluation mesurée, afin expliquer l'absence d'effet des outils d'aide à l'auto-évaluation testés. Finalement, le rôle des variables motivationnelles dans l'exactitude de l'auto-évaluation sera abordé comme une piste de travail pertinente.

*Mots-clés* : Auto-évaluation, Exactitude de l'auto-évaluation, Charge mentale, Expertise, Outils d'aide à l'auto-évaluation.

---

## ABSTRACT

---

Self-monitoring is a key metacognitive process in self-regulated learning. Decades of research, however, showed that learners are inaccurate in evaluating their own performance which led researchers to focus on understanding the self-monitoring process to ultimately improve self-monitoring accuracy. Most work on the topic was conducted in low ecological contexts and implied only easy (e.g., associative learning) or classical (i.e. comprehension of linear texts) learning tasks.

The present thesis offers three studies that focused on self-monitoring in ecological contexts. These studies explored the efficiency of several self-monitoring tools on a learning task integrated in participant's school formation setting and took into account task and learner characteristics. The first study used several self-monitoring help tools jointly (prompts aimed to make the learner aware of the strategies is used and their relevance, and self-explanations) in order to improve self-monitoring accuracy about declarative and procedural knowledge. The second study explored concept maps as a self-monitoring tool and tested their effect on learning in different conditions. The experimental manipulations were presentation mode (construction vs. consultation), moment of presentation (simultaneously to learning vs. delayed) and learners' expertise level (experts vs. novices). Finally, the third study assessed whether a long training to concept maps use was efficient enough to make this tool a relevant one to improve self-monitoring accuracy and learning.

Results of the three studies did not confirm our hypothesis about self-monitoring tools efficiency on self-monitoring accuracy in a subsequent task. Results, however, showed that participants were more accurate when they evaluated declarative knowledge than procedural knowledge (Study 1). They also replicate, in a more ecological task, the positive effect of prior knowledge level about the target learning content on self-monitoring accuracy traditionally highlighted in the literature relying on laboratory learning task (Study 2). Finally, the training duration to use concept maps did not affect concept map efficiency (Study 3).

Discussion of the present findings is centered on four main points. A first point discusses the suitability of self-monitoring tools as a function of learners' level of expertise. A second line of discussion deals with the adequacy between self-monitoring tools and the nature of knowledge which is evaluated. A third hypothesis regards the nature of self-monitoring accuracy which was measured, in order to explain the absence of effect of the

self-monitoring tools we tested. Finally, the role of motivational variables surrounding self-monitoring accuracy is discussed as a compelling avenue.

*Keywords* : Self-monitoring, Self-monitoring accuracy, Mental load, Expertise, Self-monitoring tools.

---

## REMERCIEMENTS

---

Mes premiers remerciements s'adressent à mes directrices, Nathalie Huet et Caroline Dupeyrat. Merci pour votre disponibilité, votre investissement dans ce travail, et votre soutien dans les moments de doutes. Merci également d'avoir fait en sorte que je puisse réaliser ce travail de thèse dans les meilleures conditions qui puissent être. Je suis consciente de la chance dont j'ai bénéficié et vous en suis très reconnaissante.

Je tiens à remercier très sincèrement Thérèse Bouffard, Marie Izaute et Jean-François Rouet, d'avoir accepté de faire partie de mon jury et de me faire l'honneur d'évaluer ce travail de recherche.

Je remercie toutes les personnes qui ont contribué de près ou de loin à la bonne réalisation de ce projet. Je pense notamment à Jean-Christophe Sakdavong, pour ses conseils toujours enrichissants, et sans qui toutes ces idées n'auraient jamais pu se concrétiser. Un immense merci aux responsables d'établissements et participants des lycées Saint Exupéry et de l'IUT de Toulouse, qui nous ont accueilli avec beaucoup d'enthousiasme et de disponibilité. Merci également à l'ensemble des collaborateurs de l'IRIT, membres du projet CONTINT.

Je tiens, enfin, à remercier tous les membres du laboratoire LTC avec qui j'ai pu partager mes idées, mes projets ou mes difficultés. Merci pour vos encouragements, votre bienveillance et surtout pour le partage de vos connaissances culinaires :).

Sur un plan plus personnel je tiens à exprimer tout mon amour et ma gratitude à ma famille. Je remercie bien évidemment mes parents et mon frère, pour m'avoir transmis les valeurs et offert les conditions nécessaires à la réussite d'un si long et difficile projet. Un merci tout tendre à ma mamie qui m'a portée spirituellement au cours de ces dix années d'étude, et à ma tatie dont le soutien discret est néanmoins indispensable. Enfin, je tiens également à remercier Flori, Noé, Adrien, Cassandre, Léandre, Marie, Henrique, et tous les autres membres de ma famille qui m'ont toujours soutenu, conseillé, écouté et apporté beaucoup de joie et d'amour.

Merci à tous mes amis qui ont suivi cette étape de ma vie de près, ou de loin pour la plupart. Merci pour votre amitié fidèle qui résiste à l'épreuve du temps et/ou de la distance. Vous êtes dans mon cœur à chaque instant et j'espère que nos chemins ne seront jamais loin. Merci à toutes les personnes incroyables, et désormais amis, que j'ai pu rencontrer ces quatre dernières années ; notamment, Pierre, Majdouline, Hakim, Sébastien (Puma), Sébastien (Matton), Marco, Ladis, et toute la petite troupe d'irréductibles villeneuvois(es) avec qui j'ai partagé tous ces délicieux



moments de la vie. Je fais une dédicace spéciale à Julie, Nadine et Stan qui ont partagé le coeur de mon intimité pendant quatre ans, à savoir mon bureau. J'ai tellement pris de plaisir à partager votre quotidien que j'arrive à en être triste que tout cela finisse ! Merci pour votre soutien, pour votre savoir, votre humilité, votre courage, pour les services rendus, . . . Merci pour TOUT ! Vous allez beaucoup me manquer.

Enfin, je remercie Bastien, d'un amour sincère et profond. Merci pour cet amour dont tu me remplis, pour ton soutien indéfectible, mais aussi pour toutes tes qualités qui me font avancer et qui rendent mon quotidien magique et indescriptiblement beau.

---

## VALORISATION

---

Les recherches en lien avec cette thèse nous ont permis de soumettre un article et ont fait l'objet de plusieurs communications orales et affichées dont la liste suit :

### COMMUNICATIONS ORALES

- **Maillard**, A., Dupeyrat, C., Motak L., Sakdavong, J.C. & Huet, N. (2014, Septembre). Uner which conditions concept maps are the most efficient for a more accurate monitoring ? *2nd International Conference on Metacognition*, Clermont-Ferrand, France.
- Mulet, J., **Maillard**, A., Sakdavong, J.C. & Huet, N. (2014, Septembre). Cosntruction versus consultation : How to use graphical representation for skill assessment ? *6th Biennial meeting of the European Association for Research on Learning and Instruction (EARLI) SIG 16*, Istanbul, Turquie.
- **Maillard**, A., Dupeyrat, C., Motak L., Sakdavong, J.C. & Huet, N. (2013, Juin). Effet différentiel des cartes conceptuelles en fonction de l'expertise : exploration de l'exactitude de l'auto-évaluation et de la charge cognitive. *3ème workshop : Apprendre à partir de documents numériques*, Toulouse, France.
- **Maillard**, A., Motak L., Sakdavong, J.C., Dupeyrat, C. & Huet, N. (2013, Août). How does concept map format and expertise level affect learning and self-evaluation ? *15th Biennial EARLI Conference for Research on Learning and Instruction*, Munich, Allemagne.
- **Maillard**, A., Motak L., Sakdavong, J.C., Dupeyrat, C. & Huet, N. (2013, Juin). Does timing of concept map construction increase self-evaluation accuracy and decrease cognitive overload ? *6th International Cognitive Load Theory Conference*, Toulouse, France.
- **Maillard**, A., Motak L., Sakdavong, J.C., Dupeyrat, C. & Huet, N. (2013, Juin). Concept maps for self-evaluation and learning. *2ème workshop : Apprendre à partir de documents numériques*, Toulouse, France.
- **Maillard**, A., Motak L., Sakdavong, J.C., Dupeyrat, C. & Huet, N. (2013, mai). How and when presenting a concept map for learning and an accurate self-evaluation ? *Proceedings of the 5th International Conference on Computer Supported Education*, Aachen, Allemagne.

- **Maillard**, A., Motak L., Dupeyrat, C., Catteau, O., Sakdavong, J.C. & Huet, N. (2012, Août). To train or not to train ? On restricted effects of a specific training in self-evaluation. *The 5th Biennial meeting of the European Association for Research on Learning and Instruction (EARLI) SIG 16*, Milan, Italie.

#### **COMMUNICATION AFFICHÉE**

- **Maillard**, A., Dupeyrat, C., Sakdavong, J. C. & Huet, N. (2014, Septembre). Does expertise and processes used for self-evaluation (SE) influence its accuracy ? *2nd International Conference on Metacognition*, Clermont-Ferrand, France.

#### **ARTICLE EN PRÉPARATION**

- **Maillard**, A., Motak, L., Dupeyrat, C., Sakdavong, J.C. & Huet, N. (en préparation). Factors which influence self-monitoring accuracy.

---

## TABLE DES MATIÈRES

---

1	INTRODUCTION	1
2	L' AUTO-ÉVALUATION	5
2.1	Définir l'auto-évaluation	6
2.2	L'opérationnalisation du jugement d'auto-évaluation	9
2.3	Les différents indicateurs de l'exactitude du jugement d'auto-évaluation	11
2.4	Les sources de l'exactitude de l'auto-évaluation	15
2.5	Synthèse des apports des hypothèses à l'explication de l'inexactitude de l'auto-évaluation	25
2.6	Le coût cognitif de l'auto-évaluation	27
2.7	L'auto-évaluation : une habileté transverse ou spécifique à un domaine ?	30
3	AMÉLIORER L' EXACTITUDE DE L' AUTO-ÉVALUATION	35
3.1	Agir sur la pertinence des indices utilisés	37
3.2	Agir sur l'investissement dans le processus : les incitations	48
3.3	Agir sur la temporalité	50
4	EFFET D' OUTILS D' AIDE À L' AUTO-ÉVALUATION SUR SON EXACTITUDE	53
4.1	Rappel sur le processus d'auto-évaluation	54
4.2	Le processus d'auto-évaluation est-il général ou spécifique à un domaine ?	54
4.3	Améliorer l'exactitude de l'auto-évaluation : les différentes méthodes	55
4.4	Question d'étude et hypothèses	59
4.5	Méthode	60
4.6	Résultats	70
4.7	Discussion	76
5	EFFET DU MODE ET DU MOMENT DE PRÉSENTATION D' UN OUTIL D' AIDE À L' AUTO-ÉVALUATION : LES CARTES CONCEPTUELLES ET LA BASE DE CONNAISSANCES SUR L' EXACTITUDE DE L' AUTO-ÉVALUATION	81
5.1	Les cartes conceptuelles : un outil d'aide à l'apprentissage et à l'auto-évaluation	82

5.2	Faut-il consulter ou construire les cartes conceptuelles ?	85
5.3	Base de connaissance : les effets sur l'exactitude de l'auto-évaluation	87
5.4	L'efficacité des modes d'utilisation des cartes selon la base de connaissance des apprenants	88
5.5	Différer la présentation des cartes conceptuelles pour en améliorer l'efficacité	90
5.6	Hypothèses	91
5.7	Méthode	93
5.8	Résultats	101
5.9	Discussion	120
6	UN ENTRAÎNEMENT À L'UTILISATION DES CARTES CONCEPTUELLES POUR AMÉLIORER L'EXACTITUDE DE L'AUTO-ÉVALUATION	125
6.1	Entraînement à l'utilisation des cartes conceptuelles	126
6.2	La complétion de cartes conceptuelles	129
6.3	Synthèse et hypothèses	129
6.4	Méthode	130
6.5	Résultats	137
6.6	Discussion	146
7	DISCUSSION ET PERSPECTIVES	149
7.1	Synthèse des principaux résultats	150
7.2	Pistes d'explication des résultats	151
7.3	Pistes de travail pour de futures études	153
7.4	Conclusion	156
	References	157
A	L'AUTO-ÉVALUATION DANS LES PRINCIPAUX MODÈLES DE L'APPRENTISSAGE AUTORÉGULÉ	179
B	PRÉCISIONS SUR LES MÉTHODES DE CALCUL DE L'EXACTITUDE DE L'AUTO-ÉVALUATION	183
B.1	Le Gamma de Goodman-Kruskal	183
B.2	Les courbes de calibration	183
C	INTÉGRALITÉ DU MATÉRIEL DE L'EXPÉRIENCE 1	185
c.1	Questionnaire de mesure de connaissances antérieures	185
c.2	Questionnaire similaire	187

c.3	Questions de généralisation	189
c.4	Résultats complémentaires	190
D	RÉSULTATS COMPLÉMENTAIRES DE L'EXPÉRIENCE 2	205
D.1	Le mode et le moment d'utilisation des cartes conceptuelles	205
D.2	L'effet du moment de présentation des cartes selon le niveau d'expertise de l'apprenant	208
E	COMPLÉMENTS D'INFORMATIONS SUR LE MATÉRIEL UTILISÉ DANS L'EXPÉRIENCE 3	213
E.1	Questionnaire de mesure de connaissances déclaratives (QCM 1 et 2)	213
E.2	Thèmes des cartes conceptuelles présentées lors de la phase d'entraînement	214
E.3	Exemple de carte conceptuelle présentée lors de la phase d'entraînement	215
F	LA THÉORIE DE LA CHARGE COGNITIVE : COMPLÉMENTS D'INFORMATION	217
F.1	Notions élémentaires sur la théorie de la charge cognitive	217
F.2	Effet de dissociation de l'attention	218
F.3	L'effet de redondance	219
F.4	L'effet de l'expertise inversée	220
F.5	L'effet de modalité	220

---

## LISTE DES FIGURES

---

FIGURE 1	Exemple fictif de courbe de calibration	15
FIGURE 2	Le modèle de l'auto-évaluation de Koriat	22
FIGURE 3	Routes potentielles pour réaliser un jugement d'auto-évaluation dans le cadre d'une tâche de compréhensions (adapté de Griffin et al., 2009).	24
FIGURE 4	Synthèse de la procédure utilisée dans l'Expérience 1.	70
FIGURE 5	Carte conceptuelle présentée aux apprenants devant consulter une carte parallèlement à l'apprentissage.	95
FIGURE 6	Dispositif présenté aux apprenants devant construire une carte parallèlement à l'apprentissage.	95
FIGURE 7	Copie d'écran de l'exercice d'application et du menu auquel les participants avaient accès pour résoudre l'exercice.	97
FIGURE 8	Synthèse de la procédure expérimentale de l'expérience 2. Groupe 1 = construction simultanée ; groupe 2 = construction différée ; groupe 3 = consultation simultanée ; groupe 4 = consultation différée ; groupe 5 = groupe contrôle	101
FIGURE 9	Jugement et direction du biais d'auto-évaluation en fonction du niveau d'expertise de l'apprenant et du mode d'utilisation des cartes.	115
FIGURE 10	Amplitude du biais d'auto-évaluation et performance à l'exercice d'application en fonction du niveau d'expertise de l'apprenant et du mode d'utilisation des cartes.	116
FIGURE 11	Effort mental et difficulté ressentis pour comprendre les vidéos en fonction du niveau d'expertise de l'apprenant et du mode d'utilisation des cartes.	117
FIGURE 12	Effort mental et difficulté ressentis pour utiliser les cartes en fonction du niveau d'expertise de l'apprenant et du mode d'utilisation des cartes.	118

FIGURE 13	Copie d'écran du dispositif de complétion de cartes conceptuelles tel que présenté dans la phase d'entraînement. 133
FIGURE 14	Copie d'écran du dispositif de complétion de cartes conceptuelles tel que présenté dans la phase test. 133
FIGURE 15	Synthèse de la procédure de l'étude 3 137
FIGURE 16	Effort mental et difficulté ressentie pour comprendre les vidéos en fonction du moment et du mode d'utilisation des cartes. 207
FIGURE 17	Effort mental et difficulté ressentie pour utiliser les cartes conceptuelles en fonction du moment et du mode d'utilisation des cartes. 208
FIGURE 18	Exemple de carte conceptuelle présentée lors de la phase d'entraînement 215
FIGURE 19	illustration de l'effet de dissociation de l'attention 219



---

## LISTE DES TABLEAUX

---

TABLEAU 1	Exemples de questions posées pour mesurer le jugement d'auto-évaluation selon le moment de la mesure et la granularité de la question. 10
TABLEAU 2	Synthèse des principales hypothèses explicatives de l'auto-évaluation. 26
TABLEAU 3	Illustration du calcul des indicateurs de biais d'auto-évaluation 64
TABLEAU 4	Ensemble des moyennes (et écarts-types) calculées pour le jugement d'auto-évaluation, l'amplitude du biais d'auto-évaluation, la direction du biais, et la performance en fonction de la condition expérimentale et du moment de la mesure. Le symbole * indique les variables exprimées en pourcentage. 72
TABLEAU 5	Moyennes et écarts-types de l'amplitude des biais selon le type de connaissance, le moment de la mesure et la condition expérimentale 74
TABLEAU 6	Ensemble des moyennes (et écarts-types) des cinq groupes de participants pour l'ensemble des variables dépendantes. Le symbole * indique les variables exprimées en pourcentage. 106
TABLEAU 7	Ensemble des moyennes (et écarts-types) obtenus sur les principales variables dépendantes selon le niveau d'expertise des apprenants et le mode d'utilisation des cartes. Le symbole * indique les variables exprimées en pourcentage. 114
TABLEAU 8	Ensemble des moyennes (et écart-types) pour les variables dépendantes de cette étude. Le symbole * indique les variables exprimées en pourcentage. 141
TABLEAU 9	Les modèles de l'apprentissage autorégulé. 181
TABLEAU 10	Tableau permettant le calcul du coefficient Gamma de Goodman-Kruskal 183

TABLEAU 11	Ensemble des moyennes (et écarts-types) de la performance, du jugement d'auto-évaluation et de la direction du biais pour les connaissances procédurales selon la condition expérimentale. 191
TABLEAU 12	Ensemble des moyennes (et écarts-types) de la performance, du jugement d'auto-évaluation et de la direction du biais pour les connaissances déclaratives selon la condition expérimentale 192
TABLEAU 13	Ensemble des moyennes (et écarts-types) de la performance, du jugement d'auto-évaluation et de la direction du biais pour les connaissances sur le domaine du traitement de texte selon la condition expérimentale 198
TABLEAU 14	Ensemble des moyennes (et écarts-types) de la performance, du jugement d'auto-évaluation et de la direction du biais pour les connaissances sur le domaine des styles selon la condition expérimentale 199
TABLEAU 15	Ensemble des moyennes (et écarts-types) de la performance, du jugement d'auto-évaluation et de la direction du biais pour les connaissances sur le domaine de l'automatisation selon la condition expérimentale 200



---

## PRÉAMBULE

---

Avant de commencer à présenter ce travail de recherche, je souhaiterais dédier quelques lignes à la présentation de son contexte de réalisation. En effet, dans le cadre d'une thèse à visée appliquée, certains choix méthodologiques ne se font pas seulement à la lumière de la littérature existante, mais trouvent certaines justifications dans le contexte dans laquelle la recherche se situe. Aussi, il me semble plus pertinent de définir quelques termes dès à présent afin de ne pas alourdir ce manuscrit d'explications secondaires qui affaibliraient la clarté de ce travail.

Cette thèse s'inscrit dans un programme de recherche financé par l'Agence Nationale de la Recherche (ANR) et coordonné par l'Institut de Recherche en Informatique de Toulouse (IRIT). L'objectif de ce programme était de repenser une formation proposée par l'IRIT et intitulée 'Méthodes Informatiques Appliquées à la Gestion des Entreprises' (MIAGE). Plus spécifiquement, l'objectif initial était de motiver et d'aider les étudiants de Master 1 en MIAGE à obtenir le 'Certificat Informatique et Internet niveau 2, Métiers de l'Ingénieur' (C2I 2MI), un certificat facultatif à l'obtention de leur diplôme. Cependant, lorsque le programme de recherche a débuté, le matériel nécessaire à la préparation du C2I 2MI n'était pas encore créé. Par conséquent, nous avons choisi de centrer nos études sur le matériel utilisé en vue de l'obtention de la 'Certification Informatique et Internet niveau 1' (C2I), un niveau de certification plus faible mais avec des ressources pédagogiques bien plus riches. Cette contrainte a eu pour conséquence le fait que le matériel proposé aux participants de la première étude n'était pas directement au programme de la formation qu'ils suivaient.

Le contexte 'institutionnel' justifie également les populations ciblées puisque de fait, nous étions intéressés, au départ de ce projet, par les étudiants concernés par la formation MIAGE. Plus tard, les 'faibles effectifs' de la formation <sup>1</sup> nous ont amené à nous orienter vers des populations plus variées, susceptibles d'être concernées et/ou présentant des caractéristiques spécifiques par rapport au contenu du C2I.

Enfin, ce projet ayant débuté un an avant mon arrivée, des études préliminaires non présentées ici avaient déjà été réalisées et mettaient en évidence des biais d'auto-évaluation chez la population des étudiants en MIAGE. Ces résultats justifiaient et supportaient l'objectif de dévelop-

---

1. J'entends par 'faibles' effectifs, qu'ils étaient trop faibles pour réaliser une étude empirique solide impliquant plusieurs groupes expérimentaux ( $N = 25$ ).

per l'autonomie des étudiants pour l'obtention de cette certification non obligatoire ; les études présentées dans ce travail se sont donc directement investies dans l'amélioration de l'exactitude de l'auto-évaluation.

---

## INTRODUCTION

---

Le travail de recherche présenté ici trouve son origine dans le double constat explicité dans cette introduction : d'une part, le constat que l'auto-évaluation est un processus clé dans l'apprentissage auto-régulé ; d'autre part, le constat d'une difficulté qu'ont les individus à auto-évaluer avec exactitude leurs apprentissages.

Sur le plan théorique, le courant cognitiviste a bouleversé la façon d'aborder l'activité d'apprentissage, et l'émergence de la métacognition a permis au processus d'auto-évaluation d'être progressivement conceptualisé. À l'origine, [Flavell \(1977\)](#) a développé les notions de connaissances et d'habiletés métacognitives en définissant la métacognition comme la connaissance et les compétences que possède une personne sur ses propres processus cognitifs (fonctionnement et produits qui en résultent ; [Noël, 1991](#)). Dès lors, les chercheurs se sont intéressés à ces connaissances et habiletés métacognitives et ont mis en évidence leur importance ainsi que la difficulté de leur mise en place. Plus spécifiquement, [Flavell, Friedrichs, et Hoyt \(1970\)](#) ont montré que les apprenants les plus jeunes (école élémentaire et avant) ne savent pas évaluer correctement leur niveau de connaissance : alors ces derniers déclaraient être prêts à rappeler correctement des items appris précédemment, leurs performances démontraient qu'ils ne l'étaient pas. De nombreux autres travaux ont, par la suite, mis en évidence la difficulté des apprenants à évaluer avec exactitude leurs performances à diverses tâches (e.g., [Dunlosky & Rawson, 2012](#) ; [Dunlosky & Nelson, 1992](#) ; [Glenberg & Epstein, 1987](#) ; [Koriat, 1997](#) ; [Maki, 1998](#)), y compris chez des adultes compétents ([Chi & Ceci, 1978](#) ; [Pressley & Ghatala, 1990](#)).

Si l'auto-évaluation s'avère être une activité complexe à entreprendre pour l'apprenant et à comprendre pour les psychologues, elle s'est également révélée être un processus important dans le développement de l'apprentissage autonome. L'apparition de nouveaux modèles traitant de la métamémoire et de l'apprentissage auto-régulé vont mettre en exergue le rôle clé de l'auto-évaluation dans ces processus d'acquisition de connaissances en autonomie (pour plus de détails voir [Annexe A](#) ou [Nelson & Narens, 1990](#) ; [Pintrich & De Groot, 1990](#) ; [Winne & Hadwin, 1998](#) ; [Zim-](#)

merman, 1989). Concrètement, un apprenant face à une leçon à apprendre doit commencer par planifier ses stratégies d'apprentissage pour ensuite s'engager, à proprement parlé, dans l'apprentissage. Le résultat des actions entreprises doit ensuite être évalué afin d'apprécier l'atteinte ou non de l'objectif d'apprentissage fixé. Si un écart est constaté entre l'état d'apprentissage actuel et l'état à atteindre, l'apprenant pourra se ré-engager dans un cycle d'apprentissage en s'autorégulant (par exemple, en changeant de stratégies). Toutefois, si l'auto-évaluation de l'état d'apprentissage est inexacte (par exemple, un écart existe mais n'est pas détecté), l'apprenant peut stopper son apprentissage et ne pas engager d'autorégulation. Cette relation étroite entre l'exactitude de l'auto-évaluation et l'auto-régulation a été démontrée empiriquement à de nombreuses reprises (Hines, Touron, & Hertzog, 2009 ; Huet & Mariné, 2009 ; Kornell & Bjork, 2007 ; Metcalfe & Finn, 2008 ; Thiede & Dunlosky, 1999). L'amélioration de l'exactitude de l'auto-évaluation constitue, par conséquent, un point stratégique pour aider un individu dans ses apprentissages.

Faciliter le processus d'auto-évaluation requiert, avant toute chose, d'en comprendre le fonctionnement. Nous savons aujourd'hui que l'auto-évaluation est influencée par trois grandes catégories de facteurs (Schraw, 1998) : les facteurs personnels (e.g., les caractéristiques du sujet telle que l'expertise), les facteurs liés à la tâche elle-même (e.g., la difficulté), les facteurs environnementaux (e.g., la présence d'outils d'aide). À ce jour, la majorité des études porte sur l'amélioration de l'exactitude de l'auto-évaluation dans des tâches d'apprentissage de paires de mots (e.g., Nelson & Dunlosky, 1991), sur de la compréhension de textes linéaires (e.g., M. Anderson & Thiede, 2008 ; Thiede, Dunlosky, Griffin, & Wiley, 2005), ou encore, de la résolution de problèmes (e.g., Baars, Visser, Gog, Bruin, & Paas, 2013). Pourtant, le développement des nouvelles technologies accentue la diversité des contextes d'apprentissage (e.g., apprentissage basé sur des hypermédias Azevedo & Cromley, 2004) ainsi que l'importance des compétences métacognitives (Baccino, Salmeron, & Cañas, 2008).

La considération simultanée de différents facteurs appartenant ou non à une même catégorie (i.e. facteurs personnels, environnementaux, liés à la tâche) est primordiale puisqu'il est aujourd'hui avéré qu'un outil d'aide à l'apprentissage aura des effets différents selon qu'il soit associé à une tâche simple ou complexe, donné à des individus ayant des niveaux de connaissances antérieures différents (e.g., Kalyuga, Ayres, Chandler, & Sweller, 2003 ; Stull & Mayer, 2007), ou présenté selon certaines modalités (Sweller, Van Merriënboer, & Paas, 1998). Les effets des outils d'aide à l'auto-évaluation selon la variation de ces différents facteurs sont, quant à eux, encore peu connus ; et le sont d'autant moins lorsque ces outils sont associés à une tâche d'apprentissage complexe et naturelle.

Ce travail, pour cette pluralité de raisons, a l'ambition nouvelle d'examiner l'efficacité des outils d'aide à l'auto-évaluation sur l'amélioration de son exactitude, en considérant l'influence de différents facteurs dans des conditions où la validité écologique est centrale. Dans ce cadre, la question générale sous-tendant ce travail de recherche peut être posée de la manière suivante : dans une tâche écologique, dans quelles conditions les outils d'aide à l'auto-évaluation permettent-ils d'améliorer l'exactitude de l'auto-évaluation ?

Avec pour objectif de répondre à cette question, un premier chapitre, consacré au processus d'auto-évaluation, met en avant les 'principes actifs' des outils d'aide à l'auto-évaluation. Ces outils sont décrits précisément, quant à eux, dans un second chapitre. Le troisième chapitre présente une première étude dont l'objectif est de déterminer si la complémentarité des outils d'aide à l'auto-évaluation peut s'avérer efficace sur l'exactitude de l'auto-évaluation et sur les performances d'apprentissage. Une deuxième étude vise à déterminer les conditions optimales d'utilisation des cartes conceptuelles, utilisées en tant qu'outil d'aide à l'auto-évaluation, pour améliorer l'exactitude de l'auto-évaluation selon le niveau d'expertise des apprenants. Une troisième étude a pour objectif d'évaluer l'efficacité d'un entraînement préalable et relativement long à l'utilisation de cartes conceptuelles sur l'amélioration de l'exactitude de l'auto-évaluation dans une tâche ultérieure. Le chapitre final propose une synthèse de l'ensemble des résultats obtenus, les conclusions générales qu'il est possible d'en tirer, ainsi que les perspectives pertinentes pour la suite des recherches sur le sujet.





# 2

---

## L' AUTO-ÉVALUATION

---

L'auto-évaluation est un processus clé dans l'apprentissage autorégulé. De nombreux auteurs s'accordent pour affirmer que deux processus métacognitifs sont à l'œuvre lors de l'apprentissage : un processus ascendant, l'auto-évaluation, qui informe le système métacognitif de l'état du système cognitif, et un système descendant, l'autorégulation, qui permet de modifier l'activité cognitive en fonction des informations apportées par le système ascendant (e.g., [Huet, 2004](#) ; [Nelson & Narens, 1990](#) ; [Schraw & Moshman, 1995](#)). L'auto-évaluation renvoie non seulement à la prise de conscience de l'état de notre apprentissage en fonction de la performance à atteindre mais également à l'auto-estimation de la qualité de l'apprentissage et des réponses produites. L'autorégulation renvoie, quant à elle, à la modification de l'activité cognitive lorsqu'un écart d'état d'apprentissage est constaté, ou à un arrêt si l'objectif d'apprentissage est atteint. Dans cette perspective, la pertinence de l'autorégulation dépend de la qualité de l'auto-évaluation, c'est-à-dire de son exactitude ([Nelson & Narens, 1990](#)).

Malgré l'importance d'atteindre un haut niveau d'exactitude pour le processus d'auto-évaluation, les apprenants sont rarement exacts dans l'évaluation qu'ils font de leurs performances (e.g., [Dunlosky & Rawson, 2012](#) ; [Dunlosky & Nelson, 1992](#)). Il est donc important, non seulement, de comprendre comment les apprenants s'auto-évaluent et comment améliorer l'exactitude de leur auto-évaluation mais également d'identifier les caractéristiques que doivent posséder les outils d'aide à l'auto-évaluation ainsi que les caractéristiques des apprenants qui vont utiliser ces outils. Cette thèse se donne pour objectif de répondre à ces questions. Le premier chapitre vise à retracer le soubassement théorique et méthodologique dans lequel le processus d'auto-évaluation s'inscrit.

Le concept d'auto-évaluation est utilisé par de nombreux auteurs pour désigner un processus métacognitif, mis en œuvre par l'apprenant, consistant à évaluer l'écart entre son état actuel d'apprentissage et l'état d'apprentissage à atteindre (e.g., [Nelson & Narens, 1990](#) ; [M. V. J. Veenman, 2011](#)). Ce jugement est utilisé comme base servant à la décision de poursuivre ou d'arrêter l'apprentissage. Ainsi, si l'auto-évaluation du niveau de maîtrise de l'apprentissage ne correspond pas à l'état d'apprentissage à

atteindre, l'apprentissage sera poursuivi. L'apprenant, dans ce cas, autorégule son apprentissage (ex : il change de stratégie, il alloue davantage de temps d'étude aux parties difficiles qu'aux parties faciles, etc.) jusqu'à ce que l'écart disparaisse. Si cette définition est la plus fréquemment utilisée dans la littérature, d'autres auteurs ont une définition plus large ou plus restreinte. Certains auteurs, en particulier, distinguent différents types de jugements d'auto-évaluation (e.g., Nelson & Narens, 1990). Le premier point traité dans ce chapitre fait état des définitions et des emplois du processus d'auto-évaluation tels que mentionnés dans la littérature. La mesure de ces différents types de jugement requiert diverses opérationnalisations qui sont présentées dans un second point de ce chapitre. De manière générale, l'auto-évaluation est étudiée pour en examiner son exactitude, c'est-à-dire pour déterminer si l'auto-évaluation des apprenants est exacte ou biaisée. L'auto-évaluation est biaisée (ou inexacte) lorsque l'apprenant sur-estime ou sous-estime l'état actuel d'apprentissage (comparé à l'état à atteindre). Une pluralité d'indicateurs est utilisée pour mesurer l'exactitude de l'auto-évaluation ; un troisième point de ce chapitre est ainsi consacré aux principaux indicateurs utilisés. En parallèle, la question des informations et indices utilisés par les apprenants pour produire un jugement d'auto-évaluation est centrale dans la littérature sur le sujet. Nous examinons, dans un quatrième point, les différentes hypothèses explicatives et leurs contributions dans la compréhension du processus d'auto-évaluation. Enfin, les deux derniers points de ce chapitre sont consacrés, pour l'un, au coût de mise en œuvre du processus d'auto-évaluation ; pour l'autre, à la généralisation du processus d'auto-évaluation à des contenus et tâches différentes.

## 2.1 DÉFINIR L'AUTO-ÉVALUATION

### 2.1.1 *La pluralité des terminologies*

Plusieurs termes sont présents dans la littérature pour évoquer le processus d'auto-évaluation, parmi lesquels se trouvent les appellations 'self-monitoring' (de Bruin & van Gog, 2012 ; Kauffman, Zhao, & Yang, 2011), 'self-assessment' (Boud & Falchikov, 1989 ; Kostons, van Gog, & Paas, 2010 ; Nelson-Le Gall, Kratzer, Jones, & DeCooke, 1990), ou encore 'self-evaluation' (Bouffard, Vezeau, Roy, & Lengelé, 2011 ; Gramzow, Elliot, Asher, & McGregor, 2003). Ces termes sont utilisés indifféremment les uns des autres et certains auteurs utilisent parfois les trois expressions dans le même article (i.e. assessment, monitoring, evaluation Dunlosky & Rawson, 2012). Zimmerman (1990) utilise, quant à lui, le terme de 'self-reflection' pour intégrer une dimension motivationnelle dans son modèle de l'apprentissage autorégulé. Dans ce dernier cas, le concept se

divise en deux notions distinctes : ‘l’auto-jugement’ (évaluation de ses performances) et ‘l’autoréaction’ (réaction émotionnelle et motivationnelle associée). Aucun consensus n’a réellement été trouvé pour labelliser le concept d’auto-évaluation dans la littérature anglo-saxonne. De la même façon, les termes d’auto-contrôle (e.g., [Mariné & Huet, 1998](#)), de supervision ([Bouffard-Bouchard & Pinard, 1988](#)), ou de monitoring ([Thuai, 2013](#)), sont également employés dans le langage français. Dans le présent travail, nous utiliserons le terme d’‘auto-évaluation’ pour désigner ce processus métacognitif ascendant (i.e. du niveau cognitif vers le niveau métacognitif).

Il existe une pluralité de termes pour désigner l’auto-évaluation ; mais également plusieurs types d’auto-évaluation, fonction de différentes caractéristiques, tels que le niveau de granularité ou le moment du jugement. Ces distinctions sont abordées ci-après.

### 2.1.2 *Non pas une auto-évaluation, mais plusieurs auto-évaluations*

Afin d’accéder aux processus métacognitifs et plus spécifiquement au processus d’auto-évaluation, les auteurs ont distingué plusieurs niveaux de jugement ([Mariné & Huet, 1998](#) ; [Nelson, Dunlosky, Graf, & Narens, 1994](#) ; [Nietfeld, Cao, & Osborne, 2005](#)). Les auteurs proposent les concepts de micro-prédictions ou de macro-prédictions selon que les jugements portent sur des items précis ou sur une performance globale. De même, les auteurs parlent de jugement prospectif ou rétrospectif selon que les jugements portent sur des performances futures ou passées ([Mariné & Huet, 1998](#) ; [Nelson & Narens, 1990](#)).

#### *Les micro-prédictions prospectives*

Les micro-prédictions prospectives consistent pour un apprenant à estimer la performance future qu’il pense obtenir à une question précise (ex : la probabilité qu’il rappelle ou reconnaisse un item précis). [Nelson et Narens \(1990\)](#) ont distingué différentes micro-prédictions prospectives selon le moment où elles sont produites :

1. ‘Un jugement de la facilité d’apprentissage’ (Ease Of Learning = EOL). L’EOL correspond à l’estimation que l’apprenant se fait de la facilité d’apprentissage d’un item et/ou de l’efficacité d’une stratégie pour apprendre un item. L’apprenant réalise ce jugement au moment de la prise de connaissance du contenu à apprendre et avant de procéder à l’apprentissage.
2. ‘Un jugement du sentiment de savoir’ (Feeling of Knowing = FOK). Le FOK correspond à l’évaluation que fait un apprenant de la possibilité de récupération ultérieure d’un item qui n’est pas rappelé à

un instant  $t$ . En d'autres termes, ceci correspond à l'évaluation de la probabilité de rappeler plus tard un item qui n'est pas accessible maintenant.

3. 'Un jugement d'apprentissage' (Judgment Of Learning = JOL). Le JOL correspond à la prédiction que fait un apprenant de ses performances futures pour chaque item étudié pendant et/ou immédiatement après la phase d'étude.

L'objectif de ces trois jugements n'est pas le même : plutôt, ils sont complémentaires. Dans la suite de ce travail, nous nous centrerons sur le JOL uniquement.

Le paradigme classique permettant de recueillir le JOL en tant que micro-prédiction associée à des tâches d'apprentissage de paires de mots consiste à présenter à l'apprenant un ensemble de paires de mots à apprendre. L'apprenant voit ensuite un des deux mots et doit prédire s'il pense pouvoir rappeler le second mot qui y était associé, ceci pour chaque paire apprise (e.g., [Koriat, 1997](#) ; [Nelson & Dunlosky, 1991](#)).

L'étude du JOL portait essentiellement, au départ, sur l'étude de la métamémoire (i.e. connaissance d'un individu sur le fonctionnement de sa propre mémoire, et le contrôle qu'il peut exercer sur elle [Flavell, 1979](#)). D'autres auteurs ont étendu l'étude de ce jugement à des tâches plus écologiques de compréhension ou de résolution de problème ([Baars et al., 2013](#) ; [Dunlosky & Lipko, 2007](#) ; [Griffin, Wiley, & Thiede, 2008](#) ; [Maki & Berry, 1984](#) ; [Thiede & Anderson, 2003](#) ; [Thiede et al., 2005](#)). Cet élargissement du concept de JOL à de nouvelles tâches a rendu cette notion plus globale, sans toutefois être une macro-prédiction. [Baars et al. \(2013\)](#) ont par exemple demandé à leurs participants combien d'étapes ils pensaient pouvoir résoudre lors de la présentation d'un problème similaire à celui présenté pendant un apprentissage. Les participants pouvaient ainsi estimer le nombre d'étapes auxquelles ils pensaient pouvoir répondre, sans pouvoir indiquer à quelle étape précisément ils pensaient réussir ou échouer. Ce jugement d'apprentissage constitue finalement un intermédiaire entre les micro-prédictions et les macro-prédictions.

### *Les macro-prédictions prospectives*

Les macro-prédictions sont davantage utilisées pour mesurer la prédiction de réussite d'un apprenant sur des tâches complexes, telles que la résolution de problème ([Huet, Mariné, & Escribe, 1994](#)) ou la compréhension ([Pressley & Ghatala, 1990](#)), dans lesquelles la réponse nécessite plusieurs étapes qui ne peuvent être interrompues ou décomposées sans risquer de perturber l'activité cognitive. Il est d'usage, pour recueillir les macro-prédictions, de présenter une tâche à accomplir à l'apprenant pour ensuite lui demander de prédire sa performance à l'ensemble de l'exercice.

L'apprenant peut ensuite réaliser effectivement la tâche. Ici, l'apprenant n'a pas à estimer sa performance item par item (Mariné & Huet, 1998).

### *Les jugements d'auto-évaluation rétrospectifs*

Les jugements rétrospectifs, généralement appelés 'jugement de confiance' (Dinsmore & Parkinson, 2013 ; Mengelkamp & Bannert, 2010 ; Nelson & Narens, 1990 ; Schraw, 2009), renvoient au degré de confiance qu'un apprenant a dans l'exactitude de la réponse qu'il a donné au préalable. Comme le mentionnent Mariné et Huet (1998) ou Schraw (2009), les jugements rétrospectifs se trouvent également sous forme de micro-évaluations lorsque l'apprenant exprime son niveau de confiance en sa réussite pour chaque item (Bouffard-Bouchard & Pinard, 1988 ; Mengelkamp & Bannert, 2009, 2010), ou sous forme de macro-évaluation lorsque l'apprenant exprime son niveau de confiance pour la performance globale (Glenberg, Wilkinson, & Epstein, 1982). Par ailleurs, les apprenants s'auto-évaluent avec plus d'exactitude pendant ou après la réalisation d'un test que préalablement (voir également Pressley & Ghatala, 1990 ; Mengelkamp & Bannert, 2009 ; Schraw, 1998).

Au delà de la définition stricte du concept d'auto-évaluation, différents auteurs proposent diverses manières de l'opérationnaliser.

## 2.2 L'OPÉRATIONNALISATION DU JUGEMENT D'AUTO-ÉVALUATION

La principale méthode permettant d'obtenir des informations sur le jugement d'auto-évaluation consiste à utiliser des mesures auto-rapportées (Nelson & Narens, 1990). Plusieurs outils permettent de recueillir ces mesures auto-rapportées, parmi lesquels se trouve le questionnaire (Mariné & Huet, 1998). Le questionnaire a l'avantage de permettre de collecter facilement des auto-évaluations de différentes natures (macro-prédiction, micro-prédiction, jugement prospectif ou rétrospectif) et auprès de grands groupes d'individus. Une pluralité de questions et d'échelles de réponses existent dans la littérature, fonctions des objectifs de chaque étude.

### 2.2.1 *Les questions...*

Les questions posées pour mesurer l'auto-évaluation dépendent de plusieurs paramètres tels que le matériel d'apprentissage à évaluer (e.g., des textes, des items, etc.), le type d'auto-évaluation (e.g., le JOL, le jugement de confiance) ou le moment de l'émission du jugement (prospectif vs. rétrospectif). À titre d'illustration, des exemples de questions selon le moment de la mesure et la granularité du jugement mesuré sont proposés dans le Tableau 1.

TABLEAU 1. – Exemples de questions posées pour mesurer le jugement d’auto-évaluation selon le moment de la mesure et la granularité de la question.

	Jugement prospectif		jugement rétrospectif	
	Micro-prédiction	Macro-prédiction	Micro-prédiction	Macro-prédiction
Exemple de question (auteurs)	À quel point pensez-vous être capable de définir ce concept ... ? (Dunlosky, Rawson, & Middleton, 2005)	À quel pourcentage d’items suivants pourrez-vous répondre correctement ? (Rawson & Dunlosky, 2007 ; Mengelkamp & Bannert, 2009)	Entourez un seul numéro sur l’échelle suivante pour rapporter la confiance que avez en l’exactitude de votre réponse sur la question précédente. (Glenberg & Epstein, 1987 ; Izaute, Larochelle, Morency, & Tiberghien, 1996 ; Mengelkamp & Bannert, 2009)	Estimez l’exactitude globale de vos performances au test (Schraw, 1994)

### 2.2.2 ... les réponses.

Pour recueillir le jugement d'auto-évaluation, une autre nuance concerne les échelles de réponse. Celles-ci peuvent être nominales, ordinales, ou d'intervalles.

Les échelles nominales requièrent généralement de répondre par oui ou par non (il est ainsi possible d'utiliser ces échelles pour des questions du type 'Pensez-vous pouvoir répondre correctement à la question suivante ?') (Nelson & Narens, 1990). Comme le font remarquer Dinsmore et Parkinson (2013), les échelles nominales font perdre de l'information car elles sont limitantes dans le choix de réponse qu'elles proposent ; or, un individu n'est pas nécessairement 'complètement certain' d'avoir tout juste ou tout faux.

Les échelles ordinales et d'intervalles offrent, quant à elles, plus de finesse dans la granularité de la réponse de l'apprenant, donc une plus grande sensibilité dans le recueil de l'auto-évaluation. La réponse peut se situer sur une échelle de type Likert (Pajares & Miller, 1994) en cinq ou sept points, ou sur des échelles allant de 0 à 100. Par exemple, à la question '[...] à quel point pensez-vous être capable de rappeler le second mot de la liste d'items précédente quand le premier s'affichera ?', il existe ainsi des échelles de réponses en six points telles que : '0 = je ne rappellerai pas, 20 = je suis sûr à 20%, 40... , 60... , 80... , et 100 = Je le rappellerai.' (voir Nelson & Dunlosky, 1991 ; Rawson & Dunlosky, 2007) ; ou des échelles en 100 points (e.g., Koriat, Sheffer, & Ma'ayan, 2002).

L'opérationnalisation de la mesure du jugement permet de recueillir les données qui permettent par la suite de calculer l'exactitude de l'auto-évaluation. Nous présentons ci-après différents indicateurs permettant de déterminer l'exactitude de l'auto-évaluation.

## 2.3 LES DIFFÉRENTS INDICATEURS DE L'EXACTITUDE DU JUGEMENT D'AUTO-ÉVALUATION

S'il existe plusieurs façons d'aborder le jugement d'auto-évaluation, il existe également plusieurs manières d'évaluer son exactitude (i.e. son biais). Le biais d'auto-évaluation apparaît dès lors qu'il existe une différence entre l'estimation que l'apprenant fait de sa performance et la performance elle-même. Cette différence peut être positive (l'apprenant se sur-évalue), ou négative (l'apprenant se sous-évalue), et son amplitude peut être plus ou moins grande.

L'exactitude du jugement peut être **absolue** (aussi appelée calibration, Lichtenstein, Fischhoff, & Phillips, 1982) ou **relative** (appelée résolution). Koriat et al. (2002) définissent la calibration comme 'la correspondance entre le jugement métacognitif moyen et les performances mnésiques



moyennes [...] ]. La calibration reflète l'étendue avec laquelle les jugements métacognitifs sont réalistes. Elle met en évidence, de façon globale, l'existence et l'amplitude des biais d'auto-évaluation. Mengelkamp et Bannert (2010) rappellent que les mesures de l'exactitude absolue sont préférentiellement utilisées par les chercheurs qui travaillent en situation écologique (ex : en classe, Hacker, Bol, & Bahbahani, 2008 ; Nietfeld et al., 2005). Cette mesure est également utilisée lorsqu'il est question d'améliorer de façon globale l'auto-évaluation suite à des interventions, des entraînements particuliers (Nietfeld, Cao, & Osborne, 2006) ou dans des situations où il est difficile de recueillir plusieurs jugements, notamment lorsque la tâche ne s'y prête pas (Baars et al., 2013).

L'exactitude relative (aussi appelée résolution) correspond, quant à elle, à 'l'étendue avec laquelle les jugements métacognitifs sont corrélés avec les performances mnésiques par items' (p.303 Koriat, 2007). Elle permet de déterminer avec une granularité plus fine si un apprenant est capable de distinguer les connaissances acquises de celles qui ne le sont pas. Déterminer la résolution d'un apprenant sur une activité précise nécessite de lui proposer différents contenus (e.g., plusieurs textes, plusieurs items). Il est ainsi possible d'identifier si l'apprenant s'auto-évalue avec plus d'exactitude sur certains contenus que sur d'autres. En d'autres termes, l'exactitude relative indique si un apprenant discrimine le matériel auquel il va avoir une bonne performance de celui auquel il aura une mauvaise performance (Maki, Shields, Wheeler, & Zacchilli, 2005). Contrairement à l'exactitude absolue, l'exactitude relative n'indique pas si le jugement est plus bas ou plus haut que la performance (i.e. si les apprenants se sous-évaluent ou sur-évaluent). Ainsi, un individu peut très bien avoir une mauvaise calibration, (par exemple en sur-évaluant l'état de ses connaissances), mais une très bonne résolution (il sait discriminer ce qu'il sait de ce qu'il ne sait pas) (Dunlosky & Lipko, 2007). Un apprenant peut dire, par exemple, qu'il a réussi une question et raté une autre tout en sur-estimant sa performance à la question où il pense avoir réussi et en sous-estimant sa performance à la question à laquelle il pense avoir raté.

Sur un plan opérationnel, la calibration se distingue de la résolution grâce au type d'indicateur calculé ainsi que sur la quantité de questions nécessaires au calcul de ces indicateurs. En effet, bien qu'une même question permette de calculer la résolution et la calibration, il est toutefois nécessaire pour calculer la résolution de demander à l'apprenant de s'évaluer plusieurs fois (i.e. sur plusieurs items, textes, questions, etc.). Griffin, Jee, et Wiley (2009) ont notamment mesuré l'auto-évaluation relative et absolue des participants avec une même question appliquée à plusieurs textes différents. Ils ont demandé à leurs participants de réaliser un jugement sur le nombre de questions (sur un total de 5) auxquelles ils seraient capables

de répondre pour chaque test<sup>1</sup>. Les auteurs ont ainsi pu mesurer l'exactitude de l'auto-évaluation absolue des participants en tenant compte de leurs réponses (performance et auto-évaluation) sur l'ensemble des textes ; dans un second temps ; ils ont mesuré l'exactitude de l'auto-évaluation relative en comparant l'auto-évaluation (mesurée à l'aide de la même question) à la performance texte par texte. Ces deux mesures de l'exactitude de l'auto-évaluation sont complémentaires. Toutefois, le choix de leur calcul est régi par des contraintes situationnelles et nous nous consacrerons ici à l'étude de l'exactitude absolue.

Nous présentons ci-après les principales méthodes de calcul de l'exactitude de l'auto-évaluation : deux méthodes de calcul de l'exactitude absolue et une méthode de calcul de l'exactitude relative. D'autres indicateurs existent, bien qu'ils soient, toutefois, moins utilisés du fait des nombreux inconvénients qu'ils comportent (Boekaerts & Rozendaal, 2010 ; Nelson, 1984 ; Schraw, 1995 ; Schraw, Kuch, & Gutierrez, 2013).

### 2.3.1 Calculer l'exactitude de l'auto-évaluation

#### *La différence entre le jugement et la performance*

La méthode basique pour calculer les biais d'auto-évaluation consiste à soustraire la performance au score d'auto-évaluation (sous réserve que les deux mesures partagent la même échelle). Cette méthode a été utilisée à plusieurs reprises dans diverses études (Ackerman & Lauterman, 2012 ; Baars et al., 2013 ; Baars, Gog, Bruin, & Paas, 2014 ; Bartholome & Bromme, 2009 ; Hacker et al., 2008 ; Schraw, Potenza, & Nebelsick-Gullet, 1993 ; Winne, 2004). Une différence positive reflète une sur-estimation (les individus réussissent moins bien qu'ils ne l'ont estimé) alors qu'une différence négative reflète une sous-estimation (les individus réussissent mieux qu'ils ne l'ont estimé). Par exemple, sur une échelle de 0 à 100, une performance de 100 (l'individu réussit parfaitement un exercice) associée à une auto-évaluation de 0 (l'individu est sûr de rater l'exercice) donne un biais de -100 ( $0 - 100 = -100$ ) : l'individu se sous-estime fortement.

Toutefois, ce calcul du biais moyen ne permet pas d'obtenir l'amplitude moyenne du biais étant donné que les valeurs positives et négatives peuvent s'annuler. Additionner le score d'un individu qui se sur-évalue à celui d'un individu qui se sous-évalue peut conduire à une moyenne du biais égale à 0 ; et cette moyenne refléterait une auto-évaluation exacte. Les auteurs utilisent également la valeur absolue de la différence :  $\sqrt{(0 - 100)^2}$  (Baars et al., 2013, 2014 ; Schraw et al., 1993). Sur la base de l'exemple précédent,

---

1. traduit de l'anglais : '[...] have to make a judgment about the number of questions (0-100) that they would be able to answer about each text. p.1006

cela signifie qu'une moyenne de 0 reflète une exactitude parfaite tandis qu'une moyenne de 100 reflète un biais maximal.

La première mesure permet d'avoir une idée de la direction du biais, la seconde permet d'avoir l'amplitude du biais d'auto-évaluation.

### *Les courbes de calibration*

Les courbes de calibration (cf. Figure 1) sont des graphiques représentant les biais d'auto-évaluation moyens d'un groupe d'individus afin de les comparer à une droite ( $y = x$ ) et/ou à un autre groupe (ex : comparaison de la condition 1 et 2 dans la Figure 1). La droite  $y = x$ , reflétant une calibration parfaite (représentée par la droite noire dans la Figure 1), indique que les apprenants ont rappelé 10% des items pour lesquels ils ont estimé un JOL de 10%, ont rappelé 20% des items dont le JOL était de 20% etc. Les apprenants se sous-estiment (courbe bleue sur la Figure 1) lorsque les apprenants ont rappelé plus d'items que ce que la valeur de leur JOL prédit. Les apprenants se sur-estiment (courbe orange sur la Figure 1) lorsque les apprenants ont rappelé moins d'items que ce que la valeur de leur JOL prédit. Les courbes de calibration sont utilisées pour visualiser l'amplitude du biais (i.e. l'exactitude absolue de l'auto-évaluation Dunlosky & Nelson, 1992 ; Glenberg & Epstein, 1985 ; Koriat, Lichtenstein, & Fischhoff, 1980 ; Nelson & Dunlosky, 1991). Elles sont particulièrement adaptées et utilisées lorsque les mesures de l'auto-évaluation et de la performance sont réalisées à l'aide d'une échelle d'intervalle allant généralement de 0 à 100 ou de 50 à 100 (Dunlosky & Nelson, 1992 ; Koriat et al., 2002). Le détail de leur calcul est présenté en Annexe B.

### *Le Gamma de Goodman-Kruskal*

Nelson (1984) a été le premier auteur à proposer le Gamma de Goodman-Kruskal comme mesure statistique non paramétrique pour déterminer l'exactitude relative de l'auto-évaluation. Ce coefficient de corrélation, dont le score varie de -1 à 1, permet de déterminer le degré d'association qui existe entre les mesures de la performance et de l'auto-évaluation (Dunlosky & Lipko, 2007). Le Gamma est principalement adapté à des tâches simples de mémorisation car son calcul repose sur la concordance de réponses binaires (je ne sais pas/je sais versus j'échoue/je réussis Nelson, 1984). D'autres auteurs l'ont également utilisé sur des tâches plus complexes (Glenberg & Epstein, 1987 ; Thiede et al., 2005). Cet indice représente le rapport entre le nombre de paires 'jugement-performance' en concordance sur le nombre de paires 'jugement-performance' en discordance. Un score proche de 0 signifie qu'il n'existe aucun lien entre l'auto-évaluation et la performance ; un score proche de 1 signifie que l'auto-évaluation de l'apprenant varie dans le même sens que sa per-

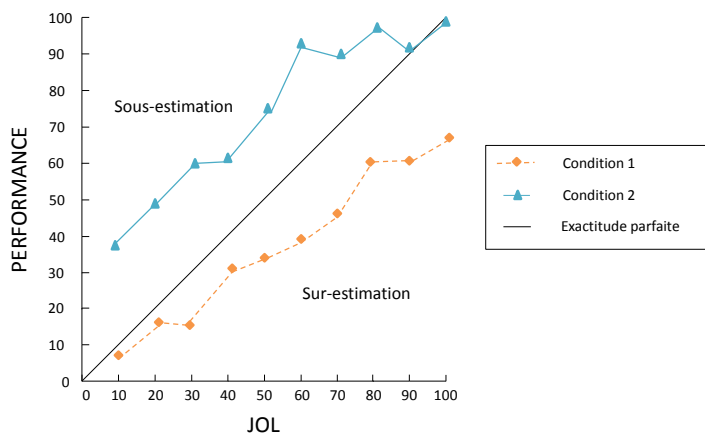


FIGURE 1. – Exemple fictif de courbe de calibration

formance (l'auto-évaluation est exacte); enfin, un score proche de -1, signifie que l'auto-évaluation varie inversement à la performance (l'auto-évaluation est inexacte). Ce score comporte des limites présentées en Annexe B. Brièvement, le Gamma ne permet pas de déterminer le sens du biais, possède une grande sensibilité aux valeurs manquantes et nécessite de disposer de plusieurs données par sujet.

#### 2.4 LES SOURCES DE L'EXACTITUDE DE L'AUTO-ÉVALUATION

S'intéresser à l'exactitude du jugement nécessite de comprendre comment les individus réalisent ce jugement, et plus précisément, d'identifier les informations qu'ils utilisent. Cette section est consacrée à la présentation de différents modèles théoriques qui se proposent d'expliquer le processus d'auto-évaluation et son exactitude.

L'un des premiers modèles explicatifs se base sur l'*hypothèse d'accès direct*, selon laquelle la force de la trace mnésique a un impact sur le jugement émis. Un second modèle se centre sur l'*hypothèse d'accessibilité*, selon laquelle la quantité d'indices utilisés détermine le jugement. Un troisième modèle se base sur le *transfert appropriate monitoring*, qui suggère que l'exactitude dépend de la similarité entre la situation de jugement et de performance. [Koriat \(1997\)](#), enfin, a proposé le *modèle d'utilisation d'indices* pour déterminer les différents indices utilisés par

les apprenants lors de la réalisation du jugement d'auto-évaluation. Ce modèle, initialement adapté aux situations d'apprentissage associatif, a été généralisé à des tâches d'apprentissage plus complexes de compréhension de texte et a donné lieu à l'*hypothèse du modèle de situation*. Ces quatre modèles sont présentés ci-après.

#### 2.4.1 *L'hypothèse d'accès direct à la trace mnésique*

L'hypothèse de l'accès direct à la trace a principalement été étudiée dans le cadre de la compréhension du Feeling of Knowing. [Hart \(1967\)](#) est le premier auteur à évoquer ce processus désigné sous le terme de MEMO (memory-monitoring). Ce processus a été mis en évidence au travers d'une étude dont le but était de montrer que des participants étaient capables de prédire relativement bien leurs performances sur une tâche de reconnaissance d'items. Les résultats montraient qu'en dessous d'un certain niveau d'encodage, les items ne laissaient pas de trace suffisamment importante en mémoire pour que le sujet puisse émettre un jugement à l'égard de cet item. Il semble donc qu'une certaine puissance de la trace mnésique soit nécessaire pour que l'apprenant puisse émettre un jugement. Selon cette hypothèse, les processus de récupération et d'évaluation de l'existence d'une information se distinguent. Toutefois, si les deux processus sont indépendants, ils doivent en théorie être influencés par le même facteur qui est la force de la trace en mémoire ([Koriat, 1993](#)). En effet, plus la trace mnésique laissée par une information est importante, plus l'apprenant sera en mesure de dire si cette information existe, mais également de la rappeler. [Mazzoni et Nelson \(1995\)](#), cependant, ont invalidé cette hypothèse. Les auteurs ont manipulé la profondeur du traitement de l'information, en faisant générer ou visualiser de l'information à des apprenants et ont montré que la condition de génération d'information améliorait le JOL mais pas la mémorisation. Les auteurs ont conclu que l'auto-évaluation et son exactitude n'étaient pas seulement issues de la force de la trace mnésique mais que d'autres facteurs expliquaient également l'exactitude du jugement.

#### 2.4.2 *L'hypothèse d'accessibilité*

L'hypothèse d'accessibilité a également été initialement proposée pour expliquer le Feeling Of Knowing ([Koriat, 1993, 1995](#)) et a par la suite été utilisée pour expliquer le Judgment Of Learning. Cette hypothèse repose sur le postulat que les jugements métacognitifs sont inférentiels et se basent sur la quantité d'information accessible par l'apprenant, ainsi que sur la vitesse d'accès à l'information ([Dunlosky et al., 2005](#)). Lors de la réalisation d'un jugement métacognitif, l'individu simule le processus

de récupération de l'information (Koriat, 1993). Cette simulation donne accès à une certaine quantité d'information, à un sentiment de connaître ou non la réponse. C'est sur la base de cette quantité d'information que l'apprenant va inférer un jugement. Koriat (1995) montre que le niveau de Feeling Of Knowing dépend de l'accessibilité des informations partielles (justes ou erronées) liées à la cible, indépendamment de leur exactitude. Dans cette perspective, l'exactitude du jugement dépend essentiellement de l'exactitude de ces réponses activées.

Selon cette approche, l'individu ne se base pas sur le contenu de l'information puisqu'il n'accède pas directement à l'information cible. En d'autres termes, l'exactitude de l'auto-évaluation ne repose pas sur la capacité de l'apprenant à accéder à l'information cible car celui-ci n'utilise que des indices associés. Par exemple, un apprenant peut, pour estimer sa performance à une question de questionnaire à choix multiple, ne pas se rappeler de la réponse immédiatement et de façon précise mais savoir que la réponse est liée à telle information, qu'elle est écrite à tel endroit, qu'elle a été apprise à tel moment, que le concept principal est phonétiquement proche de tel autre mot, etc. Plus la quantité d'indices associés à la réponse sera importante, plus le jugement pourra être élevé.

En résumé, cette hypothèse postule que l'apprenant utilise la quantité d'information disponible au moment du jugement pour le réaliser. Néanmoins, utiliser la quantité d'information sans en évaluer la qualité, ne permet pas de s'assurer que l'auto-évaluation sera exacte. L'apprenant, en effet, peut rappeler beaucoup d'informations peu prédictives de la performance et par conséquent se sur-estimer très fortement. La qualité de l'information rappelée est donc primordiale pour garantir un niveau d'exactitude élevé ; cette hypothèse postule toutefois que la qualité ne serait pas prise en compte par l'apprenant.

Dans ce cadre, le biais d'auto-évaluation a deux origines possibles. La première renvoie au fait que l'inspection est basée sur la simulation rapide et globale du processus de récupération, ce qui peut empêcher l'individu de considérer toute l'information à évaluer (Dunlosky et al., 2005). La deuxième renvoie au fait que les indices utilisés sont activés temporairement et peuvent être erronés. Ces informations ne seront donc plus disponibles quelques minutes après<sup>2</sup>. Le jugement est basé sur une quantité d'information disponible à un temps  $t$  mais qui ne le sera peut-être pas à un temps  $t+1$ .

Cette hypothèse explique finalement l'existence de biais d'auto-évaluation même s'il a été montré qu'elle n'expliquait pas à elle seule

---

2. Cette hypothèse explique l'effet positif du jugement différé sur l'exactitude de l'auto-évaluation. L'effet du jugement différé renvoie au fait que l'auto-évaluation est plus exacte quelques minutes après l'apprentissage (i.e. jugement différé) plutôt qu'immédiatement après.

l'exactitude de l'auto-évaluation associée à de la compréhension de texte (i.e. métacompréhension, voir [Dunlosky & Rawson, 2005](#)).

#### 2.4.3 *Le Transfert Approprié d'Évaluation (Transfert Appropriate Monitoring)*

Le 'Transfer Appropriate Monitoring' (ou encore appelé TAM) est dérivé de l'hypothèse des 'transferts appropriés des processus mnésiques' ([Weaver III & Kelemen, 2003](#)). Cette dernière postule que les performances mnésiques d'un individu sont meilleures dès lors qu'il se remémore les processus qu'il a utilisés pendant l'encodage au moment de la récupération de l'information ([Morris, Bransford, & Franks, 1977](#)). Par extension, le TAM stipule que l'exactitude du JOL devrait augmenter à mesure que la correspondance entre les conditions de réalisation d'un JOL et de la performance associée augmentent, tant au niveau du contexte que des processus engagés. En d'autres termes, l'augmentation de la similarité du contexte ou des processus utilisés pour réaliser le JOL et la performance doit permettre d'augmenter l'exactitude de l'auto-évaluation. Par exemple, à une tâche de rappel indicé, le JOL différé avec présentation d'un indice seul est une condition dans laquelle le JOL et la performance à la tâche sont très proches. Dans le cas d'une tâche de reconnaissance, la proximité est attestée à condition que les mesures du JOL différé et de la performance incluent toutes deux l'item cible accompagné de l'item indice.

Les résultats de diverses études, toutefois, ne permettent pas de retenir cette hypothèse pour expliquer l'exactitude du JOL dans le cadre d'un apprentissage de paires de mots, quelle que soit la similarité invoquée (des processus ou du contexte, voir [Dunlosky & Nelson, 1997](#) ; [Weaver III & Kelemen, 2003](#)). [Dunlosky et Nelson \(1997\)](#) ont mis en évidence, dans le cas d'une tâche de reconnaissance de paires de mots, que l'exactitude de l'auto-évaluation était meilleure lorsqu'il était demandé à un apprenant de s'évaluer en ne voyant que l'item indice (ex : chou - ...) comparativement à la situation où l'apprenant voyait la paire de mots (ex : chou - neige), suite à un apprentissage. La similarité des contextes entre le JOL et la performance, cependant, était plus importante dans la condition où les apprenants voyaient la paire de mots.

[Weaver III et Kelemen \(2003\)](#) ont remis en cause les résultats de [Dunlosky et Nelson \(1997\)](#) qui, à travers la simple manipulation des indices (et non des tâches), ont testé la similarité des contextes et non des processus. Si [Weaver III et Kelemen \(2003\)](#) reconnaissent la difficulté de distinguer la manipulation du contexte avec la manipulation du processus, ils argumentent qu'un moyen d'isoler les deux consiste à comparer une tâche de rappel à une tâche de reconnaissance. Les auteurs ont comparé l'exactitude de l'auto-évaluation de participants dans une condition de

rappel et de reconnaissance. Dans leur étude, les participants devaient apprendre des paires de mots puis, quelques minutes après, évaluer la probabilité de rappeler (ou de reconnaître) le mot cible (ou la paire de mots). Les participants étaient répartis selon cinq conditions différentes : (1) seul le mot indice était présenté, (2) la paire de mots était présentée (indice + réponse), (3) le mot indice était présenté seul parmi d'autres propositions où il était accompagné de réponses incorrectes, (4) la paire de mots était présentée parmi d'autres paires présentant le mot indice accompagné de réponses incorrectes, (5) cette condition était similaire à la 4 mais la paire correcte était marquée d'une astérisque. Pour valider le TAM, l'exactitude de l'auto-évaluation devrait être la plus exacte dans la condition 1 pour une tâche de rappel, et dans la condition 4 pour une tâche de reconnaissance. En effet, les conditions 1 et 4 sont respectivement exactement les mêmes que ce qui est demandé aux participants lors des tests de rappel et de reconnaissance. Les résultats, toutefois, vont à l'encontre des hypothèses issues du TAM, car si le jugement d'auto-évaluation pour une tâche de rappel est effectivement plus exact dans la première condition, il l'est également dans les conditions 3 et 4. En ce qui concerne la reconnaissance, les conditions 1 et 3 (qui nécessitent un processus de récupération propre au rappel) se sont avérées plus efficaces sur l'exactitude de l'auto-évaluation que la quatrième.

Par conséquent, il apparaît que le TAM n'est pas une hypothèse valide pour expliquer l'exactitude de l'auto-évaluation sur des tâches d'apprentissage simples. [Weaver III et Kelemen \(2003\)](#) préservent l'éventualité, néanmoins, que cette hypothèse puisse expliquer l'exactitude de l'auto-évaluation sur des tâches plus complexes.

[Dunlosky et Rawson \(2005\)](#) ont par la suite cherché à déterminer l'hypothèse la plus à même d'expliquer l'exactitude de la métacompréhension (i.e. auto-évaluation associée à une tâche de compréhension), en confrontant l'hypothèse d'accessibilité et le TAM. Si le TAM est une hypothèse valide, l'exactitude de l'auto-évaluation doit s'améliorer lorsqu'il est demandé à l'apprenant de réaliser un processus similaire au moment d'émettre un jugement et au moment de réaliser le test. Si l'hypothèse d'accessibilité est valide, la similarité des processus ne doit pas suffire à améliorer l'exactitude de l'auto-évaluation. En effet, les apprenants pourraient s'engager dans un processus similaire, rappeler beaucoup d'informations lors de la réalisation du jugement et être très confiants dans leurs capacités à donner une bonne réponse, alors même que leur performance serait basse. Selon l'hypothèse d'accessibilité, le fait que l'individu utilise la quantité d'information rappelée ne suffit pas à produire une auto-évaluation exacte. Pour que celle-ci soit exacte, il faut que l'information qui émane des processus réalisés lors du jugement et de la performance le soit également.



Afin de tester ces deux hypothèses, Dunlosky et Rawson (2005) ont demandé à la moitié des participants prenant part à leur étude de réaliser un pré-rappel avant d'émettre un jugement spécifique sur leur capacité à rappeler une définition apprise dans un texte. Ce pré-rappel simule les conditions de test et engage les apprenants dans un processus similaire lors de la réalisation du jugement métacognitif. L'ensemble des participants prenait ensuite part à un test visant à mesurer leurs performances finales sur le rappel des définitions apprises. Les résultats montrent que les apprenants qui ont effectué un pré-rappel avant de réaliser un JOL spécifique à chaque définition s'auto-évaluaient avec plus d'exactitude (bien que leur degré d'exactitude ne soit pas parfait) que les participants n'ayant pas effectué le pré-rappel. Toutefois, la qualité de l'information rappelée avant de réaliser un JOL n'influçait pas son exactitude. Il semble finalement que l'analyse de la qualité de l'information rappelée n'ait pas permis d'améliorer l'exactitude de l'auto-évaluation.

Les auteurs ont conclu que l'hypothèse d'accessibilité apparaît comme plus adéquate pour expliquer ces résultats que le TAM. En effet, les performances au pré-rappel (réalisé avant le JOL) et au test final étaient très corrélées ( $r = .9$ ), ce qui signifie que les processus et/ou le contexte utilisé au moment du jugement métacognitif et ceux utilisés au moment de la performance étaient très semblables. Bien que les auteurs aient constaté une amélioration dans l'exactitude du JOL, cette amélioration n'est toutefois pas aussi importante que peut l'être la correspondance entre les processus de jugement et de test. Aussi, les auteurs critiquent le TAM du fait qu'il ne précise pas la manière dont l'apprenant traite l'information pour auto-évaluer ses performances. De même, les causes de l'amélioration observée lorsque les processus effectués lors du jugement et de la performance se ressemblent ne sont pas expliquées. L'hypothèse d'accessibilité, au contraire, explique la manière dont les apprenants utilisent l'information pour générer un jugement. Ainsi, la quantité, et non la qualité de l'information, serait utilisée comme indice. L'exactitude de l'auto-évaluation s'améliore lorsque des informations extraites sont valides (i.e. prédictives de la performance). Enfin, il est à noter qu'une conséquence négative de l'incitation à générer de l'information est que si l'apprenant génère de l'information fautive, l'exactitude de son auto-évaluation diminue.

#### 2.4.4 *L'hypothèse d'utilisation d'indices*

Après avoir étudié l'hypothèse d'accessibilité (Koriat, 1993, 1995), Koriat (1997) a proposé le modèle d'utilisation d'indices (cf. Figure 2) pour expliquer comment un apprenant émet un jugement métacognitif de type JOL. Selon ce modèle, le JOL repose sur des inférences et des heuris-

tiques elles-mêmes basées sur l'utilisation de divers indices intrinsèques, extrinsèques, ou mnémoniques.

- Les indices intrinsèques relèvent des caractéristiques relatives aux items à étudier, notamment leur difficulté perçue. L'apprenant peut utiliser la difficulté de l'information à apprendre pour s'auto-évaluer (ex : cette leçon est difficile, j'ai peu de chance de m'en rappeler).
- Les indices extrinsèques renvoient aux facteurs associés à la condition d'apprentissage ou aux opérations d'encodage appliquées par l'apprenant (e.g., Des perceptions mnésiques issues de la situation d'apprentissage, nombre d'essais d'apprentissage, stratégies d'encodage ...). L'apprenant peut utiliser les informations relatives à la situation d'apprentissage pour émettre un jugement d'auto-évaluation (ex : Je n'étais pas très concentré et il y avait du bruit, je n'ai pas bien appris ma leçon).
- Les indices mnémoniques indiquent à quel point un item a été appris et sera rappelé dans le futur (e.g., la rapidité de l'accès à l'information pertinente, la facilité de traitement des indices présentés, la mémoire des performances de précédentes tentatives de rappel ...). L'apprenant peut utiliser des informations relatives à la force de la trace mnésique pour émettre un jugement d'auto-évaluation (ex : j'ai déjà réussi à rappeler cette information plusieurs fois, je réussirais à la rappeler à nouveau).

Les indices intrinsèques et extrinsèques peuvent avoir une influence directe sur le JOL si les apprenants les utilisent pour procéder à un traitement analytique, conscient, coûteux et s'auto-évaluer sur la base de règles, de théories ou de croyances qui leurs sont propres. Par exemple, l'apprenant sait qu'il a généralement une bonne mémoire ; qu'une tâche de rappel indicé est plus simple qu'une tâche de rappel libre ; qu'un mot qu'il ne connaît pas sera moins facile à retenir qu'un mot qu'il connaît, etc. Ces indices peuvent également avoir une influence médiée par les indices mnémoniques. Le processus utilisé est alors qualitativement différent puisque les indices mnémoniques sont issus d'heuristiques basées sur l'expérience ou l'intuition des apprenants. En d'autres termes, l'utilisation de processus mnémoniques se fait de façon non consciente, peu coûteuse, à l'aide d'inférences implicites.

L'exactitude du jugement d'auto-évaluation dépend alors de la 'corrélation empirique' qui existe entre les indices utilisés pour ces inférences et les critères impliqués dans la performance, c'est-à-dire de la validité prédictive des indices envers la performance. En d'autres termes, le jugement d'auto-évaluation sera exact si l'apprenant utilise des indices qui affectent effectivement la performance. À l'inverse, l'utilisation d'un indice peu prédictif de la performance devrait conduire à une faible correspondance entre cet indice et la performance. [Koriat \(1997\)](#) précise également que

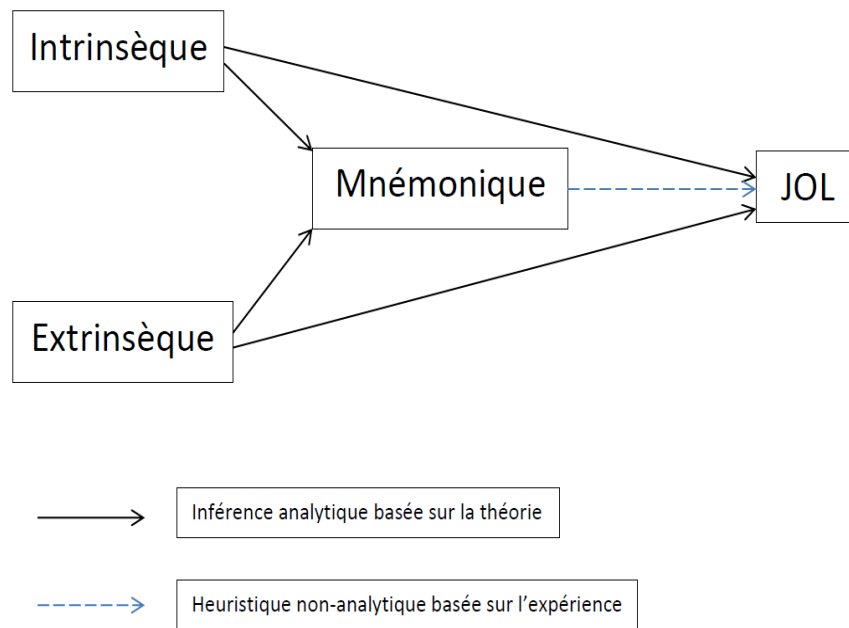


FIGURE 2. – Le modèle de l’auto-évaluation de Koriat

les inexactitudes dans le jugement peuvent être liées à une ‘théorie’ ou croyance fautive qu’a l’apprenant sur les situations d’apprentissage (ex : il suffit de lire une fois un contenu pour le maîtriser) ainsi qu’à l’existence de biais ‘inhérents à l’application d’heuristiques particulières’ (p.350). Par exemple, les apprenants utilisent parfois les informations les plus présentes à l’esprit pour prendre une décision ou dans le cas de l’auto-évaluation, pour juger un état d’apprentissage, reflétant le biais de disponibilité. Ces informations, cependant, ne sont peut-être pas prédictives de la performance. Par exemple, dans le cas de l’auto-évaluation suite à la lecture d’un texte long mais simple, un apprenant peut estimer avoir de mauvaise performance s’il a en tête des informations sur la longueur d’un texte plutôt que sur sa difficulté. Si le texte est simple, toutefois, l’apprenant peut obtenir une bonne performance.

Afin de déterminer les indices utilisés par les apprenants, Koriat (1997) a manipulé ces indices, à savoir le nombre de répétition d’apprentissage (indice extrinsèque), et la difficulté des items (indices intrinsèques) puis a mesuré l’exactitude de l’auto-évaluation et son évolution. Par exemple, l’auteur émettait l’hypothèse, dans une première étude, que l’apprenant ne va pas considérer le nombre de répétition d’apprentissage pour réaliser son jugement métacognitif si les indices extrinsèques sont peu utilisés, comparativement aux indices intrinsèques. Dès lors, si l’apprenant révise une liste déjà apprise sans considérer ce facteur pour prédire ses perfor-

mances, le jugement ne doit pas évoluer mais les performances doivent s'améliorer. En conséquence, l'exactitude absolue de l'auto-évaluation va diminuer. Les résultats obtenus confirmaient la diminution de l'exactitude absolue de l'auto-évaluation avec la répétition de l'apprentissage d'une même liste de mots ; les apprenants utiliseraient davantage les indices intrinsèques qu'extrinsèques. En parallèle, l'auteur postule également que la pratique répétée d'une même liste de mots doit augmenter la sensibilité de l'apprenant à la différence d'apprentissage inter-items et ainsi améliorer l'exactitude relative. Selon l'auteur, l'amélioration de l'exactitude relative reflèterait une modification des indices utilisés. Les apprenants baseraient initialement leur jugement sur l'évaluation directe de la difficulté des items (indices intrinsèques). Avec la pratique, les apprenants seraient de plus en plus sensibles à la facilité de traitement ou l'accessibilité de l'item (indices mnémoniques), plus prédictifs de la performance. Par conséquent l'exactitude relative du jugement d'apprentissage devrait s'améliorer. L'auteur confirme son hypothèse, observant une amélioration de l'exactitude relative avec la répétition de l'apprentissage d'une même liste de mots. Il conclut que 'ce pattern corrélationnel entre les JOLs, le rappel et la difficulté des items a changé systématiquement avec la pratique d'une façon qui suggère une dépendance décroissante aux indices intrinsèques et une dépendance croissante aux indices mnémoniques, internes' (traduit de l'anglais : 'the correlational pattern between JOLs, recall, and item difficulty changed systematically with practice in a manner suggesting decreased reliance on intrinsic cues and increased reliance on internal, mnemonic cues.' p.363).

Le modèle d'utilisation d'indices fait encore à ce jour office de modèle de référence pour rendre compte de l'exactitude de l'auto-évaluation (Dunlosky & Matvey, 2001). Ce succès peut s'expliquer par le fait que le modèle autorise une grande diversité de déterminants au JOL et de son exactitude selon le poids respectif de chacun des indices. Ce modèle a en premier lieu été testé avec des tâches d'apprentissage simples telles que des tâches d'apprentissage de paires de mots, pour être par la suite adapté et généralisé à des tâches plus complexes de compréhension en se basant sur le modèle de la compréhension de Kintsch (1988), (Griffin et al., 2009 ; Rawson, Dunlosky, & Thiede, 2000 ; Thiede, Griffin, Wiley, & Anderson, 2010 ; Zaromb, Karpicke, & Roediger III, 2010). Cette adaptation du modèle est présentée ci-après.

#### 2.4.5 *L'hypothèse du modèle de situation*

Afin d'adapter le modèle de Koriat à des tâches de compréhension, Rawson et al. (2000) l'ont associé au modèle de la compréhension de Kintsch (1988). Le modèle de Kintsch (1988) décompose la compréhens-

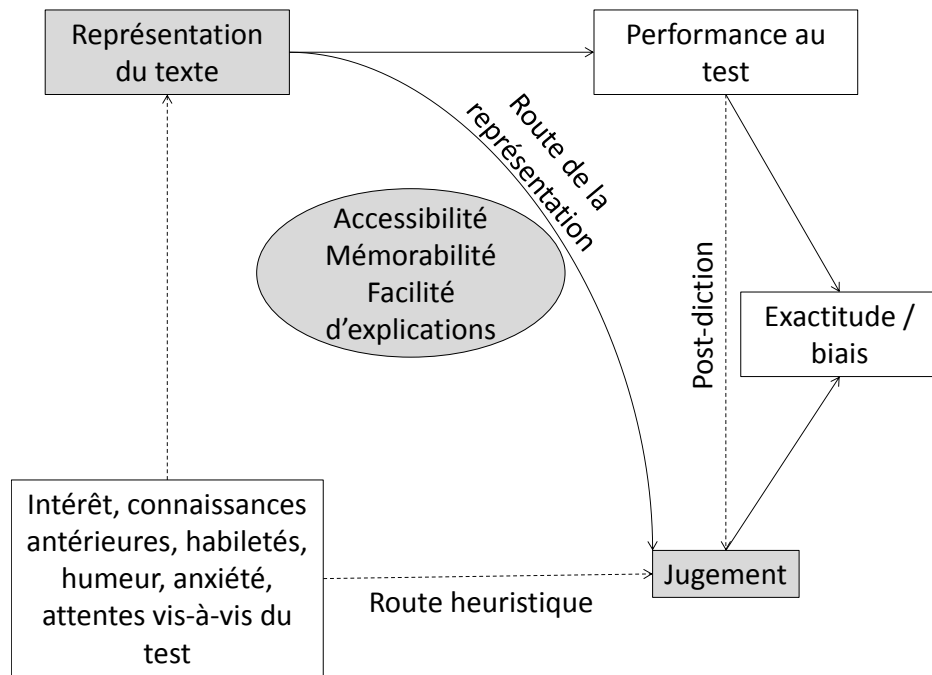


FIGURE 3. – Routes potentielles pour réaliser un jugement d’auto-évaluation dans le cadre d’une tâche de compréhensions (adapté de Griffin et al., 2009).

sion en trois niveaux : le niveau lexical, le niveau de base et le niveau du modèle de situation. Le niveau lexical correspond au décodage des mots et des phrases. Au niveau de base, les mots et phrases sont regroupés pour former des propositions (i.e. des ensembles signifiants), et les relations entre les concepts sont explicitement présentées dans le texte. Le niveau du modèle de situation constitue, quant à lui, le niveau de compréhension le plus élevé. Il requiert de l’apprenant qu’il engage des connaissances antérieures afin de faire des inférences implicites dans le texte. Les individus possédant un haut niveau de connaissances antérieures sont ainsi avantagés, étant en mesure de faire plus d’inférences (McNamara, Kintsch, Songer, & Kintsch, 1996). Lors de la mesure des performances de compréhension, le niveau du modèle de situation est le plus souvent évalué.

En lien avec le modèle de Koriat, il est nécessaire, lors du jugement d’auto-évaluation, que l’apprenant utilise des indices qui soient basés sur ce modèle de situation puisque ce niveau de compréhension prédit le mieux les performances finales (Griffin et al., 2009 ; Rawson et al., 2000). Griffin et al. (2009) ont proposé une schématisation (cf. Figure 3) qui illustre l’idée que la représentation que l’individu se construit du texte détermine directement la performance et le jugement d’auto-évaluation.

Dans cette perspective, des études ont été réalisées afin de développer des méthodes et outils pour améliorer et rendre plus saillant le modèle de situation, afin que celui-ci puisse être utilisé par les apprenants pour estimer leur compréhension (M. Anderson & Thiede, 2008 ; Redford, Thiede, Wiley, & Griffin, 2012 ; Thiede, Anderson, & Therriault, 2003 ; Thiede & Anderson, 2003 ; Thiede et al., 2005, 2010) . Ces méthodes et outils font l'objet d'un développement dans le chapitre suivant.

## 2.5 SYNTHÈSE DES APPORTS DES HYPOTHÈSES À L'EXPLICATION DE L'INEXACTITUDE DE L'AUTO-ÉVALUATION

Plusieurs hypothèses ont été élaborées afin d'expliquer la façon dont les apprenants émettent un jugement d'auto-évaluation. L'hypothèse de l'accès direct à la trace a été étudiée en lien avec le jugement d'auto-évaluation sur des paires de mots. Cette hypothèse, stipulant que le jugement repose sur la force de la trace mnésique d'une information en mémoire, a rapidement été invalidée du fait que des stratégies d'encodage avaient un effet différentiel sur le jugement d'auto-évaluation et la performance (Mazzoni & Nelson, 1995). Le 'transfer appropriate monitoring' (TAM) suppose que l'exactitude de l'auto-évaluation peut être améliorée lorsque le contexte et/ou les processus utilisés lors du jugement d'auto-évaluation sont les mêmes que ceux utilisés pour réaliser la performance. Cette hypothèse s'est également révélée inadaptée pour expliquer le jugement d'auto-évaluation sur une tâche simple d'apprentissage (Weaver III & Kelemen, 2003) ainsi que sur une tâche plus complexe de compréhension (Dunlosky & Rawson, 2005). En effet, bien que le TAM permette d'expliquer l'amélioration de l'exactitude de l'auto-évaluation lorsque les conditions de réalisation du jugement d'auto-évaluation sont similaires aux conditions de réalisation des performances, il n'explique pas comment cette amélioration est possible, ni comment le jugement est réalisé.

L'hypothèse d'accessibilité, quant à elle, postule que le jugement d'auto-évaluation repose sur la quantité d'information disponible. L'exactitude de l'auto-évaluation dépend alors de la qualité de l'information disponible. Si l'apprenant dispose d'informations valides prédictives des performances futures, l'auto-évaluation sera exacte sinon elle sera inexacte. L'hypothèse d'utilisation d'indices de Koriat (1997) est consistante avec cette notion de qualité des indices utilisés, tout en étant plus précise dans la description du processus cognitif d'évaluation des performances. Cette hypothèse définit trois catégories d'indices (intrinsèques, extrinsèques et mnémoniques) qui reposent sur l'utilisation de deux processus distincts (analytique et heuristique). L'auto-évaluation est d'autant plus exacte que les apprenants utilisent des indices prédictifs de la performance. L'hypothèse d'utilisation d'indices de Koriat (1997) est adaptée à des tâches simples d'apprentissage

TABLEAU 2. – Synthèse des principales hypothèses explicatives de l'auto-évaluation.

Nom du modèle	Type de tâche concernée	Principe
L'hypothèse d'accès à la trace	apprentissage de paires de mots	Le jugement se base sur la force de la trace mnésique (e.g., <a href="#">Koriat, 1993</a> )
L'hypothèse d'accessibilité	apprentissage de paires de mots / compréhension de textes	Le jugement se base sur la quantité d'indices disponibles (e.g., <a href="#">Dunlosky et al., 2005</a> )
Le transfer appropriate monitoring	compréhension de textes	l'exactitude dépend de la similarité des processus/contextes entre le jugement et la performance. Pas d'explication sur ce qui est à l'origine du jugement (e.g., <a href="#">Weaver III &amp; Kelemen, 2003</a> )
L'hypothèse d'utilisation d'indices	apprentissage de paires de mots	le jugement se base sur des indices intrinsèques/extrinsèques/mnémoriques ( <a href="#">Koriat, 1997</a> )
L'hypothèse du modèle de situation	compréhension de textes	le jugement se base sur des niveaux de compréhension du textes (e.g., <a href="#">Baars et al., 2013</a> ; <a href="#">Griffin et al., 2008</a> )

mais a été ensuite associée au modèle de la compréhension de [Kintsch \(1988\)](#) pour créer l'hypothèse du modèle de la situation, adaptée à des tâches plus complexes de compréhension. Cette hypothèse avance que l'apprenant doit se baser sur le modèle de situation du texte (niveau le plus élaboré de compréhension) pour s'auto-évaluer avec exactitude.

Le modèle d'utilisation d'indices pourrait également être généralisé à d'autres tâches d'apprentissage complexes qui n'impliquent pas uniquement la compréhension de textes linéaires. Quel que soit le contenu et/ou le support d'apprentissage, un apprenant peut réaliser des apprentissages superficiels ou en profondeur, tels qu'[Ausubel \(1960\)](#) les a distingués dans sa théorie de l'apprentissage significatif (en anglais 'meaningful learning'). L'apprentissage profond suppose que les connaissances nouvelles ont été reliées aux plus anciennes pour être intégrées à la mémoire à long terme de l'apprenant, tandis que les connaissances apprises par cœur sont stockées indépendamment des autres connaissances existantes. Solliciter un apprentissage profond permettrait à l'apprenant d'obtenir de meilleures performances, comme l'ont montré [Paas \(1992\)](#) ; [J. J. Van Merriënboer \(1990\)](#) ; [J. Van Merriënboer, Schuurman, De Croock, et Paas \(2002\)](#) sur des tâches de résolution de problème. Aussi, de la même façon qu'encourager la construction du modèle de situation permet à l'apprenant d'en extraire des indices pour produire une auto-évaluation plus exacte, inciter l'apprentissage profond devrait également permettre à l'apprenant d'utiliser des indices plus prédictifs de la performance pour améliorer l'exactitude de son auto-évaluation.

## 2.6 LE COÛT COGNITIF DE L'AUTO-ÉVALUATION

Différents points de vue s'opposent, qui concernent le coût cognitif de l'auto-évaluation ([Ehrlich & Cahour, 1991](#) ; [M. V. J. Veenman, Van Hout-Wolters, & Afflerbach, 2006](#)). Selon certains auteurs, les processus métacognitifs doivent être conscients (délibérés) pour être considérés comme tels (e.g., [Nelson, 1996](#)). D'autres auteurs, à l'inverse, pensent que les processus métacognitifs peuvent sortir du champ de la conscience lorsque tout se déroule sans encombre et redevenir conscients dès lors qu'une difficulté est rencontrée ([M. Veenman, Prins, & Elshout, 2002](#) ; [M. V. J. Veenman et al., 2006](#)). Enfin, une troisième catégorie de chercheurs avance l'idée qu'il n'existe pas de processus métacognitifs distincts des processus cognitifs ([Kalyuga, 2009](#)). Plusieurs arguments semblent converger vers l'idée que le processus d'auto-évaluation pourrait nécessiter des ressources dans certaines situations uniquement, notamment les situations les plus complexes, ou lorsque l'apprenant est novice d'un domaine.

Premièrement, une étude récente a montré que les performances d'un individu qui réalisait des Sudokus étaient plus faibles lorsque celui-ci devait



auto-évaluer ses performances simultanément à la résolution de la tâche, et ce uniquement lorsque que le Sudoku était difficile (Van Gog, Kester, & Paas, 2011). Ce résultat, basé sur le paradigme de la double tâche, montre que des individus invités à évaluer leurs performances simultanément à une tâche principale éprouvent des difficultés à allouer leurs ressources cognitives à cette tâche ; les ressources étant accaparées par le processus d'auto-évaluation. L'auto-évaluation, cependant, ne demande pas suffisamment de ressources pour perturber les apprenants lors de la résolution d'une tâche simple. Dans cette étude, l'exactitude de l'auto-évaluation n'a pas été mesurée et il n'est pas possible de déterminer si la disponibilité des ressources cognitives affecte ou non l'exactitude de l'auto-évaluation.

En second lieu, Koriat (1997) avance que le processus d'auto-évaluation est un jugement métacognitif qui repose sur l'évaluation d'indices de différentes natures (situation, matériel, ressenti, etc.). Plus les indices utilisés sont prédictifs de la performance, plus l'auto-évaluation sera exacte. Selon cette hypothèse, les indices sont extraits pendant l'apprentissage même si leur pondération n'a lieu qu'au moment de répondre à la question du jugement d'auto-évaluation. L'auteur met en évidence le fait que certains indices (intrinsèques et extrinsèques) sont issus de règles utilisées par l'apprenant (ex : des lettres qui forment un mot connu de l'apprenant sont plus faciles à mémoriser que des lettres qui ne forment pas un mot connu) et que l'utilisation de ces règles requiert l'engagement d'un processus analytique, coûteux en ressources cognitives. D'autres indices dits 'mnémoniques' ne demanderaient au contraire pas ou peu de ressources car ils émergeraient de processus heuristiques basés sur des intuitions (Koriat, 1997).

Si l'utilisation des indices a lieu pendant l'apprentissage et nécessite des ressources cognitives, il est probable que l'apprenant, dont les ressources sont accaparées par l'activité de traitement de l'information à apprendre, ne soit pas en mesure d'extraire ou d'utiliser des indices pertinents. Son jugement sera dès lors biaisé. Les arguments présentés ci-dessus tendent à confirmer l'idée que l'auto-évaluation est un processus coûteux qui peut être biaisé par nature, du simple fait qu'il soit concomitant à l'apprentissage. Il existe, toutefois, des éléments suggérant que le coût de l'auto-évaluation peut diminuer avec l'acquisition d'une certaine expertise dans un domaine spécifique.

Koriat (1997) a montré, pour des tâches d'apprentissage associatif, que le passage de l'utilisation d'indices intrinsèques et extrinsèques (i.e. issus de processus analytiques coûteux) à l'utilisation d'indices mnémoniques (i.e. issus de processus heuristiques non coûteux) se produit suite à la pratique répétée de l'apprentissage d'un même contenu. Cet argument est cohérents avec les théories de l'expertise qui avancent qu'un processus répété à de multiples reprises va progressivement s'automatiser et demander de moins en moins de ressources cognitives (Leplat, 2005). Selon

Flavell (1979), les connaissances métacognitives auraient les mêmes caractéristiques que les connaissances, ce qui laisse supposer que les habiletés métacognitives pourraient également s'automatiser. Ce changement de processus signifierait que le processus d'auto-évaluation chez des apprenants experts d'un domaine serait moins coûteux et plus efficace que celui de novices, pour un contenu équivalent.

Le constat d'une meilleure exactitude de l'auto-évaluation chez les apprenants experts que chez les novices a été fait à de multiples reprises. Griffin et al. (2009) ont comparé les performances et l'exactitude absolue et relative de l'auto-évaluation d'étudiants avec différents niveaux d'expertise sur le domaine du baseball qui avaient pour instruction de lire des textes relatifs au baseball ou à un autre domaine. Les auteurs ont montré que le niveau d'expertise était lié à l'exactitude absolue mais pas à l'exactitude relative. Les apprenants les plus experts étaient plus exacts dans l'évaluation globale de leurs performances mais ne discriminaient pas mieux que les novices ce qu'ils savaient de ce qu'ils ne savaient pas. Il n'apparaît pas que les experts étaient plus exacts du fait que le processus d'auto-évaluation est moins coûteux cognitivement (il n'existe, à notre connaissance, aucune étude ayant montré ce phénomène). Les auteurs proposent, en revanche, que les experts auraient plus de facilité à accéder au modèle de situation. L'exactitude de l'auto-évaluation plus élevée chez les apprenants avec un niveau élevé de connaissances antérieures s'expliquerait potentiellement par un accès plus facile aux indices pertinents issus du modèle de situation. Dans une étude plus ancienne, Chi et Ceci (1978) ont comparé l'exactitude de l'auto-évaluation chez des joueurs d'échec ayant différents niveaux d'expertise. Les joueurs devaient mémoriser des patterns de pièces d'échec qui formaient une combinaison cohérente ou non selon les principes d'une partie d'échec. Les résultats indiquaient que les joueurs les plus experts ont prédit avec plus d'exactitude leurs performances pour les patterns cohérents que pour les patterns non cohérents. Chez les moins experts, la différence d'exactitude entre les patterns cohérents et non cohérents était moins marquée. Les résultats montraient également que l'exactitude de l'auto-évaluation augmentait avec l'expertise bien que les participants avaient toujours tendance à surestimer leurs performances.

En résumé, si la question du coût cognitif de l'auto-évaluation n'a pas de réponse établie, il semblerait que le coût soit indéniable chez les apprenants novices d'un domaine et moins important chez les individus experts d'un domaine d'apprentissage. Les experts, du fait de leur structure de connaissances enrichie, ont potentiellement le double avantage de voir leur processus d'auto-évaluation ainsi que leur processus de compréhension leur demander moins de ressources cognitives (Chi, 2006). Ces arguments expliqueraient pourquoi les experts d'un domaine s'avèrent être plus exacts dans l'auto-évaluation qu'ils font de leurs performances (Chi

& Ceci, 1978 ; Griffin et al., 2009 ; Kostons et al., 2010). Cette conclusion, néanmoins, est à considérer avec prudence, étant donné que certaines études rapportent une auto-évaluation plus exacte chez des novices que chez des experts (e.g., [Glenberg & Epstein, 1987](#)).

Les éléments présentés ici laissent penser que les capacités d'auto-évaluation sont fortement dépendantes du domaine d'expertise et seraient donc spécifiques à un domaine délimité. Diverses recherches, toutefois, suggèrent que les processus d'auto-évaluation et plus généralement les habiletés métacognitives sont transférables à divers domaines de compétence. Cette question de la transférabilité des habiletés peut également avoir un impact sur la conception des outils d'aide à l'auto-évaluation. Il est important de savoir, en effet, si les outils doivent être conçus de manière à permettre l'acquisition de compétences métacognitives générales qui pourront se transférer d'un domaine à un autre où s'ils doivent être spécifiés à un domaine précis. Cette question est abordée dans la section suivante.

## 2.7 L'AUTO-ÉVALUATION : UNE HABILITÉ TRANSVERSE OU SPÉCIFIQUE À UN DOMAINE ?

[Schraw \(1998\)](#) traitent la question de la transférabilité des habiletés métacognitives sous l'angle de la similarité entre les habiletés métacognitives et cognitives. Les auteurs distinguent au moins deux hypothèses distinctes. La première est que les habiletés cognitives et métacognitives ne sont pas différentes les unes des autres et sont intrinsèquement liées. Dans ce cas, les habiletés métacognitives ne sont pas transférables à d'autres domaines de compétences. La seconde est que les habiletés cognitives et métacognitives sont de nature différente. Les habiletés métacognitives peuvent donc se transférer à plusieurs domaines de compétence. Un troisième point de vue existe, qui suggère qu'une partie des habiletés métacognitives est spécialisée aux domaines d'expertises mais qu'il existe également des habiletés métacognitives qui sont générales et qui permettent à l'apprenant d'assurer un processus d'autorégulation minimum quel que soit son niveau de compétence dans un domaine.

### 2.7.1 *Les habiletés métacognitives sont spécifiques à un domaine de compétence*

La première hypothèse est celle de 'l'encapsulation stricte' (de l'anglais 'strict encapsulation hypothesis' ; pour une revue voir [Schraw, 1998](#)). Cette hypothèse avance que les habiletés cognitives et métacognitives ne sont qualitativement pas différentes les unes des autres. Les habiletés métacognitives sont par conséquent spécifiques à un domaine de connaissance et

non transférables à un autre. Comme le souligne (Huet, 2004), les auteurs qui adoptent ce point de vue expliquent leur propos sur la base des travaux sur l'expertise. Ainsi, les habiletés métacognitives des experts seraient très spécifiques à un domaine et, par conséquent, non transférables à d'autres domaines (Glaser, Chi, & Farr, 1988). Cette possibilité expliquerait qu'un individu expert dans un domaine s'auto-évalue avec plus d'exactitude sur son domaine d'expertise que sur un autre domaine dans lequel il n'est pas expert (e.g., Griffin et al., 2009). D'autres études suggèrent, à l'inverse, que les compétences métacognitives sont généralisables à plusieurs domaines (Schraw & Nietfeld, 1998 ; Schraw, 1998).

### 2.7.2 *Les habiletés métacognitives sont générales à plusieurs domaines de compétences*

Ces travaux s'appuient sur la seconde hypothèse, diamétralement opposée à la première, selon laquelle les habiletés métacognitives sont générales (i.e. non spécifiques à un domaine), et qu'un individu n'a pas besoin de connaissances spécifiques dans un domaine pour s'auto-évaluer avec exactitude. Cette hypothèse, soutenue par plusieurs auteurs pour qui les stratégies métacognitives sont transsituationnelles (e.g., Schraw, Dunkle, Bendixen, & Roedel, 1995), suppose que la nature des habiletés cognitives et métacognitives est différente. Par conséquent, ces dernières peuvent s'appliquer à plusieurs domaines quel que soit le niveau de compétence de l'apprenant. Cette approche s'est vu confirmée par des travaux qui ont mis en évidence des corrélations positives entre les valeurs de biais d'auto-évaluations dans des domaines mobilisant aussi bien l'intelligence fluide que cristallisée (Schraw & Nietfeld, 1998), ou des biais similaires sur des domaines différents (Schraw et al., 1995). Les résultats de ces travaux tendent à démontrer la transférabilité des habiletés métacognitives par les corrélations positives entre jugement d'auto-évaluation obtenus dans divers domaines de compétence ; toutefois, ils ne permettent pas d'expliquer pourquoi les auto-évaluations des individus peuvent être plus exactes dans leur domaine d'expertise (Griffin et al., 2009).

### 2.7.3 *Du général au spécifique*

L'existence de corrélations entre les biais d'auto-évaluation sur plusieurs domaines différents (Schraw et al., 1995), parallèlement au fait que les experts s'auto-évaluent avec plus d'exactitude sur leur domaine de compétence que sur d'autres domaines (Griffin et al., 2009), contredisent les deux points de vue abordés précédemment.

Une troisième approche concevrait que non seulement les apprenants disposent d'habiletés métacognitives générales leur permettant de réguler

leurs processus cognitifs en toutes circonstances mais également que les experts, en se spécialisant dans un domaine de compétence, développent des habiletés métacognitives spécifiques. Cette approche concorde avec de nombreux travaux (e.g., Schraw, 1998 ; M. V. Veenman, Elshout, & Meijer, 1997) qui ont permis de montrer que les habiletés générales seraient propres aux novices d'un domaine et que les habiletés métacognitives continuent de se développer avec l'acquisition de l'expertise (M. Veenman & Elshout, 1999).

Finalement, aucune étude ne permet véritablement de trancher sur la pertinence d'une hypothèse spécifique. Il semblerait, néanmoins, que l'apprenant dispose de règles générales et transférables lui permettant d'auto-évaluer ses performances quelle que soit la situation d'apprentissage. L'acquisition d'une expertise dans un domaine permettrait par la suite de spécialiser ses habiletés métacognitives (et/ou connaissances métacognitives) afin de rendre le processus plus efficace (Chi & Ceci, 1978). Toutefois, en considérant l'hypothèse d'un coût cognitif du processus d'auto-évaluation, il est également envisageable que les experts aient autant de connaissances et/ou habiletés métacognitives que les novices, mais que la qualité de leur structure cognitive leur permette de libérer des ressources cognitives pour les utiliser plus efficacement et ainsi améliorer l'exactitude de leur auto-évaluation.

Nous avons présenté, dans ce chapitre, le processus d'auto-évaluation et les caractéristiques qui le définissent. Il est désormais important d'explorer les solutions existantes permettant d'améliorer l'exactitude de l'auto-évaluation. Le chapitre suivant est consacré à cette question.

Nous avons vu que :

- L'auto-évaluation est un processus métacognitif lié à l'autorégulation de l'apprentissage. Du processus d'auto-évaluation résulte un jugement métacognitif dont l'exactitude influence les comportements d'apprentissage ainsi que les performances. Améliorer l'exactitude de l'auto-évaluation constitue donc un enjeu majeur pour les différents acteurs du monde éducatif.
- L'auto-évaluation peut s'étudier sur différents types de tâches parmi lesquelles la mémorisation d'item, la compréhension de texte, la résolution de problème.
- Différentes hypothèses ont été formulées pour tenter d'expliquer la formation du jugement d'auto-évaluation. Aucune n'explique l'auto-évaluation dans sa globalité, toutefois, l'hypothèse du modèle de situation est la plus adaptée aux situations écologiques les tâches de compréhension de textes.
- L'auto-évaluation semble nécessiter des ressources cognitives pour être réalisée correctement, plus particulièrement lorsque la situation d'apprentissage est complexe ou que l'apprenant ne dispose pas de connaissances antérieures sur le domaine.
- Il existerait des connaissances métacognitives générales, indépendantes des connaissances cognitives qui permettraient aux apprenants d'auto-évaluer leurs apprentissages quel que soit le domaine mais également des habiletés métacognitives encapsulées dans les habiletés cognitives qui permettraient aux experts d'être plus exacts que les novices.



# 3

---

## AMÉLIORER L'EXACTITUDE DE L'AUTO-ÉVALUATION

---

Nous l'avons vu précédemment, de nombreuses études ont mis en évidence la difficulté qu'ont les apprenants à évaluer avec exactitude l'état de leurs connaissances, principalement dans des tâches complexes de compréhension de textes (Glenberg & Epstein, 1985, 1987 ; Maki, 1998 ; Miller & Geraci, 2011 ; Redford et al., 2012) et de résolution de problèmes (e.g., Baars et al., 2013). L'auto-évaluation étant un processus clé dans l'apprentissage autorégulé, améliorer son exactitude constitue un enjeu important pour les experts de l'apprentissage (Hines et al., 2009 ; Huet & Mariné, 2009 ; Kornell & Bjork, 2007 ; Metcalfe & Finn, 2008 ; Nelson & Narens, 1990 ; Thiede & Dunlosky, 1999). Une plus grande exactitude de l'auto-évaluation est associée à une autorégulation plus adaptée, c'est-à-dire, à des comportements d'apprentissage en adéquation avec le niveau de connaissance de l'apprenant. Cette amélioration de l'autorégulation se traduit, par la suite, par une amélioration des performances.

Dans le but de fournir les fondements théoriques nécessaires à l'élaboration de recommandations pratiques, le premier chapitre répertoriait les hypothèses explicatives du processus d'auto-évaluation, les principales caractéristiques du processus ainsi qu'une partie des facteurs qui affectent son exactitude. Sur la base de ces éléments théoriques, ce deuxième chapitre présente différentes méthodes proposées afin d'améliorer l'exactitude de l'auto-évaluation. Les trois principales méthodes recensées sont brièvement introduites ci-après, avant d'être présentées plus profondément dans des sections spécifiques.

Afin d'introduire la première méthode d'aide à l'auto-évaluation, il est important de rappeler que l'hypothèse d'utilisation d'indices de Koriat (1997) reste la référence majeure pour expliquer la façon dont les apprenants réalisent un jugement d'auto-évaluation. Selon cette hypothèse, rappelons-nous, l'auto-évaluation repose sur l'utilisation de divers indices, et son niveau d'exactitude dépend de la capacité des indices utilisés à prédire les performances évaluées. Des auteurs (e.g., Griffin et al., 2009) ont généralisé ce modèle à des tâches de compréhension de textes en le couplant au modèle de la compréhension de Kintsch (1998). Dans le cadre



d'une tâche de compréhension de textes, le modèle de situation est le niveau le plus élaboré de compréhension qu'un apprenant puisse atteindre. Aussi, les indices basés sur ce niveau de représentation sont les plus valides pour réaliser un jugement d'auto-évaluation exact (e.g., Baars et al., 2013 ; Griffin et al., 2009 ; Redford et al., 2012). Plus généralement, l'objectif d'un apprenant en situation d'apprentissage est de réaliser un apprentissage profond plutôt que superficiel (Ausubel, 1960). Par extension, les indices issus d'un apprentissage profond seront plus valides pour réaliser un jugement d'auto-évaluation précis que les indices issus d'un apprentissage superficiel. Une première méthode possible pour améliorer l'exactitude de l'auto-évaluation consiste à rendre plus saillants pour l'apprenant des éléments du modèle de situation dans le cas d'une tâche de compréhension (ou, plus généralement, les éléments associés à un apprentissage profond). Les éléments plus saillants pourront, dès lors, être utilisés par l'apprenant lors de la réalisation du jugement d'auto-évaluation.

Par ailleurs, les indices intrinsèques et extrinsèques du modèle de Koriat (1997) sont assimilables à des règles transférables d'un domaine à un autre (par exemple, la 'règle' selon laquelle 'un contenu avec beaucoup d'informations est plus difficile à apprendre qu'un contenu avec peu d'informations' est valable quel que soit le domaine d'apprentissage, Koriat & Bjork, 2006). Ces règles permettent à l'apprenant d'évaluer l'état de ses connaissances quel que soit son niveau d'expertise sur le domaine étudié. Aussi, une deuxième méthode d'aide à l'auto-évaluation consiste à inciter les apprenants à mettre en œuvre des processus métacognitifs. Plus spécifiquement, l'objectif est d'aider les apprenants à utiliser, dans un premier temps, des règles générales (i.e. des indices intrinsèques et extrinsèques) avant qu'ils puissent, avec la pratique, utiliser des indices mnémoniques plus efficaces (Koriat & Bjork, 2006).

L'exactitude de l'auto-évaluation étant sensible au moment de réalisation du jugement (Nelson & Dunlosky, 1991), une dernière méthode, enfin, consiste à différer la réalisation du jugement d'auto-évaluation de la tâche d'apprentissage de quelques secondes ou minutes.

Dans la suite de ce chapitre, nous revenons sur chacune des trois méthodes introduites ci-dessus et nous présentons les différents outils qui s'en inspirent. Nous présenterons, en premier lieu, les méthodes qui permettent de rendre plus saillants les indices pertinents pour l'auto-évaluation. Nous présenterons, dans un second temps, les incitations qui visent à aider l'apprenant à s'engager dans un processus d'auto-évaluation. Nous terminerons le chapitre en décrivant les effets de la temporalité du jugement d'auto-évaluation selon le type de tâche visé.

### 3.1 AGIR SUR LA PERTINENCE DES INDICES UTILISÉS

Selon [Koriat \(1997\)](#), améliorer l'exactitude de l'auto-évaluation requiert des apprenants qu'ils utilisent des indices qui prédisent du mieux possible les performances. Par exemple, si la difficulté du matériel à apprendre prédit les performances des apprenants (i.e. plus un matériel est difficile, plus les performances sont basses), il est alors pertinent pour l'apprenant d'utiliser le niveau de difficulté du matériel pour évaluer ses capacités à réaliser un exercice.

Les méthodes présentées ici visent à aider l'apprenant à accéder au modèle de situation ou à un apprentissage profond, afin qu'il en extrait des indices pertinents pour l'évaluation des performances futures. La plupart des méthodes présentées dans cette partie ont été testées sur des tâches de compréhension, justifiant le fait que nous évoquerons principalement l'accès au modèle de situation et non la profondeur d'apprentissage. Nous présentons ci-après l'activité de relecture, puis nous traitons des différentes stratégies génératives par la suite.

#### 3.1.1 *La relecture*

La relecture est une stratégie d'étude qui, dans un cadre expérimental et contrôlé, consiste à présenter une deuxième fois un texte après que l'apprenant l'ait lu une première fois ([Minguela, Solé, & Pieschl, 2015](#) ; [Hegarty & Just, 1993](#) ; [Wigent, 2013](#)). Différentes recherches ont prouvé l'efficacité de cette méthode, en montrant que des apprenants qui relisent un texte s'auto-évaluent avec plus d'exactitude que des apprenants qui ne le relisent pas ('effet de la relecture') ([Dunlosky & Rawson, 2005](#) ; [Griffin et al., 2008](#) ; [Rawson et al., 2000](#)).

L'effet de la relecture s'explique partiellement par l'hypothèse du niveau d'interruption, selon laquelle le jugement métacognitif est basé sur le nombre et le niveau des interruptions faites au cours de la lecture d'un texte ([Dunlosky & Rawson, 2005](#)). Si les interruptions ont lieu lorsque l'individu traite le niveau du modèle de situation (i.e. l'apprenant interrompt sa lecture pour se créer une représentation mentale de la situation décrite), l'exactitude de l'auto-évaluation de la compréhension devrait être plus exacte ([Dunlosky & Rawson, 2005](#)). La relecture, en encourageant le traitement au niveau du modèle de situation, doit donner à l'apprenant davantage d'occasions d'interrompre sa lecture pour créer son modèle de situation et ainsi améliorer l'exactitude de l'auto-évaluation sur des questions de compréhension et pas seulement de mémorisation.

Afin de tester cette hypothèse, [Dunlosky et Rawson \(2005\)](#) ont demandé à des étudiants de lire six textes différents. Une partie des participants lisait un texte et auto-évaluait sa performance future immédiatement après

cette première lecture. Une autre partie des participants relisait les textes une deuxième fois et ne prédisait sa performance qu'immédiatement après cette deuxième lecture. Les participants devaient, pour s'auto-évaluer, estimer à quel point ils pensaient pouvoir répondre à des questions posées sur les textes, à l'aide d'une échelle en six points, allant de 0 à 100 (i.e. 0 ; 20 ; 40 ; 60 ; 80 ; 100). Ce n'est qu'après avoir prédit leurs performances pour chacun des textes que les participants complétaient les questionnaires de mesure des performances, composés de six questions à choix multiples associées à chacun des textes. Trois questions mesuraient la mémorisation des participants (questions factuelles) et trois mesuraient la compréhension (questions inférentielles). Les résultats ont montré que les participants qui relisaient les textes étaient plus exacts sur les questions inférentielles que les participants qui n'avaient pas eu l'opportunité de relire ces textes. Toutefois, aucune différence sur les questions de mémorisation n'a été observée. Dans le but d'explorer si les résultats observés sont liés aux interruptions, les auteurs ont conduit des analyses de régression du temps de lecture des phrases sur différentes variables prédictives du niveau de traitement du texte (ex : le nombre de propositions par phrases renvoie à la mesure d'un traitement au niveau de base du texte ; l'importance de chaque phrase par rapport au sens global du texte renvoie à la mesure d'un traitement au niveau du modèle de situation). Leurs résultats valident partiellement leur hypothèse puisque les relecteurs passent moins de temps à lire des éléments relatifs au niveau de base (ce résultat est consistant avec leur hypothèse) comparativement aux participants qui ne réalisent qu'une lecture ; néanmoins, les relecteurs ne passent pas plus de temps sur des éléments relatifs au niveau du modèle de situation (ce qui ne va à l'encontre de leur hypothèse). [Dunlosky et Rawson \(2005\)](#) ont proposé des hypothèses alternatives pour expliquer leurs résultats, dont 'l'hypothèse de l'effort optimal' et 'l'hypothèse de la disponibilité des ressources'. La première de ces deux hypothèses, basée sur les travaux de [Weaver et Bryant \(1995\)](#), postule que la compréhension d'un texte trop difficile nécessite un traitement qui va mettre en péril le processus d'auto-évaluation et par conséquent son exactitude. Un texte trop facile, par contre, ne va pas permettre aux lecteurs d'engager suffisamment de ressources pour les investir dans un processus d'auto-évaluation. Il serait ainsi nécessaire de trouver un texte dont la difficulté soit adaptée au niveau de l'apprenant, sans être ni trop facile ni trop difficile. [L.-M. Lin, Zabrusky, et Moore \(2002\)](#), cependant, ont montré que, contrairement à ce que suggère 'l'hypothèse de l'effort optimal', des apprenants avec un niveau de lecture leur permettant de traiter des textes difficiles étaient meilleurs sur des textes de difficulté modérée. Aussi, [Dunlosky et Rawson \(2005\)](#) trouvent l'hypothèse de disponibilité des ressources cognitives plus appropriée pour expliquer leurs résultats. Cette hypothèse, plus générale que

celle de l'effort optimal, affirme que l'auto-évaluation sera d'autant plus exacte qu'un individu dispose de ressources cognitives pour auto-évaluer sa compréhension. En accord avec cette hypothèse, de précédentes études ont montré que l'apprenant dispose de davantage de ressources cognitives lors d'une relecture que lors d'une première lecture (e.g., [Raney, 1993](#)). Les auteurs concluent que des études supplémentaires sont nécessaires afin de déterminer laquelle des deux hypothèses ('hypothèse de la disponibilité des ressources' ou 'hypothèse du niveau d'interruption') est la plus pertinente.

[Griffin et al. \(2008\)](#) ont, eux, mis en évidence que la relecture permettait à des étudiants en psychologie ayant de faibles habiletés de lecture, et par conséquent une auto-évaluation moins exacte, d'égaliser le niveau d'exactitude des apprenants ayant de bonnes habiletés de lecture. La relecture, en revanche, n'a pas d'effet sur l'exactitude de l'auto-évaluation des apprenants disposant d'habiletés de lecture ou des capacités de mémoire de travail élevées. Dans leur étude, le niveau d'habileté en lecture était mesuré à l'aide d'un score composite de la performance des participants à des questions inférentielles sur les quatre textes cibles. Les capacités de mémoire de travail, quant à elles, étaient mesurées à l'aide de mesures d'empan mnésiques. Pour expliquer ces résultats, les auteurs basent leur raisonnement sur le postulat que le processus d'auto-évaluation a lieu simultanément au processus de compréhension. Selon eux, la relecture aide les apprenants avec des habiletés de lecture ou des capacités de mémoire de travail faibles car certains processus nécessaires au traitement du texte à un niveau de base n'ont plus à être réalisés lors d'une seconde lecture. Les apprenants peuvent, par conséquent, non seulement allouer leurs ressources au processus d'auto-évaluation et au traitement des informations métacognitives qui en sont issues, mais également traiter le texte à un niveau plus conceptuel.

La relecture ne semble pas être suffisamment engageante pour permettre aux apprenants qui disposent d'habiletés de lecture et des capacités de mémoire de travail élevées d'améliorer l'exactitude de leur auto-évaluation. D'autres stratégies, où l'apprenant se doit d'être plus actif, ont été testées et sont présentées ci-après.

### 3.1.2 *Les stratégies génératives*

L'effet de génération (appelé 'effet de production' dans l'article de [Taconnat, 2005](#)), initialement mis en évidence par [Slamecka et Graf \(1978\)](#), consiste à augmenter les performances de rappel de mots lorsque ceux-ci sont présentés sous une forme incomplète (e.g., shadow - win. . ., [Mulligan, 2002](#)) ou différente (e.g., anagrammes [Taconnat & Isingrini, 2004](#)). L'efficacité des stratégies de génération sur les performances s'expliquerait par

l'effort sollicité plus important en phase d'apprentissage que celui sollicité lors d'une simple lecture. Cet effort plus important permet d'activer un réseau lexical plus large qu'en condition de simple lecture, facilitant ainsi la récupération de l'information lors du rappel.

L'effet des stratégies génératives sur l'exactitude de l'auto-évaluation (Mazzoni & Nelson, 1995) explique le fait que l'auto-évaluation est plus exacte lorsque l'apprentissage nécessite de l'apprenant qu'il génère des mots à partir d'anagrammes comparativement à une condition où l'apprentissage nécessite que l'apprenant lise des mots. L'efficacité des stratégies génératives a, par la suite, été testée et démontrée sur l'amélioration de l'exactitude de l'auto-évaluation liée à des tâches de compréhension de textes linéaires (e.g., M. Anderson & Thiede, 2008 ; Thiede et al., 2003, 2010) ou de résolution de problème (Baars et al., 2013).

Les stratégies génératives se divisent en deux catégories, selon la temporalité par rapport à l'apprentissage avec laquelle elles sont utilisées. Les stratégies de 'génération différée' (e.g., la production de résumés ou de mots-clés) sont réalisées un peu après l'apprentissage. Les stratégies de 'génération immédiate' (e.g., les cartes conceptuelles, les auto-explications), sont accomplies pendant l'apprentissage (Baars et al., 2013). L'objectif de ces stratégies différées ou immédiates est double. Les stratégies génératives permettent dans un premier temps de favoriser la focalisation de l'attention de l'apprenant sur le niveau du modèle de situation, ou, plus généralement, sur des caractéristiques profondes de la tâche d'apprentissage afin que l'individu puisse utiliser des indices prédictifs de la performance. Elles permettent également d'aider l'apprenant à se créer une représentation plus cohérente du modèle de situation dans le cas d'une tâche de compréhension de texte (M. Anderson & Thiede, 2008 ; Griffin et al., 2008), ou d'améliorer l'apprentissage profond dans le cas de tâche de résolution de problème.

### *Stratégie générative différée*

LA PRODUCTION DE MOTS-CLÉS ET DE RÉSUMÉS Cette méthode consiste à demander à l'apprenant de générer des mots-clés ou un résumé d'un contenu d'apprentissage étudié au préalable (M. Anderson & Thiede, 2008 ; Thiede & Anderson, 2003 ; Thiede et al., 2005).

M. Anderson et Thiede (2008) ont testé l'hypothèse d'accessibilité (selon laquelle les apprenants se basent sur la quantité d'information disponible pour estimer leurs performances futures) ainsi que l'hypothèse du modèle de situation, afin d'identifier laquelle était la plus à même d'expliquer l'amélioration de l'auto-évaluation lorsque la production d'un résumé est réalisée en différé de l'apprentissage. Pour comparer ces deux hypothèses, M. Anderson et Thiede (2008) ont réparti des étudiants en

psychologie dans trois groupes différents. Un groupe lisait cinq textes et évaluait ensuite leur compréhension pour chacun des textes, sans écrire de résumé (groupe contrôle). Un groupe lisait les textes et écrivait un résumé après la lecture de chaque texte (groupe 'résumé immédiat'). Après avoir lu et résumé l'ensemble des textes, les participants du groupe de 'résumé immédiat' évaluaient leur niveau de compréhension sur chacun des textes. Le dernier groupe lisait l'ensemble des textes puis produisait dans un second temps un résumé pour chaque texte (groupe 'résumé différé') et s'auto-évaluait sur l'ensemble des textes. L'ensemble des participants des trois groupes a également eu pour consigne d'indiquer les indices utilisés pour prédire leur performance sur l'ensemble des textes et tous ont répondu à des questions de compréhension relatives à chaque texte. Leurs résultats répliquent l'effet positif de la production différée de résumé sur l'exactitude de l'auto-évaluation, précédemment mis en évidence par [Thiede et Anderson \(2003\)](#). Les résultats indiquent également que l'hypothèse du modèle de situation explique mieux l'exactitude de l'auto-évaluation que l'hypothèse d'accessibilité. En effet, bien que les participants en condition de résumé immédiat rapportent davantage d'indices que les participants en condition de résumé différé, les premiers (groupe 'résumé immédiat') utilisent davantage d'indices basés sur le détail tandis que les apprenants en condition de résumé différé utilisent davantage d'indices basés sur le sens général du texte. Ces derniers étant plus exacts dans l'auto-évaluation qu'ils font de leurs performances, il semblerait bien que c'est la nature et non la quantité d'indices utilisés qui importe. Il semble important, finalement, que les outils d'aide à l'auto-évaluation incitent les apprenants à mettre l'accent sur le modèle de situation.

Des résultats comparables ont été obtenus concernant la production de mots-clés ([Thiede et al., 2005, 2003](#)). [Thiede et al. \(2003\)](#) ont mis en évidence un effet positif de la génération de mots-clés sur l'exactitude de l'auto-évaluation à condition que ceux-ci soient également générés avec un délai par rapport à la phase d'apprentissage. Le fait de générer des mots-clés immédiatement après avoir lu un texte ne produisait aucun effet. Cette première étude a permis de mettre en évidence qu'une amélioration de l'exactitude de l'auto-évaluation à l'aide de mots-clés conduisait les apprenants à mieux s'autoréguler (i.e. choisir de revoir les textes les moins bien compris) et, en conséquence, à améliorer les performances lors de la tâche de compréhension. [Thiede et al. \(2005\)](#) ont voulu comprendre, dans un second temps, le mécanisme lié à cet effet positif des 'mots-clés différés'. Les auteurs ont identifié trois distinctions majeures entre les apprenants qui créaient des mots-clés immédiatement après la lecture des textes et ceux qui créaient des mots-clés quelques minutes après la lecture des textes. Ces distinctions concernent (1) le délai entre la lecture du texte et l'activité de génération de mots-clés associés à ce texte (le groupe

immédiat produisait des mots-clés associés à un texte immédiatement après l'avoir lu, le groupe différé produisait des mots-clés associé à un texte longtemps après l'avoir lu); (2) le délai entre la génération de mots-clés d'un texte à un autre (le groupe différé produisait les mots-clés pour tous les textes à la suite, le groupe immédiat lisait un texte, générait les mots-clés associés puis recommençait pour chaque texte); (3) le délai entre l'activité de génération de mots et le jugement de compréhension (le groupe différé devait réaliser un jugement peu de temps après la génération de mots clés car la génération de mots-clés pour tous les textes prenait peu de temps, le groupe immédiat réalisait un jugement longtemps après avoir généré des mots-clés car les participants devaient lire les autres textes entre temps). Sur la base de ces différences, les auteurs ont comparé dans une étude quatre groupes expérimentaux où le délai (immédiat versus différé) entre la lecture et la production de mots-clés ainsi qu'entre la production de mots-clés et le jugement variait. Les résultats obtenus indiquaient que les participants des groupes où la production de mots-clés était réalisée en différée de l'apprentissage s'auto-évaluaient avec plus d'exactitude que les participants des groupes où la production de mots-clés était réalisée immédiatement après la lecture. Thiede et al. (2005) concluaient que dans le cas de la génération différée de mots-clés, les informations précises et relatives aux caractéristiques de surface ne sont plus disponibles en mémoire de travail. Les mots-clés générés après un délai nécessitent que l'apprenant accède à des indices plus généraux et plus prédictifs de la performance.

Thiede et al. (2005) ont également exploré lequel du processus de génération ou de la simple visualisation des mots-clés permettait d'améliorer l'exactitude de l'auto-évaluation. Les auteurs ont, pour cela, comparé une condition où les mots-clés étaient pré-élaborés et présentés aux apprenants à une condition où les apprenants devaient les générer par eux-mêmes. Les résultats indiquaient que l'apprenant doit bien s'engager dans le processus d'élaboration de mots-clés pour que cette stratégie soit efficace.

### *Les stratégies génératives immédiates*

**LES AUTO-EXPLICATIONS** Selon Chi (2000), les auto-explications renvoient à la génération d'explications pour soi-même dans le but de donner du sens au contenu d'apprentissage et pas seulement de le mémoriser (voir aussi Butcher, 2006). Elles permettent également de concentrer l'attention des lecteurs sur le niveau du modèle de situation dans le cas d'une tâche de compréhension de texte (Wiley & Voss, 1999). Les auto-explications permettent de produire ce que Chi (2000) appelle des 'inférences d'auto-explication' et qui renvoient à des 'morceaux de connaissances' qui explicitent des informations non présentes dans le contenu à apprendre. Les

auto-explications peuvent également contenir des informations relatives à l'activité métacognitive de l'apprenant. L'apprenant peut, par exemple, évaluer la reformulation qu'il a faite d'une phrase, ou se questionner sur les inférences à en extraire.

Les auto-explications sont adaptées à divers contextes d'apprentissage et pas seulement à celui d'un apprentissage de textes (Chi, 2000). Des effets positifs des auto-explications ont été observés sur l'apprentissage de divers domaines tels que la physique (Chi, Bassok, Lewis, Reimann, & Glaser, 1989), la biologie (Chi, Leeuw, Chiu, & LaVancher, 1994), mais également les probabilités (Renkl, 1997), ou encore la résolution de problème de programmation (Bielaczyc, Pirolli, & Brown, 1995). Dans les études de Chi et al. (1989) et Chi et al. (1994), plus les apprenants généraient des auto-explications, meilleures étaient leurs performances d'apprentissage. Les résultats sont plus contrastés en ce qui concerne l'efficacité des auto-explications dans des environnements informatisés (Atkinson, Renkl, & Merrill, 2003). Certaines études ont mis en évidence des effets positifs sur l'apprentissage (Alevan & Koedinger, 2002) tandis que d'autres n'ont pas observé d'effet (Conati & Vanlehn, 2000). Alevan et Koedinger (2002) ont montré que des apprenants générant des auto-explications à l'aide d'un tuteur cognitif étaient meilleurs pour résoudre des problèmes de transfert et produisaient des explications de meilleure qualité. De manière consistante, Hausmann et Chi (2002) ont montré que les incitations étaient aussi efficaces lorsqu'elles étaient générées par un ordinateur que par un intervenant humain. Les auteurs ont précisé qu'il est particulièrement important de guider l'apprenant dans la production d'auto-explications, sous réserve qu'il ne le fasse pas correctement, voire pas du tout.

Peu de travaux ont étudié l'effet des auto-explications sur l'exactitude de l'auto-évaluation. Parmi ces dernières, Griffin et al. (2008) ont mis en évidence que, dans le cadre d'une tâche de compréhension de textes linéaires, la production d'auto-explications au cours d'une phase de re-lecture améliorerait l'exactitude de l'auto-évaluation. Les auto-explications aideraient les apprenants à accéder à des indices plus spécifiques liés au modèle de situation.

Les auto-explications se doivent d'être de bonne qualité pour améliorer l'apprentissage (Renkl, 1997). Toutefois, cette variable n'a pas d'effet sur l'auto-évaluation (Griffin et al., 2008). Griffin et al. (2008) précisent que la qualité de l'auto-explication importe peu. Il suffit que l'individu tente de générer une auto-explication pour que des indices prédictifs des performances soient accessibles. La production d'auto-explications possède donc l'avantage de ne pas nécessiter de surveillance spécifique, ce qui facilite leur implémentation.



Pour conclure sur les auto-explications, il apparaît que ces dernières améliorent l'apprentissage ainsi que l'exactitude de l'auto-évaluation. Il ressort, néanmoins, qu'il est nécessaire de proposer à l'apprenant des incitations à l'auto-explication sous peine qu'il n'en produise pas dans un contexte numérisé. Cet inconvénient n'est toutefois pas particulièrement problématique pour l'auto-évaluation. Il semblerait en effet que ce n'est pas tant la qualité de l'auto-explication qui importe pour améliorer l'exactitude de l'auto-évaluation, mais plutôt l'engagement de l'apprenant dans le processus explicatif (Griffin et al., 2008). En conséquence, malgré la faible quantité d'études ayant mis en évidence l'efficacité des auto-explications sur l'amélioration de l'exactitude de l'auto-évaluation, cette dernière méthode semble avantageuse dans un contexte écologique d'apprentissage assisté par ordinateur, de par sa simplicité d'implémentation et son efficacité.

**LES CARTES CONCEPTUELLES** Les cartes conceptuelles sont des représentations graphiques qui permettent d'organiser et de représenter les connaissances (Novak & Cañas, 2008). Plusieurs termes existent dans la littérature pour évoquer des outils plus ou moins similaires et dont les éléments de différenciation sont plus ou moins établis. Se retrouvent notamment les concepts de cartes conceptuelles (Amadiou, Tricot, & Mariné, 2010 ; Novak, 1984 ; Redford et al., 2012 ; Thiede et al., 2010), de cartes cognitives (Kitchin, 1994), de cartes de connaissances (Amer, 1994 ; O'donnell, Dansereau, & Hall, 2002), ou encore d'organiseurs graphiques (Nesbit & Adesope, 2006 ; Stull & Mayer, 2007). Nesbit et Adesope (2006) proposent une distinction entre les organisateurs graphiques (traduit de l'anglais : 'graphic organizer'), les cartes conceptuelles (traduit de l'anglais : 'concept maps') et les cartes de connaissances (traduit de l'anglais : 'knowledge maps') à l'aide d'une relation d'inclusion. Les organisateurs graphiques intègrent toutes formes de représentations graphiques possibles (e.g., des frises chronologiques, des organigrammes, des tableaux, etc.). Les cartes conceptuelles n'utilisent que des concepts représentés sous forme de nœuds reliés par des liens qui représentent les relations entre les concepts, que celles-ci soient labellisées ou non, directionnelles ou non directionnelles. Les cartes de connaissances sont, quant à elles, des cartes conceptuelles dont les relations sont représentées de façon uni-directionnelle et labellisées par un vocabulaire précis. Dans la littérature propre à la psychologie cognitive, le terme de carte conceptuelle est souvent privilégié. Par souci de clarté, ce terme sera conservé tout au long de ce manuscrit.

Les cartes conceptuelles sont utilisées dans de nombreuses situations telles que l'apprentissage collaboratif, l'évaluation de la compréhension, l'amélioration de la compréhension de texte linéaire ou non linéaires

(pour une revue sur les hypertextes, voir [Amadiou & Salmeron, 2014](#)), ou encore l'auto-évaluation. Par souci de cohérence avec le but de la recherche présente, seuls les travaux en lien avec la compréhension et l'auto-évaluation seront traités ci-après.

[Amadiou et Salmeron \(2014\)](#) recensent plusieurs modèles pour expliquer l'efficacité des cartes conceptuelles sur la compréhension. Le premier modèle est 'l'hypothèse d'assimilation de l'encodage' ([Mayer, 1979](#)). Ce modèle prend sa source dans les travaux d'[Ausubel \(1960\)](#) et ne mentionne pas encore le terme de carte conceptuelle, mais privilégie plutôt celui 'd'organisateur avancé' (traduit de l'anglais 'advanced organizer'). Cet outil faciliterait l'encodage de l'information en donnant une structure visible au texte lu. Plus tard, [Paivio \(1986\)](#) ; [Clark et Paivio \(1991\)](#) a proposé le modèle du double codage. Les cartes amélioreraient l'apprentissage en optimisant l'exploitation des ressources cognitives d'un individu, en utilisant à la fois le 'magasin' verbal pour le traitement du texte ainsi que le 'magasin' visuel pour le traitement des cartes. Enfin, [Schnotz \(2002\)](#) avance l'idée que les cartes, grâce à un traitement graphique de l'information, permettraient la création d'un modèle mental différent du texte qui reposerait, quant à lui, sur un traitement verbal. Les deux modèles (issu du texte ou issu de la carte) s'influenceraient mutuellement et partageraient les informations. Ainsi, les cartes permettraient d'aider l'organisation conceptuelle de l'information.

Ces trois modèles présentés mettent en avant l'idée que l'efficacité des cartes conceptuelles se traduit principalement par la possibilité qu'elles offrent de structurer et d'organiser l'information. Ainsi, elles facilitent la construction d'une représentation mentale ou d'un modèle de situation. Cet argument est renforcé par une étude récente qui a mis en évidence la supériorité d'un apprentissage accompagné de cartes conceptuelles comparativement à un apprentissage associé à une activité de prise de note ou de relecture ([Ponce & Mayer, 2014](#)). Dans cette étude, les auteurs ont montré, grâce à l'oculométrie, que les apprenants qui utilisaient des cartes semblaient utiliser davantage de stratégies génératives que des apprenants qui utilisaient un outil de prise de note. Les mouvements oculaires des premiers révélaient une tentative de connexion entre les différentes informations qui nécessitaient d'être comparées. Au contraire, les mouvements oculaires des seconds révélaient une stratégie de lecture linéaire. Les deux derniers modèles mettent également en avant la propension des cartes à optimiser l'utilisation des ressources cognitives et ainsi augmenter le traitement des informations à intégrer.

Sur la base de ces explications et en lien avec l'hypothèse du modèle de situation, les cartes conceptuelles sont apparues comme un outil pertinent pour améliorer l'auto-évaluation et son exactitude. Peu d'études, cependant, se sont intéressées à cette question ([Chularut & DeBacker,](#)

2004 ; Redford et al., 2012 ; Thiede et al., 2010). Chularut et DeBacker (2004) ont montré que l'utilisation de cartes conceptuelles améliorait les performances (mesurées à l'aide d'un questionnaire à choix multiples) et les stratégies d'auto-évaluation (mesurées à l'aide d'un questionnaire composé de neuf items), principalement chez des étudiants avec peu de connaissances antérieures. Toutefois, l'exactitude de l'auto-évaluation n'a pas été mesurée dans cette étude. Ce n'est que plus tard que Thiede et al. (2010) ont mis en évidence que la construction de cartes conceptuelles parallèlement à la lecture d'un texte améliorait l'exactitude de l'auto-évaluation d'étudiants dont les capacités de lecture n'étaient pas très élevées. Enfin, Redford et al. (2012) ont comparé des élèves de cinquième qui construisaient des cartes simultanément à la lecture d'un texte linéaire à des élèves de cinquième qui ne construisaient pas de cartes conceptuelles. Les résultats ont mis en évidence un effet faible mais positif des cartes sur l'exactitude de l'auto-évaluation. Afin d'obtenir un effet plus prononcé, les auteurs ont réalisé une seconde étude dans laquelle les élèves ont été entraînés plus longuement à l'utilisation des cartes. Dans cette étude, Redford et ses collègues ont également comparé une condition de consultation à une condition de construction de cartes conceptuelles. Les résultats de cette seconde étude ont montré une efficacité plus grande de la construction comparativement à la consultation sur l'exactitude de l'auto-évaluation, même si ces résultats n'atteignaient pas l'amplitude attendue.

Si les cartes conceptuelles et principalement leur construction ont prouvé leur efficacité sur l'amélioration de l'exactitude de l'auto-évaluation, elles ont également montré des limites en termes de coût cognitif. Stull et Mayer (2007) ont discuté de l'efficacité des cartes sur l'amélioration de la compréhension selon le mode de leur présentation (construction versus consultation). Les auteurs ont proposé que la consultation aurait pour inconvénient de ne pas obliger l'apprenant à s'investir dans le processus de traitement de l'information. En contrepartie, elle permet à l'apprenant d'utiliser l'intégralité des ressources pour le traitement de l'information à intégrer lorsqu'il s'y investit. En revanche, si la construction oblige plus explicitement l'apprenant à manipuler et donc à traiter l'information, une partie de ses ressources peut être potentiellement gaspillée par le simple fait de devoir manipuler cet outil d'aide. Aussi, certains auteurs recommandent d'entraîner les apprenants à utiliser les cartes conceptuelles afin de limiter l'utilisation excessive de ressources par leur construction (Chang, Sung, & Chen, 2001 ; Hilbert & Renkl, 2009 ; O'donnell et al., 2002). Un entraînement n'est toutefois pas toujours possible. Dans ces conditions, il paraît pertinent de savoir si la consultation ne serait pas plus efficace que la construction.

L'étude de [Stull et Mayer \(2007\)](#) nous amène effectivement à penser que la consultation de cartes conceptuelles pourrait parfois être plus efficace que la construction. Les auteurs ont réalisé trois études au cours desquelles la complexité des textes diminuait progressivement. Dans la première étude, les apprenants devaient consulter ou construire 27 cartes conceptuelles, 18 dans la deuxième et 10 dans la dernière. Dans chaque étude, les participants ayant de faibles connaissances dans le domaine devaient soit (1) lire un texte accompagné de plusieurs petites cartes conceptuelles à consulter, (2) lire le texte et construire eux-mêmes les cartes, ou (3) lire simplement le texte. Les résultats obtenus vont en faveur d'une plus grande efficacité de la consultation des cartes conceptuelles. Les participants qui ont eu à consulter les cartes sont ceux qui ont obtenu de meilleures performances dans les épreuves de transfert. Aucune différence n'a été observée sur les épreuves de rétention. La plus grande efficacité de la consultation de cartes sur les tâches de transfert laisse dès lors naturellement penser que la consultation de carte pourrait s'avérer plus efficace que leur construction sur l'amélioration de l'exactitude de l'auto-évaluation. Par ailleurs, les auteurs se sont intéressés à savoir si des apprenants possédant un meilleur niveau de connaissances antérieures auraient pu bénéficier davantage de la condition de construction, du fait qu'ils disposent de plus de ressources cognitives pour utiliser les cartes. La modalité d'utilisation des cartes pourrait donc potentiellement interagir avec le niveau d'expertise de l'apprenant.

### 3.1.3 *Synthèse sur les outils qui agissent sur la pertinence des indices à utiliser*

Ce chapitre avait pour objectif de présenter différentes façons d'améliorer l'exactitude de l'auto-évaluation, en augmentant la saillance des indices pertinents. Si ces méthodes se sont toutes avérées efficaces, certaines le sont à un plus grand degré que d'autres. Les stratégies génératives semblent plus efficaces que les stratégies où l'apprenant est davantage passif. En effet, la relecture est moins efficace que les auto-explications qui améliorent l'exactitude de l'auto-évaluation d'un panel d'individu plus large ([Griffin et al., 2008](#)). Aussi, les apprenants aidés de cartes conceptuelles s'auto-évaluent avec plus d'exactitude que les apprenants qui relisent un texte ([Redford et al., 2012](#)). En parallèle, des études ont montré qu'il est préférable de produire des mots-clés ou des cartes conceptuelles plutôt que de consulter ces informations faites par des experts ([Thiede et al., 2005](#) ; [Redford et al., 2012](#)).

Les stratégies génératives, notamment l'utilisation de cartes conceptuelles, semble néanmoins requérir une grande quantité de ressources cognitives ([Chang et al., 2001](#) ; [Hilbert & Renkl, 2009](#) ; [O'donnell et al.,](#)

2002 ; Stull & Mayer, 2007). D'après la théorie de la charge cognitive (Sweller, 1988) ou la théorie de l'allocation des ressources (Kanfer & Ackerman, 1989), les ressources utilisées à des fins autres que destinées au traitement des informations ne seraient plus disponibles pour l'apprentissage profond et/ou la mise en application de processus métacognitifs. Dès lors, les conditions favorables à l'utilisation des ressources cognitives dans d'autres processus que celui de l'intégration des connaissances pourraient s'avérer défavorables à l'apprentissage et son auto-évaluation.

Enfin, la mise en place de nouveaux protocoles expérimentaux semble nécessaire afin de déterminer sous quelles conditions les stratégies génératives sont les plus efficaces et/ou plus efficaces que des stratégies non génératives, afin d'améliorer l'exactitude de l'auto-évaluation.

### 3.2 AGIR SUR L'INVESTISSEMENT DANS LE PROCESSUS : LES INCITATIONS

L'action sur l'investissement des apprenants dans le processus d'auto-évaluation trouve son origine dans le constat que les individus disposent d'habiletés métacognitives ou d'outils d'aide mais ne les utilisent pas toujours spontanément et/ou correctement (Azevedo, Moos, Greene, Winters, & Cromley, 2008 ; Renkl, 1997 ; Hausmann & Chi, 2002). Si Bannert et Reimann (2012) ont montré que les apprenants pouvaient spontanément utiliser des processus métacognitifs, ceux-ci ne sont pas aussi fréquents et/ou efficaces qu'accompagnés d'incitations. L'objectif d'utiliser des incitations est double. Il permettrait (1) d'aider les apprenants à prendre conscience (explicitement ou implicitement) de l'état de leurs connaissances mais surtout des processus métacognitifs qu'ils utilisent ou non et de leur efficacité respective. Il encouragerait les apprenants (2) à davantage s'engager dans des activités métacognitives efficaces et pertinentes. L'approche utilisée peut-être motivationnelle (i.e. récompenser un comportement adapté) ou métacognitive (i.e. assister la prise de conscience).

Bannert et Reimann (2012) ont travaillé sur les incitations relatives à une tâche d'apprentissage et les définissent comme '[... ] des procédures qui induisent et stimulent les activités cognitives, métacognitives, motivationnelles, volontaires et/ou coopératives pendant l'apprentissage [... ]' (p.195, traduit de l'anglais '[... ] instructional prompting are procedures to induce and stimulate cognitive, metacognitive, motivational, volitional and/or cooperative activities during learning [... ]'). Selon Ifenthaler (2012), les incitations vont faciliter et induire l'utilisation de stratégies et de compétences de résolution de problème existantes en focalisant l'attention de l'apprenant sur des aspects spécifiques de la tâche afin qu'il prenne conscience des activités engagées. La prise de conscience des processus utilisés permet à l'apprenant de transférer ces compétences à d'autres types

de tâches, conférant à cet outil un bénéfice sur le long terme. D'autres termes ont été utilisés pour évoquer les incitations, comme le terme de 'réflexion' (Van den Boom, Paas, Van Merriënboer, & Van Gog, 2004). Les auteurs utilisent des incitations de type 'réflexion' dans le but d'aider l'acquisition de compétences métacognitives plutôt que l'exploitation de compétences acquises. Hacker et al. (2008) distinguent, quant à eux, les 'incitations' des 'réflexions'. Les auteurs s'éloignent de la définition d'Ifenthaler (2012) et entendent 'incitations' comme un synonyme de récompense. Selon eux, une incitation est une bonification de la note de l'étudiant s'il s'auto-évalue avec plus d'exactitude. Schraw et al. (1993) ou encore Miller et Geraci (2011) se situent dans la continuité d'Hacker et al. (2008). Dans leurs études, l'incitation ne consistait pas en une phrase visant à demander à l'apprenant d'utiliser des processus d'auto-évaluation mais plutôt en une bonification de crédit si l'auto-évaluation était exacte. L'objectif de cette technique était d'inciter indirectement l'apprenant à s'engager dans un traitement plus actif et/ou à encourager les individus à générer eux-mêmes des feedbacks internes. Ces incitations ont prouvé leur efficacité (voir Yates, 1990), mais pas de façon systématique (Miller & Geraci, 2011).

La plupart des dispositifs ayant intégré des incitations ont mis en exergue les répercussions positives sur l'utilisation de stratégies métacognitives (X. Lin & Lehman, 1999), les capacités d'autorégulation et/ou les performances d'apprentissage des apprenants (Azevedo et al., 2008 ; Bannert & Reimann, 2012). À notre connaissance, et selon (Mihalca, 2014), peu d'études traitent de l'efficacité directe des incitations sur l'exactitude de l'auto-évaluation. Gama (2004) a, toutefois, créé et étudié un dispositif d'assistance à la réflexion dans le cadre de tâches de résolution de problème d'algèbre. Cet outil était notamment composé d'un système spécifique permettant d'évaluer les compétences d'auto-évaluation des apprenants. Ce système indiquait à l'apprenant si celui-ci s'évaluait avec exactitude ou non, afin de l'aider à prendre conscience de son niveau de connaissance. Ce dispositif s'est avéré efficace puisqu'il a notamment permis d'aider les apprenants à mieux gérer leur temps de résolution des problèmes. Les apprenants acquéraient également une meilleure conscience des stratégies d'apprentissage. Toutefois, Gama (2004) n'a pas pu mettre en évidence d'amélioration significative de l'exactitude de l'auto-évaluation.

En conclusion, l'effet des incitations sur l'amélioration de l'exactitude de l'auto-évaluation a été très peu exploré. Les effets bénéfiques reconnus ont été mis en évidence sur l'utilisation des habiletés métacognitives (X. Lin & Lehman, 1999) de façon générale et sur les performances de transfert le plus souvent (X. Lin & Lehman, 1999 ; Bannert & Reimann, 2012). L'utilisation d'incitations, toutefois, serait potentiellement coûteuse en ressources cognitives (Bannert & Reimann, 2012 ; Fiorella, Vogel-

Walcutt, & Fiore, 2012). De manière consistante avec cette idée, le niveau de connaissances antérieures semble être un facteur déterminant pour l'utilisation d'incitations. Les apprenants possédant plus de connaissances ont ainsi plus de facilité, probablement du fait qu'ils aient besoin d'allouer moins de ressources cognitives au traitement du contenu d'apprentissage (Bannert & Reimann, 2012).

### 3.3 AGIR SUR LA TEMPORALITÉ

L'action sur la temporalité ne renvoie pas à un outil spécifique mais fait référence à une méthode basée sur la découverte de l'effet du JOL différé (de l'anglais 'delayed JOL effect' Nelson & Dunlosky, 1991 ; Dunlosky & Nelson, 1992). Selon cet effet, un jugement d'auto-évaluation (type JOL) réalisé peu de temps après l'apprentissage est plus exact qu'un jugement d'auto-évaluation réalisé immédiatement après l'apprentissage. Cette méthode s'est avérée être parmi les plus efficaces pour améliorer l'exactitude de l'auto-évaluation, et l'effet du JOL différé est particulièrement robuste (Rhodes & Tauber, 2011). Cet effet a été montré à l'aide d'un paradigme simple qui consiste à faire apprendre des listes de paires de mots à des participants pour leur demander ensuite de prédire leurs performances futures, le plus souvent à une tâche de rappel indicé. Dans la condition où le jugement métacognitif est émis immédiatement après l'apprentissage, l'apprenant prédit ses performances après chaque paire de mot. Dans la condition où le jugement métacognitif est émis en différé de l'apprentissage, l'apprenant doit mémoriser l'ensemble des paires de mots avant d'émettre un jugement sur chacune de ces paires.

Si plusieurs hypothèses sont susceptibles d'expliquer cet effet, la plus pertinente semble être celle proposée par Nelson et Dunlosky (1991). Selon cette hypothèse, l'individu dispose, immédiatement après la tâche d'apprentissage, d'un grand nombre d'informations en mémoire de travail. Lors d'un jugement immédiat, ces informations seront utilisées pour prédire les performances futures. Toutefois, la performance étant réalisée quelques minutes plus tard, la plupart des informations disponibles au moment du jugement ne le sont plus. L'individu va alors puiser dans ses connaissances en mémoire à long terme pour réaliser la tâche. Les informations utilisées pour l'auto-évaluation et la performance ne sont, par conséquent, plus les mêmes, ce qui crée un biais. Au contraire, si l'individu évalue sa performance quelques instants après l'apprentissage, son jugement se base sur des connaissances en mémoire à long terme, plus probablement similaires à celles utilisées pour réaliser la performance. L'auto-évaluation est en conséquence plus exacte.

Cet effet a été répliqué de nombreuses fois, principalement sur l'auto-évaluation de tâches d'apprentissage de paires de mots (pour une méta-

analyse voir Rhodes & Tauber, 2011). Les tentatives de réplique sur des tâches plus écologiques telle que l'auto-évaluation de la compréhension ont eu les effets escomptés, sous réserve qu'une stratégie générative soit intégrée (M. Anderson & Thiede, 2008 ; Thiede et al., 2005 ; Thiede & Anderson, 2003). Sans ces tâches génératives, reporter la réalisation du jugement d'auto-évaluation associé à une tâche de compréhension ne suffit pas à en améliorer l'exactitude (Dunlosky et al., 2005 ; Maki, 1998). Toutefois, le moment de la présentation de la stratégie de génération est également important. À l'aide d'un paradigme similaire à celui utilisé pour montrer l'effet du JOL différé, Thiede et Anderson (2003) ont montré que les participants qui ont produit un résumé après un délai ont vu l'exactitude de leur auto-évaluation s'améliorer significativement comparativement à des étudiants qui n'ont pas produit de résumé, ou qui l'ont produit immédiatement après la lecture. Des résultats similaires ont été obtenus lorsqu'il s'agissait de produire des mots-clés (Thiede et al., 2003).

Ifenthaler (2012) défend néanmoins l'idée que le meilleur moment pour présenter une incitation à l'auto-évaluation est simplement d'attendre que l'individu en ait besoin. L'auteur propose que la présentation d'une incitation à un moment non approprié provoquerait une surcharge cognitive dommageable pour l'apprentissage et l'auto-évaluation de l'apprenant (voir Thillmann, Künsting, Wirth, & Leutner, 2009).

Si l'effet du JOL différé semble être la méthode la plus efficace pour améliorer l'exactitude de l'auto-évaluation, il est important de noter que celui-ci n'a pas été répliqué sur la mesure de l'exactitude absolue de l'auto-évaluation (Van Overschelde & Nelson, 2006). Dans cette étude, les auteurs ont même montré que l'exactitude absolue diminuait au fur et à mesure que le délai augmentait. Ainsi, l'effet du JOL différé nécessite de plus amples investigations en lien avec l'exactitude absolue, afin de déterminer les conditions qui pourraient le rendre efficace quelle que soit la mesure de l'exactitude de l'auto-évaluation.

Plus généralement, ce résultat invite à la plus grande prudence quant à la généralisation des effets des outils d'aide à l'auto-évaluation sur l'exactitude relative aux effets des outils d'aide sur l'exactitude absolue.



Nous avons vu que :

- Dans un but d'améliorer l'exactitude de l'auto-évaluation, il est possible d'aider l'apprenant à accéder à des indices pertinents, basés sur un apprentissage profond, d'inciter l'apprenant à s'engager dans des activités métacognitives, et de différer le jugement d'auto-évaluation de la phase d'apprentissage.
- Les méthodes génératives semblent plus efficaces ; pourtant, le coût cognitif de leur utilisation pourrait les rendre moins pertinentes dans certaines conditions.
- L'effet des incitations sur l'exactitude de l'auto-évaluation a été très peu étudié.
- Certaines méthodes efficaces pour améliorer l'exactitude relative de l'auto-évaluation ne le sont plus quand il s'agit d'améliorer l'exactitude absolue.

# 4

---

## EFFET D'OUTILS D'AIDE À L'AUTO-ÉVALUATION SUR SON EXACTITUDE

---

Comme il l'a été évoqué dans le chapitre s'y consacrant, le processus d'auto-évaluation est particulièrement important pour un apprentissage efficace. Celui-ci, cependant, n'en est pas moins difficile, et les apprenants tendent généralement à surestimer leurs performances dans des tâches complexes, telles que la compréhension ou la résolution de problème (Dunlosky & Lipko, 2007 ; García, Rodríguez, González-Castro, González-Pienda, & Torrance, 2015). Cette difficulté à s'auto-évaluer avec exactitude a de nouveau été mise en évidence lors de deux précédentes études réalisées dans le cadre du contrat ANR-10-CORD- 011 (Vidal, Huet, Miravette, & Marquié, 2015). Dans ces deux études, des étudiants de Master 1 'Méthodes Informatiques Appliquées à la Gestion des Entreprises' (MIAGE) devaient estimer leur degré de maîtrise de savoir ou savoir-faire liés à différents domaines du référentiel du C2I (Certificat Informatique et Internet). Ils répondaient dans un second temps à un questionnaire à choix multiples composé de questions relatives à chaque domaine. Les résultats ont révélé que les étudiants tendaient à sur-évaluer leurs connaissances dans différents domaines du C2I, dont ceux relatifs aux fonctionnalités avancées du traitement de texte et aux tableurs.

L'inexactitude de l'auto-évaluation est donc un problème fréquent, qui concerne aussi des étudiants qui, selon leur niveau d'étude, sont fréquemment amenés à s'auto-évaluer. Soutenir la réalisation de cette activité s'avère donc être un enjeu majeur dans le développement de l'autonomie dans l'apprentissage, et ce, quel que soit le niveau d'étude de l'apprenant. Dans cette perspective, l'ensemble du travail de recherche présenté ici s'attache à déterminer les caractéristiques que doivent avoir les outils d'aide à l'auto-évaluation pour que ces derniers soient mieux adaptés à celles de l'apprenant.

Dans cette première étude, l'objectif est de combiner les différentes méthodes d'aide à l'auto-évaluation recensées, afin d'améliorer l'exactitude de l'auto-évaluation et d'en faciliter la généralisation à d'autres domaines. La population visée est les étudiants, et notamment, dans le cadre du contrat ANR dans lequel s'inscrit cette thèse, les étudiants de

Master MIAGE. Dans l'introduction théorique de cette première étude, nous commencerons par rappeler brièvement les notions utiles à la compréhension du processus d'auto-évaluation en situation d'apprentissage complexe, ainsi que des éléments de réponses relatifs à la question de la spécificité de l'auto-évaluation à un domaine ou de sa généralisation potentielle. Nous aborderons ensuite les différentes méthodes ou outils efficaces qui permettent d'améliorer l'exactitude de l'auto-évaluation, avant de présenter l'étude.

#### 4.1 RAPPEL SUR LE PROCESSUS D'AUTO-ÉVALUATION

Rappelons que lors de la réalisation d'un jugement d'auto-évaluation, l'apprenant peut utiliser des indices de différentes natures (Koriat, 1997). L'auto-évaluation de la performance est exacte si elle est en accord avec la performance effective à un test. Le modèle d'utilisation d'indices (cf. section 2.4), initialement élaboré pour des tâches d'apprentissage associatif, a été spécifié aux tâches de compréhension de textes au travers de l'hypothèse du modèle de situation. Cette hypothèse repose également sur le modèle de la compréhension de Kintsch (1988) qui stipule que la compréhension possède plusieurs niveaux dont le plus haut est le niveau du modèle de situation. Ce niveau est atteint lorsqu'un individu est capable de faire des inférences entre les différents concepts du texte lu et que les connaissances nouvelles s'intègrent aux anciennes. La mesure des performances de compréhension repose bien souvent sur des questions relatives au modèle de situation. L'hypothèse du modèle de situation postule que l'apprenant s'auto-évaluera avec plus d'exactitude s'il utilise le modèle de situation (i.e. équivalent des indices extrinsèques, qui renvoient notamment à la profondeur de traitement de l'information Koriat, 1997) pour baser son jugement plutôt que les niveaux inférieurs de compréhension.

#### 4.2 LE PROCESSUS D'AUTO-ÉVALUATION EST-IL GÉNÉRAL OU SPÉCIFIQUE À UN DOMAINE ?

Parallèlement, la question de la généralisation du processus d'auto-évaluation d'un domaine à un autre n'est pas clairement résolue (e.g. Huet, 2004) et les conceptions s'opposent (e.g. Meijer, Veenman, & van Hout-Wolters, 2012). Pour certains, les habiletés métacognitives sont complètement dépendantes d'un domaine de connaissance (e.g. Kelemen, Frost, & Weaver, 2000). Ce point de vue permet d'expliquer que, bien souvent, les experts sont plus exacts que les novices dans l'auto-évaluation qu'ils font de leurs performances (e.g. Griffin et al., 2009 ; Kostons et al., 2010 ; Maki & Serra, 1992 ; Schneider, 1996). En revanche, d'autres travaux (e.g. Schraw et al., 1995 ; Schraw & Nietfeld, 1998 ; M. V. Veenman et

al., 1997) ont montré que les apprenants étaient capables d'auto-évaluer leurs performances dans différents domaines avec une exactitude similaire, suggérant que les apprenants disposent d'habiletés d'auto-évaluation générales sans lien avec un domaine de compétence. Ces habiletés générales permettraient aux apprenants d'évaluer leurs performances quel que soit leur niveau de compétence ou de connaissance dans un domaine.

Entre les deux approches, un compromis semble pouvoir être trouvé. Par exemple, M. Veenman et Elshout (1999) ont non seulement montré que des apprenants sans connaissance sur un domaine disposaient d'habiletés métacognitives générales mais également que le développement de connaissances spécifiques à un domaine permettait d'améliorer la qualité des habiletés métacognitives engagées. Schraw et Robinson (2011) quant à eux, considèrent que l'auto-évaluation spécifique à un domaine dépend de processus métacognitifs généraux auxquels s'ajoutent des connaissances spécifiques au domaine.

En conclusion, il semblerait qu'il existe des habiletés et des connaissances métacognitives transverses à plusieurs domaines de compétences qui permettraient à l'individu d'avoir des comportements d'apprentissage adaptés quel que soit le domaine d'apprentissage. Ces compétences et connaissances seraient complétées et/ou développées avec d'autres plus spécifiques à un domaine ce qui expliquerait que les experts s'auto-évaluent avec plus d'exactitude que les novices.

Cette question interroge également sur la différence éventuelle des habiletés métacognitives pour juger l'exactitude de l'état des connaissances déclaratives et procédurales. À notre connaissance, ce dernier point n'a été que très peu voire pas du tout traité dans la littérature.

#### 4.3 AMÉLIORER L'EXACTITUDE DE L'AUTO-ÉVALUATION : LES DIFFÉRENTES MÉTHODES

Basés principalement sur les modèles explicatifs de l'auto-évaluation mais également sur les caractéristiques globales du processus, trois types de méthodes, présentées précédemment dans le Chapitre 3, visent à améliorer l'exactitude de l'auto-évaluation. Ces méthodes sont brièvement rappelées dans cette partie.

##### 4.3.1 *Le jugement différé*

La première méthode consiste à agir sur la temporalité de la production du jugement en décalant l'activité d'auto-évaluation de celle de l'apprentissage. L'exactitude relative d'un jugement d'auto-évaluation réalisé quelques secondes ou minutes après l'apprentissage est meilleure que l'exactitude relative d'un jugement réalisé immédiatement après l'appren-

tissage. Cet effet mis en évidence par Nelson et Dunlosky (1991) sur une tâche d'apprentissage associatif a été répliqué sur des tâches plus complexes de compréhension de textes sous réserve que le décalage temporel existe non pas entre le jugement et la performance mais entre le jugement et la réalisation d'une stratégie générative (cf. paragraphe suivant) qui agit sur la compréhension du contenu d'apprentissage (M. Anderson & Thiede, 2008 ; Thiede et al., 2003). L'amélioration de l'auto-évaluation d'une tâche complexe nécessite donc de coupler une action sur la temporalité avec une action sur la compréhension qu'a l'apprenant du contenu d'apprentissage (voir ci-dessous). Les effets du délai sur l'exactitude absolue ont été beaucoup moins étudiés et les premiers résultats ont montré que sur une tâche d'apprentissage associatif le délai détériore l'exactitude absolue de l'auto-évaluation (Van Overschelde & Nelson, 2006). Sur une tâche plus écologique de résolution de problème Baars et al. (2014) n'ont pas observé de différences dans l'exactitude absolue de l'auto-évaluation lorsque le jugement était réalisé immédiatement ou quelques minutes après l'utilisation de tâches génératives.

Les effets bénéfiques du délai entre l'apprentissage et l'émission du jugement d'auto-évaluation sont donc manifestes sur l'exactitude relative de l'auto-évaluation tandis qu'aucun effet ne semble apparaître dans le cadre de l'auto-évaluation d'une tâche écologique entre un jugement réalisé immédiatement après ou en différé d'une tâche générative.

#### 4.3.2 *Faciliter l'accès aux indices pertinents*

Le deuxième type de moyen consiste à agir sur la compréhension qu'a l'apprenant du contenu d'apprentissage. Dans le cadre d'une tâche de compréhension, cela revient à aider l'apprenant à atteindre un modèle de situation du contenu d'apprentissage (Griffin et al., 2009), tandis que dans une situation de résolution de problème, l'objectif serait d'aider l'apprenant à se focaliser sur le processus de résolution (Baars et al., 2013). Globalement, il s'agit donc d'aider l'apprenant à traiter en profondeur l'information afin qu'il puisse utiliser des indices issus d'un traitement profond (i.e. des indices extrinsèques selon Koriat, 1997), plus prédictifs de la performance, plutôt qu'issus d'un traitement superficiel. Pour atteindre cet objectif, les stratégies génératives visent à amener l'apprenant à élaborer de l'information à partir de ce qu'il a lu ou étudié lors d'une phase d'apprentissage afin qu'il s'approprié les connaissances et que l'intégration de ces connaissances nouvelles avec celles antérieures soit facilitée. Ces stratégies ont prouvé leur efficacité sur l'apprentissage associé à des tâches de compréhension de textes (M. Anderson & Thiede, 2008 ; Thiede et al., 2010, 2005, 2003) et dans une moindre mesure sur l'apprentissage associé à des tâches de résolution de problème (Baars et al., 2013). Thiede a mon-

tré à plusieurs reprises (M. Anderson & Thiede, 2008 ; Thiede et al., 2010, 2005, 2003) que les différents outils basés sur les stratégies génératives (i.e. résumés, mots-clés, cartes conceptuelles) amélioraient l'exactitude de l'auto-évaluation à des tâches de compréhension de textes. De leur côté, dans une tâche de résolution de problème, Baars et al. (2013) ont révélé que lorsque les exemples résolus (de l'anglais 'worked-examples') étaient présentés sous forme de stratégie générative (i.e. des exemples à trous que l'apprenant doit compléter), ils modifiaient le biais d'auto-évaluation (de la surestimation vers la sous-estimation) mais pas son amplitude.

Parmi les stratégies génératives, les auto-explications se sont également avérées efficaces pour améliorer l'exactitude de l'auto-évaluation (Griffin et al., 2008). Les auto-explications consistent à demander à l'apprenant d'explicitier les raisons pour lesquelles il procède de telle façon et avec tel principe sous-jacent. Plus précisément, chez Griffin et al. (2008) et Bielaczyc et al. (1995), les auto-explications consistaient à expliquer les relations entre des concepts présentés (ex : 'Identifier et élaborer les relations entre les idées principales' ou 'déterminer la forme et la signification du code Lips' ; notre traduction de : 'identify and elaborate relations between the main ideas' ; 'determine both the form and meaning of the Lips code', p.237 Bielaczyc et al., 1995).

Les auto-explications permettent d'enrichir et d'améliorer la structure du réseau de connaissances, mais également de simplifier la production d'inférences utilisées pour former des principes généraux (Chi et al., 1989). La plupart des études qui ont étudié les auto-explications se sont penchées sur leurs effets sur les performances d'apprentissage ou de résolution de problèmes (e.g. Atkinson et al., 2003 ; Bielaczyc et al., 1995 ; Conati & Vanlehn, 2000 ; Hausmann & Chi, 2002). D'autres auteurs ont montré le lien entre les auto-explications et la qualité de l'autorégulation. Ainsi, Berardi-Coletta, Buyer, Dominowski, et Rellinger (1995) ont montré que des apprenants devant très concrètement expliquer à voix haute les stratégies et procédures métacognitives utilisées pour résoudre un problème (i.e. de produire des auto-explications) obtenaient une meilleure représentation du problème, et par conséquent, régulaient plus efficacement le processus de résolution du problème que des apprenants qui devaient simplement dire ce qu'ils faisaient pendant la résolution de problème. Ici le lien entre auto-explications et auto-évaluation nécessite de faire référence au modèle de Nelson et Narens (1990) qui annonce que la qualité de l'autorégulation dépend de la qualité de l'auto-évaluation. Néanmoins, l'amélioration des performances et de l'autorégulation suggère que les auto-explications pourraient améliorer l'auto-évaluation. Cette hypothèse est en accord avec l'étude de Griffin et al. (2008), dans laquelle les auteurs ont montré que les participants qui ont eu à produire des auto-explications tout en lisant un texte ont obtenu une auto-évaluation relative plus exacte que des appre-

nants qui ont simplement eu à relire le texte à comprendre. De leur côté [Chi et al. \(1989\)](#) ont mis en évidence que les apprenants les plus exacts dans l'évaluation de leurs performances produisaient des auto-explications de meilleure qualité.

En résumé, il ressort que faciliter l'accès aux indices pertinents permet de rendre l'auto-évaluation plus exacte. Plusieurs études ont montré l'efficacité des stratégies génératives et des auto-explications sur l'amélioration de l'exactitude de l'auto-évaluation relative en association avec des tâches de compréhension. En revanche, l'étude de [Baars et al. \(2013\)](#) a montré que certaines tâches génératives (les exemples résolus) permettent de modifier, sans l'améliorer, l'exactitude absolue à une tâche de résolution de problème tandis, qu'à notre connaissance, l'efficacité des auto-explications sur l'exactitude absolue n'ait pas encore été testé.

#### 4.3.3 *Inciter l'engagement dans le processus d'auto-évaluation*

Enfin, le troisième mode d'action consiste à inciter les apprenants à s'engager, de façon consciente ou non, dans l'utilisation de processus métacognitifs et plus particulièrement dans celui d'auto-évaluation. Ces incitations peuvent prendre la forme de récompenses et ainsi agir sur la motivation des apprenants à utiliser des processus métacognitif (i.e. bonification des notes des apprenants dont l'auto-évaluation est la plus exacte, [Miller & Geraci, 2011](#) ; [Schraw et al., 1993](#)). Elles peuvent également favoriser la prise de conscience des processus utilisés et inviter, de façon plus ou moins explicite, l'apprenant à s'engager dans des processus métacognitifs qui ne seraient pas toujours spontanés ([Azevedo et al., 2008](#) ; [Renkl, 1997](#)) ; ceci, en focalisant l'attention des apprenants sur les processus et stratégies d'apprentissage pertinents ([Bannert & Reimann, 2012](#)). [Bannert et Reimann \(2012\)](#) nomment le second type d'incitations 'incitations éducatives' (en anglais 'instructional prompts') et les définissent comme 'des procédures qui induisent et stimulent les activités cognitives, métacognitives, motivationnelles, personnelles et/ou coopératives pendant l'apprentissage' (notre traduction, p.195). Ces incitations peuvent également s'appeler explicitations ou auto-réflexions en référence au processus décrit par [Zimmerman \(2000\)](#), équivalent de l'auto-évaluation. [Gama \(2004\)](#) rappelle que ce type d'incitation aide les apprenants à analyser les stratégies qu'ils utilisent pour réaliser une tâche d'apprentissage et peuvent prendre la forme de questions fermées ou ouvertes.

Les effets des incitations métacognitives intégrées dans des systèmes d'apprentissage informatisés ont été assez peu étudiés sur l'amélioration de l'exactitude de l'auto-évaluation ([Mihalca, 2014](#)). Des études ont cependant mis en évidence les effets positifs des incitations sur l'apprentissage. Par exemple, [X. Lin et Lehman \(1999\)](#) ont montré qu'expliciter les raisons

des actions mises en place lors de l'apprentissage par l'apprenant était particulièrement efficace pour améliorer l'apprentissage notamment sur des tâches de transfert éloigné. Dans leur étude, les auteurs ont initialement confronté des étudiants en biologie à des problèmes qu'une étudiante fictive ne pouvaient pas résoudre seule. Les apprenants devaient, pour l'aider, identifier le type de problème, comprendre l'idée générale du sujet, etc. Lors de la confrontation des participants aux problèmes de l'étudiante fictive, les auteurs ont administré des incitations de différentes natures. Ces prompts étaient basés soit sur l'analyse des émotions ressenties, soit sur la nature de la tâche étudiée, soit sur la compréhension des stratégies et procédures utilisées. Un groupe contrôle ne recevait aucune incitation. Les auteurs ont ensuite mesuré les performances des étudiants dans la résolution de problèmes dont la thématique et le format des questions étaient similaires (transfert proche) puis sur des problèmes dont la thématique et le format des questions étaient différents (transfert éloigné). Les étudiants ont tous réussi les problèmes similaires, quel que soit leur groupe d'appartenance (groupe sans incitations inclus). En revanche, les étudiants incités à expliquer les stratégies et procédures utilisées ont été plus performants que les autres sur les problèmes contextuellement différents. Ainsi, en agissant sur la prise de conscience, il est possible que ces incitations aient permis aux apprenants d'extraire des règles générales (i.e. indices intrinsèques et extrinsèques selon Koriat) applicables à d'autres tâches.

Des conclusions tranchées concernant l'efficacité des incitations à l'utilisation de processus d'auto-évaluation sur l'exactitude de cette dernière ne peuvent être données. Il est, néanmoins, intéressant de vérifier si l'efficacité de ces incitations sur les performances d'apprentissage est liée à une amélioration de l'exactitude de l'auto-évaluation.

#### 4.4 QUESTION D'ÉTUDE ET HYPOTHÈSES

L'objectif de cette étude est de tester l'effet des auto-explications couplées à des incitations présentées sous forme de questions ouvertes et fermées et en différé de la tâche d'apprentissage sur l'exactitude de l'auto-évaluation. Les trois méthodes ont été combinées afin de viser une complémentarité et donc une efficacité maximale, le tout dans un environnement informatisé. En effet, les environnements numériques sont de plus en plus présents dans le paysage éducatif et pourtant, l'efficacité des outils d'aide à l'auto-évaluation sur l'amélioration de l'exactitude de l'auto-évaluation n'y est pas toujours univoque et mérite d'être approfondie. Les outils d'aides à l'auto-évaluation sont ici présentés a posteriori de la tâche d'apprentissage afin de créer un délai entre les tâches d'apprentissage et d'auto-évaluation. Ce délai a également permis d'éviter les inconvénients liés au traitement simultané du contenu d'apprentissage et de l'outil d'aide



à l'auto-évaluation qui occasionnerait une charge cognitive supplémentaire et délétère (Atkinson et al., 2003).

Dans cette étude, nous nous attendons à ce que les apprenants qui utilisent le dispositif d'aide à l'auto-évaluation s'auto-évaluent ensuite à un test de connaissances avec plus d'exactitude que ceux qui n'en bénéficient pas. De plus, la question de la généralisation de l'exactitude de l'auto-évaluation ayant été peu étudiée, nous explorerons l'effet du dispositif d'aide sur l'exactitude de l'auto-évaluation à d'autres contenus. Par ailleurs, rappelons que le contenu de connaissances à apprendre ciblé par le projet ANR dans lequel s'inscrit cette thèse est celui lié au C2I. Le référentiel du C2I est composé de questions de différentes natures et distingue notamment le type de connaissances évaluées (procédural versus déclaratif). La question de l'exactitude de l'auto-évaluation selon le type de connaissance n'ayant semble-t-il pas encore été examinée, nous nous proposons d'explorer également cette question.

Concernant l'exactitude de l'auto-évaluation, il est d'usage de distinguer l'exactitude relative et absolue (Koriat et al., 2002 ; Koriat, 2007). L'exactitude relative, mesurée généralement à l'aide du coefficient gamma de Goodman-Kruskal (Nelson, 1984), renseigne sur la capacité d'un apprenant à discriminer les informations apprises de celles qui ne le sont pas. Cependant le gamma ne dit rien sur la relation entre la performance prédite et réelle par item ; ce que permet de faire l'exactitude absolue (Baars et al., 2013 ; Schraw, 2009). De plus, l'exactitude absolue distingue le biais d'auto-évaluation, qui donne la direction du biais (i.e. si l'individu se sous-estime ou se surestime) et l'amplitude absolue du biais (i.e. l'écart absolu entre le jugement et la performance sans tenir compte du sens du biais). La direction du biais constitue une information intéressante puisqu'elle aura des conséquences en termes de régulation (Baars et al., 2013). Par exemple, Dunlosky et Rawson (2012) indiquent que la surestimation peut amener l'apprenant à stopper l'apprentissage avant que les connaissances ne soient parfaitement acquises. Pour ces raisons, nous ne mesurons dans notre étude, que l'exactitude absolue de l'auto-évaluation (direction et amplitude des biais). En effet, nous nous intéressons aux potentialités des dispositifs à aider les apprenants à améliorer l'exactitude de leur auto-évaluation de façon générale plutôt qu'à aider les apprenants à discriminer ce qu'ils savent de ce qu'ils ne savent pas.

## 4.5 MÉTHODE

### 4.5.1 *Participants et plan d'expérience*

Cinquante étudiants en Master 1 'Méthodes Informatiques Appliquées à la Gestion des Entreprises' (MIAGE) âgés de 20 à 25 ans ( $M = 22.08$  ans,

SD = 1.05 ans ; 9 filles et 41 garçons) ont été recrutés pour cette étude. Seuls 48 sujets ont été intégrés dans les traitements statistiques pour cause de données manquantes. Une mesure préalable des connaissances antérieures sur le domaine du traitement de texte a permis d'apparier les participants dans le groupe expérimental (dispositif avec aide à l'auto-évaluation, N = 24) et contrôle (dispositif contrôle sans aide à l'auto-évaluation, N = 24) de façon à ce que les niveaux de connaissances antérieures moyens soient équivalents pour les deux groupes.

#### 4.5.2 *Matériel et mesures*

Les participants ont rempli trois questionnaires à choix multiples dont les questions, présentées dans leur intégralité en Annexe C, étaient issues d'une base de données nationale délivrant le Certificat Informatique et Internet (C2I) niveau 1. Le choix de ce contenu est justifié par les études précédentes issues du contrat ANR (Vidal et al., 2015).

##### *Performance*

Pour chacun des trois questionnaires présentés ci-dessous, toutes les questions comprenaient plusieurs propositions de réponses parmi lesquelles (1) une ou plusieurs étaient correctes, (2) la proposition 'je ne sais pas', (3) la proposition 'aucune des réponses précédentes'. Les réponses fausses et 'je ne sais pas' étaient cotées 0 ; les réponses complètement justes valaient 1 ; les réponses partiellement justes étaient calculés comme suit : nombre de réponses données/nombres de réponses totales. Pour chaque question, le score était ensuite multiplié par 100 afin de pouvoir faire la différence avec la valeur du jugement d'auto-évaluation émis sur la même question.

Par exemple, pour une question, si trois propositions étaient à cocher mais que l'individu n'en cochait qu'une, le score de ce participant à cette question était  $(1/3) \times 100 = 33$ .

##### *Mesure des connaissances antérieures sur le traitement de texte*

Le premier questionnaire mesurait le niveau de connaissances initial des participants dans le domaine du traitement de texte. Le questionnaire était composé de douze questions découpées en trois blocs : quatre questions portaient sur les fonctionnalités de base du traitement de texte, quatre questions portaient sur les feuilles de styles et quatre questions étaient relatives à l'automatisation des styles. Dans chaque bloc de questions, deux questions faisaient appel à des connaissances déclaratives et deux autres à des connaissances procédurales.

Par exemple une question déclarative du bloc 'feuilles de styles' était 'Quelles sont les propriétés que l'on peut affecter à un style de paragraphe ? (1) L'en-tête de page ; (2) Les tabulations ; (3) L'alignement horizontal ; (4) les retraits ; (5) Les marges ; (6) Aucune des réponses précédentes ; (7) Je ne sais pas.' Une question procédurale du bloc 'feuille de style' était 'Dans un document, pour modifier la mise en forme de tous les titres de style 'Titre2', quelle est la méthode la plus efficace ? (1) Il faut créer un nouveau style basé sur le style 'Titre 2' ; (2) Il faut modifier le style 'Titre 2' et l'appliquer de nouveau aux titres de même niveau ; (3) Il faut utiliser pour chaque titre la commande 'Copier' puis 'Coller' ; (4) Il faut utiliser pour chaque titre le pinceau pour reproduire la mise en forme ; (5) Il faut seulement modifier le style 'Titre 2' ; (6) Aucune des réponses précédentes ; (7) Je ne sais pas.

Pour l'ensemble des indicateurs issus de ce questionnaire (i.e. performance, jugement d'auto-évaluation, direction du biais d'auto-évaluation, amplitude du biais d'auto-évaluation), les moyennes ont été calculées, pour chaque participant, selon les sous-ensembles de questions suivantes : les moyennes globales des indicateurs ont été calculées à partir des scores obtenus aux douze questions. Les scores aux connaissances déclaratives et procédurales ont été calculés en réalisant la moyenne aux six questions de natures déclaratives puis procédurales présentées.

#### *Mesure des connaissances sur le traitement de texte après dispositif : les questions similaires*

Le deuxième questionnaire visait à mesurer les connaissances après l'usage du dispositif d'aide à l'auto-évaluation ou du dispositif contrôle. Il était composé de six questions différentes du premier questionnaire mais néanmoins similaires car elles portaient sur les trois mêmes domaines. Deux questions portaient sur fonctionnalités de base du traitement de texte, deux questions portaient sur les feuilles de style, et deux questions portaient sur l'automatisation des styles. Pour chaque domaine, une question faisait plutôt appel à des connaissances déclaratives (exemple pour les ' fonctionnalités de base' : Quelle(s) propriété(s) n'a aucun sens pour formater un caractère dans un traitement de texte ? (1) Le retrait ; (2) La proportionnalité ; (3) la casse ; (4) la police ; (5) la couleur ; (6) Aucune des réponses précédentes ; (7) Je ne sais pas) ; l'autre question référait aux connaissances procédurales (exemple : 'Dans un traitement de texte, quelle est la meilleure façon d'avoir un saut de ligne au début de chaque paragraphe d'un groupe de paragraphes ? (1) Appliquer un format de paragraphe 'double interligne' en ayant le curseur dans le groupe de paragraphes ; (2) Appliquer un format de paragraphe 'saut de ligne avant' en sélectionnant le groupe de paragraphes ; (3) Appliquer un double interligne

en sélectionnant le groupe de paragraphes ; (4) Appliquer un format de paragraphe ‘saut de ligne avant’ en ayant le curseur dans le groupe de paragraphes ; (5) Aucune des réponses précédentes ; (6) Je ne sais pas’).

Comme précédemment, les moyennes pour chaque participant ont été calculées à partir du score obtenu aux six questions, et ce pour chaque indicateur (performance, jugement d’auto-évaluation, direction et amplitude du biais). Les scores aux connaissances déclaratives et procédurales ont été calculés en réalisant la moyenne aux trois questions de natures déclaratives puis procédurales présentées.

#### *Mesure des connaissances sur les tableurs après dispositif : question de généralisation*

Le troisième questionnaire, composé de sept questions, mesurait les connaissances sur les tableurs, un contenu différent des deux précédents. Celui-ci permettait d’investiguer la généralisation possible du processus d’auto-évaluation à différents domaines, c’est pourquoi ce questionnaire est nommé questionnaire de généralisation. Comme pour les deux questionnaires précédents, des questions faisaient appel soit à des connaissances déclaratives (par exemple : ‘À quel type de référence fait appel la notation suivante :  $\$B\$3$  ? (1) La référence monétaire ; (2)(\*) La référence absolue ; (3) La référence relative ; (4) La référence symbolique ; (5) La référence canadienne ; (6) Aucune des réponses précédentes ; (7) Je ne sais pas’). Soit à des connaissances procédurales (par exemple : ‘Dans un tableur, quelle formule est incorrecte pour calculer la somme des cellules A1, A2, A3 et A4 ? (1) =SOMME(A1 :A2 ;A3 :A4) ; (2) =SOMME(A1-A4) ; (3) =SOMME(A1 :A4) ; (4) =A1+A2+A3+A4 ; (6) Aucune des réponses précédentes ; (7) Je ne sais pas’.)

Pour ce questionnaire, les moyennes pour chaque participant ont été calculées à partir du score obtenu aux sept questions, et ce pour chaque indicateur (performance, jugement d’auto-évaluation et direction et amplitude du biais). Les scores aux connaissances déclaratives et procédurales ont été calculés en réalisant la moyenne des scores aux quatre questions de natures déclaratives et aux trois questions procédurales présentées.

#### *Mesure de l’auto-évaluation et de son exactitude*

Dans cette étude, l’objectif était d’évaluer principalement l’exactitude absolue de l’auto-évaluation sur des connaissances relatives au traitement de texte avant (1er questionnaire) et après usage du dispositif (2ème questionnaire). La généralisation du jugement et de l’exactitude de l’auto-évaluation a été évaluée sur des questions portant sur les tableurs (3ème questionnaire).

Une mesure du jugement d'auto-évaluation et deux indicateurs de son exactitude ont été utilisées.

LE JUGEMENT D'AUTO-ÉVALUATION. La mesure du jugement d'auto-évaluation permet de repérer à quel point les participants se sentent capables ou non de réussir la tâche proposée.

A chaque questionnaire et pour toutes les questions présentées, les sujets évaluaient a priori leurs performances sur une échelle de 0% (pas du tout sûr de réussir) à 100% (tout à fait sûr de réussir) en répondant à la question suivante 'estimez à quel point vous pensez réussir à répondre à la question suivante'. Un score du jugement moyen d'auto-évaluation a été calculé pour chaque participant sur chacun des trois questionnaires en réalisant la moyenne des jugements d'auto-évaluation émis pour chaque item d'un questionnaire.

DIRECTION DU BIAIS D'AUTO-ÉVALUATION. La direction du biais d'auto-évaluation était issu du calcul de la moyenne des différences entre l'auto-évaluation et la performance (sur 100) pour chacune des questions. Si la différence moyenne entre le jugement d'auto-évaluation et la performance est négative, l'apprenant a globalement sous-évalué ses performances ; si la différence moyenne est positive, l'apprenant a globalement sur-évalué ses performances.

Un exemple concret est donné dans le Tableau 3 sur la base d'un questionnaire fictif composé de quatre questions.

TABLEAU 3. – Illustration du calcul des indicateurs de biais d'auto-évaluation

Numéro de la question	Score de performance sur 100	Score au jugement d'auto-évaluation sur 100	Direction du biais d'auto-évaluation	Amplitude du biais d'auto-évaluation
1	100	0	100	100
2	33	50	-17	17
3	0	100	-100	100
4	100	100	0	0
Moyenne des différents scores	58.25	62.5	-4.25	54.25

La mesure de la 'direction du biais d'auto-évaluation' permet de déterminer si les apprenants ont globalement tendance à sous-évaluer ou sur-évaluer leurs performances. Ce score est intéressant mais il ne permet

pas de définir l'ampleur du biais (i.e. l'écart absolue entre le jugement et la performance) lorsque des moyennes sont calculées pour un groupe de participants ou pour un même sujet sur la base de plusieurs items. En effet, un participant qui sous-évalue sa performance (score égal à -50) peut s'équilibrer avec un participant qui se sur-évalue (score égal à +50). La moyenne sera égale à 0 ce qui laisserait penser que les apprenants s'auto-évaluent avec une exactitude parfaite. Pour contrer cette limite, il est nécessaire de calculer l'amplitude du biais (Baars et al., 2013). De cette façon, les biais de directions différentes ne peuvent plus s'annuler. Par exemple, dans l'exemple précédent, la moyenne pour les deux sujets serait de 50, ce qui reflèterait bien que les apprenants ne s'auto-évaluent pas avec exactitude. Ainsi, l'amplitude du biais se réduit si le score des apprenants se rapproche de 0 (i.e. lorsqu'ils tendent vers une auto-évaluation exacte Baars et al., 2013, 2014).

AMPLITUDE DU BIAIS D'AUTO-ÉVALUATION. Pour déterminer l'amplitude de l'auto-évaluation (cf. Tableau 3), nous avons calculé la racine carré des écarts entre le jugement d'auto-évaluation et la performance élevée au carré pour chaque question des questionnaires (e.g. Baars et al., 2013). Le score peut varier de 0 à 100. Plus le score est proche de 0, plus l'apprenant sait évaluer avec exactitude ses performances puisqu'il y a peu d'écart entre les performances prédites et les performances réelles. Plus le score tend vers 100 moins l'apprenant sait évaluer ses performances puisque l'écart entre les performances prédites et réelles est important.

#### 4.5.3 *Le dispositif d'aide à l'auto-évaluation*

Les douze questions du premier questionnaire sur les fonctionnalités de traitement de texte ont été reprises et utilisées comme support au dispositif d'aide à l'auto-évaluation.

Le dispositif se composait de trois aides à l'auto-évaluation aux fonctions complémentaires dont une copie d'écran est présentée dans la Figure 4.5.3. Deux types d'incitations à la prise de conscience ont été utilisés. (1) Les 'explicitations', sous forme de questions ouvertes visaient à aider les étudiants à prendre conscience des stratégies qu'ils avaient utilisées pour répondre aux questions du premier questionnaire. Les apprenants disposaient d'un espace pour expliquer et justifier à l'écrit les stratégies utilisées pour répondre aux questions et s'auto-évaluer (X. Lin & Lehman, 1999). (2) Un outil de prise de conscience des stratégies utilisées et des stratégies pertinentes présenté sous forme de question fermées (Gama, 2004). Les participants devaient cocher parmi une liste de six stratégies (trois étaient pertinentes, i.e. efficaces pour un apprentissage en profondeur, trois étaient non pertinentes) celles qu'ils avaient utilisées lors des

auto-évaluations de leurs connaissances mais également celles qui leur semblaient être pertinentes par la suite. Ces questions fermées offraient un guidage supplémentaire aux apprenants en leur proposant un support pré-existant pour réfléchir aux stratégies d'apprentissage qu'ils avaient utilisées et/ou qui étaient pertinentes pour l'avenir. Ceci venait compléter les questions ouvertes qui pouvaient ne pas être traitées par un apprenant qui ne comprenait pas ce qui lui était demandé, ou ne savait pas quoi répondre. (3) Les auto-explications ont été construites sous le modèle de celles utilisées par Griffin et al. (2008) ou Bielaczyc et al. (1995). L'objectif étant de faire générer de l'information aux participants sur les concepts importants présents dans la consigne et/ou les réponses proposées. Par exemple l'apprenant devait répondre sur l'ordinateur, dans un encadré prévu à cet effet à la question suivante : 'Essayez d'explicitier les concepts clés présents dans la question et les réponses proposées.' Ainsi, l'apprenant pouvait accéder à ses connaissances antérieures et se rendre compte d'éventuels écarts entre ce qu'il savait réellement et ce qu'il pensait savoir.

1. Outil d'explicitation, l'apprenant doit expliquer pourquoi il a répondu ce qu'il a répondu.

Dans un traitement de texte, afin de réaliser une table des matières à deux niveaux, que doit-on mettre en place?

- Un tableau avec tous les titres des deux premiers niveaux hiérarchiques
- Un signet pour chacun des titres
- Des styles pour les niveaux de titre 1 et 2
- Un index des titres
- Aucune des réponses précédentes

**Veillez justifier votre réponse à la question précédente.**

2. Outil de choix de stratégies.

Nous vous proposons ci-dessous une liste de stratégies que vous auriez pu utiliser et qui justifieraient l'évaluation produite. Veuillez cocher les stratégies que vous avez réellement utilisées, de même que les stratégies qui vous semblent pertinentes maintenant.

Stratégies	Stratégies réellement utilisées	Stratégies pertinentes
J'ai fait ma prédiction de réussite au hasard.	<input type="checkbox"/> OUI <input type="checkbox"/> NON	<input type="checkbox"/> OUI <input type="checkbox"/> NON
J'ai essayé de répondre à la question avant de lire les propositions de réponses.	<input type="checkbox"/> OUI <input type="checkbox"/> NON	<input type="checkbox"/> OUI <input type="checkbox"/> NON
J'ai basé mon auto-évaluation sur un de mes résultats précédents.	<input type="checkbox"/> OUI <input type="checkbox"/> NON	<input type="checkbox"/> OUI <input type="checkbox"/> NON
J'ai découpé la question en différentes sous-questions.	<input type="checkbox"/> OUI <input type="checkbox"/> NON	<input type="checkbox"/> OUI <input type="checkbox"/> NON
J'ai relu la question pour voir les parties qui ne me semblaient pas claires.	<input type="checkbox"/> OUI <input type="checkbox"/> NON	<input type="checkbox"/> OUI <input type="checkbox"/> NON
J'ai essayé de répondre à la question sans évaluer ma compréhension des concepts associés.	<input type="checkbox"/> OUI <input type="checkbox"/> NON	<input type="checkbox"/> OUI <input type="checkbox"/> NON



3. Outil d'auto-explication des connaissances, l'apprenant doit expliquer les principaux concepts présentés.

**Question 2/61**

Quelles sont les propriétés que l'on peut affecter à un style de paragraphes ?

- Les tabulations
- L'en-tête de page
- L'alignement horizontal
- Les retraits
- Les marges
- Aucune des réponses précédentes

**Essayez d'explicitier les concepts clés présents dans la question et les réponses proposées.**

#### 4.5.4 *Dispositif contrôle*

Le dispositif contrôle correspond à ce qui a été donné aux participants qui n'ont pas bénéficié du dispositif d'aide à l'auto-évaluation. Il a été conçu afin que le temps de réalisation de l'expérience ainsi que le format de réponses des questions soient approximativement les mêmes pour les participants du groupe contrôle et ceux du groupe expérimental.

Trois questions conçues pour ne solliciter aucune activité métacognitive et portant sur cinq domaines du quotidien (le sport, la musique, la cuisine, le cinéma et les voyages) étaient posées. Un exemple de chaque type de question est donné ci-après (le thème est ici la cuisine). (1) 'Quel est votre plat préféré (la choucroute, les lasagnes, le cassoulet...)? Veuillez indiquer votre réponse dans l'espace prévu à cet effet.' (2) 'Vous préférez plutôt : les pâtisseries, les plats en sauce, les salades,...' (3) 'Dans la catégorie choisie précédemment, quel(s) plat(s) préférez-vous? Veuillez indiquer votre réponse dans l'espace prévu à cet effet.'

#### 4.5.5 *Procédure*

L'expérience avait lieu dans une salle informatique où chaque étudiant disposait d'un ordinateur pour répondre seul aux questions.

Lors d'une première session les étudiants répondaient uniquement au questionnaire mesurant leur niveau de connaissances antérieures (questionnaire 1). Pour chacune des douze questions, il prédisait a priori la performance qu'ils pensaient obtenir.

Dans un second temps, la répartition des apprenants selon la condition expérimentale était faite (dispositifs d'aide à l'auto-évaluation ou contrôle). Pour les étudiants du groupe expérimental, les trois outils d'aide à l'auto-évaluation étaient mis en lien avec les questions du premier questionnaire présenté lors de la première session. Pour chaque question, l'apprenant devait expliciter les principaux concepts, expliciter les stratégies utilisées pour répondre aux questions, s'auto-évaluer et enfin, choisir les stratégies qu'il avait utilisées et qui lui semblait pertinentes. Les étudiants du groupe contrôle répondaient aux questions du dispositif contrôle. Ensuite, ils répondaient tous au deuxième questionnaire à choix multiple sur le traitement de texte ainsi qu'au questionnaire sur les tableurs. La passation se clôturait avec un débriefing afin d'expliquer aux étudiants le but de cette expérimentation. Une schématisation de la procédure est présentée dans la Figure 4.

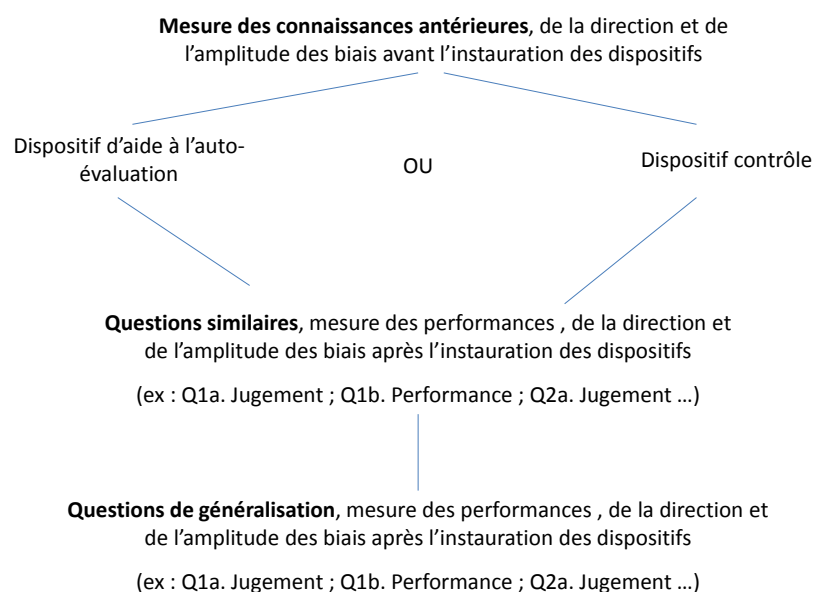


FIGURE 4. – Synthèse de la procédure utilisée dans l'Expérience 1.

## 4.6 RÉSULTATS

### 4.6.1 Remarques préliminaires

L'objectif de l'étude étant de détecter l'effet d'un dispositif d'aide à l'auto-évaluation sur l'exactitude de l'auto-évaluation, des analyses de la variance (ANOVA) à mesures mixtes  $3 \times 2$  ont été conduites. La variable indépendante en intra-sujet était le moment de la mesure (avant le dispositif versus après sur des questions similaires et après sur des questions de généralisation). La variable 'condition expérimentale' était quant à elle en inter-sujet et comportait deux modalités : (a) dispositif d'aide à l'auto-évaluation (i.e. groupe expérimental) (b) dispositif contrôle (i.e. groupe contrôle). Les variables dépendantes analysées étaient successivement la valeur du jugement d'auto-évaluation, la direction des biais d'auto-évaluation, l'amplitude des biais d'auto-évaluation et les performances. Afin d'examiner la question de l'exactitude de l'auto-évaluation selon le type de connaissances (déclaratives vs. procédurales), les analyses globales ont été complétées par des analyses distinguant les connaissances procédurales et déclaratives (pour les trois questionnaires). Pour ces dernières analyses, et dans le but de ne pas alourdir inutilement la section résultat, seule l'amplitude des biais qui constitue le centre de notre intérêt sera présentée dans le corps du texte. Les mêmes analyses conduites sur

la direction du biais, la performance et le jugement ainsi que les analyses distinguant les trois sous-domaines de compétence (i.e. les fonctionnalités de base du traitement de texte ; les feuilles de styles ; l'automatisation des styles, pour les questionnaires de connaissances initiales et similaires) seront présentées en Annexe D. Lorsqu'un effet d'interaction et/ou un effet principal de la variable indépendante 'moment de la mesure' ont été observés, des t-tests étaient conduits afin de distinguer les conditions qui différaient les unes des autres.

#### 4.6.2 *Statistiques descriptives*

Les moyennes et écart types des quatre variables dépendantes selon le moment de la mesure et la condition expérimentale sont présentés dans le Tableau 4. Les moyennes et écarts-types concernant l'amplitude des biais relatifs à la distinction entre connaissances déclaratives et procédurales sont présentés dans le Tableau 5.

Concernant le groupe expérimental, les apprenants estiment généralement bien réussir l'ensemble des questionnaires ( $M = 67.37$ ) et ce jugement est plus haut pour le questionnaire initial ( $M = 70.69$ ) que similaire ( $M = 65.69$ ) ou du questionnaire de généralisation ( $M = 63.1$ ). Les performances correspondent à ces jugements puisque les performances des participants sont globalement hautes (70% de réussite) et sont meilleures lors du premier questionnaire ( $M = 72.1$ ) que lors des deux questionnaires qui ont suivi la présentation du dispositif d'aide à l'auto-évaluation ( $M = 68.58$  pour le questionnaire similaire ;  $M = 69.41$  pour le questionnaire de généralisation). Au regard de la direction des biais, les trois moyennes sont négatives et celle pour le premier questionnaire est presque égale à 0 ( $M = -1.22$ ) tandis que pour le questionnaire de généralisation, le biais est légèrement plus élevé ( $M = -6.31$ ) sans pour autant être très haut. Ces premières observations laisseraient penser que les biais d'auto-évaluation sont faibles. Pourtant, les données relatives à l'amplitude des biais d'auto-évaluation montrent que les apprenants s'auto-évaluent avec une exactitude qui n'est pas si parfaite puisque l'amplitude est comprise entre 28 et 37. L'amplitude du biais est quasiment égale pour le questionnaire initial ( $M = 27.66$ ) et celui de généralisation ( $M = 28.06$ ) tandis que l'exactitude de l'auto-évaluation est plus faible sur le questionnaire similaire ( $M = 37.63$ ). Pour ce qui est des résultats sur l'amplitude du biais selon le type de connaissances, les résultats suivent la tendance générale puisque l'amplitude du biais est un petit peu plus élevée pour le questionnaire similaire tant pour les connaissances déclaratives ( $M = 36.78$ ) que procédurales ( $M = 38.47$ ). Il est toutefois possible de remarquer que les apprenants sont globalement plus exacts dans l'évaluation de leurs connaissances déclaratives ( $M = 29.23$ ) que procédurales ( $M = 32.74$ ).

TABLEAU 4. – Ensemble des moyennes (et écarts-types) calculées pour le jugement d'auto-évaluation, l'amplitude du biais d'auto-évaluation, la direction du biais, et la performance en fonction de la condition expérimentale et du moment de la mesure. Le symbole \* indique les variables exprimées en pourcentage.

	Min ; Max	Groupe expérimental (N = 24)	Groupe contrôle (N = 24)	Total (N = 48)	
<b>Jugement</b>	Connaissances antérieures	0 ; 100	70.69 (13.11)	70 (12.75)	70.35 (12.8)
	Questions similaires	0 ; 100	65.69 (16.55)	69.99 (16.86)	67.84 (16.67)
	<b>d'auto-évaluation *</b>	Questionnaire de généralisation	0 ; 100	63.1 (26.56)	73.52 (14.81)
	Total	0 ; 100	67.37 (13.55)	70.98 (8.65)	69.17 (11.39)
<b>Amplitude du biais</b>	Connaissances antérieures	0 ; 100	27.66 (6.53)	29.43 (7.87)	28.55 (7.2)
	Questions similaires	0 ; 100	37.63 (10.44)	34.19 (9.97)	35.91 (10.25)
	<b>d'auto-évaluation *</b>	Questions de généralisation	0 ; 100	28.06 (12.23)	27.29 (11.41)
	Total	0 ; 100	30.16 (5.75)	29.97 (5.92)	30.07 (5.77)
<b>Direction du biais</b>	Connaissances antérieures	-100 ; 100	-1.22 (13.98)	-4.38 (14.82)	-2.8 (14.3)
	Questions similaires	-100 ; 100	-2.89 (22.38)	-3.52 (20.52)	.32 (21.49)
	<b>d'auto-évaluation</b>	Questions de généralisation	-100 ; 100	-6.31 (19.38)	-5 (16.1)
	Total	-100 ; 100	-3.05 (14.08)	-2.66 (11.62)	-2.85 (12.78)
<b>Performance *</b>	Connaissances antérieures	0 ; 100	72.1 (10.35)	74.5 (9.33)	73.28 (9.82)
	Questions similaires	0 ; 100	68.58 (19.33)	66.46 (17.74)	67.52 (18.39)
	<b>d'auto-évaluation *</b>	Questions de généralisation	0 ; 100	69.41 (26)	78.53 (11.88)
	Total	0 ; 100	70.48 (13.04)	73.7 (8.37)	72.09 (10.96)

Concernant le groupe contrôle, les tendances sont sensiblement les mêmes au moins sur les indicateurs de l'exactitude des biais d'auto-évaluation. En effet, la direction des biais indique à nouveau que les participants ont eu globalement tendance à se sous-évaluer légèrement ( $M = -2.66$ ) et l'amplitude montre que ce biais était plus important pour le questionnaire similaire ( $M = 34.19$ ) que pour les deux autres questionnaires ( $M = 29.43$  pour le questionnaire initial ;  $M = 27.29$  pour le questionnaire de généralisation). En revanche, concernant les jugements d'auto-évaluation, les résultats indiquent que les apprenants ont estimé plus fortement leurs performances au questionnaire de généralisation ( $M = 73.52$ ) plutôt qu'au questionnaire initial ( $M = 70$ ) et leurs performances ont suivi ces tendances ( $M = 78.53$  au questionnaire de généralisation ;  $M = 74.5$  au questionnaire initial). Les résultats distinguant le type de connaissance indiquent ici pour les connaissances déclaratives que l'amplitude des biais au second questionnaire ( $M = 36.97$ ) est plus importante comparativement à l'amplitude des biais au deux autres questionnaire ( $M = 27.81$  pour le questionnaire initial ;  $M = 22.48$  pour le questionnaire de généralisation). Sur les connaissances procédurales, les biais sont sensiblement plus importants sur le questionnaire de généralisation ( $M = 33.7$ ). Enfin, l'auto-évaluation est à nouveau globalement plus exacte pour les questions déclaratives ( $M = 29.09$ ) que procédurales ( $M = 32.06$ ).

#### 4.6.3 *Le jugement d'auto-évaluation*

L'Anova n'a révélé aucun effet du moment du jugement (avant dispositif, questions similaires puis questions de généralisation),  $F(2, 92) = .35$ ,  $p = .7$ ,  $\eta_p^2 = .01$ , aucun effet du groupe,  $F(1, 46) = 1.85$ ,  $p = .2$ ,  $\eta_p^2 = .04$ , et aucun effet d'interaction,  $F(2, 92) = 1.54$ ,  $p = .2$ ,  $\eta_p^2 = .03$ .

#### 4.6.4 *La direction des biais d'auto-évaluation*

L'Anova réalisée sur la direction des biais d'auto-évaluation n'a révélé aucun effet du moment,  $F(2, 92) = 1.88$ ,  $p = .2$ ,  $\eta_p^2 = .04$ , aucun effet principal du dispositif,  $F(1, 46) = .16$ ,  $p = .7$ ,  $\eta_p^2 = .003$  et aucun effet d'interaction,  $F(2, 92) = 1.2$ ,  $p = .3$ ,  $\eta_p^2 = .03$ .

#### 4.6.5 *L'amplitude des biais d'auto-évaluation*

L'Anova a révélé un effet principal du moment de la mesure,  $F(2, 92) = 11.27$ ,  $p < .001$ ,  $\eta_p^2 = .2$ . Mais contrairement à ce qui était attendu, les t-tests montrent une différence significative entre l'amplitude des biais d'auto-évaluation observés avant l'introduction du dispositif et les biais

TABLEAU 5. – Moyennes et écarts-types de l'amplitude des biais selon le type de connaissance, le moment de la mesure et la condition expérimentale

		<b>Groupe Expérimental</b>	<b>Groupe contrôle</b>	<b>Total</b>
<b>Connaissances antérieures</b>	Connaissances déclaratives	28,39 (6,63)	27,81 (9,08)	28,1 (7,87)
	Connaissances procédurales	26,94 (10,35)	31,05 (9,75)	29 (10,16)
<b>Questions similaires</b>	Connaissances déclaratives	36,78 (14,37)	36,97 (14,03)	36,88 (14,05)
	Connaissances procédurales	38,47 (12,94)	31,42 (11,52)	34,94 (12,63)
<b>Questions de généralisation</b>	Connaissances déclaratives	22,97 (13,55)	22,48 (4,89)	22,72 (14,09)
	Connaissances procédurales	34,85 (18,22)	33,7 (14,35)	34,28 (16,23)
<b>Total</b>	Connaissances déclaratives	29,38 (8,02)	29,09 (7,51)	29,23 (7,69)
	Connaissances procédurales	33,42 (8,79)	32,06 (7,99)	32,74 (8,34)

d'auto-évaluation observés sur des questions similaires,  $t(47) = -4.1, p < .001, d = .8$ , suggérant que les participants se sont auto-évalués avec plus d'exactitude avant l'introduction du dispositif (quel qu'il soit,  $M = 28.55, SD = 7.2$ ) qu'après sur un contenu similaire ( $M = 35.91, SD = 10.25$ ). Les t-tests ont indiqué également une différence significative entre l'amplitude des biais d'auto-évaluation obtenus sur des questions similaires et les biais d'auto-évaluation obtenus sur des questions de généralisation,  $t(47) = -4.3, p < .001, d = .75$ , indiquant que les participants se sont auto-évalués avec plus d'exactitude sur des questions de généralisation ( $M = 27.67, SD = 11.71$ ) que pour des questions similaires. Aucune différence significative n'a été observée entre les biais d'auto-évaluation observés avant l'introduction du dispositif (d'aide ou contrôle) et les biais d'auto-évaluation observés pour des questions de généralisation,  $t(47) = .44, p > .7, d = .09$ .

L'Anova n'a en revanche pas révélé d'effet principal du type de dispositif,  $F(1, 46) = .19, p = .7, \eta_p^2 = .004$ , ni d'effet d'interaction entre nos deux variables indépendantes,  $F(2, 92) = .93, p = .4, \eta_p^2 = .02$ .

#### 4.6.6 *Les performances*

L'Anova n'a pas révélé d'effet du dispositif d'aide à l'auto-évaluation sur les performances,  $F(1, 46) = .8, p = .4, \eta_p^2 = .02$ , ni d'effet d'interaction entre les deux variables indépendantes,  $F(2, 92) = 1.90, p = .2, \eta_p^2 = .04$ . Un effet marginal du moment de la mesure a été détecté,  $F(2, 92) = 2.99, p = .06, \eta_p^2 = .06$ . Les t-tests indiquent une différence significative entre les performances mesurées avant l'introduction du dispositif et les performances mesurées sur des questions similaires,  $t(47) = 2.28, p = .03, d = .4$ ; les participants ont été plus performants sur le questionnaire initial ( $M = 73.28, SD = 9.82$ ) que sur les questions similaires ( $M = 67.52, SD = 18.39$ ). Aussi, les participants tendent à mieux réussir les questions de généralisation ( $M = 73.97, SD = 20.53$ ) que les questions similaires,  $t(47) = -1.97, p = .06, d = .33$ .

#### 4.6.7 *Analyses complémentaires*

L'observation des données initiales nous a permis de constater qu'il existait souvent chez un individu des écarts importants dans l'exactitude de l'auto-évaluation d'une question à une autre. Aussi, nous avons distingué les moyennes de l'amplitude des biais selon que les questions portaient sur un contenu déclaratif ou procédural. Les résultats portant sur nos autres variables dépendantes sont présentés en Annexe D, de même que



les analyses portant sur la distinction entre les différents sous-domaines des questionnaires 1 et 2.

Une Anova  $3 \times 2 \times 2$  a été conduite sur l'amplitude des biais avec pour variables indépendantes le moment de la mesure (avant le dispositif versus après sur des questions similaires et des questions de généralisation); la 'condition expérimentale' (dispositif d'aide à l'auto-évaluation versus dispositif contrôle) et le type de connaissances (déclaratives versus procédurales).

L'Anova à trois facteurs n'a révélé d'interaction ni entre le groupe expérimental, le moment de la mesure et le type de connaissance,  $F(2, 92) = 1.76, p = .18, \eta_p^2 = .04$ , ni entre le moment et le groupe,  $F(2, 92) = .92, p = .4, \eta_p^2 = .02$  ou entre le type de connaissance et le groupe,  $F(2, 92) = .14, p = .7, \eta_p^2 = .003$ . Seul un effet d'interaction entre le type de connaissance et le moment de la mesure a été détecté,  $F(2, 92) = 9.98, p < .001, \eta_p^2 = .18$ .

Des t-test à mesures appariées ont été conduits afin de d'identifier les paires significativement différentes. Six paires de t-tests ont été réalisées, le seuil de signification a donc été abaissé à .008 (Dancey & Reidy, 2007). Une différence significative a été observée sur les connaissances déclaratives entre l'amplitude des biais d'auto-évaluation obtenue au questionnaire initial et au questionnaire similaire,  $t(47) = -3.884, p < .001, d = .77$  et l'amplitude des biais obtenue au questionnaire similaire et au questionnaire de généralisation,  $t(47) = 5.26, p < .001, d = 1$ . Ces résultats suggèrent que, pour les connaissances déclaratives, les participants étaient plus exacts au questionnaire initial ( $M = 28.1, SD = 7.87$ ) et au questionnaire de généralisation ( $M = 22.72, SD = 14.08$ ) qu'au questionnaire similaire ( $M = 36.88, SD = 14.05$ ). Aucune différence significative n'est apparue entre l'exactitude des biais pour les connaissances procédurales au seuil .008.

Notons également qu'un effet principal des connaissances a été révélé grâce à un t-test à mesures appariées,  $t(47) = .248, p < .02$  suggérant que les participants s'auto-évaluent avec plus d'exactitude sur des connaissances déclaratives ( $M = 29.23, SD = 7.69$ ) que sur des connaissances procédurales ( $M = 32.74, SD = 8.34$ ).

## 4.7 DISCUSSION

### 4.7.1 Rappel des hypothèses et synthèse des principaux résultats

L'objectif de cette première étude était de tester l'efficacité de plusieurs méthodes d'aides à l'auto-évaluation cumulées sur l'amélioration de son exactitude. Pour cela, l'exactitude de l'auto-évaluation de participants ayant utilisé un dispositif d'aide à l'auto-évaluation était comparée

à l'exactitude de l'auto-évaluation d'étudiants n'ayant pas utilisé ce dispositif. Afin de tester la généralisation des compétences métacognitives éventuellement apprises à des nouveaux contenus, les individus devaient s'auto-évaluer sur un domaine identique à celui utilisé lors du dispositif puis sur un domaine différent. Nous nous attendions non seulement à observer un effet principal du dispositif, les participants avec le dispositif d'aide à l'auto-évaluation devaient s'auto-évaluer avec plus d'exactitude que les participants du dispositif contrôle ; mais également un effet principal du moment, l'auto-évaluation des participants devait être globalement plus exacte après l'introduction du dispositif d'aide qu'avant. L'interaction entre ces deux variables aurait également dû mettre en évidence une amélioration de l'exactitude de l'auto-évaluation chez les participants du dispositif d'aide à l'auto-évaluation uniquement.

Les résultats n'ont pas mis en évidence les effets positifs attendus de l'outil d'aide à l'auto-évaluation puisque seul un effet principal du moment indiquait que les participants réussissaient moins bien et avaient un biais d'amplitude plus grande après l'utilisation des dispositifs qu'avant, quel que soit la nature de celui-ci. Cet effet n'a été observé que sur les questions similaires et non sur les questions de généralisation puisque lors du dernier questionnaire, les performances et l'auto-évaluation ne différaient pas du niveau de base. Ce résultat relève à la fois un manque d'efficacité du dispositif d'aide à l'auto-évaluation ainsi que l'impact délétère d'un facteur commun aux groupes expérimental et contrôle (puisque les scores des deux groupes ont été affectés). En parallèle, nous avons pu observer que les apprenants ne s'auto-évaluaient pas avec la même exactitude selon le type de connaissances. L'auto-évaluation en lien avec des connaissances déclaratives s'est révélée significativement plus exacte que l'auto-évaluation liée à des connaissances procédurales.

Pour expliquer successivement l'ensemble de ces résultats, nous proposerons dans un premier lieu une justification en termes de manque de familiarisation au dispositif puis nous nous pencherons sur la difficulté des questions qui aurait pu causer la baisse des performances et l'augmentation des biais d'auto-évaluation aux questions similaires. Enfin, nous évoquerons l'expertise afin d'expliquer les différences observées entre les connaissances déclaratives et procédurales.

#### 4.7.2 *L'absence de familiarisation à l'utilisation du dispositif d'aide à l'auto-évaluation*

L'absence d'effet du dispositif d'aide à l'auto-évaluation peut s'expliquer potentiellement du fait que les participants ont manqué d'investissement cognitif dans l'utilisation des outils d'aide à l'auto-évaluation du fait de l'absence de phase de familiarisation aux différentes questions propo-

sées. En effet, [Berardi-Coletta et al. \(1995\)](#) et [Bannert et Reimann \(2012\)](#) précisent que l'efficacité des aides dépend de l'investissement cognitif des apprenants pour leur traitement. En parallèle, [Berry et Broadbent \(1987\)](#) avaient spécifié au moins pour les auto-explications qu'il était nécessaire d'instaurer préalablement au dispositif une tâche de familiarisation visant à indiquer aux étudiants comment répondre aux questions. De la même façon [Bielaczyc et al. \(1995\)](#) ont mis en évidence que des étudiants qui avaient reçu un entraînement explicite à l'utilisation des auto-explications et de différentes stratégies métacognitives utilisaient plus efficacement ces outils que les participants pour qui l'entraînement était tacite. Ainsi, les apprenants n'ont peut-être pas su comment répondre de façon pertinente aux questions et par conséquent, ils n'ont pas investi la tâche autant qu'ils auraient dû.

Pour la suite, il sera nécessaire de proposer une phase de familiarisation relative à l'outil utilisé et de contrôler que les apprenants aient bien utilisé les outils.

#### 4.7.3 *Effet de la difficulté des questions*

[Weaver et Bryant \(1995\)](#) ont mis en évidence que l'auto-évaluation d'apprenants était plus exacte quand la difficulté de la tâche d'apprentissage était intermédiaire comparativement à une tâche difficile où l'exactitude diminue. Pour le second questionnaire, les résultats concernant les performances chutent ce qui reflèterait une plus grande difficulté des questions malgré notre volonté de les avoir conçues de difficulté équivalente et que leur construction ait été effectuée sur la base du référentiel du C2I et en collaboration avec des enseignants de la discipline. Aussi il est probable que, comme pour l'évaluation de la compréhension de textes narratifs et expositifs dans l'étude de [Weaver et Bryant \(1995\)](#), l'exactitude de l'auto-évaluation diminue lorsque le contenu des questions se complexifie.

#### 4.7.4 *L'expertise pour expliquer l'effet du type de connaissance*

Au regard des performances obtenues, les individus qui ont participé à cette étude semblaient être relativement experts dans les domaines sur lesquels ils ont été interrogés. L'observation d'une auto-évaluation d'amplitude différente selon la nature des connaissances (déclaratives versus procédurales) pourrait être liée au fait que comme le rappelle [Chi \(2006\)](#) les experts, dans différents domaines sont particulièrement dépendants du contexte de la situation pour réaliser une tâche. Les individus extraient implicitement des indices de la situation pour accomplir correctement cette tâche. Par exemple, dans une tâche de diagnostic médical, les médecins experts, comparativement aux novices, utilisent des informations contex-

tuelles qui sont à leur disposition (ex : âge, antécédents médicaux, etc. [Hobus, Schmidt, Boshuizen, & Patel, 1987](#)).

L'auto-évaluation repose sur l'utilisation d'indices, il est possible que les experts en évaluant leurs performances relatives à des connaissances procédurales sans être dans un contexte de mise en application de ces connaissances, ne soient pas en capacité d'utiliser des indices pertinents, prédictifs de la performance.

Aussi dans les expériences suivantes, une tâche d'apprentissage avec un exercice procédural seront mise en place afin que les conditions d'auto-évaluation soient davantage contextualisées.

Par ailleurs, l'absence d'effet du dispositif peut refléter une inadéquation de l'outil présenté au niveau d'expertise des apprenants. En effet, les experts et les novices semblent utiliser des indices différents pour prédire leurs performances. Par exemple, sur une tâche complexe de compréhension, [Griffin et al. \(2008\)](#) mettent en évidence que les experts tendent à utiliser davantage d'indices basés sur la représentation qu'ils ont du texte. En parallèle, la théorie de la charge cognitive a mis en évidence l'effet de l'expertise inversée ([Kalyuga et al., 2003](#)) qui stipule que des outils d'aide à l'apprentissage utiles pour des novices peuvent s'avérer inutiles voire néfastes pour des experts. Cette inversion de l'efficacité d'un outil selon le niveau d'expertise d'un apprenant s'explique du fait que les novices, possédant peu de connaissances dans un domaine peuvent bénéficier d'informations complémentaires explicitant certaines relations entre des concepts. Au contraire, les experts possèdent ces connaissances dans leur domaine d'expertise. Leur présenter des connaissances qu'ils possèdent déjà deviendrait alors une source de perturbation. Par extension, il est possible que l'outil d'aide à l'auto-évaluation, en demandant à des experts d'utiliser des indices qu'ils utilisent déjà spontanément n'a pas eu les effets escomptés.

Dans l'étude suivante, des individus de différents niveaux d'expertise sur le domaine du C2I seront recrutés. Aussi, un nouvel outil permettant de proposer différents formats d'utilisation sera également exploité. L'utilisation d'un même outil déclinable sous différents formats permet d'explorer la question de l'expertise inversée appliquée à l'exactitude de l'auto-évaluation tout en évitant des biais qui seraient causés par la comparaison d'outils différents.

Nous avons vu que :

- L'exactitude de l'auto-évaluation des participants qui ont bénéficié d'un outil d'aide à l'auto-évaluation ne s'est pas améliorée.
- Les participants ont vu l'amplitude de leur auto-évaluation diminuer sur un contenu identique et revenir au niveau initial sur un contenu différent.
- Les participants auto-évaluent avec plus d'exactitude leurs connaissances déclaratives que procédurales lorsque les deux types de connaissances sont présentées sous un même format (type QCM).

# 5

---

## EFFET DU MODE ET DU MOMENT DE PRÉSENTATION D'UN OUTIL D'AIDE À L'AUTO-ÉVALUATION : LES CARTES CONCEPTUELLES ET LA BASE DE CONNAISSANCES SUR L'EXACTITUDE DE L'AUTO-ÉVALUATION

---

L'objectif de la précédente étude était d'apporter une première réponse à la question relative aux conditions qui permettent d'améliorer l'exactitude de l'auto-évaluation d'un apprenant. Pour cela, nous avons testé l'efficacité de plusieurs outils d'aides à l'auto-évaluation cumulés (incitations à la justification des réponses, à la prise de conscience de l'efficacité des stratégies cognitives utilisées et des stratégies pertinentes et demande d'auto-explications), sur l'amélioration de l'exactitude de l'auto-évaluation.

Les résultats n'ont pas mis en évidence les effets positifs attendus de l'outil d'aide à l'auto-évaluation puisque l'ensemble des participants (avec ou sans dispositif d'aide à l'auto-évaluation) se sont évalués avec moins d'exactitude après le dispositif qu'avant. Ce résultat peut s'expliquer du fait que les participants de l'expérience précédente avaient un niveau de connaissance préalable relativement élevé. En référence à la théorie de l'expertise, il est possible qu'un dispositif d'aide à l'auto-évaluation ne soit pas adapté à un tel niveau de connaissance, tout comme certains outils d'aide à l'apprentissage n'aident pas les apprenants plus expérimentés à améliorer leurs performances d'apprentissage (Kalyuga et al., 2003). Néanmoins, cet outil aurait pu être bénéfique à des apprenants avec un niveau moins élevé de connaissances antérieures. Pour cette raison, nous allons comparer des apprenants novices à des apprenants plus expérimentés dans la seconde étude.

Par ailleurs, dans la première étude le choix des outils s'est basé sur leur efficacité, démontrée au préalable, sur l'amélioration de l'exactitude de l'auto-évaluation et/ou les performances d'apprentissage (e.g. Chi et al., 1989 ; Griffin et al., 2009 ; X. Lin & Lehman, 1999 ; Nelson & Dunlosky, 1991). Cependant, les effets bénéfiques de ces outils n'ont pas été répliqués dans la précédente étude. Nous nous orientons par conséquent vers les cartes conceptuelles, un autre outil plus récemment testé comme

outil d'aide à l'auto-évaluation. Peu d'études ont démontré l'efficacité des cartes conceptuelles à améliorer l'exactitude de l'auto-évaluation (Thiede et al., 2010 ; Redford et al., 2012). C'est pourquoi celles-ci méritent d'être étudiées plus précisément notamment selon le mode et le moment de leur présentation ainsi que selon le niveau de connaissances antérieures (i.e. le niveau d'expertise). En effet, la littérature sur l'apprentissage présente des résultats différents selon ces variables et nous conduisent à penser que des effets différentiels pourraient apparaître sur l'auto-évaluation également.

Enfin, les résultats de notre première étude ont également révélé que les apprenants s'auto-évaluaient avec plus d'exactitude sur des connaissances déclaratives que sur des connaissances procédurales. Par conséquent, la deuxième étude portera plus spécifiquement sur l'auto-évaluation de connaissances procédurales et la tâche consistera à mobiliser principalement des connaissances procédurales afin de réaliser un exercice portant sur les fonctionnalités avancées du traitement de texte. Aussi, afin de s'assurer que tous les participants, y compris les novices, disposent d'une base de connaissances suffisante pour réaliser cet exercice, tous les participants bénéficieront d'une phase d'apprentissage préalable.

La présente étude se propose donc non seulement d'évaluer l'efficacité d'un autre outil susceptible d'améliorer l'exactitude de l'auto-évaluation : la carte conceptuelle ; mais également d'approfondir la question des conditions de l'efficacité de cet outil. Ainsi, nous évaluerons l'effet du mode de présentation des cartes conceptuelles sur l'exactitude de l'auto-évaluation selon le moment de leur présentation mais également selon le niveau de connaissances antérieures des apprenants. Nous présenterons tout d'abord comment les cartes conceptuelles améliorent l'apprentissage et l'exactitude de l'auto-évaluation. Ensuite, nous rappellerons quel est l'effet des connaissances antérieures sur l'exactitude de l'auto-évaluation. Enfin, nous examinerons si l'effet du mode de présentation des cartes sur l'exactitude de l'auto-évaluation varie en fonction du moment de leur présentation et du niveau de connaissances antérieures des apprenants.

## 5.1 LES CARTES CONCEPTUELLES : UN OUTIL D'AIDE À L'APPRENTISSAGE ET À L'AUTO-ÉVALUATION

L'efficacité des cartes conceptuelles sur l'apprentissage a été mis en évidence à de nombreuses reprises sur des tâches d'apprentissage linéaire classique (e.g. Amer, 1994 ; BouJaoude & Attieh, 2008 ; Novak & Cañas, 2008 ; O'donnell et al., 2002), sur des tâches de mémorisation et d'apprentissage non linéaires (e.g. Amadiou & Salmeron, 2014 ; Vörös, Rouet, & Pléh, 2011) ; en lien avec l'acquisition de connaissances issues de différents domaines tels que la physique (Gurlitt & Renkl, 2008) ou la biologie (Adesope & Nesbit, 2013) ; dans des environnements informatisés (Chang

et al., 2001 ; Tsai, Lin, & Yuan, 2001) ; en situation de classe (Vanides, Yue, Tomita, & Ruiz-Primo, 2005) ; sur des apprenants de niveaux scolaires variés (Nesbit & Adesope, 2006).

Plus spécifiquement, les cartes conceptuelles (cf. section 3.1.2), améliorent et facilitent l'apprentissage profond (Novak, 2010) pour plusieurs raisons potentielles. En schématisant les informations essentielles d'un contenu à apprendre, elles permettraient à l'apprenant de se construire une représentation mentale plus précise (Vörös et al., 2011), de donner une structure visible et ainsi servir de modèle pour organiser et structurer les informations telles qu'elles sont susceptibles de l'être en mémoire (Novak & Cañas, 2008) et d'optimiser l'utilisation des ressources mentales (Clark & Paivio, 1991).

D'autre part, rappelons que pour auto-évaluer ses performances futures avec exactitude, l'apprenant doit utiliser des indices qui soient prédictifs de la performance (Koriat, 1997). Étant donné que l'objectif des tâches d'apprentissage est que des connaissances nouvelles soient intégrées aux anciennes, les performances mesurent fréquemment la qualité du modèle de situation dans le cas d'une tâche de compréhension ou l'apprentissage profond dans le cas de tâches plus générales (ex : résolution de problème). Le but de l'apprenant est donc d'accéder au modèle de situation ou à un apprentissage profond, afin non seulement d'améliorer la qualité de son apprentissage, mais également de pouvoir en extraire des indices pertinents pour s'auto-évaluer avec exactitude. Pour aider les apprenants dans cette tâche difficile, nous avons vu qu'il existait plusieurs méthodes et outils (cf. Chapitre 3). Parmi ces méthodes, certaines visent à aider l'apprenant à accéder au modèle de situation ou à un apprentissage profond afin de rendre les indices pertinents plus 'saillants' pour l'apprenant. Les cartes conceptuelles, en facilitant l'apprentissage en profondeur, devraient également aider l'apprenant à se baser sur des indices pertinents pour s'auto-évaluer avec plus d'exactitude (Thiede et al., 2010 ; Redford et al., 2012).

Les premières études qui ont testé l'effet des cartes conceptuelles sur l'exactitude de l'auto-évaluation tendent à confirmer cette hypothèse. Par exemple, Thiede et al. (2010) ont réalisé une étude auprès d'étudiants universitaires nécessitant une remédiation en lecture. Lors de cette étude, les auteurs ont comparé l'exactitude de l'auto-évaluation lorsque les étudiants devaient construire une carte simultanément à la lecture d'un texte et lorsqu'ils n'avaient pas de cartes à construire. Les résultats ont mis en évidence que l'auto-évaluation était plus exacte lorsque les participants étaient aidés d'une carte que lorsqu'ils ne l'étaient pas. Des analyses corrélationnelles entre la qualité des cartes et les performances de compréhension tendent à montrer que la qualité des cartes corrélait directement avec les performances au test : plus le nombre de connexions entre les concepts construits



correctement dans la carte était important, plus les performances au test étaient élevées. De même, la qualité des cartes corrèle positivement avec le jugement d'auto-évaluation des apprenants. Les auteurs concluent que les cartes conceptuelles augmentent la saillance de la qualité du modèle de situation, ce qui fournit une base prédictive de la performance.

De façon similaire, Redford et al. (2012) ont testé l'efficacité de la construction de cartes conceptuelles simultanément à une tâche de lecture sur l'exactitude de l'auto-évaluation comparativement à une condition de simple relecture. Plus précisément, les apprenants devaient soit lire trois textes une seule fois en construisant des cartes simultanément à la lecture de chaque texte, soit lire les textes puis les relire une seconde fois sans construire de carte pendant les lectures. Les élèves prédisaient ensuite les performances qu'ils pensaient obtenir à un questionnaire à choix multiples composé de cinq questions qui évaluaient la mémorisation de faits contenus dans le texte mais également la compréhension profonde (nécessitant des inférences). Dans une seconde étude, les apprenants qui construisaient une carte étaient comparés à des apprenants d'un groupe contrôle qui ne lisait le texte qu'une fois. Dans les deux études, les apprenants qui avaient réalisé une carte pouvaient l'utiliser pour la phase de jugement et de performance. Les résultats de la première étude, bien que marginalement significatifs, ont révélé que l'auto-évaluation était plus exacte pour les participants qui avaient construit une carte que pour ceux qui avaient relu le texte. Les résultats de la seconde étude ont montré que les participants aidés d'une carte conceptuelle s'auto-évaluaient significativement avec plus d'exactitude que ceux qui n'avaient pas eu l'aide de cartes conceptuelles.

Van Loon, de Bruin, van Gog, Van Merriënboer, et Dunlosky (2014) ont étudié les effets de la complétion de diagrammes (activité similaire à la construction d'une carte conceptuelle) sur l'exactitude de l'auto-évaluation. La complétion de diagrammes, tout comme la construction de cartes conceptuelles, est associée à une amélioration de la compréhension en promouvant l'apprentissage profond et en permettant à l'apprenant de se construire un modèle mental plus riche (Gobert & Clement, 1999). Dans leur étude, les auteurs ont présenté le diagramme après une tâche de lecture de textes. Les apprenants devaient lire des textes puis compléter ou non des diagrammes (selon le groupe expérimental) dont la structure (cases et flèches) était donnée et une des cinq cases était complétée. Les apprenants devaient remplir les cases vides et ne pouvaient pas revenir au texte associé. Après avoir complété ou non les diagrammes, les participants évaluaient le nombre de questions auquel ils pensaient pouvoir répondre correctement sur une échelle allant de 0 à 100%. Les résultats de cette étude ont mis en évidence que les apprenants qui devaient compléter un diagramme s'auto-évaluaient avec plus d'exactitude que les apprenants qui n'avaient pas eu à en compléter.

Si l'utilisation des cartes conceptuelles s'est avérée efficace pour améliorer les performances d'apprentissage et l'exactitude de l'auto-évaluation, un débat concernant leur mode d'utilisation (construction vs. consultation) subsiste.

## 5.2 FAUT-IL CONSULTER OU CONSTRUIRE LES CARTES CONCEPTUELLES ?

Ce débat a été souligné à de nombreuses reprises (Gurlitt & Renkl, 2008 ; Stull & Mayer, 2007 ; Tsai et al., 2001) et comme le résume Stull et Mayer (2007), l'enjeu est de savoir si la consultation est suffisamment engageante et la construction pas trop coûteuse pour être chacune respectivement efficaces.

La consultation, en effet, procure l'avantage d'aider l'apprenant à centrer son attention sur les relations entre les concepts plutôt que sur des faits non connectés et à comprendre comment il est possible de structurer l'information de façon à transformer un texte linéaire en une structure non linéaire. De cette façon, les apprenants peuvent se construire une représentation correcte et cohérente du contenu à apprendre et ne gaspillent pas de ressources dans des traitements autres que ceux nécessaires à l'intégration du contenu (Stull & Mayer, 2007). En revanche la simplicité apparente de la tâche de consultation peut faire courir le risque que l'apprenant n'investisse pas suffisamment d'effort mental (Salomon, 1984).

Par ailleurs, la tâche de construction s'est avérée pertinente (et nécessaire selon Larkin & Simon, 1987) car elle offre une marge de liberté plus importante pour se créer sa propre structure de connaissance (Gurlitt & Renkl, 2008). Elle permet également à l'apprenant de s'engager dans des processus cognitifs utiles à l'intégration des connaissances (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Les cartes rassemblent les informations concernant un même élément, évitant ainsi des processus de recherche et de mises en correspondance de l'information (Larkin & Simon, 1987). Néanmoins, la construction même de la carte nécessite des ressources cognitives et peut parfois créer une surcharge cognitive (Hilbert & Renkl, 2008) au détriment de l'intégration des connaissances (Mayer & Moreno, 2003). L'hypothèse selon laquelle la nécessité d'utiliser des ressources pour la construction des cartes perturberait l'apprentissage et l'auto-évaluation se justifie notamment avec la théorie de l'allocation des ressources (Kanfer & Ackerman, 1989) pour l'auto-évaluation. Kanfer et Ackerman (1989) considèrent que les ressources utilisées par la gestion des processus métacognitifs peut entrer en conflit avec la gestion des processus liés à la réalisation de la tâche. La théorie de la charge cognitive (Sweller, 1988) a également mis en évidence que l'utilisation de plusieurs

sources d'informations pouvait créer une charge superflue délétère pour l'apprentissage profond.

Afin de trancher cette question, [Stull et Mayer \(2007\)](#) ont comparé l'efficacité des conditions de construction et de consultation de cartes conceptuelles sur l'apprentissage. Dans leur étude, les auteurs ont montré l'efficacité de la consultation de cartes conceptuelles pour l'apprentissage profond comparativement à la construction de cartes. L'absence de groupe contrôle dans cette étude ne permet pas de conclure sur les apports de la construction de cartes dans l'amélioration de l'apprentissage.

[Redford et al. \(2012\)](#) ont testé l'efficacité de la construction de cartes conceptuelles comparativement à une condition de consultation sur l'exactitude de l'auto-évaluation. Les apprenants devaient lire trois textes et simultanément soit construire des cartes conceptuelles, soit consulter des cartes conceptuelles. Un groupe contrôle ne disposait d'aucune carte. Leurs résultats ont révélé que la consultation de cartes était moins efficace que la construction de cartes et que par conséquent il était nécessaire de générer de l'information issue du contenu à apprendre plutôt que de simplement la visualiser pour améliorer l'exactitude de l'auto-évaluation. Néanmoins, les auteurs rapportent dans leur étude que les performances n'ont pas été affectées et bien que la construction de cartes conceptuelles ait réduit l'illusion de savoir, l'auto-évaluation n'était pas aussi exacte que ce qui avait été précédemment rapporté dans la littérature (e.g. [Dunlosky & Lipko, 2007](#)).

Les effets opposés de la construction et de la consultation de cartes sur l'apprentissage et l'exactitude de l'auto-évaluation constatés dans les études de ([Stull & Mayer, 2007](#)) et de ([Redford et al., 2012](#)) montrent que cette question mérite toujours d'être approfondie. En effet, pour la réalisation de ces deux activités, l'objectif est de faciliter la construction et l'accès au modèle de situation. Par conséquent, le même mode d'utilisation des cartes devrait être efficace dans les deux cas.

Par ailleurs, de nombreuses études ([Gurlitt & Renkl, 2008](#) ; [Lambiotte & Dansereau, 1992](#) ; [Nesbit & Adesope, 2006](#)) ont également fait état que l'efficacité du mode d'utilisation des cartes (construction versus consultation) sur l'apprentissage dépendait fortement du niveau d'expertise des apprenants dans un domaine tandis qu'aucune n'aborde les effets sur l'exactitude de l'auto-évaluation. Avant d'aborder l'effet différentiel du mode d'utilisation des cartes selon le niveau de connaissances antérieures des apprenants, nous allons présenter les effets de la base de connaissance sur l'exactitude de l'auto-évaluation.

### 5.3 BASE DE CONNAISSANCE : LES EFFETS SUR L'EXACTITUDE DE L'AUTO-ÉVALUATION

Rappelons (cf. section 2.6 du chapitre 2) que la plupart des études sur la question montrent que l'expertise affecte positivement l'exactitude de l'auto-évaluation. En effet, bien qu'au moins une étude ait montré un effet néfaste de l'expertise sur l'exactitude de l'auto-évaluation (Glenberg & Epstein, 1987), la majorité des travaux a mis en évidence que le niveau de connaissances antérieures conduit à une plus grande exactitude (i.e. les experts sont plus exacts que les novices Griffin et al., 2009 ; Kostons et al., 2010 ; Maki & Serra, 1992 ; Schneider, 1996).

Dans les études qui ont mis en évidence un effet facilitateur des connaissances antérieures sur l'exactitude de l'auto-évaluation, les auteurs ont proposé plusieurs pistes explicatives. Par exemple, Nietfeld et Schraw (2002) rappellent que les connaissances antérieures peuvent servir de base pour évaluer les performances (Pressley, 1995) mais également que les apprenants avec un niveau suffisamment élevé de connaissances antérieures peuvent disposer de davantage de ressources pour évaluer leurs performances (Allen & Casbergue, 1997). Cette dernière hypothèse est renforcée par le modèle de l'allocation des ressources (Kanfer & Ackerman, 1989) qui avance non seulement que les activités d'autorégulation nécessitent des ressources, mais également que l'apprentissage et la pratique répétée d'une activité transforme des connaissances déclaratives en connaissances procédurales pour lesquelles le besoin attentionnel est moindre. De plus, ces hypothèses sont cohérentes avec les travaux de Griffin et al. (2008) qui supposent à l'issue de leur étude que la possibilité pour un apprenant de s'engager dans un processus d'auto-évaluation parallèlement à l'apprentissage dépend des ressources cognitives disponibles, et de Van Gog et al. (2011) qui mettent en évidence que le processus d'auto-évaluation est un processus qui nécessite des ressources principalement lorsque la tâche est complexe. Aussi, les experts ayant une base de connaissances riche et bien structurée dans leur domaine de connaissance ont plus de facilité à résoudre des problèmes (Chi, Glaser, & Rees, 1981) traiter des textes complexes linéaires (McNamara et al., 1996) ou non linéaires (Amadiou et al., 2010) et pourraient par conséquent disposer de davantage de ressources que les novices pour s'engager dans des processus d'auto-évaluation.

Ainsi, si les experts disposent de davantage de ressources pour apprendre et s'auto-évaluer, la construction de cartes conceptuelles, a priori plus coûteuse, aurait-elle davantage d'effets bénéfiques sur l'amélioration de leur apprentissage et l'exactitude de leur auto-évaluation que sur celle des novices ? Au contraire, les novices qui utiliseraient davantage de ressources pour l'intégration des connaissances et leur évaluation seraient-ils favorisés par une condition de consultation ?

#### 5.4 L'EFFICACITÉ DES MODES D'UTILISATION DES CARTES SELON LA BASE DE CONNAISSANCE DES APPRENANTS

À notre connaissance, il n'existe pas d'étude qui ait traité des effets des cartes conceptuelles sur l'exactitude de l'auto-évaluation selon le niveau d'expertise des apprenants. Les études (e.g. [Lambiotte & Dansereau, 1992](#) ; [Gurlitt & Renkl, 2008](#)) traitant de l'effet des cartes conceptuelles sur l'apprentissage selon le niveau de connaissance tendent à indiquer un effet différentiel selon le mode d'utilisation des cartes. En effet, les cartes faciliteraient davantage l'apprentissage des individus avec le moins de connaissances antérieures lorsqu'elles sont à consulter tandis qu'elles seraient plus utiles aux individus avec le plus de connaissances antérieures lorsqu'elles sont à construire. Ainsi, [Lambiotte et Dansereau \(1992\)](#) ont comparé trois techniques d'aide à l'apprentissage (la consultation de cartes conceptuelles, de résumé ou de mots-clés) en association à une situation d'apprentissage sur cassette audio sur le rappel chez des étudiants en biologie de différents niveaux d'expertise. Les résultats indiquent que les apprenants avec un faible niveau de connaissances initiales ont obtenu de meilleures performances d'apprentissage lorsqu'ils ont été assistés d'une carte conceptuelle tandis que les experts ont davantage bénéficié des mots-clés. [Nesbit et Adesope \(2006\)](#) rappellent que ce résultat peut s'expliquer en référence à la théorie de l'assimilation de [Mayer \(1979\)](#) du fait que la macrostructure proposée par la carte guiderait la construction des connaissances des novices, tandis que la structure proposée par la carte entrerait en conflit avec la structure de connaissances pré-établie des experts.

Dans une autre étude, [Gurlitt et Renkl \(2008\)](#) comparent l'efficacité des cartes à construire versus à consulter avant l'apprentissage pour faciliter l'activation des connaissances antérieures chez des lycéens peu expérimentés en physique à des étudiants en physique (plus expérimentés). Les participants devaient ensuite apprendre un contenu à partir d'un hypertexte puis répondre à des questions ouvertes. Leurs résultats montrent que les lycéens tirent davantage de bénéfices d'une carte où les liens sont fournis tandis que les étudiants en physique tirent davantage de bénéfices dans la condition où les liens sont à créer et à nommer. Pour expliquer le fait que la carte à consulter ne convienne pas aux plus expérimentés, [Gurlitt et Renkl \(2008\)](#) avancent que celle-ci les empêche non seulement d'exprimer leur propre modèle conceptuel mais également que la simplicité apparente de la tâche n'incite pas les apprenants experts à investir leurs ressources cognitives dans le traitement de la carte. Aussi, il semblerait ici que la carte à consulter désengage les apprenants avec un niveau de connaissances antérieures élevé qui n'utiliseraient alors pas l'intégralité des ressources cognitives dont ils disposent. La consultation serait donc une condition *trop*

peu coûteuse pour les apprenants avec un niveau élevé de connaissances antérieures. Cependant, cette explication entre en contradiction avec la théorie de la charge cognitive et plus spécifiquement l'effet de l'expertise inversée (Kalyuga et al., 2003) qui postule que des aides adaptées aux novices sont délétères pour les experts car elles sont redondantes pour un individu expert qui possède déjà ces informations et l'oblige à réaliser un traitement non nécessaire. Ce traitement superflu ferait gaspiller aux experts des ressources qu'ils pourraient autrement utiliser pour traiter des informations qu'ils ne possèdent pas.

Ainsi, il semblerait que la consultation de cartes conceptuelles perturbe l'apprentissage des experts mais si d'après les hypothèses de Gurlitt et Renkl (2008) cette perturbation est liée à une sous-utilisation des ressources, la théorie de la charge cognitive (Kalyuga et al., 2003) soutient que ce problème est lié à une surcharge cognitive.

Dans la mesure où l'exactitude de l'auto-évaluation dépend de la profondeur de traitement, si les experts ne disposent pas de suffisamment de ressources pour traiter correctement l'information dans une condition de consultation, ils ne devraient pas avoir les ressources nécessaires pour s'auto-évaluer avec exactitude.

Ainsi, nous nous attendons à ce que les performances et l'exactitude de l'auto-évaluation des apprenants avec un niveau élevé de connaissances soient plus élevées en condition de construction plutôt qu'en condition de consultation des cartes conceptuelles. Nous mesurerons la charge cognitive ressentie par les apprenants afin d'évaluer si les différences d'efficacité des cartes en condition de construction et de consultation sont liées à une charge cognitive trop faible ou trop élevée.

Chez les novices, les travaux sur l'apprentissage précédemment cités tendent à montrer qu'une condition de consultation leur serait plus bénéfique qu'une condition de construction pour l'apprentissage et l'auto-évaluation. En effet, la consultation guide davantage la construction des connaissances tandis que la construction tend à créer une surcharge chez les apprenants qui ne possèdent pas un haut niveau de connaissances antérieures (Gurlitt & Renkl, 2008). De plus, comme le rappellent Huet et Mariné (2009), lorsqu'une tâche est particulièrement coûteuse, les novices ne parviennent pas à s'auto-évaluer avec exactitude contrairement aux experts. Ainsi, la surcharge créée par la construction de la carte perturberait le processus d'auto-évaluation.

En référence aux travaux de Thiede et al. (2010) et Redford et al. (2012) qui ont montré que la construction de cartes conceptuelles pouvait être efficace chez des étudiants qui n'ont a priori pas un niveau de connaissances antérieures suffisamment élevé, nous pouvons nous attendre à ce que la condition de construction soit plus efficace que la condition de consultation sur l'exactitude de l'auto-évaluation. Selon les travaux sur l'apprentissage

(Gurlitt & Renkl, 2008 ; Lambiotte & Dansereau, 1992) indiquant que la construction peut s'avérer plus coûteuse que la consultation, nous pouvons, au contraire, nous attendre à ce que la condition de consultation soit plus efficace que la condition de construction sur l'exactitude de l'auto-évaluation. En effet, les travaux sur l'auto-évaluation indiquent que les apprenants novices en surcharge s'auto-évaluent avec moins d'exactitude que les apprenants qui ne sont pas en surcharge (Huet & Mariné, 2009). Une mesure de la charge cognitive ressentie nous permettra de vérifier quel mode d'utilisation des cartes est le moins coûteux.

Nous venons de voir que le mode d'utilisation des cartes semble interagir avec le niveau d'expertise des apprenants. Le manque d'efficacité des cartes serait lié à une difficulté de l'apprenant à gérer l'utilisation de ses ressources cognitives. Il semblerait donc pertinent de trouver des solutions qui permettraient à l'apprenant de mieux gérer ses ressources. L'une de ces solutions consisterait à présenter les cartes conceptuelles après la présentation du contenu à apprendre.

#### 5.5 DIFFÉRER LA PRÉSENTATION DES CARTES CONCEPTUELLES POUR EN AMÉLIORER L'EFFICACITÉ

Dans les travaux de Thiede et al. (2010) et Redford et al. (2012) présentés précédemment, les apprenants utilisaient les cartes simultanément à la lecture de textes. De même, Stull et Mayer (2007) ont montré un effet positif de la consultation simultanée des cartes conceptuelles sur l'apprentissage. Néanmoins, Redford et al. (2012) constatent que l'efficacité de la construction de carte n'a pas eu l'ampleur attendue puisque l'amélioration de l'exactitude de l'auto-évaluation n'était pas aussi importante que ce que d'autres auteurs avaient pu mettre en évidence à l'aide d'autres méthodes (e.g. Maki, 1998 ; Thiede, Griffin, Wiley, & Redford, 2009), et que la consultation n'avait pas d'effets significatifs.

L'efficacité supérieure de l'utilisation simultanée des cartes conceptuelles pour l'apprentissage et l'auto-évaluation peut se justifier par les principes de modalité (Mayer, 2001 ; Sweller et al., 1998) et de contiguïté (Mayer, 2001). Selon le principe de modalité, il est plus efficace pour l'apprentissage de présenter de l'information sous deux modalités différentes (visuelle et auditive) que sous une seule. La condition de double modalité serait moins coûteuse que la condition où les informations sont présentées sous une même modalité (Tabbers, Martens, & Van Merriënboer, 2001 ; Tabbers, Martens, & Merriënboer, 2004). Selon le principe de contiguïté, il est préférable que l'apprenant utilise une information visuelle simultanément à une information auditive plutôt qu'en différé (Mayer & Sims, 1994). Ces effets ont été démontrés sur des tâches de compréhension de matériels simples (Moreno & Mayer, 1999) et complexes (voir Rouet,

2005). Dans son article, Rouet (2005) rapporte les résultats d'une étude dans laquelle des apprenants devaient apprendre un contenu accompagné de commentaires écrits ou audios. L'apprenant pouvait ou non gérer la vitesse de présentation des informations. Les résultats montrent une supériorité des performances de rappel dans la condition de double modalité que dans la condition de simple modalité. Cette supériorité est effective lorsque la vitesse de présentation de l'information est imposée mais ne l'est plus si l'apprenant gère lui-même la vitesse.

En revanche, l'effet de modalité n'a pas toujours été répliqué. Tabbers et al. (2004) ont ainsi montré, dans une situation écologique d'apprentissage en classe, que des apprenants qui apprenaient un modèle psychologique sur la base d'un diagramme accompagné d'un texte visuel obtenaient de meilleures performances de rappel et de transfert que les apprenants qui utilisaient un diagramme accompagné d'un texte audio. En lien avec l'exactitude de l'auto-évaluation, Van Loon et al. (2014) montre les effets bénéfiques de la complétion différée d'un diagramme (i.e. complétion après la lecture de textes) sur l'exactitude de l'auto-évaluation. D'autres études ont également mis en évidence les effets positifs sur l'exactitude de l'auto-évaluation des stratégies génératives telle que l'élaboration de mots-clés ou de résumés réalisées après l'apprentissage (e.g. M. Anderson & Thiede, 2008 ; Thiede et al., 2005).

En référence aux travaux de Thiede et al. (2010), de Redford et al. (2012) et des effets de modalité et de contiguïté de la théorie de l'apprentissage multimédia, la présentation des cartes simultanément à un contenu à apprendre serait plus efficace pour l'apprentissage et l'exactitude de l'auto-évaluation que la présentation différée. En revanche, au vu des travaux de Van Loon et al. (2014) et Tabbers et al. (2004), l'utilisation différée des cartes pourrait s'avérer autant voire plus efficace que l'utilisation simultanée. Cette question mérite donc d'être explorée, et la mesure de la charge mentale ressentie nous permettra, à nouveau, de distinguer la condition la plus coûteuse.

## 5.6 HYPOTHÈSES

D'après les premiers travaux sur l'effet des cartes conceptuelles sur l'exactitude de l'auto-évaluation (Redford et al., 2012 ; Thiede et al., 2010) et l'apprentissage (e.g. Nesbit & Adesope, 2006), nous supposons que les apprenants qui bénéficient d'une carte conceptuelle seront plus exacts dans l'évaluation qu'ils feront de leurs performances et auront de meilleures performances d'apprentissage que les participants qui ne bénéficient pas de cartes conceptuelles (H1).

Par ailleurs, les travaux sur l'expertise nous permettent de prédire que les experts auront, de façon générale, de meilleures performances d'ap-



prentissage et seront plus exacts dans l'évaluation qu'ils feront de leurs connaissances que les novices (H2).

D'après la théorie de l'expertise inversée (Kalyuga et al., 2003), les plus expérimentés ressentiront davantage de charge mentale en condition de consultation qu'en condition de construction (H3a). Au contraire, d'après Gurlitt et Renkl (2008), les plus expérimentés ressentiront moins de charge mentale en condition de consultation qu'en condition de construction (H3b). Dans les deux cas, les apprenants les plus expérimentés devraient obtenir de meilleures performances d'apprentissage et s'auto-évaluer avec plus d'exactitude lorsqu'ils construisent les cartes que lorsqu'ils les consultent (H3c).

Chez les novices les hypothèses sont moins claires. D'après les travaux de Thiede et al. (2010) et de Redford et al. (2012), les novices devraient s'auto-évaluer avec plus d'exactitude lorsqu'ils construisent les cartes que lorsqu'ils les consultent (H4a). Au contraire, d'après Gurlitt et Renkl (2008) ; Lambiotte et Dansereau (1992) la construction nécessite davantage de ressources que la consultation. D'après Kanfer et Ackerman (1989), une charge mentale trop importante nuit à l'apprentissage et à l'auto-évaluation des novices. Ainsi, les novices, devraient avoir des performances d'apprentissage moins élevées et devraient ressentir plus de charge mentale et, par conséquent, s'auto-évaluer avec moins d'exactitude lorsqu'ils construisent des cartes conceptuelles que lorsqu'ils les consultent (H4b).

D'après les résultats de Redford et al. (2012) ; Stull et Mayer (2007) ; Thiede et al. (2010), et selon le principe de contiguïté et de modalité (Mayer & Moreno, 2002) la présentation simultanée des cartes serait moins coûteuse que la présentation différée et devrait être plus efficace pour améliorer l'exactitude de l'auto-évaluation et les performances d'apprentissage (H5). Cependant, d'après les résultats de l'étude de Van Loon et al. (2014) la présentation différée des cartes conceptuelles pourrait être plus efficace que la présentation simultanée des cartes conceptuelles. Nous explorerons quel moment de présentation des cartes est plus efficace sur l'exactitude de l'auto-évaluation et la performance.

Finalement, toujours dans l'objectif de déterminer la condition d'utilisation des cartes la plus efficace pour l'apprentissage et l'auto-évaluation, nous examinerons, de façon globale, les effets de la consultation et de la construction ainsi que la qualité des cartes conceptuelles construites.

## 5.7 MÉTHODE

### 5.7.1 *Participants et Plan d'expérience*

Parmi les 272 participants recrutés, 136 participants étaient étudiants en première année de licence de GEA 'Gestion des Entreprises et des Administrations' de l'Université Toulouse III (82 filles et 51 garçons ; âge moyen,  $M = 18.5, SD = 1.2$  ; les informations démographiques manquaient pour trois étudiants en GEA), et 136 étaient élèves de seconde dans un lycée de la région toulousaine (41 filles et 95 garçons ; âge moyen,  $M = 15.5, SD = .6$ ).

Avant de participer à cette étude, les étudiants en GEA ont suivi, dans le cadre de leur formation, quatre heures de travaux pratiques sur l'utilisation spécifique des styles de mise en forme à l'aide d'un logiciel de traitement de texte. Cet enseignement délivré par nos collaborateurs de l'Institut de Recherche en Informatique de Toulouse leur a permis d'acquérir des connaissances antérieures sur le domaine ciblé par la phase d'apprentissage de cette étude ; ils constituaient donc notre groupe dit 'expérimenté'. Au contraire, les lycéens n'ont reçu aucune formation sur l'utilisation des styles dans un logiciel de traitement de texte, ils constituaient notre groupe dit 'novice'.

Les participants ont été répartis aléatoirement dans cinq groupes expérimentaux (cf. Figure 8). Les participants du premier groupe avaient pour consigne de suivre le contenu d'apprentissage et de construire simultanément une carte conceptuelle (construction simultanée,  $N = 49$ ) ; les participants du deuxième groupe devaient suivre l'apprentissage puis construire une carte conceptuelle (construction différée,  $N = 45$ ) ; les participants du troisième groupe apprenaient et consultaient simultanément une carte conceptuelle construite préalablement par un enseignant (consultation simultanée,  $N = 66$ ) ; les participants du quatrième groupe apprenaient puis consultaient a posteriori la carte conceptuelle (consultation différée,  $N = 63$ ). Enfin, les participants du cinquième groupe apprenaient à l'aide du support d'apprentissage proposé sans avoir de cartes conceptuelles à construire ou à consulter (groupe contrôle,  $N = 49$ ).

### 5.7.2 *Matériel et tâche*

L'ensemble du matériel d'apprentissage, des mesures de l'auto-évaluation et de l'exercice d'application ont été élaborés en étroite collaboration avec un maître de conférences en informatique et chercheur en psychologie cognitive, qui s'est chargé de l'intégralité de la mise en place informatique. Le dispositif informatique interconnecté à la plateforme pédagogique de notre université a été réalisé en Java JEE [1]. Outre la mise

à disposition des ressources, ce dispositif a aussi permis l'enregistrement, le stockage et le pré-traitement des traces.

### *Questionnaire de mesure de connaissances antérieures*

Un questionnaire à choix multiples (cf. Annexe), composé de onze questions extraites du Certificat Informatique et Internet (C2I niveau 1) et portant sur le traitement de texte, a été utilisé afin de mesurer le niveau de connaissances antérieures des participants sur ce domaine. Quatre à sept propositions étaient données pour chaque question, avec une seule réponse correcte.

### *Les vidéos d'apprentissage*

Trois vidéos successives, complémentaires, dont la durée variait entre 3 minutes et 30 secondes et 6 minutes et 30 secondes et ayant pour thème l'automatisation de la mise en forme de document grâce à l'utilisation des 'styles' constituait la séquence d'apprentissage. La première avait pour objectif de rappeler les notions élémentaires nécessaires à la compréhension des deux autres vidéos, qui présentaient quant à elles, comment modifier et appliquer des styles. Ces vidéos, réalisées par un enseignant du C2I, consistaient à montrer les procédures d'utilisation du traitement de texte associées à des explications (Sakdavong, Amadiou, Huet, & Dupeyrat, 2011).

Une consigne préalable spécifiait aux participants (1) que les vidéos étaient montées de sorte à ce qu'ils les voient toutes dans l'ordre défini sans pouvoir revenir en arrière, (2) qu'ils pouvaient mettre en pause les vidéos quand ils le souhaitaient, (3) qu'il leur était interdit de prendre des notes ou de consulter Internet, et (4) que la suite de l'expérimentation se composait d'un exercice d'application basé sur les vidéos. Ces consignes nous permettaient de contrôler que les participants aient tous accès à la même quantité d'information.

### *Les cartes conceptuelles*

Une carte conceptuelle à consulter ou à construire était proposée aux apprenants. La carte à consulter (présentée dans la Figure 5) a été réalisée par un enseignant du C2I (Maillard, Motak, Sakdavong, Dupeyrat, & Huet, 2013) et portait sur les concepts et leurs relations présentées dans les vidéos. Lorsque la carte était à construire, les participants disposaient seulement des concepts et des relations et devaient les placer et les relier entre eux dans un cadre réservé (cf. Figure 6). Le logiciel COGUI (<http://www.lirmm.fr/cogui/>) a été utilisé pour les cartes à consulter et à construire.

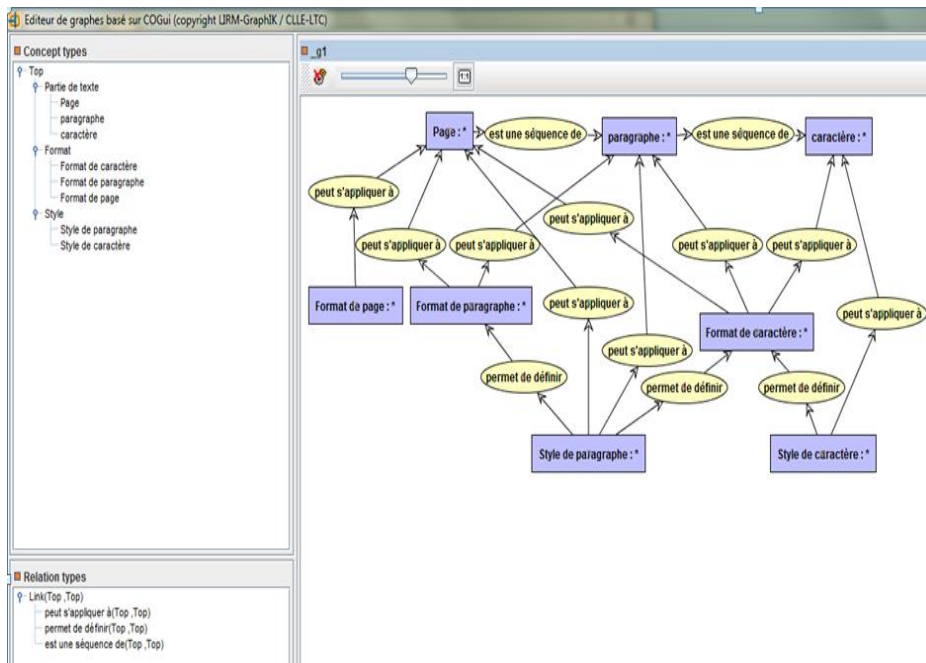


FIGURE 5. – Carte conceptuelle présentée aux apprenants devant consulter une carte parallèlement à l'apprentissage.

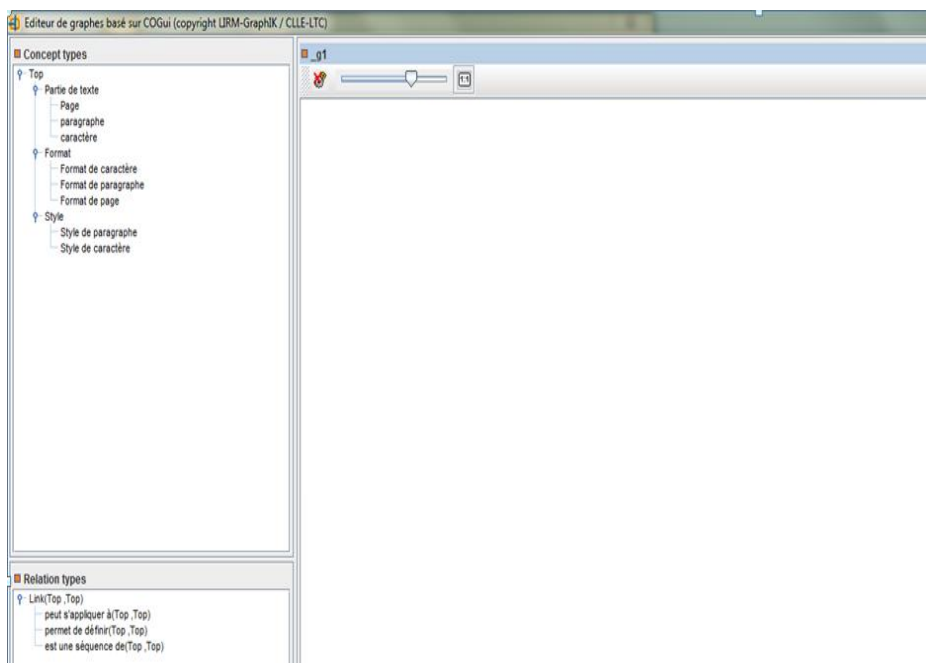


FIGURE 6. – Dispositif présenté aux apprenants devant construire une carte parallèlement à l'apprentissage.

### *La tâche : l'exercice d'application*

L'exercice d'application (Maillard et al., 2013 ; Sakdavong et al., 2011) visait à évaluer le niveau d'apprentissage des apprenants et consistait à mettre en forme un texte à l'aide des styles. L'exercice était présenté sur deux fenêtres distinctes, la première présentait la consigne et la seconde le texte à mettre en forme selon les conditions imposées par la consigne (cf. point 1 et 2 dans la Figure 7).

Pour réussir l'exercice, les apprenants devaient avoir compris la notion et le but des styles. Ils commençaient alors par modifier le style 'Titre 1' en sélectionnant l'onglet 'format' puis 'styles'. Il sélectionnait le type de section à modifier ('Titre 1'). Les apprenants devaient également avoir intégré les notions d'attribut de caractère et de paragraphe pour créer les styles comme demandé par la consigne. Concrètement, cela nécessitait que les participants cliquent sur 'modifier les attributs de caractère du style sélectionné' pour modifier la taille de la police, puis sur 'modifier les attributs de paragraphes du style sélectionné' pour modifier l'alignement et le retrait. Les participants devaient produire des opérations similaires pour modifier les sous-titres ('Titre 2') et les paragraphes ('normal') selon les consignes affichées. Suite à la modification des styles, les apprenants devaient sélectionner correctement tous les titres de niveau 1 pour y appliquer le style 'Titre 1', les titres de niveau 2 pour y appliquer le style 'Titre2' et les paragraphes pour y appliquer une mise en forme 'normal'. Les apprenants pouvaient également commencer par appliquer les styles pour les modifier ensuite.

L'interface de l'exercice d'application différait de celle présentée dans les vidéos pour ne pas que les participants puissent simplement répliquer de mémoire la procédure (cf. point 3 dans la Figure 7). En résumé, la réussite de l'exercice nécessitait que les participants aient correctement assimilé les notions et procédures présentées dans les vidéos. Afin d'éviter une surcharge associée à la prise en main de cette nouvelle interface, un exercice de familiarisation, sur un contenu différent, précédait l'exercice d'application.

### *5.7.3 Mesures*

#### *Les connaissances antérieures*

La performance au questionnaire de mesure des connaissances antérieures était notée sur 11 points, une bonne réponse comptant pour 1 point et une mauvaise réponse comptant pour 0. Le score sur 11 points a ensuite été transformé sur 100. Par exemple, un participant qui répondait à huit questions sur onze obtenait un score sur 100 égal à  $(8 * 100) / 11 = 72.73$ .

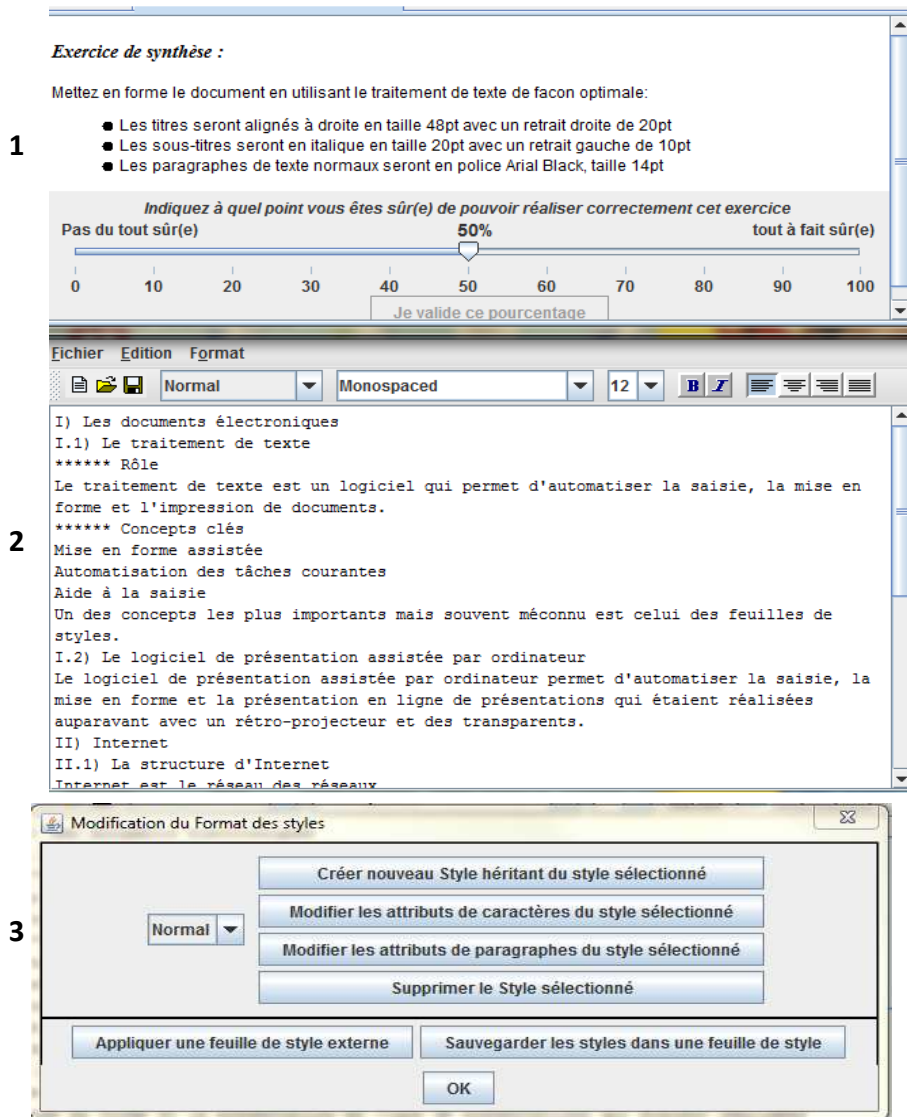


FIGURE 7. – Copie d'écran de l'exercice d'application et du menu auquel les participants avaient accès pour résoudre l'exercice.

### *Performance à l'exercice d'application*

La performance à l'exercice était évaluée sur quinze points, chaque point correspondant à une action à réaliser pour modifier et appliquer les styles selon la consigne et ainsi terminer correctement l'exercice. Cette note sur 15 était ensuite transformée en score sur 100 afin de pouvoir comparer ce score à celui du jugement d'auto-évaluation. Par exemple, un apprenant qui obtenait un score de 10, obtenait une note sur 100 égale à  $(10 * 100) / 15 = 66.66$ .

### *L'auto-évaluation et de son exactitude*

L'auto-évaluation et son exactitude ont été calculées à deux reprises. Les participants évaluaient une première fois leurs performances au questionnaire de mesure des connaissances antérieures, puis ils évaluaient leurs performances à l'exercice d'application.

**LE JUGEMENT D'AUTO-ÉVALUATION** Comme dans la précédente expérience, les participants évaluaient leurs performances à chacune des questions du questionnaire de mesure des connaissances antérieures. Pour cela, les participants répondaient, sur une échelle de 0 (pas du tout sûr) à 100 (tout à fait sûr), à la question 'Estimez à quel point vous pensez réussir à répondre à la question suivante sachant que des propositions à cocher vous seront fournies : [...]'. Le jugement par participant correspondait à la moyenne des onze jugements émis.

Pour l'exercice d'application, une seule mesure du jugement a été faite. Les apprenants prédisaient leur performance en répondant sur une échelle de 0 à 100 à la question 'Indiquez à quel point vous êtes sûr(e) de pouvoir réaliser correctement cet exercice [...]?' (0 = pas du tout sûr ; 100 = tout à fait sûr). L'énoncé de la question était présenté afin que le participant puisse émettre son jugement en toute connaissance de cause, en prenant le temps nécessaire.

**LA DIRECTION DU BIAIS D'AUTO-ÉVALUATION.** La direction du biais d'auto-évaluation sur les connaissances antérieures a été calculée comme dans l'étude précédente (pour un exemple, cf. Tableau 3), en réalisant pour chaque question, la différence entre le score (sur 100) du jugement d'auto-évaluation à la question et le score (sur 100) de performance à la question. La moyenne des onze scores de la direction du biais était ensuite calculée afin d'obtenir une unique valeur par participant.

La mesure de la direction des biais d'auto-évaluation à l'exercice d'application a été calculée en soustrayant le score de performance à l'exercice d'application sur 100 au score du jugement d'auto-évaluation à l'exercice d'application (Baars et al., 2013, 2014).

Ce calcul permettait de voir si les apprenants tendaient à se sous-évaluer (score inférieur à 0) ou à se sur-évaluer (score supérieur à 0).

**L'AMPLITUDE DU BIAIS D'AUTO-ÉVALUATION.** Comme pour le calcul de la direction du biais, l'amplitude du biais d'auto-évaluation associée au questionnaire de mesure des connaissances antérieures a été calculée pour chaque question en réalisant la racine carrée de l'écart entre le jugement d'auto-évaluation à une question et la performance obtenue à cette même question, élevée au carré. L'amplitude moyenne était ensuite calculée sur la base des onze scores obtenus afin d'obtenir une unique valeur par participant.

Concernant l'amplitude du biais à l'exercice d'application, ce score a été mesuré en calculant la racine carrée des écarts entre le jugement d'auto-évaluation à l'exercice d'application et la performance à l'exercice d'application élevée au carré.

Une note proche de 0 indiquait que l'auto-évaluation est exacte, une note proche de 100 reflétait un écart maximum entre la performance et le jugement d'auto-évaluation (Baars et al., 2013, 2014).

#### *Mesure de la qualité de la carte construite*

Une carte complète et correcte, telle que l'était la carte experte à consulter, était composée de quatorze liens. En condition de construction, le participant pouvait construire autant de liens qu'il le souhaitait. Le dispositif informatique calculait le nombre de liens justes construits par l'apprenant (i.e. liens présents dans la carte experte). La qualité de la carte construite a pu être calculée en effectuant le pourcentage du rapport entre le nombre de liens justes et le nombre total de liens justes qui pouvaient être construits.

Par exemple, l'apprenant effectuait dix relations entre des concepts. Parmi ces dix relations, seules quatre étaient justes. Alors, la carte de l'apprenant est correcte à  $(4/14) \times 100 = 28.6\%$ .

#### *Mesure de la charge mentale*

Dans notre étude, deux indicateurs de la charge cognitive ont été utilisés : l'effort mental et la difficulté ressentis (Amadiou, Van Gog, Paas, Tricot, & Mariné, 2009 ; Amadiou, Mariné, & Laimay, 2011 ; Paas, 1992)

**EFFORT MENTAL RESSENTI POUR COMPRENDRE LES VIDÉOS.** L'ensemble des participants devait répondre sur une échelle de 1 à 9 à la question suivante : 'l'effort mental produit pour comprendre les vidéos était : 1 (très très peu important) ; 9 (très très important)'.



DIFFICULTÉ RESENTIE POUR COMPRENDRE LES VIDÉOS. De la même façon, l'ensemble des participants devait répondre sur une échelle de 1 à 9 à la question : 'Indiquez à quel point il vous a été difficile de comprendre les vidéos : 1 (très très facile) ; 9 (très très difficile)'.

EFFORT MENTAL RESENTI POUR UTILISER LES CARTES. Les participants ayant utilisé des cartes conceptuelles devaient en plus évaluer : 'l'effort mental produit pour construire/consulter les cartes conceptuelles étaient : 1 (très très peu important) ; 9 (très très important)'

DIFFICULTÉ RESENTIE POUR UTILISER LES CARTES. Ces mêmes participants devaient également répondre à cette dernière question : 'Indiquez à quel point il vous a été difficile de construire/consulter les cartes conceptuelles : 1 (très très facile) ; 9 (très très difficile)'.

#### 5.7.4 Procédure

L'ensemble des passations ont été réalisées dans les salles de cours habituellement utilisées par les élèves et intégrées à leur cursus. Ainsi, les étudiants en GEA suivaient l'expérimentation à la place d'un de leurs Travaux Dirigés, dans leur salle de cours habituelle. Pour les lycéens, la passation avait lieu dans les salles informatiques du lycée pendant les heures d'accompagnement personnalisé. Chaque participant disposait d'un casque et d'un ordinateur afin de visualiser les vidéos et de réaliser l'ensemble de la passation de façon individuelle.

Les participants commençaient par lire et valider un formulaire de consentement. Ils répondaient dans un second temps au questionnaire à choix multiples mesurant leur niveau de connaissances antérieures. Ils visualisaient ensuite une vidéo qui présentait l'objectif et le déroulement de l'expérience. Cette vidéo avait aussi pour objectif de familiariser les participants aux cartes conceptuelles Elle comportait la définition d'une carte conceptuelle et son objectif, un exemple de carte conceptuelle et la façon dont on la lit, son intérêt pour l'apprentissage et l'auto-évaluation. Enfin, elle présentait une démonstration de l'utilisation de l'outil de conception de cartes conceptuelles. S'ensuivaient deux exercices d'entraînement qui consistaient à compléter puis re-construire une carte conceptuelle dont le thème était relatif à l'éducation civique. Un feedback expliquant quelques différences entre une carte conceptuelle correcte et incorrecte séparait les deux exercices. Suite à cette familiarisation aux cartes conceptuelles, les participants visionnaient les trois vidéos d'apprentissage accompagnées ou non de la carte conceptuelle (selon la condition expérimentale à laquelle ils étaient affectés). Les participants qui avaient utilisé une carte devaient ensuite estimer la charge cognitive ressentie (effort mental et difficulté

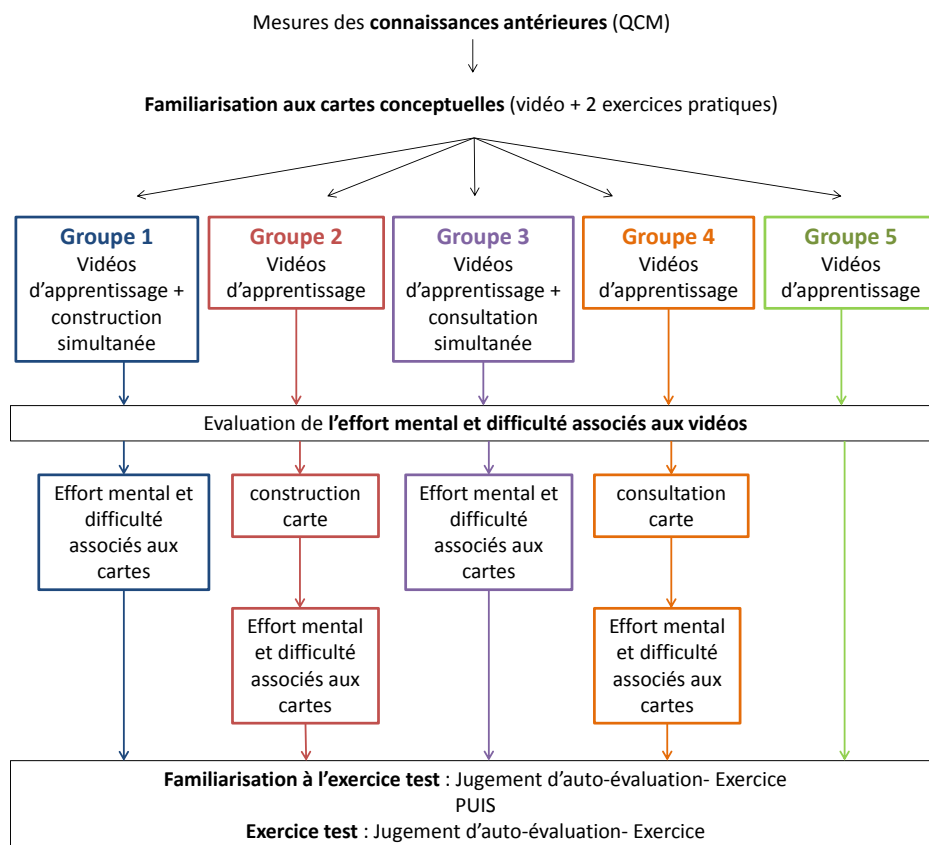


FIGURE 8. – Synthèse de la procédure expérimentale de l'expérience 2. Groupe 1 = construction simultanée ; groupe 2 = construction différée ; groupe 3 = consultation simultanée ; groupe 4 = consultation différée ; groupe 5 = groupe contrôle

ressentis) pour comprendre les vidéos et utiliser les cartes tandis que les participants du groupe contrôle estimaient uniquement la charge cognitive ressentie (effort mental et difficulté) pour comprendre les vidéos. Finalement, tous les participants étaient familiarisés à l'exercice d'application avant de répondre à la question mesurant l'auto-évaluation et de réaliser cet exercice. La procédure est synthétisée dans la Figure 8.

## 5.8 RÉSULTATS

L'objectif général de cette étude est d'identifier sous quelles conditions de présentation les cartes conceptuelles sont les plus efficaces pour l'apprentissage et l'auto-évaluation. Nous nous sommes également intéressés, de façon plus spécifique, aux effets des cartes selon le niveau de connaissances antérieures de l'apprenant. L'ensemble des analyses présentées ici ont été réalisées sur huit variables dépendantes à savoir : le jugement

d'auto-évaluation, la direction et l'amplitude du biais d'auto-évaluation, la performance à l'exercice d'application, l'effort mental et la difficulté ressentis pour comprendre les vidéos et l'effort mental et la difficulté ressentis pour utiliser les cartes.

Notre première hypothèse porte sur l'efficacité globale de l'utilisation d'une carte pour l'amélioration de l'exactitude de l'auto-évaluation et l'apprentissage. Des t-tests comparant les apprenants qui ont utilisé les cartes avec ceux qui ne les ont pas utilisées ont été réalisés. Ces t-tests ont été complétés par une analyse de la variance (ANOVA) à un facteur visant à comparer plus précisément nos cinq groupes expérimentaux afin de déterminer la condition la plus efficace pour aider l'apprentissage et l'auto-évaluation.

Les trois hypothèses suivantes intègrent le niveau d'expertise de l'apprenant. Plus particulièrement, les hypothèses 3 et 4 portent sur l'effet du mode de présentation des cartes selon le niveau de connaissances antérieures des apprenants. Pour vérifier ces hypothèses, nous avons réalisé une ANOVA à mesures indépendantes  $2 \times 2$ , avec en variable indépendante le niveau d'expertise de l'apprenant (expérimenté vs. novices) et le mode d'utilisation des cartes (consultation vs. construction).

La dernière hypothèse concerne l'effet du moment de la présentation des cartes, nous terminerons donc l'analyse des résultats avec un t-test permettant de comparer la condition de présentation simultanée à la condition de présentation différée des cartes conceptuelles.

L'analyse de la qualité des cartes selon le moment de leur construction et le niveau d'expertise de l'apprenant viendront compléter ces résultats. Des résultats statistiques complémentaires sont également intégrés en Annexe D. Ces résultats concernent, non seulement, l'interaction entre le mode et le moment d'utilisation des cartes conceptuelles, mais également l'interaction entre le moment de présentation des cartes et le niveau d'expertise de l'apprenant.

Concernant la description des effets principaux et d'interaction, des tests post-hoc ont été réalisés suite aux Anovas quand cela était possible (le test de Tukey lorsque les variances étaient homogènes, le T3 de Dunnett lorsqu'elles ne l'étaient pas). Lorsqu'il y avait moins de deux groupes concernés par les traitements, des t-tests à mesures indépendantes ont été conduits. L'analyse des résultats a systématiquement été précédée par l'exclusion des valeurs extrêmes associées à chacune des variables dépendantes, ce qui explique les différences de degré de liberté.

### 5.8.1 *Effet de la carte conceptuelle : statistiques descriptives*

Le Tableau 6 présente l'ensemble des conditions de présentation des cartes conceptuelles et identifie les conditions d'utilisation des cartes les

plus efficaces pour l'apprentissage et l'auto-évaluation. Il décrit, dans une première partie, les informations relatives aux connaissances antérieures des participants (les performances, le jugement d'auto-évaluation et l'exactitude du jugement d'auto-évaluation obtenus par les participants au questionnaire de mesure des connaissances antérieures). Dans une seconde partie, les informations présentées concernent les performances, l'auto-évaluation et l'exactitude de l'auto-évaluation des participants à la tâche d'évaluation de leur apprentissage (i.e à l'exercice d'application). Les statistiques descriptives présentées ci-après suivent cette même structure.

### 1. Niveau de connaissances antérieures

- Les performances au questionnaire de mesure des connaissances antérieures sont relativement faibles puisque les participants ont répondu correctement au mieux à 36% des questions (groupe de construction différée,  $M = 36.03$ ). Les participants du groupe de consultation différée ont obtenu la performance la plus basse au niveau de connaissances antérieures ( $M = 27.99$ ). Les moyennes des autres groupes sont comprises entre ces deux extrêmes.
- Le jugement moyen émis au questionnaire de mesure des connaissances antérieures était relativement moyen. Les participants du groupe de construction différée ont initialement émis le jugement de prédiction le plus élevé ( $M = 46.72$ ). Les participants en condition de consultation différée ont émis le jugement le plus faible ( $M = 37.16$ ). L'ensemble des moyennes des autres groupes se situent entre ces deux valeurs.
- L'ensemble des participants sustimaient légèrement leur performance au questionnaire de mesure de leurs connaissances antérieures. Initialement, les participants du groupe contrôle sont ceux qui se sustimaient le moins ( $M = 5.67$ ). Au contraire, les participants du groupe de consultation simultanée se sustimaient initialement le plus, ( $M = 12.59$ ).
- L'amplitude du biais d'auto-évaluation initiale est globalement peu élevée. En effet, les participants les moins exacts dans leur jugement d'auto-évaluation ont un biais moyen égal à 22.58 (construction simultanée). Les participants du groupe contrôle ont été initialement les plus exacts dans leurs prédiction de réussite ( $M = 16.02$ ).

### 2. Exercice d'application

- Globalement, les participants font une prédiction de réussite à l'exercice d'application plutôt modérément élevée (jugements d'auto-évaluation supérieurs à 57%). Les prédictions de réussite sont plus élevées lorsque les apprenants construisent la carte

après les vidéos d'apprentissage (condition construction différée ;  $M = 70.58$ ) que lorsque les participants construisent les cartes simultanément à la tâche d'apprentissage ( $M = 59.73$ ) ou les consultent après l'apprentissage ( $M = 57.78$ ).

- Globalement, l'amplitude des biais d'auto-évaluation est relativement élevée puisqu'elle est comprise entre 40 et 45% pour les groupes qui ont construit ou consulté une carte conceptuelle. Ceci suggère que les participants ne sont pas exacts dans l'évaluation qu'ils font de leur réussite future. En pensant mieux réussir, les participants des groupes de construction ont également des biais dont l'amplitude est plus importante ( $M = 44.74$  pour la construction différée ;  $M = 46.42$  pour la construction simultanée) que celle des groupes en consultation ( $M = 42.65$  pour la consultation différée ;  $M = 40.69$  pour la consultation simultanée). L'amplitude des biais des participants qui ont consulté les cartes est moindre mais reste supérieure à celle des participants du groupe contrôle qui ont l'auto-évaluation absolue la plus exacte ( $M = 37.81$ ).
- Ces différences inter-groupes sont similaires mais plus marquées au regard de la direction du biais d'auto-évaluation. L'ensemble des participants, tous groupes confondus, se sur-évalue. Ceux qui ont construit les cartes sont les plus confiants ( $M = 36.65$  pour la construction différée ;  $M = 36.33$  pour la construction simultanée) comparativement aux participants qui les ont consultées ( $M = 31$  pour la consultation différée ;  $M = 29.7$  pour la consultation simultanée). Enfin, les participants du groupe contrôle, en accord avec les résultats sur l'amplitude des biais d'auto-évaluation, sont ceux qui se sur-évaluent le moins ( $M = 13.71$ ).
- Concernant la performance, celles des participants des groupes expérimentaux est plutôt faible avec un pourcentage d'actions correctes moyen compris entre 23.4 (pour le groupe de construction simultanée) et 33.93% (pour le groupe de construction différée). Les performances du groupe contrôle sont supérieures mais restent relativement faibles ( $M = 46.53$ ).
- Enfin, la charge mentale ressentie était modérée qu'elle soit mesurée au travers de l'effort mental ou de la difficulté ressentis et tant pour comprendre les vidéos que pour utiliser les cartes. Ainsi, l'effort mental ressenti pour comprendre les vidéos était pour tous les participants compris entre 5.05 (pour le groupe de consultation différée) et 5.35 (pour le groupe de construction simultanée) à l'exception des participants du groupe de construction différée qui ont ressenti moins d'effort ( $M = 3.64$ ).

Similairement, la difficulté ressentie pour comprendre les vidéos était moindre pour les participants du groupe de construction différée ( $M = 3.02$ ) que pour les participants des autres groupes (moyennes comprises entre 4.18 pour le groupe contrôle et 4.59 pour le groupe de construction simultanée).

Pour ce qui est de l'utilisation des cartes, les participants du groupe de consultation simultanée ont trouvé que l'effort mental ( $M = 4.89$ ) et la difficulté ( $M = 4.73$ ) associés à cette activité étaient moins importants que pour les participants des trois autres groupes (moyennes comprises entre 5.24 et 5.79 pour la difficulté et l'effort mental confondus).

- La qualité des cartes, reflétée par le pourcentage de liens justes, était relativement faible qu'elles aient été construites pendant ( $M = 8.66$ ) ou après l'apprentissage ( $M = 11.03$ ).

### 5.8.2 *Effet des cartes conceptuelles : Analyses inférentielles*

L'ensemble des participants a été considéré dans ces premières analyses, quel qu'était leur niveau d'expertise. Aussi, afin de s'assurer que d'éventuels effets observés entre les différents groupes expérimentaux ne puissent pas être attribués à des écarts dans le niveau de connaissances antérieures, un premier t-test permettant de comparer les participants qui ont utilisé des cartes et ceux qui n'en ont pas utilisées a été conduit. Cette analyse n'a pas révélé de différence significative entre les deux groupes de participants,  $t(266) = -1.1, p = .3, d = .17$ . En complément, une Anova a été conduite avec le groupe expérimental en variable indépendante et la moyenne de connaissances antérieures dans l'utilisation du traitement de texte en variable dépendante. Celle-ci n'a pas révélé de différence significative entre nos cinq groupes,  $F(4, 263) = 1.56, p = .2, \eta_p^2 = .02$ . Les participants sont donc équitablement répartis dans les cinq groupes selon leur niveau de connaissances.

Ces mêmes analyses ont été conduites sur les indicateurs de l'exactitude de l'auto-évaluation. Le t-test comparant la direction initiale du biais d'auto-évaluation (i.e. mesurée lors du questionnaire de mesure des connaissances antérieures) des participants ayant utilisé des cartes de ceux n'ayant pas utilisé de cartes ne révèle aucune différence significative entre les deux groupes,  $t(270) = 1.23, p = .2, d = .2$ . De la même façon, l'Anova conduite sur la même variable dépendante mais avec le groupe expérimental en variable indépendante n'a pas révélé d'effet significatif du groupe sur la direction initiale du biais d'auto-évaluation,  $F(4, 267) = .68, p = .6, \eta_p^2 = .01$ .

Enfin, ces mêmes analyses ont été conduites sur l'amplitude du biais initial d'auto-évaluation. Le t-test comparant les apprenants qui ont utilisé

TABLEAU 6. – Ensemble des moyennes (et écarts-types) des cinq groupes de participants pour l'ensemble des variables dépendantes. Le symbole \* indique les variables exprimées en pourcentage.

	Construction		Consultation			
	Min ; Max	Différée (N = 45)	Simultanée (N = 49)	Différée (N = 63)	Simultanée (N = 66)	Contrôle (N = 49)
<b>1. Niveau de connaissances antérieures :</b>						
Performance *	0 ; 100	36.03 (20.79)	33.95 (20.81)	27.99 (21.68)	29.02 (21.2)	34.88 (22.14)
Jugement d'auto-évaluation *	0 ; 100	46.72 (18.37)	42.27 (18.38)	37.16 (20)	41.6 (19.22)	40.64 (16.65)
Direction du biais d'auto-évaluation *	-100 ; 100	10.84 (21.70)	8.23 (26.83)	9.15 (21.95)	12.59 (23.59)	5.67 (20.41)
Amplitude du biais d'auto-évaluation *	0 ; 100	19.29 (14.24)	22.58 (16.49)	19.32 (13.69)	21.85 (15.64)	16.02 (13.58)
<b>2. Exercice d'application :</b>						
Jugement d'auto-évaluation *	0 ; 100	70.58 (25.89)	59.73 (24)	57.78 (25)	63.33 (27.44)	60.24 (24.38)
Direction du biais d'auto-évaluation	-100 ; +100	<b>36.65 (43.05)</b>	<b>36.33 (39.8)</b>	31 (38.84)	29.7 (39.57)	<b>13.71 (44.17)</b>
Amplitude du biais d'auto-évaluation *	0 ; 100	44.74 (34.35)	46.42 (27.1)	42.65 (25.24)	40.69 (27.94)	37.81 (26.15)
Performance *	0 ; 100	33.93 (43.6)	<b>23.4 (35.75)</b>	26.77 (39.3)	33.64 (42.83)	<b>46.53 (45.29)</b>
Effort mental ressenti pour comprendre les vidéos	1 ; 9	<b>3.64 (2.56)</b>	<b>5.35 (2.18)</b>	<b>5.05 (2.43)</b>	<b>5.21 (2.2)</b>	<b>5.33 (2.58)</b>
Difficulté ressentie pour comprendre les vidéos	1 ; 9	<b>3.02 (2.08)</b>	<b>4.59 (2.11)</b>	<b>4.57 (2.42)</b>	<b>4.32 (2.26)</b>	4.18 (2.35)
Effort mental ressenti pour utiliser les cartes	1 ; 9	5.24 (2.04)	5.65 (1.56)	<b>5.79 (2.1)</b>	<b>4.89 (2)</b>	X
Difficulté ressentie pour utiliser les cartes	1 ; 9	5.27 (2.2)	5.51 (1.79)	5.29 (2.07)	4.73 (2.04)	X
Qualité des cartes construites *	0 ; 100	11.03 (13.2)	8.66 (10.94)	X	X	X

des cartes conceptuelles pendant l'apprentissage à ceux qui n'en ont pas utilisé a révélé un effet significatif du groupe,  $t(266) = 2.05, p = .04, d = .33$ . Ce résultat suggère que les participants qui ont utilisé une carte avaient initialement un biais d'amplitude plus importante ( $M = 20.81, SD = 15.01$ ) que les participants n'ayant pas utilisé de cartes ( $M = 16.02, SD = 13.58$ ). Le  $d$  de Cohen étant égal à  $.33$ , cet effet est considéré comme faible, ce résultat est donc à considérer avec prudence. L'anova conduite avec le groupe expérimental sur la même variable a révélé que les cinq groupes expérimentaux ne différaient pas significativement les uns des autres,  $F(4, 263) = 1.57, p = .2, \eta_p^2 = .02$ . L'amplitude initiale du biais d'auto-évaluation était similaire pour l'ensemble des groupes de participants.

Dans la suite des analyses, des t-tests ont permis de comparer l'auto-évaluation, les performances et la charge mentale des apprenants qui ont utilisé des cartes avec celles des apprenants qui n'en ont pas utilisées. Ces t-test sont complétés d'Analyses de la Variance (ANOVA), réalisées afin de comparer l'effet du groupe expérimental (i.e. construction simultanée, construction différée, consultation simultanée, consultation différée, et le groupe contrôle) sur chacune des variables dépendantes de notre étude. Les statistiques descriptives relatives aux variables dépendantes sont présentées dans le Tableau 6. Les moyennes significativement différentes les unes des autres sont en gras.

#### *Le jugement d'auto-évaluation*

Le t-test conduit sur la mesure du jugement d'auto-évaluation n'a pas révélé de différence significative entre les participants qui ont utilisé une carte conceptuelle ( $M = 62.44, SD = 25.96$ ) et les participants qui n'ont pas utilisé de cartes conceptuelles ( $M = 60.24, SD = 24.38$ ),  $t(270) = .54, p = .59, d = .09$ .

L'Anova conduite entre nos cinq groupes expérimentaux sur la valeur du jugement d'auto-évaluation émis à l'exercice d'application n'a pas révélé d'effet significatif,  $F(4, 261) = 1.91, p = .1, \eta_p^2 = .03$ , les participants des cinq groupes ont globalement exprimé une prédiction de réussite à l'exercice relativement élevée .

#### *La direction du biais d'auto-évaluation*

Le t-test conduit sur la direction du biais d'auto-évaluation a révélé une différence significative entre les participants qui ont utilisé une carte et ceux qui n'en ont pas utilisée,  $t(270) = 2.99, p = .003, d = .45$ . Les participants qui ont utilisé des cartes conceptuelles ( $M = 32.93, SD = 39.99$ ) se sur-évaluent davantage que les participants du groupe contrôle qui n'en ont pas utilisées ( $M = 13.71, SD = 44.17$ ).



L'Anova réalisée avec la direction du biais d'auto-évaluation en variable dépendante a révélé l'existence d'une différence significative entre nos cinq groupes expérimentaux,  $F(4, 267) = 2.53, p = .04, \eta_p^2 = .04$ . L'analyse des tests post-hoc (test de Tukey) suggère que les participants des groupes de construction différée ( $M = 36.65, SD = 43.05$ ) et de construction simultanée ( $M = 36.33, SD = 39.8$ ) ont eu tendance à plus sur-évaluer leur capacité à réussir l'exercice d'application que les participants du groupe contrôle ( $M = 13.71, SD = 44.17; ps = .05$ ). Les groupes de consultation différée ( $M = 31, SD = 38.84$ ) et simultanée ( $M = 29.7, SD = 39.57$ ) ne différaient pas significativement des trois autres groupes ( $ps > .18$ ).

#### *L'amplitude du biais d'auto-évaluation*

Le t-test réalisé sur l'amplitude du biais d'auto-évaluation n'a pas révélé de différence significative entre les participants qui ont utilisé une carte ( $M = 43.32, SD = 28.34$ ) et ceux qui n'en ont pas utilisées ( $M = 37.81, SD = 26.15$ ),  $t(270) = 1.25, p = .21, d = .2$ .

L'Anova réalisée sur l'amplitude du biais d'auto-évaluation en fonction du groupe expérimental n'a pas révélé de différences significatives entre nos cinq groupes,  $F(4, 267) = .72, p = .58, \eta_p^2 = .01$ . L'amplitude du biais est relativement élevée dans tous les groupes.

#### *La performance à l'exercice d'application*

Le t-test conduit sur la mesure des performances à l'exercice d'application a révélé une différence significative entre les participants qui ont utilisé une carte et ceux qui n'ont pas utilisé de carte,  $t(270) = -2.61, p = .01, d = .4$ . Ce résultat ne valide pas notre hypothèse car il montre que les participants qui n'ont pas utilisé de carte ( $M = 46.53, SD = 45.29$ ) ont obtenu des performances plus élevées que les participants qui ont utilisé une carte ( $M = 29.51, SD = 40.51$ ).

L'Anova réalisée sur les performances à l'exercice d'application a révélé un effet marginalement significatif du groupe,  $F(4, 267) = 2.32, p = .06, \eta_p^2 = .03$ . Contrairement à ce que nous supposions, les post-hoc du T3 de Dunnett ont montré que les participants du groupe contrôle ( $M = 46.53, SD = 45.29$ ) avaient des performances plus élevées que les participants du groupe de construction simultanée ( $M = 23.4, SD = 35.75; p = .06$ ). Aucune autre différence significative n'a été révélée ( $ps > .16$ ).

### *La charge mentale associée à la compréhension des vidéos*

L'EFFORT MENTAL RESSENTI. Le t-test conduit sur l'effort mental ressenti pour comprendre les vidéos n'a pas mis en évidence de différence significative entre les participants qui ont utilisé des cartes ( $M = 4.88$ ,  $SD = 2.41$ ) et ceux qui n'ont pas utilisé de cartes ( $M = 5.33$ ,  $SD = 2.58$ ),  $t(270) = -1.16$ ,  $p = .25$ ,  $d = .18$ .

L'anova réalisée sur l'effort mental ressenti pour comprendre les vidéos a mis en évidence un effet significatif du groupe,  $F(4, 267) = 4.24$ ,  $p = .002$ ,  $\eta_p^2 = .06$ . Les analyses post-hoc de Tukey ont mis en évidence une différence significative entre les participants du groupe de construction différée et les participants des quatre autres groupes ( $ps < .05$ ) suggérant que les participants du groupe de construction différée ont ressenti moins d'effort, ( $M = 3.64$ ,  $SD = 2.56$ ), que les participants des autres groupes.

LA DIFFICULTÉ RESSENTIE. Aucune différence de difficulté ressentie pour comprendre les vidéos n'a été révélée entre les participants qui ont utilisé des cartes ( $M = 4.19$ ,  $SD = 2.31$ ) et ceux qui n'en ont pas utilisées ( $M = 4.18$ ,  $SD = 2.35$ ),  $t(270) = .01$ ,  $p = .99$ ,  $d = 0$ .

L'anova réalisée sur la difficulté ressentie pour comprendre les vidéos a, en revanche, révélé un effet significatif du groupe,  $F(4, 267) = 3.88$ ,  $p = .004$ ,  $\eta_p^2 = .06$ . Les tests post-hoc de Tukey ont révélé une différence significative entre les participants du groupe de construction différée qui ont ressenti moins de difficulté que les participants de construction simultanée ( $p = .01$ ), de consultation différée ( $p = .01$ ) et de consultation simultanée ( $p = .03$ ), mais il n'y a pas de différence avec les participants du groupe contrôle, ( $p = .1$ ).

### *La charge mentale associée à l'utilisation des cartes*

Pour le traitement de ces deux derniers indicateurs, les t-tests n'ont pas pu être réalisés et le groupe contrôle n'apparaît pas dans les anovas puisque ces participants n'ont pas eu de cartes à utiliser.

L'EFFORT MENTAL RESSENTI. L'Anova réalisée sur l'effort mental ressenti pour utiliser les cartes a mis en évidence un effet significatif du groupe expérimental,  $F(3, 219) = 2.69$ ,  $p = .05$ ,  $\eta_p^2 = .04$ . Les tests post-hoc de Tukey ont mis en évidence que les participants qui ont consulté les cartes pendant l'apprentissage, ( $M = 4.89$ ,  $SD = 2$ ) ont ressenti moins d'effort à utiliser les cartes que les participants qui ont consulté les cartes après l'apprentissage, ( $M = 5.79$ ,  $SD = 2.1$ ;  $p = .05$ ).

LA DIFFICULTÉ RESSENTIE. Enfin, l'Anova réalisée sur la difficulté ressentie pour utiliser les cartes n'a pas permis de mettre en évidence d'effet du groupe,  $F(3, 219) = 1.61, p = .19, \eta_p^2 = .02$ .

Ces premiers résultats n'ont pas confirmé notre première hypothèse selon laquelle les individus avec des cartes apprennent mieux et s'auto-évaluent avec plus d'exactitude que les individus sans cartes puisque les participants du groupe contrôle tendent à avoir de meilleures performances ainsi qu'un biais d'amplitude moins important que les participants des quatre autres groupes. Leur niveau de charge ressenti pour comprendre les vidéos était seulement supérieur à celui des participants qui ont construit les cartes après l'apprentissage.

### 5.8.3 *Effet du mode d'utilisation des cartes selon le niveau d'expertise de l'apprenant*

Rappelons que notre seconde hypothèse portait sur la supériorité des apprenants expérimentés sur les novices en termes de performance d'apprentissage et d'exactitude de l'auto-évaluation. Dans nos hypothèses 3 et 4, deux alternatives étaient proposées concernant le coût cognitif des conditions de consultation et de construction et les répercussions en termes d'exactitude de l'auto-évaluation chez les apprenants selon leur niveau d'expertise. Ainsi, la littérature nous a conduit à penser que les apprenants expérimentés auraient de meilleures performances et une auto-évaluation plus exacte dans la condition de consultation que dans la condition de construction (H3) sans savoir si cet effet serait associé à une charge mentale trop faible ou trop élevée en condition de consultation. Chez les novices à l'inverse, nous pensons que la condition de consultation pourrait être plus efficace que la condition de construction (H4).

Un premier t-test conduit sur le niveau de connaissances antérieures a permis de vérifier que les participants expérimentés possédaient plus de connaissances antérieures que les participants novices. Cinq novices ayant obtenu des valeurs extrêmes au questionnaire de mesure des connaissances antérieures ont été ôtés des analyses car leur niveau d'expertise était trop élevé. Le t-test a révélé une différence significative entre nos deux groupes de participants,  $t(266) = -21.21, p < .001, d = 2.59$ , suggérant que les participants expérimentés ( $M = 49.05, SD = 14.03$ ) avaient un niveau de connaissances antérieures plus élevé que les participants novices ( $M = 15.11, SD = 12.13$ ).

Pour vérifier les hypothèses 3 et 4, des Anovas  $2 \times 2$  ont été conduites avec le niveau d'expertise (expérimentés vs. novices) et le mode d'utilisation des cartes (consultation vs. construction) en variables indépendantes. Ces analyses ont été réalisées sur les trois mesures de l'auto-évaluation et de son exactitude, la performance et les quatre mesures de la charge

mentale. L'analyse de l'effet principal de l'expertise nous permettra de répondre à la seconde hypothèse tandis que les effets d'interaction nous aideront à répondre à nos hypothèses 3 et 4. L'effet principal du mode d'utilisation des cartes sera également présenté. Les interactions significatives ont été détaillées à l'aide de t-tests.

#### 5.8.4 *Effet du mode d'utilisation des cartes selon le niveau d'expertise de l'apprenant : statistiques descriptives*

Comme pour les statistiques descriptives du tableau précédent, la présentation a, de nouveau, été scindée en deux parties dans le Tableau 7 et dans la description qui l'accompagne. Les statistiques présentent les moyennes des participants selon leur niveau d'expertise et le mode d'utilisation des cartes. Dans un premier temps les informations relatives aux connaissances antérieures des participants sont décrites. Dans un second temps, les informations concernent les performances, le jugement et l'exactitude de ce jugement obtenus par les participants à l'exercice d'application.

L'observation des données descriptives, présentées dans le Tableau 7 montrent que pour :

##### 1. Les connaissances antérieures

- Le niveau de connaissances antérieures des participants expérimentés est relativement élevé ( $M = 49.05$ ), tandis qu'il est plutôt faible pour les novices ( $M = 15.11$ ).
- Les prédictions de réussite au questionnaire de mesures des connaissances antérieures sont moyennes chez les participants expérimentés ( $M = 49.65$ ) et relativement basses chez les apprenants novices ( $M = 34.26$ ).
- La direction du biais d'auto-évaluation indique que, dans l'ensemble, les participants expérimentés sont exacts dans l'évaluation de leur performances ( $M = .54$ ) tandis que les novices tendent à se surévaluer ( $M = 18.8$ ).
- L'amplitude moyenne des biais d'auto-évaluation au questionnaire de mesure des connaissances antérieures est relativement basse pour les deux populations. Les participants expérimentés sont, toutefois, plus exacts ( $M = 16.38$ ) que les participants novices ( $M = 24.67$ ).

##### 2. L'exercice d'application

- Les prédictions de réussite sont élevées pour l'ensemble des participants mais les participants expérimentés évaluent plus fortement leurs performances futures ( $M = 70.25$ ) que les novices ( $M = 55.1$ ).

- Quel que soit leur niveau d'expertise, les apprenants ont tendance à se surévaluer. Cependant, cette tendance est moins forte chez les participants expérimentés : en condition 'construction' la direction moyenne du biais indique une tendance plus forte à la surestimation ( $M = 23.12$ ) qu'en condition de 'consultation' où la moyenne indique une surestimation très faible ( $M = 5.26$ ). Chez les novices, les moyennes relativement élevées pour les conditions 'construction' ( $M = 53.94$ ) et 'consultations' ( $M = 46.79$ ), indiquent une tendance assez forte à se surestimer. Les écarts-types indiquent cependant une forte variation autour de ces moyennes pour les deux groupes.
- L'exactitude absolue, exprimée par l'amplitude des biais d'auto-évaluation, est relativement faible pour l'ensemble des participants, et l'est davantage pour les novices ( $M = 51.36$ ) que pour les expérimentés ( $M = 33.98$ ). Chez les novices, l'amplitude des biais est relativement homogène quelle que soit la condition d'utilisation des cartes conceptuelles ( $M = 53.94$  pour la construction ;  $M = 50.14$  pour la consultation). Les expérimentés, quant à eux, tendent à être plus exacts dans la condition de consultation ( $M = 28.88$ ) que dans la condition de construction ( $M = 38.72$ ).
- Bien que les performances des expérimentés soient relativement moyennes ( $M = 55.72$ ), ces derniers ont beaucoup mieux réussi que les novices ( $M = 6.01$ ) qui ont des performances très faibles. La condition de consultation de cartes a permis aux deux populations d'avoir des performances plus élevées ( $M = 65.36$  pour les expérimentés ;  $M = 7.54$  pour les novices) que la condition de construction de cartes ( $M = 46.79$  pour les expérimentés ;  $M = 2.78$  pour les novices).
- La charge mentale ressentie pour comprendre les vidéos, exprimée au travers de l'effort mental et de la difficulté ressentie, était moyenne et sensiblement plus faible pour les apprenants expérimentés ( $M = 4.49$  pour l'effort mental ;  $M = 3.76$  pour la difficulté ressentie) que pour les apprenants novices ( $M = 5.21$  pour l'effort mental ;  $M = 4.5$  pour la difficulté ressentie). La charge ressentie était similaire quelle que soit la condition d'utilisation des cartes à l'exception de la difficulté ressentie par les participants expérimentés qui était sensiblement plus basse en condition de construction ( $M = 3.35$ ) qu'en condition de consultation ( $M = 4.22$ ).
- La charge mentale ressentie pour utiliser les cartes était moyenne et comparable pour les participants expérimentés et les novices tant en termes d'effort mental ressenti ( $M = 5.38$  pour les expé-

rimentés ;  $M = 5.38$  pour les novices) qu'en termes de difficulté ressentie ( $M = 5.14$  pour l'effort mental ;  $M = 5.21$  pour la difficulté ressentie). De même, la charge mentale ressentie selon les différentes conditions d'utilisation était homogène.

- Le pourcentage de liens correctement construits est faible, quel que soit le niveau d'expertise des apprenants. Néanmoins, les experts ont construit des cartes de qualité supérieure ( $M = 13.87$ ) à celles des novices ( $M = 4.4$ ).

### 5.8.5 Effet du mode d'utilisation des cartes selon le niveau d'expertise de l'apprenant : statistiques inférentielles

#### *Le jugement d'auto-évaluation*

L'Anova réalisée sur le jugement d'auto-évaluation n'a pas révélé d'interaction entre le niveau d'expertise et le mode d'utilisation des cartes,  $F(1, 214) = .20, p = .7, \eta_p^2 = .001$  (cf. Figure 9). Aucun effet du mode n'a été détecté,  $F(1, 214) = .06, p = .8, \eta_p^2 = 0$ .

Un effet principal de l'expertise a été révélé,  $F(1, 214) = 17.48, p < .001, \eta_p^2 = .08$ , indiquant que les participants expérimentés ( $M = 70.25, SD = 22.43$ ) ont un jugement d'auto-évaluation plus élevé que les participants novices ( $M = 55.1, SD = 27.27$ ).

#### *La direction du biais d'auto-évaluation*

L'Anova réalisée sur la direction du biais d'auto-évaluation, n'a pas révélé d'interaction entre le niveau d'expertise et le mode d'utilisation des cartes,  $F(1, 214) = 1.13, p = .29, \eta_p^2 = .01$ .

Seuls des effets principaux de l'expertise,  $F(1, 214) = 51.82, p < .001, \eta_p^2 = .2$ , et du mode,  $F(1, 214) = 6.2, p = .01, \eta_p^2 = .03$  ont été mis en évidence (cf. Figure 9). L'effet de l'expertise indiquait, conformément à nos prédictions, que les participants expérimentés ( $M = 14.53, SD = 40.64$ ) se surévaluaient moins que les participants novices ( $M = 49.09, SD = 31.59$ ). L'effet du mode indiquait que la consultation ( $M = 30.11, SD = 39.33$ ) a conduit les participants à moins se sur-évaluer qu'en condition de construction ( $M = 35.32, SD = 41.22$ ).

#### *L'amplitude du biais d'auto-évaluation*

L'Anova réalisée sur l'amplitude du biais d'auto-évaluation n'a pas confirmé l'hypothèse d'un effet différent de l'utilisation des cartes selon le niveau d'expertise puisqu'aucun effet d'interaction n'a été mis à jour,  $F(1, 214) = .64, p = .43, \eta_p^2 = .003$ .

TABLEAU 7. – Ensemble des moyennes (et écarts-types) obtenus sur les principales variables dépendantes selon le niveau d'expertise des apprenants et le mode d'utilisation des cartes. Le symbole \* indique les variables exprimées en pourcentage.

	Expérimentés			Novices			
	min ; max	Construction (N = 55)	Consultation (N = 51)	Total (N = 106)	Construction (N = 39)	Consultation (N = 78)	Total (N = 117)
<b>1. Niveau de connaissances antérieures :</b>							
Performance*	0 ; 100	48.4 (13.66)	49.86 (13.81)	49.13 (13.69)	17.25 (13.96)	14.57 (11.66)	15.46 (12.48)
Jugement d'auto-évaluation*	0 ; 100	49.36 (18.56)	49.94 (18.25)	49.65 (18.32)	37.68 (16.16)	32.56 (17.47)	34.26 (17.15)
Direction du biais d'auto-évaluation*	-100 ; 100	1 (21.13)	.08 (19.67)	.54 (20.32)	20.43 (24.54)	17.99 (21.97)	18.80 (22.78)
Amplitude du biais d'auto-évaluation*	0 ; 100	16.62 (12.77)	16.13 (11.72)	16.38 (12.2)	26.91 (16.95)	23.55 (15.78)	24.67 (16.18)
<b>2. Exercice d'application :</b>							
Jugement d'auto-évaluation*	0 ; 100	69.91 (23.73)	70.63 (21.17)	<b>70.25 (22.43)</b>	56.72 (26.63)	54.33 (27.72)	<b>55.1 (27.27)</b>
Direction du biais d'auto-évaluation	-100 ; 100	23.12 (44.23)	5.26 (34.45)	<b>14.53 (40.64)</b>	53.94 (27.56)	46.79 (33.26)	<b>49.09 (31.59)</b>
Amplitude du biais d'auto-évaluation*	0 ; 100	38.72 (31.21)	28.88 (19.1)	33.98 (26.44)	53.94 (27.56)	50.14 (27.87)	<b>51.36 (27.71)</b>
Performance*	0 ; 100	46.79 (42.97)	65.36 (37.26)	<b>55.72 (41.2)</b>	2.78 (9.61)	7.54 (24.01)	<b>6.01 (20.58)</b>
Effort mental ressenti pour comprendre les vidéos	1 ; 9	4.16 (2.28)	4.84 (2.21)	4.49 (2.26)	5.11 (2.75)	5.26 (2.37)	<b>5.21 (2.49)</b>
Difficulté ressentie pour comprendre les vidéos	1 ; 9	3.35 (1.89)	4.22 (2.07)	<b>3.76 (2.02)</b>	4.53 (2.49)	4.49 (2.44)	<b>4.5 (2.45)</b>
Effort mental ressenti pour utiliser les cartes	1 ; 9	5.42 (1.67)	5.33 (2.1)	5.38 (1.88)	5.47 (1.98)	5.34 (2.13)	<b>5.38 (2.07)</b>
Difficulté ressentie pour utiliser les cartes	1 ; 9	5.45 (1.94)	4.8 (1.99)	<b>5.14 (1.98)</b>	5.36 (2.02)	5.13 (2.14)	<b>5.21 (2.09)</b>
Qualité des cartes construites	0 ; 100	13.87 (13.6)	X	13.87 (13.6)	4.4 (6.68)	X	4.4 (6.68)

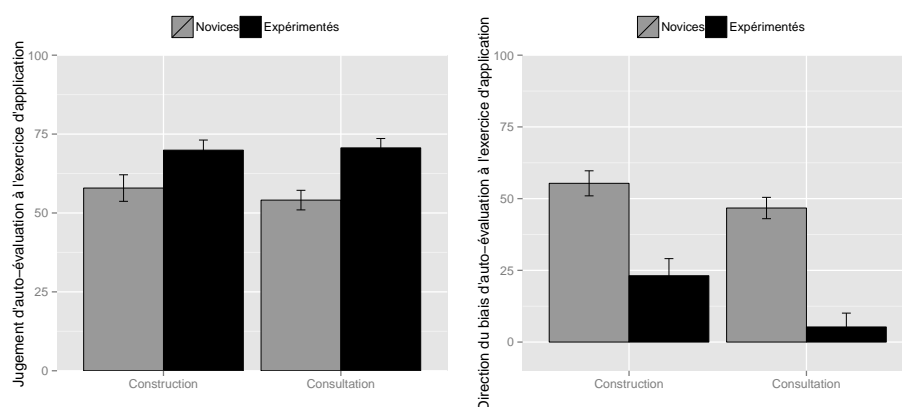


FIGURE 9. – Jugement et direction du biais d'auto-évaluation en fonction du niveau d'expertise de l'apprenant et du mode d'utilisation des cartes.

À nouveau, un effet de l'expertise a été révélé,  $F(1, 214) = 23.24, p < .001, \eta_p^2 = .1$ , indiquant, en accord avec notre deuxième hypothèse, que les participants expérimentés ( $M = 33.98, SD = 26.44$ ) étaient plus exacts que les participants novices ( $M = 51.36, SD = 27.71$ ). L'effet principal du mode n'était que marginal,  $F(1, 214) = 3.26, p = .07, \eta_p^2 = .02$  et indiquait que consultation ( $M = 41.6, SD = 26.77$ ) entraînait un biais d'une amplitude moins importante que la construction ( $M = 44.74, SD = 30.59$ ).

#### *La performance à l'exercice d'application*

Sur les performances obtenues à l'exercice d'application, l'Anova conduite n'a pas révélé d'interaction entre le niveau d'expertise et le mode d'utilisation des cartes,  $F(1, 214) = 2.4, p = .12, \eta_p^2 = .01$ .

En revanche, en accord avec notre prédiction, un effet de l'expertise a été révélé,  $F(1, 214) = 130.7, p < .001, \eta_p^2 = .4$ , indiquant que les participants expérimentés ( $M = 55.72, SD = 41.2$ ) ont obtenu des performances plus élevées à l'exercice d'application que les participants novices ( $M = 6.01, SD = 20.58$ ).

Un effet principal du mode d'utilisation des cartes a également émergé,  $F(1, 214) = 6.87, p < .01, \eta_p^2 = .03$ , suggérant que les participants qui ont construit les cartes ( $M = 29.38, SD = 40.15$ ) ont moins bien réussi que les participants qui ont consulté les cartes ( $M = 30.76, SD = 41.28$ ).

#### *L'effort mental et la difficulté associés à la compréhension des vidéos*

L'EFFORT MENTAL. Aucune interaction entre le niveau d'expertise et le mode d'utilisation des cartes n'a été obtenue sur l'effort mental ressenti



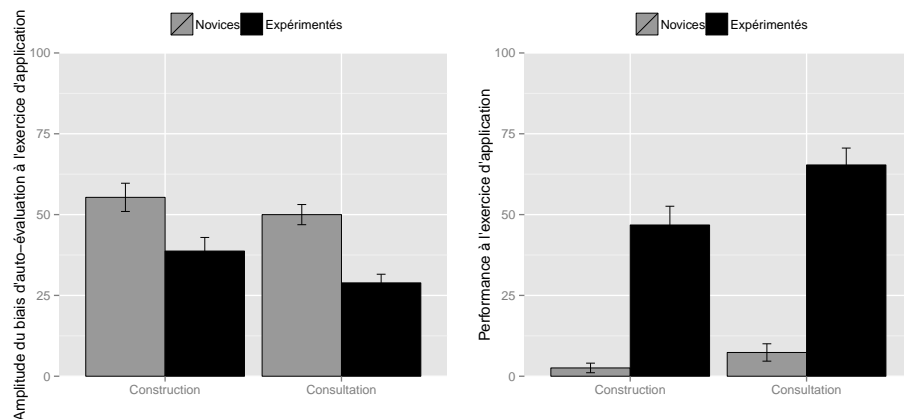


FIGURE 10. – Amplitude du biais d’auto-évaluation et performance à l’exercice d’application en fonction du niveau d’expertise de l’apprenant et du mode d’utilisation des cartes.

pour comprendre les vidéos,  $F(1, 214) = .62, p = .4, \eta_p^2 = .003$ . Seul un effet significatif de l’expertise,  $F(1, 214) = 4.19, p = .04, \eta_p^2 = .02$ , a mis en évidence que les novices ( $M = 5.21, SD = 2.49$ ) ont ressenti plus d’effort pour comprendre les vidéos que les participants expérimentés ( $M = 4.49, SD = 2.26$ ).

Aucun effet du mode d’utilisation des cartes n’a émergé sur l’effort mental,  $F(1, 214) = 1.55, p = .2, \eta_p^2 = .01$ .

**LA DIFFICULTÉ.** Les résultats obtenus sur la difficulté pour comprendre les vidéos sont similaires à ceux obtenus sur l’effort. Aucun effet d’interaction n’a été révélé,  $F(1, 214) = 2.11, p = .15, \eta_p^2 = .01$ . Un effet de l’expertise,  $F(1, 214) = 5.36, p = .02, \eta_p^2 = .02$ , a montré que les participants expérimentés ( $M = 3.76, SD = 2.02$ ) ont trouvé les vidéos moins difficiles à comprendre que les novices ( $M = 4.5, SD = 2.45$ ). Aucun effet du mode d’utilisation des cartes n’est apparu,  $F(1, 214) = 1.75, p = .19, \eta_p^2 = .01$ .

#### *L’effort mental et la difficulté associés à l’utilisation des cartes*

**L’EFFORT MENTAL.** L’Anova conduite sur l’effort mental n’a pas révélé d’interaction entre l’expertise et le mode d’utilisation des cartes,  $F(1, 214) = .01, p = .94, \eta_p^2 = 0$ , pas plus qu’un effet de l’expertise,  $F(1, 214) = .01, p = .91, \eta_p^2 = 0$ , ou qu’un effet du mode d’utilisation des cartes,  $F(1, 214) = .15, p = .7, \eta_p^2 = .001$ . Les participants expérimentés ( $M = 5.38, SD = 1.88$ ) ont donc ressenti autant d’effort mental dans l’utilisation des cartes que les participants novices ( $M = 5.38, SD = 2.07$ ).

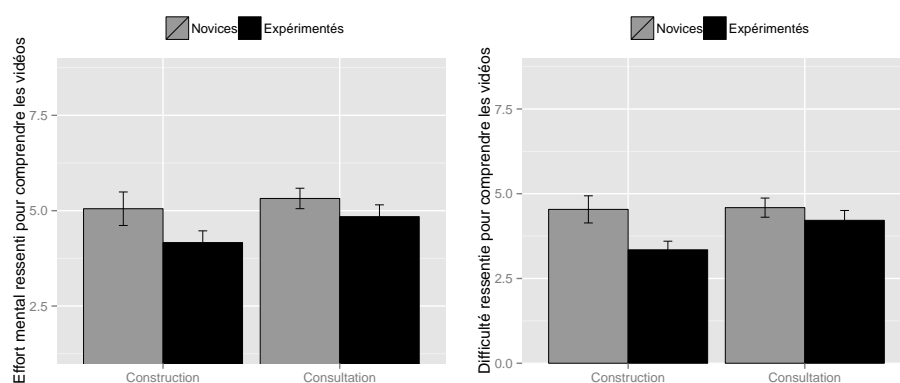


FIGURE 11. – Effort mental et difficulté ressentis pour comprendre les vidéos en fonction du niveau d'expertise de l'apprenant et du mode d'utilisation des cartes.

**LA DIFFICULTÉ.** Ces résultats sont confirmés par l'analyse de la difficulté ressentie pour utiliser les cartes puisqu'à nouveau, aucun effet d'interaction,  $F(1, 214) = .54, p = .46, \eta_p^2 = .003$ , aucun effet du mode d'utilisation des cartes,  $F(1, 214) = 2.38, p = .13, \eta_p^2 = .01$ , et aucun effet de l'expertise,  $F(1, 214) = .17, p = .68, \eta_p^2 = .001$ , n'ont été révélés. Les participants expérimentés ( $M = 5.14, SD = 1.98$ ) ont trouvé l'utilisation des cartes aussi difficile que les participants novices ( $M = 5.21, SD = 2.09$ ).

### 5.8.6 La qualité des cartes conceptuelles

Un t-test a été conduit afin de comparer la qualité de la carte construite selon le niveau d'expertise des apprenants. Ce t-test a révélé une différence significative,  $t(89) = -4.0, p < .001, d = .88$  suggérant que les participants expérimentés produisaient un pourcentage de liens justes plus important ( $M = 13.87, SD = 13.6$ ) que les novices ( $M = 4.4, SD = 6.68$ ). Toutefois, le pourcentage de liens corrects même chez les plus expérimentés reste très faible.

### 5.8.7 Effet du moment d'utilisation des cartes conceptuelles

Afin de répondre à notre cinquième hypothèse et déterminer le moment de l'utilisation des cartes le plus utile pour améliorer l'apprentissage et l'exactitude de l'auto-évaluation, des t-tests ont été conduits. Ces analyses comparent les participants qui ont utilisé les cartes pendant l'apprentissage à ceux qui les ont utilisées après l'apprentissage, indépendamment du mode d'utilisation des cartes sur chacune des huit variables dépendantes.

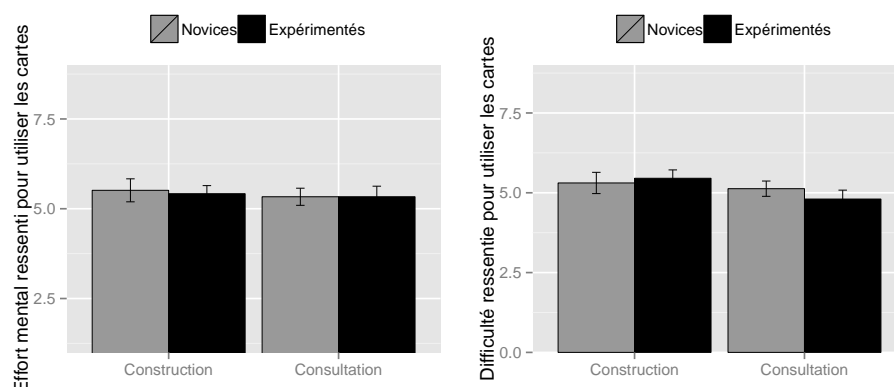


FIGURE 12. – Effort mental et difficulté ressentis pour utiliser les cartes en fonction du niveau d’expertise de l’apprenant et du mode d’utilisation des cartes.

#### *Le jugement d’auto-évaluation*

Le test conduit sur le jugement d’auto-évaluation ne montre aucune différence significative entre les participants qui ont utilisé les cartes après l’apprentissage et ceux qui les ont utilisées pendant,  $t(221) = .38, p = .7, d = .05$ .

#### *La direction du biais d’auto-évaluation*

Le t-test conduit sur la direction du biais n’a révélé aucune différence significative entre les participants qui ont utilisé les cartes pendant ou après l’apprentissage,  $t(221) = .16, p = .88; d = .02$ .

#### *L’amplitude du biais d’auto-évaluation*

Le t-test conduit sur l’amplitude du biais d’auto-évaluation n’a révélé aucune différence significative entre les participants qui ont utilisé les cartes pendant l’apprentissage ( $M = 43.13, SD = 27.59$ ) et les participants qui ont utilisé les cartes après ( $M = 43.52, SD = 29.25$ ),  $t(221) = .10, p = .92, d = .01$ .

#### *La performance à l’exercice d’application*

Le t-test conduit sur la performance n’a pas révélé de différences significatives entre les participants qui ont utilisé les cartes pendant l’apprentissage, ( $M = 29.28, SD = 40.12$ ), et ceux qui les ont utilisés après l’apprentissage, ( $M = 29.75, SD = 41.10$ ),  $t(221) = .09, p = .9, d = .01$

### *L'effort mental et la difficulté associés à la compréhension des vidéos*

Le t-test réalisé sur l'effort mental associé à la compréhension des vidéos a révélé une différence significative selon que les participants aient utilisé les cartes pendant ou après l'apprentissage,  $t(221) = -2.53, p = .01, d = .35$ . Les participants qui ont utilisé les cartes après l'apprentissage ont ressenti moins d'effort pour comprendre les vidéos, ( $M = 4.46, SD = 2.57$ ) que les participants qui ont utilisé les cartes pendant l'apprentissage ( $M = 5.27, SD = 2.19$ ).

Le résultat précédent se retrouve sur la difficulté ressentie pour comprendre les vidéos. Ainsi, le t-test a révélé que les participants qui utilisaient les cartes après l'apprentissage trouvaient, de façon significative,  $t(221) = -1.65, p = .1, d = .22$ , les vidéos moins difficiles ( $M = 3.92, SD = 2.4$ ) que les participants qui les ont utilisées pendant l'apprentissage, ( $M = 4.43, SD = 2.19$ ).

### *L'effort mental et difficulté associés à l'utilisation des cartes*

Le t-test n'a pas révélé d'effet du moment d'utilisation des cartes sur l'effort mental ressenti pour utiliser les cartes,  $t(221) = 1.31, p = .19, d = .18$ . Les participants qui ont utilisé les cartes pendant l'apprentissage, ( $M = 5.22, SD = 1.86$ ), ont ressenti autant d'effort mental que les participants qui ont utilisé les cartes après l'apprentissage, ( $M = 5.56, SD = 2.08$ ).

Le t-test n'a pas, non plus, révélé d'effet du moment d'utilisation des cartes sur la difficulté ressentie pour utiliser les cartes,  $t(221) = .79, p = .43, d = .11$ . À nouveau, les participants qui ont utilisé les cartes pendant l'apprentissage, ( $M = 5.06, SD = 1.97$ ) ont ressenti autant de difficulté que les participant qui ont utilisé les cartes après l'apprentissage, ( $M = 5.28, SD = 2.11$ ).

### *5.8.8 La qualité des cartes conceptuelles*

Afin d'évaluer la qualité des cartes construites par les participants des groupes de construction différée et simultanée, un t-test à mesures indépendantes a été conduit. Celui-ci n'a révélé aucune différence significative entre les deux groupe de construction,  $t(89) = .94, p = .35, d = .2$ . Les participants qui ont construit les cartes après l'apprentissage ont produit aussi peu de liens justes ( $M = 1.55, SD = 1.85$ ) que les participants qui ont construit les cartes pendant l'apprentissage ( $M = 1.21, SD = 1.53$ ). La carte des participants de la condition de construction différée était en moyenne juste à 11% et celle des participants de la condition de construction simultanée était juste en moyenne à 8.7%.

## 5.9 DISCUSSION

Face à des tâches d'apprentissage complexes, les apprenants ont tendance à mal évaluer l'état de leurs connaissances (e.g. [Maki, 1998](#)). Pour améliorer l'exactitude de l'auto-évaluation, la construction de cartes conceptuelles s'est précédemment avérée efficace ([Redford et al., 2012](#) ; [Thiede et al., 2010](#)). La construction de cartes conceptuelles aide, en effet, l'apprenant à réaliser un apprentissage profond et à accéder à des indices plus pertinents pour l'auto-évaluation ([Baars et al., 2013](#) ; [Novak & Cañas, 2008](#) ; [Redford et al., 2012](#)). Néanmoins, il a également été montré que l'utilisation des cartes pouvait créer une charge superflue, c'est-à-dire qu'elles peuvent potentiellement consommer des ressources déjà limitées et nécessaires pour un apprentissage profond et son évaluation (e.g. [Stull & Mayer, 2007](#)). La présente étude avait pour objectif d'identifier sous quelles conditions d'utilisation les cartes conceptuelles sont les plus efficaces pour améliorer l'exactitude de l'auto-évaluation et l'apprentissage selon le niveau d'expertise de l'apprenant.

### 5.9.1 *Rappel des hypothèses et synthèse des principaux résultats*

Le principal objectif de cette étude était de tester l'efficacité de différentes présentations des cartes conceptuelles sur l'exactitude de l'auto-évaluation et l'apprentissage

Nous prédisions, en référence à [Redford et al. \(2012\)](#) ; [Thiede et al. \(2010\)](#) ; [Van Loon et al. \(2014\)](#), que les apprenants ayant utilisé des cartes auraient une auto-évaluation plus exacte et des performances d'apprentissage plus élevées que les apprenants n'ayant pas utilisé cet outil (H1). L'analyse des résultats ne valide pas cette hypothèse. En effet, bien que des résultats significatifs aient été observés sur les performances d'apprentissage et la direction du biais d'auto-évaluation, ceux-ci indiquent que les participants du groupe contrôle ont obtenu des performances d'apprentissage plus élevées et se sont moins sur-évalués que les apprenants aidés de cartes conceptuelles. Une explication potentielle serait que les participants qui ont construit les cartes ont eu l'illusion de comprendre le contenu proposé. L'absence de feedback sur l'exactitude de la carte construite ne leur a pas permis d'ajuster leur évaluation.

Dans une seconde hypothèse, nous prédisions que les experts seraient plus exacts et plus performants que les novices (H2). En accord avec ces prédictions et les précédentes études (e.g., [Chi & Ceci, 1978](#) ; [Griffin et al., 2008](#)), nos résultats ont confirmé cette hypothèse. Les participants les plus expérimentés ont ainsi obtenu des performances d'apprentissage beaucoup plus élevées que les novices ainsi qu'une auto-évaluation plus exacte. En outre, la charge mentale ressentie par les expérimentés était également plus

faible que celle ressentie par les novices, mais uniquement celle associée à la compréhension des vidéos. Expérimentés et novices ressentait une charge mentale similaire pour utiliser les cartes conceptuelles.

Par ailleurs, nous prédisions également que les experts s'auto-évalueraient avec plus d'exactitude en condition de construction qu'en condition de consultation (H3c). Cette supériorité de la construction pouvait être liée à une charge cognitive trop importante ressentie en condition de consultation (H3a) ou à une charge trop peu importante ressentie en condition de consultation (H3b). Aucune de ces hypothèses n'a été confirmée par nos résultats. Les apprenants expérimentés n'ont, en effet, pas obtenu de performances significativement meilleures, ni une auto-évaluation plus exacte en condition de construction qu'en condition de consultation. Les résultats sur la charge mentale ressentie pour comprendre les vidéos ou pour utiliser les cartes ne permettent pas, quant à eux, de conclure sur la validité de l'hypothèse 3a ou 3b. Les charges mentales ressenties pour comprendre les vidéos d'apprentissage et utiliser les cartes étaient, en effet, similaires selon les deux modes d'utilisation des cartes. Comme le soulignaient [Redford et al. \(2012\)](#), suite à leur première expérience, les participants ont pu utiliser les cartes conceptuelles comme outil de synthèse au lieu de les utiliser comme support à l'auto-évaluation. Dès lors, leur jugement de réussite ne se serait pas basé sur des indices issus de l'utilisation des cartes mais sur d'autres indices, tel que leur niveau d'expertise.

Concernant les novices, des hypothèses contradictoires avaient été émises. Certains travaux nous avait conduit à penser que l'auto-évaluation des novices serait plus exacte après avoir consulté des cartes conceptuelles qu'après les avoir construites (H4b). D'autres nous amenaient à penser que la construction serait plus efficace que la consultation (H4a). Les résultats obtenus ne valident aucune de ces deux hypothèses. En effet, aucun mode d'utilisation des cartes ne s'est révélé plus efficace que l'autre. La charge mentale ressentie pour comprendre les vidéos ou utiliser les cartes est également la même quelle que soit la condition d'utilisation. Pour cette population, le niveau de connaissances antérieures, particulièrement peu élevé (15% de réussite au questionnaire de mesure des connaissances antérieures), peut en partie expliquer l'absence d'effet des cartes conceptuelles. Potentiellement, les novices disposaient de trop peu de connaissances antérieures pour pouvoir tirer profit des cartes conceptuelles. [Novak et Cañas \(2008\)](#) déclarait que l'apprenant doit posséder des connaissances antérieures pertinentes et que les informations données doivent être mises en relation avec les connaissances antérieures des apprenants pour qu'un apprentissage profond puisse avoir lieu. De plus, similairement à l'explication donnée précédemment pour les participants expérimentés, les novices ont également pu utiliser les cartes conceptuelles proposées comme outil de

synthèse plutôt qu'outil d'aide à l'auto-évaluation, et ainsi utiliser d'autres indices (par exemple : leur niveau d'expertise) pour l'auto-évaluation.

Nous prédisions finalement que l'utilisation simultanée pouvait être moins coûteuse et par conséquent plus efficace pour l'apprentissage et l'auto-évaluation que l'utilisation différée (H5). Les travaux de [Van Loon et al. \(2014\)](#) nous conduisaient à penser, à l'inverse, que l'utilisation différée des cartes conceptuelles pouvait être plus efficace pour l'apprentissage et l'auto-évaluation que l'utilisation simultanée. Les résultats obtenus indiquent que l'utilisation différée des cartes permettrait à l'apprenant de ressentir moins de charge pour comprendre un contenu d'apprentissage comparativement à une utilisation simultanée. Toutefois, le moment de l'utilisation n'a pas d'effet sur la charge mentale associée à l'utilisation des cartes qui était moyenne dans les deux conditions. De plus, le moment de l'utilisation des cartes n'a pas eu de répercussion sur l'exactitude de l'auto-évaluation, les performances d'apprentissage ou la qualité des cartes construites.

Nous nous intéressions finalement, à titre exploratoire, à l'effet principal du mode d'utilisation des cartes ainsi qu'à la qualité des cartes construites selon le moment de présentation des cartes et le niveau d'expertise des apprenants. L'analyse des effets principaux indiquait que les participants qui ont consulté les cartes conceptuelles tendaient à s'auto-évaluer avec plus d'exactitude que les participants qui ont construit les cartes. Similairement, les performances d'apprentissage des participants qui ont consulté les cartes étaient supérieures à celles des participants qui ont construit les cartes. Aucun effet du mode n'a, toutefois, été détecté sur la charge ressentie pour comprendre les vidéos ou utiliser les cartes. Ainsi, contrairement à ce qu'ont montré [Redford et al. \(2012\)](#), nos résultats indiquent que la consultation est plus utile que la construction. En désaccord avec [Stull et Mayer \(2007\)](#), il s'avère que la consultation et la construction de cartes sont perçues comme étant aussi coûteuses l'une que l'autre en ressources cognitives. Par ailleurs, l'analyse de la qualité des cartes a révélé que l'ensemble des participants avait globalement construit des cartes peu élaborées (peu de liens) et de faible qualité (peu de liens justes). Ce résultat est valable quel que soit le moment de production de la carte ou le niveau d'expertise de l'apprenant.

En conclusion, les cartes conceptuelles se sont, précédemment, révélées efficaces pour l'apprentissage et l'auto-évaluation (e.g., [Nesbit & Adesope, 2006](#) ; [Redford et al., 2012](#)). En contrepartie, leur utilisation, et principalement leur construction, peut parfois surcharger les capacités cognitives de l'apprenant (e.g., [Chang, Sung, & Chen, 2002](#) ; [Gurlitt & Renkl, 2008](#) ; [Hilbert & Renkl, 2008](#)). Les résultats de notre étude n'ont pas permis de vérifier directement l'hypothèse supposant que les apprenants aidés de cartes se trouvaient en surcharge cognitive comparativement aux ap-

prenants non aidés. Il semblerait, par ailleurs, que les cartes ont été peu et/ou mal utilisées par les apprenants. En condition de construction la qualité des cartes s'est, en effet, avérée très faible (i.e. les apprenants ont construit peu de liens justes) ; en condition de consultation, le temps de consultation moyen était relativement court (1 minute et 25 secondes)<sup>1</sup>. Aussi, les apprenants n'ont probablement pas utilisé la carte correctement, et n'ont peut-être pas fait de lien entre l'usage de cet outil et l'évaluation de leur niveau de compréhension. Le manque d'entraînement à l'utilisation des cartes conceptuelles pourrait expliquer que les apprenants n'aient pas intégré les méthodes d'utilisation des cartes et les aient peu ou mal utilisées.

### 5.9.2 *L'entraînement à l'utilisation de la carte conceptuelle*

Une façon d'aider les apprenants à utiliser plus efficacement les cartes conceptuelles serait de les y entraîner. Dans la discussion de leur étude, [Stull et Mayer \(2007\)](#) supposaient qu'un entraînement à l'utilisation des cartes pourrait aider les apprenants à interpréter et/ou à construire des cartes et ainsi obtenir de meilleures performances. Les résultats de plusieurs travaux vont dans le sens de cette explication. Par exemple, [Hilbert et Renkl \(2008\)](#) ont démontré que les apprenants qui avaient déjà manipulé les cartes conceptuelles investissaient davantage d'effort dans la planification du processus de construction de cartes conceptuelles et construisaient des cartes plus exactes. Ils réalisaient également les processus nécessaires à l'apprentissage aidé de cartes conceptuelles (e.g. l'évaluation de la progression) avec plus de facilité et obtenaient ainsi de meilleurs résultats. Dans une autre étude, [Thiede et al. \(2010\)](#) avaient proposé huit cours de cinquante minutes aux apprenants avant de leur proposer des cartes à construire parallèlement à l'apprentissage. Enfin, lors d'une première étude où les apprenants étaient familiarisés, mais pas entraînés, à la construction de cartes, [Redford et al. \(2012\)](#) n'avaient pas obtenu d'effet significatif de la construction des cartes sur l'exactitude de l'auto-évaluation. Dans leur seconde étude, les auteurs avaient donné aux apprenants huit leçons de quinze minutes chacune avant d'utiliser les cartes. Ils ont, alors, obtenu un effet positif et significatif de la construction de cartes conceptuelles sur l'exactitude de l'auto-évaluation.

Finalement, afin d'augmenter l'efficacité des cartes conceptuelles en évitant la surcharge cognitive de la construction et en aidant l'apprenant à

---

1. Ce temps correspond au temps de consultation moyen des cartes uniquement pour les apprenants qui ont consulté les cartes après l'apprentissage. En effet, le temps moyen de consultation pour la condition de consultation simultanée intègre le temps de visualisation des vidéos et, par conséquent, ne reflète pas le temps de consultation de la carte à proprement parlé.



s'engager dans la tâche, [Chang et al. \(2002\)](#) rappellent que la complétion de cartes conceptuelles peut être une alternative intéressante. Bien que les auteurs aient privilégié l'étude d'une autre méthode (i.e. la correction de cartes conceptuelles) que la complétion de cartes, [Van Loon et al. \(2014\)](#) ont démontré que l'activité, similaire, de complétion de diagramme améliorerait l'exactitude de l'auto-évaluation.

Dans l'étude suivante, l'effet de l'entraînement à l'utilisation des cartes sera examiné. Aussi, afin de pallier les inconvénients de la construction et de la consultation de cartes conceptuelles, nous opterons pour une activité de complétion de cartes conceptuelles. L'objectif sera d'étudier les effets de la durée d'un entraînement à l'utilisation des cartes conceptuelles sur l'efficacité des cartes en tant qu'outil d'aide à l'auto-évaluation.

Nous avons vu que :

- Contrairement à ce qui a été montré dans de précédentes études, nos résultats ne confirment pas l'efficacité des cartes sur l'amélioration de l'exactitude de l'auto-évaluation et de l'apprentissage.
- Les apprenants ayant davantage de connaissances antérieures s'auto-évaluent avec plus d'exactitude que les apprenants avec peu de connaissances antérieures.
- L'utilisation différée de la carte conceptuelle tend à diminuer la charge mentale ressentie associée à la compréhension du contenu d'apprentissage. Cela n'a toutefois pas d'effet sur la qualité de l'apprentissage ou sur l'exactitude de l'auto-évaluation.
- La consultation de cartes conceptuelles tend à être plus efficace que la construction de cartes conceptuelles sur l'exactitude de l'auto-évaluation et l'apprentissage.

# 6

---

## UN ENTRAÎNEMENT À L'UTILISATION DES CARTES CONCEPTUELLES POUR AMÉLIORER L'EXACTITUDE DE L'AUTO-ÉVALUATION

---

L'objectif principal de la seconde étude était de tester l'efficacité des cartes conceptuelles sur l'amélioration de l'exactitude de l'auto-évaluation et l'apprentissage. Plus spécifiquement, nous nous sommes intéressés à l'efficacité du mode de présentation des cartes (construction/consultation) selon le niveau d'expertise de l'apprenant (expérimentés/novices). Nos résultats ont, d'un côté, montré que la condition de consultation de cartes conduisait les participants à moins se surévaluer et à obtenir de meilleures performances que la condition de construction de cartes ; d'un autre côté, il est apparu que les experts étaient plus exacts et plus performants que les novices quel que soit le mode d'utilisation des cartes. Nos résultats ont, cependant, également mis en évidence que les apprenants qui n'utilisaient pas de cartes se surévaluaient moins et avaient des performances d'apprentissage plus élevées que les apprenants aidés de cartes. Ce dernier résultat ne semble pas pouvoir s'expliquer en termes de surcharge mentale associée à la compréhension des vidéos d'apprentissage. En effet, aucune différence n'a été constatée sur cette variable entre les participants ayant bénéficié d'une carte conceptuelle et les participants du groupe contrôle n'en ayant pas bénéficié. En revanche, il est possible que l'absence d'effet bénéfique des cartes sur l'exactitude de l'auto-évaluation soit due à un traitement cognitif de surface des cartes par les apprenants qui les ont utilisées. Ce niveau de traitement n'aurait pas permis aux participants d'accéder à un apprentissage profond afin qu'ils en extraient des indices plus pertinents pour l'auto-évaluation. Ce traitement cognitif de surface est au moins attesté par le faible nombre de liens corrects produits par les participants en condition de construction, quel que soit leur niveau d'expertise. De ce fait, les cartes conceptuelles n'ont probablement pas été utilisées comme un outil d'aide à l'auto-évaluation, ni comme un outil d'aide à l'apprentissage.

Le traitement cognitif des cartes à un niveau superficiel pourrait s'expliquer par un manque d'expérience dans l'utilisation des cartes de la part de nos apprenants. En effet, l'expérience précédente intégrait une

présentation de l'utilité des cartes conceptuelles pour l'apprentissage et l'auto-évaluation, suivie d'une familiarisation au dispositif de construction des cartes d'une durée d'une dizaine de minutes. L'observation d'un bénéfice des cartes conceptuelles sur l'exactitude de l'auto-évaluation nécessiterait un entraînement plus long que celui dont bénéficiaient les apprenants de l'étude 2. Par ailleurs, dans la deuxième étude, aucun feedback n'était donné à l'apprenant sur l'exactitude des cartes construites. Cette absence de feedback a pu contribuer au fait que, lors de l'auto-évaluation de sa réussite future à l'exercice d'application, l'apprenant n'utilise pas les informations issues de la carte construite ou qu'il utilise des informations erronées.

L'objectif de cette troisième étude est d'améliorer l'efficacité des cartes conceptuelles sur l'apprentissage et l'exactitude de l'auto-évaluation. Pour cela, deux moyens seront mis en œuvre. Le premier consistera à entraîner les apprenants à utiliser des cartes. Cet entraînement nous permettra de vérifier si la durée de l'entraînement à l'utilisation de cartes conceptuelles influence leur efficacité en tant qu'outil d'aide à l'exactitude de l'auto-évaluation. Le second moyen mis en œuvre sera de proposer une tâche de complétion de cartes, qui aurait le double avantage d'être moins coûteuse que la construction mais plus engageante que la consultation (Chang et al., 2002).

Dans cette introduction théorique, nous présenterons, les conditions nécessaires à la réalisation d'un entraînement à l'utilisation des cartes conceptuelles. Nous discuterons ainsi de la durée d'entraînement, de la présence de feedback sur la justesse des cartes et de l'activité de complétion de cartes.

## 6.1 ENTRAÎNEMENT À L'UTILISATION DES CARTES CONCEPTUELLES

Les cartes conceptuelles sont des outils complexes dont l'utilisation n'est pas simple, principalement pour des apprenants novices (Álvarez-Montero, Sáenz-Pérez, & Vaquero-Sánchez, 2015 ; Chang et al., 2001 ; Weinerth, Koenig, Brunner, & Martin, 2014). L'utilisateur doit non seulement comprendre la façon dont fonctionne le dispositif informatique permettant de créer la carte ; mais également comment générer une carte pour représenter au mieux l'information (Weinerth et al., 2014). Selon McCagg et Dansereau (1991), un manque de familiarité avec la technique d'utilisation des cartes conceptuelles peut créer une frustration chez les apprenants. De même, pour que les cartes soient utilisées, il est important que l'apprenant comprenne le fonctionnement et l'utilité des cartes (Santhanam, Leach, & Dawson, 1998). Ainsi, un entraînement à l'utilisation des cartes permettrait à l'apprenant de maîtriser ces tâches avant de les associer à une tâche d'apprentissage.

### 6.1.1 *La durée de l'entraînement*

Dans les études sur l'effet des cartes conceptuelles en situation d'apprentissage, la durée de l'entraînement à l'utilisation des cartes, lorsqu'elle est rapportée, varie entre 10 et 140 minutes (Weinerth et al., 2014). Toutefois, les études qui comparent les effets d'entraînements de différentes durées sur l'efficacité des cartes pour l'apprentissage et l'auto-évaluation sont rares.

Ainsi, Hilbert et Renkl (2008) ont mis en évidence que les apprenants qui avaient déjà manipulé les cartes conceptuelles étaient plus investis dans le processus de construction de cartes conceptuelles et produisaient des cartes plus exactes. Les apprenants entraînés réalisaient également plus facilement les processus nécessaires à l'apprentissage aidé de cartes conceptuelles, telle que l'évaluation de la progression. Les apprenants obtenaient ainsi de meilleures performances. Dans une autre étude, Scevak, Moore, et Kirby (1993) ont comparé les performances d'apprentissage de participants entraînés à l'utilisation de cartes conceptuelles aux performances d'apprenants non entraînés. Dans leur étude, les apprenants entraînés suivaient trois sessions de quarante minutes chacune. Lors des deux premières sessions, réalisées le même jour, les apprenants voyaient comment organiser les informations sur une carte. Ils lisaient ensuite un texte et devaient placer avec l'aide d'un enseignant les concepts importants sur une carte. La troisième session avait lieu le lendemain et les apprenants utilisaient la carte de la veille pour rappeler un maximum d'informations en lien avec le texte. Les apprenants étaient également incités à réfléchir à l'utilité des cartes dans le rappel d'information. Les apprenants non entraînés lisaient le même texte mais rédigeaient une dissertation au lieu de construire une carte. Trois semaines après l'entraînement l'ensemble des participants lisaient un texte nouveau accompagné d'une carte conceptuelle à consulter. Tous avaient pour consigne d'utiliser la carte. Les résultats ont montré que les performances de rappel des apprenants entraînés à l'utilisation des cartes étaient supérieures à celles des apprenants non entraînés. Bien que leurs résultats aient été concluants, les auteurs pensent qu'un entraînement plus long aurait pu augmenter le bénéfice des cartes.

En lien avec l'exactitude de l'auto-évaluation, les résultats des études de Redford et al. (2012) mettent également en exergue l'importance d'un entraînement à l'utilisation des cartes. Les auteurs n'ont pas observé d'effet significatif de la condition construction des cartes comparativement à une condition de relecture lorsque les participants de leur étude ne recevaient qu'un entraînement court (15 minutes) à l'utilisation des cartes. Dans une seconde expérience, l'entraînement à l'utilisation des cartes conceptuelles s'étendait sur huit semaines à raison d'une séance de quinze minutes par semaine. Les auteurs ont alors observé un effet significatif de la construc-

tion de carte sur l'exactitude de l'auto-évaluation comparativement à une condition contrôle (i.e. sans utilisation de cartes). Auparavant, [Thiede et al. \(2010\)](#) avaient également obtenu un effet significatif des cartes sur l'exactitude de l'auto-évaluation comparativement à une condition contrôle dans laquelle les apprenants relisaient uniquement le texte. Dans leur étude, les apprenants n'utilisaient les cartes pour la situation d'apprentissage qu'après avoir suivi huit séances d'entraînement de cinquante minutes chacune.

Par ailleurs, les travaux sur l'expertise ont montré que les individus sont capables d'automatiser des procédures lorsque celles-ci sont répétées de nombreuses fois (pour une synthèse sur les compétences des experts, voir [Chi, 2006](#)). Ainsi, la répétition d'une activité permet d'en diminuer le coût cognitif ([Leplat, 2005](#)). Il est donc possible que l'entraînement à l'utilisation de carte conceptuelle permette de diminuer la quantité de ressources attentionnelles allouée à cette tâche. Les apprenants pourraient alors consacrer ces ressources aux activités d'apprentissage et d'auto-évaluation.

#### 6.1.2 *Feedback dans l'utilisation des cartes conceptuelles*

Des auteurs soulignent l'importance du feedback fourni à l'apprenant sur la qualité de la carte qu'il est en train de construire, notamment chez les novices (e.g., [Álvarez-Montero et al., 2015](#) ; [Chang et al., 2001](#)). Un feedback permet d'aider les apprenants à réviser et à compléter les cartes conceptuelles. Une absence de feedback sur l'exactitude des concepts placés, des relations entre les concepts au cours de la construction peut conduire un apprenant novice à une frustration et à une démotivation ([Chang et al., 2001](#)). Les feedback peuvent être proposés à différentes étapes de l'apprentissage mais certains auteurs recommandent l'usage de feedback immédiats dans l'utilisation de cartes conceptuelles (i.e. au moment où l'apprenant commet une erreur, e.g., [Shute, 2008](#)). [Álvarez-Montero et al. \(2015\)](#) rappellent qu'utiliser ce type de feedback pendant la construction ou la complétion de cartes a deux avantages : il est particulièrement efficace pour les novices ([Shute, 2008](#)) ; en signalant les erreurs au fur et à mesure que l'apprenant les produit, il évite que les apprenants ne construisent une carte complètement fautive ([Cimolino, Kay, & Miller, 2003](#)). Par ailleurs, il semble également que présenter des aides à la complétion (ex : définitions de concepts) lors de l'apprentissage à la construction de cartes permette à l'apprenant de développer une pensée auto-réflexive ([Collins & Brown, 1988](#)).

## 6.2 LA COMPLÉTION DE CARTES CONCEPTUELLES

Dans l'optique de diminuer la charge cognitive associée à l'utilisation des cartes conceptuelles, il est également possible de proposer une alternative à la construction de cartes. D'après [Chang et al. \(2002\)](#) ou [Stull et Mayer \(2007\)](#), l'activité de construction peut provoquer une charge mentale superflue, néfaste pour l'apprentissage, surtout pour les novices. Cependant, la consultation n'engage pas concrètement l'apprenant qui peut opter pour de la passivité et ne pas utiliser les processus adaptés à un traitement profond de l'information présentée. De même, lors de la consultation de cartes conceptuelles, les apprenants tendraient à se focaliser sur les concepts plutôt que sur les relations entre ces concepts ([McCagg & Danseureau, 1991](#)). La complétion de cartes permettrait de palier l'ensemble de ces problèmes. La complétion de cartes conceptuelles consiste à proposer à l'apprenant une carte incomplète pré-construite par un expert. Certaines parties peuvent être laissées vides ou le squelette de la carte peut-être présenté avec la liste des concepts à placer ([Chang et al., 2002](#)). Une tâche de complétion apparaît comme une solution intermédiaire adéquate qui engage l'apprenant dans la tâche et favorise l'utilisation de charge essentielle en limitant la charge superflue ([Paas, 1992](#)). L'efficacité de la complétion de cartes conceptuelles a notamment été démontrée sur l'apprentissage de textes de biologie ([Chang et al., 2001](#)) et l'amélioration de l'exactitude de l'auto-évaluation en association avec des textes traitant de relations de causalité ([Van Loon et al., 2014](#)).

## 6.3 SYNTHÈSE ET HYPOTHÈSES

Sur la base de la revue de la littérature présentée ci-dessus, nous faisons l'hypothèse qu'un entraînement à l'utilisation des cartes conceptuelles permettrait d'améliorer l'exactitude de l'auto-évaluation et les performances d'apprentissage en diminuant la charge cognitive associée à l'utilisation des cartes lors de l'apprentissage d'un nouveau contenu. Dit autrement, les apprenants entraînés à l'utilisation des cartes conceptuelles devraient ressentir moins de charge mentale pour utiliser les cartes conceptuelles que les apprenants peu entraînés. Aussi, face à un contenu nouveau d'apprentissage, les apprenants entraînés à l'utilisation des cartes conceptuelles devraient investir davantage de ressources dans le traitement du contenu proposé par les cartes conceptuelles. En accordant plus de ressources au traitement de l'information, les apprenants entraînés devraient voir leurs performances d'apprentissage ainsi que l'exactitude de leur auto-évaluation s'améliorer comparativement à celles d'individus peu entraînés ainsi qu'à celles d'individus n'ayant pas utilisé de cartes. Par ailleurs, dans l'étude précédente, nous avons également vu que la présentation

différée des cartes conceptuelles permettait de diminuer la charge mentale ressentie pour comprendre les vidéos. Aussi, dans cette étude, les cartes conceptuelles seront présentées systématiquement après le visionnage des vidéos. Par conséquent, nous n'attendons aucune différence sur la charge mentale ressentie pour comprendre les vidéos entre les différents groupes.

Enfin, à titre exploratoire, nous analyserons également le temps de complétion des cartes conceptuelles. Nous nous intéresserons également à l'évolution de l'exactitude de l'auto-évaluation et des performances d'apprentissage afin de déterminer l'effet de la durée de l'entraînement sur celles-ci.

## 6.4 MÉTHODE

### 6.4.1 *Participants et plan d'expérience*

Les quarante-sept participants recrutés sur la base du volontariat (Âge moyen = 22.84,  $SD = 7.8$ ; 37 femmes, 10 hommes) étaient inscrits en Licence de sciences humaines et sociales dans diverses disciplines à l'Université Toulouse Jean Jaurès ainsi qu'au Certificat Informatique et Internet (C2I). Lors du premier semestre, un dispositif d'aide à la révision des enseignements du C2I basé sur la complétion de cartes conceptuelles a été mis à la disposition des étudiants et servait de dispositif d'entraînement à la complétion de cartes conceptuelles. Ce dispositif présentait, sur douze semaines, des cartes à compléter résumant les différentes parties du cours du C2I. La séance traitant de l'utilisation des fonctions avancées du traitement de texte, présentée dans la tâche d'apprentissage de notre étude, n'était pas abordée lors de ces séances mais était prévue lors du semestre suivant. Le choix de l'utilisation ou non de ce dispositif était laissé à l'apprenant.

Parmi notre échantillon de participants, les étudiants qui avaient réalisé au minimum cinq cartes conceptuelles à partir du dispositif d'entraînement étaient assignés au groupe 'entraîné' (les participants entraînés utilisaient une carte conceptuelle pendant l'apprentissage de la phase test,  $N = 17$ ). Les étudiants qui n'avaient pas utilisé le dispositif d'entraînement aux cartes conceptuelles étaient répartis aléatoirement dans le groupe 'peu entraîné' (les étudiants peu entraînés utilisaient une carte conceptuelle pendant l'apprentissage de la phase test,  $N = 13$ ) ou le groupe 'contrôle' (les étudiants peu entraînés n'utilisaient pas de cartes conceptuelles pendant l'apprentissage de la phase test,  $N = 17$ ).

#### 6.4.2 *Tâche et matériel*

La tâche principale des participants était similaire à celle de l'expérience précédente : après avoir visualisé une vidéo sur le thème de la mise en forme automatisée de document grâce à l'utilisation des 'styles', les participants réalisaient le même exercice d'application que dans l'expérience précédente (cf. section 5.7.2). Rappelons que cet exercice consistait à mettre en forme un texte à l'aide des styles.

##### *La vidéo d'apprentissage*

Les vidéos d'apprentissage de la précédente expérience (cf. section 5.7.2) ont été fusionnées pour n'en constituer, qu'une seule de 15 minutes et 30 secondes. L'information délivrée était néanmoins la même qu'auparavant (i.e. comment modifier et appliquer des styles avec un logiciel de traitement de texte). Comme pour la précédente étude, il était impossible pour l'apprenant de revenir en arrière, de revisualiser la vidéo ou de consulter internet pendant toute la durée de la phase test. L'apprenant pouvait cependant effectuer des pauses quand il le souhaitait lors de la diffusion de la vidéo.

##### *Questionnaire de mesure des connaissances*

Les connaissances des apprenants ont été évaluées à deux reprises à l'aide d'un seul et même questionnaire. Ce questionnaire était composé de dix questions à choix multiples (cf. Annexe E), extraites du Certificat Informatique et Internet (C2I niveau 1). Chaque question était accompagnée de quatre propositions de réponse dont une seule était correcte. Le premier questionnaire visait à mesurer le niveau de connaissances antérieures de l'apprenant (QCM 1) ; le second permettait de déterminer la progression de l'apprentissage (QCM 2).

##### *Exercice d'application*

L'exercice d'application visant à mesurer l'apprentissage des participants était identique à celui de l'expérience précédente (cf. section 5.7.2). L'objectif était de mettre en forme un document textuel en utilisant les fonctions d'automatisation de la mise en forme de document, enseignées dans la vidéo. Cet exercice était précédé d'un exercice de familiarisation.

##### *Cartes conceptuelles et outil de complétion*

Un outil de complétion de carte a été créé spécifiquement pour cette étude par un chercheur du laboratoire Cognition, Langues, Langage, Ergonomie, maître de conférences en Informatique et enseignant du C2I. Cet



outil a été développé en Java sous la forme d'une application connectée à un serveur de cartes/QCM/Traces s'exécutant sur un serveur JEE. Ce dispositif s'intégrait dans la plateforme Moodle de l'université Toulouse Jean-Jaurès. Il présentait les cartes conceptuelles dans les phases d'entraînement et test, avec des modifications mineures entre les deux qui sont expliquées ci-dessous.

LES CARTES ET L'OUTIL DE COMPLÉTION DE CARTES DANS LA PHASE D'ENTRAÎNEMENT. Douze cartes de révision du contenu du C2I étaient proposées aux étudiants. Parmi ces douze cartes, la première avait pour objectif de familiariser le participant avec le dispositif tandis que les autres portaient sur divers thèmes traités lors des enseignements (e.g. le droit numérique, les licences informatiques, le téléchargement, l'identité numérique, la recherche d'information, etc.).

Le dispositif de complétion de cartes était composé de trois zones possédant chacune une fonction bien précise (cf. Figure 13). La zone principale présentait la carte conceptuelle pré-remplie. Dans certaines cases un concept était déjà affiché et toutes les cases étaient reliées par des traits explicitant les relations entre les concepts. Les cases à compléter étaient vides et signalées par un point d'interrogation au centre. Les concepts à placer étaient proposés au fur et à mesure à l'apprenant, mais celui-ci pouvait les faire défiler afin de les placer dans l'ordre qui l'accommodait. Un feedback sur l'exactitude du placement du concept était donné à l'apprenant dans une seconde zone de la carte conceptuelle. Cette seconde zone affichait les informations requises pour la bonne réalisation de l'exercice (consignes, feedback positifs ou négatifs, définitions éventuelles des concepts, etc.). Enfin, une troisième zone présentait trois types d'aides à la complétion, dans le but d'aider l'apprenant à utiliser les cartes et à apprendre les concepts. L'apprenant pouvait choisir d'utiliser ou non, une ou plusieurs aides successivement.

LES CARTES ET L'OUTIL DE COMPLÉTION DANS LA PHASE TEST. Le format de la carte conceptuelle de la phase test présentée dans la séquence d'apprentissage (cf. Figure 14) était identique à celui des cartes présentées en entraînement, excepté que celle-ci portait sur 'les styles' et qu'elle ne comportait pas l'aide dans la zone (3). Le participant pouvait quitter la carte sans l'avoir complétée entièrement, à partir du moment où il avait fait défiler au moins une fois l'ensemble des concepts à placer.

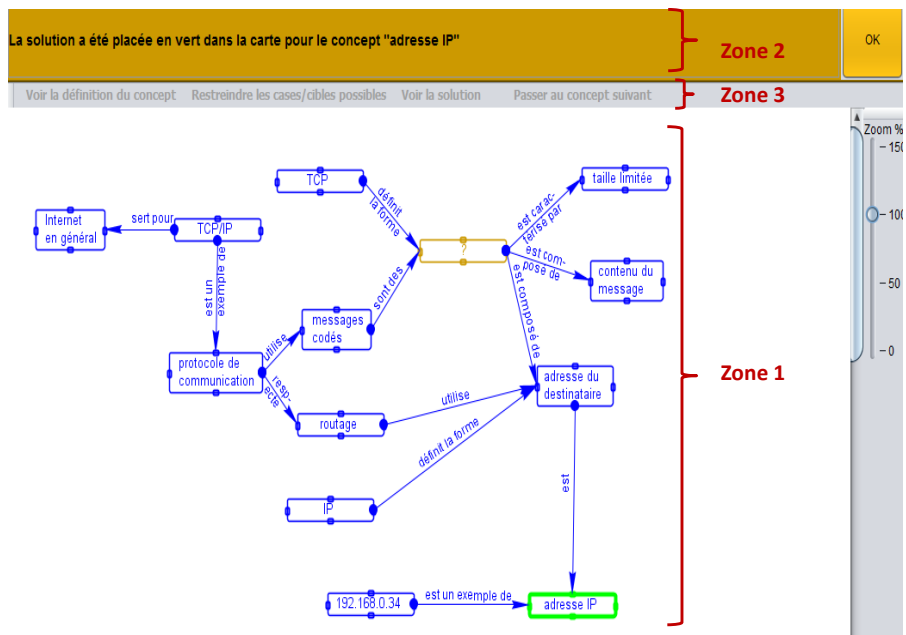


FIGURE 13. – Copie d’écran du dispositif de complétion de cartes conceptuelles tel que présenté dans la phase d’entraînement.

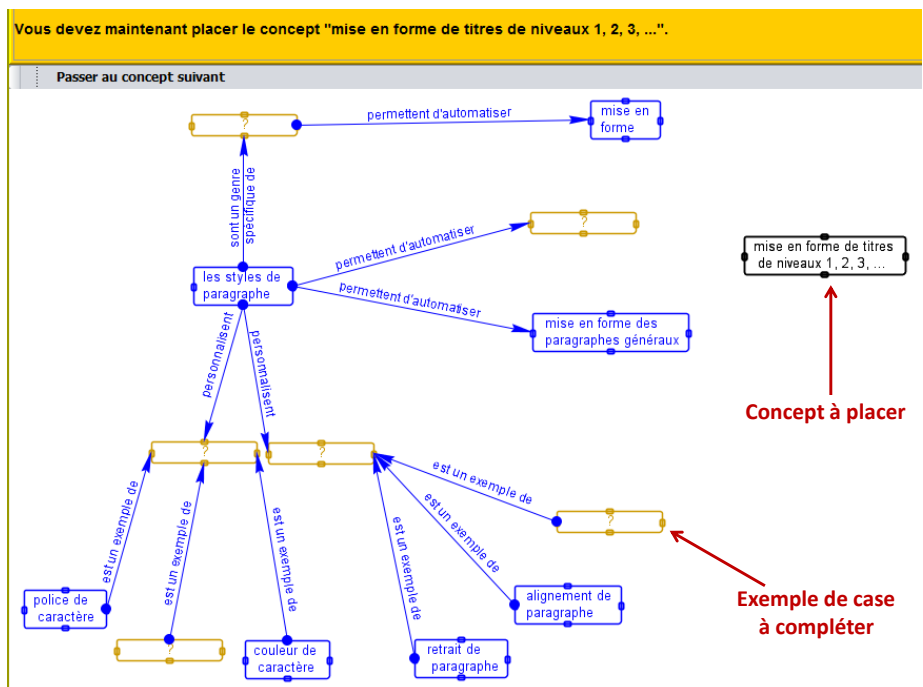


FIGURE 14. – Copie d’écran du dispositif de complétion de cartes conceptuelles tel que présenté dans la phase test.

### 6.4.3 Mesures

#### *Les performances*

Les performances ont été mesurées à trois reprises : une première fois à l'aide du questionnaire proposé avant l'apprentissage afin de contrôler le niveau des connaissances antérieures (QCM 1) et une seconde fois à l'aide du même questionnaire (QCM 2) après l'apprentissage afin de mesurer la progression d'apprentissage de l'apprenant. Les performances ont été mesurées une troisième fois, après la passation du QCM 2 à l'aide de l'exercice d'application.

**PERFORMANCES AUX QUESTIONNAIRES QCM 1 ET QCM 2** Les performances aux deux questionnaires étaient calculées de façon identique. Pour chaque question, la réponse juste valait 1 point, une réponse fautive valait 0. Le score total à chaque questionnaire variait donc entre 0 et 10 points. Ce score sur 10 était ensuite converti sur 100 afin d'être comparable au score de l'apprenant au jugement d'auto-évaluation.

**PERFORMANCES À L'EXERCICE D'APPLICATION** De la même façon que dans l'expérience précédente, la performance à l'exercice d'application était notée initialement sur quinze points, en correspondance aux quinze manipulations nécessaires à réaliser pour que l'exercice soit complètement réussi. Ce score était ensuite transformé sur 100 afin de qu'il puisse être soustrait au score du jugement d'auto-évaluation.

#### *La charge mentale*

La charge mentale, comme dans l'expérience précédente, a été évaluée à l'aide de deux indicateurs : la mesure de l'effort mental et de la difficulté ressentis. Ces deux indicateurs ont été mesurés sur une échelle allant de 1 à 9 et portent à la fois sur la compréhension de la vidéo d'apprentissage et la complétion de la carte conceptuelle (cf. section F).

#### *Mesures de l'auto-évaluation et de son exactitude*

L'auto-évaluation et son exactitude ont été mesurées à trois reprises : avant le QCM 1, avant le QCM 2 et avant l'exercice d'application.

**LE JUGEMENT D'AUTO-ÉVALUATION.** Concernant le jugement d'auto-évaluation aux QCM 1 et QCM 2, les participants devaient répondre, sur une échelle allant de 0 (je ne pourrais pas répondre correctement aux questions) à 100 (je pourrais répondre correctement aux questions), à la question : 'Dix questions, de type QCM, sur le traitement de texte vous

sont présentées ci-dessous. Après avoir lu ces questions et sachant que quatre propositions vous seront systématiquement proposées, estimez à quel point vous êtes sûr de pouvoir répondre correctement à ces questions.’

À l’exercice d’application, les participants devaient répondre sur une échelle allant de 0 (pas du tout sûr) à 100 (tout à fait sûr) à la question : ‘Indiquez à quel point vous êtes sûr(e) de pouvoir réaliser correctement cet exercice [...]?’

**LA DIRECTION DU BIAIS D’AUTO-ÉVALUATION.** Contrairement aux études précédentes et comme spécifié ci-dessus, pour les questionnaires à choix multiples, les apprenants n’avaient pas à réaliser un jugement pour chaque question mais uniquement un jugement global pour l’ensemble des questions. Par conséquent, la direction du biais d’auto-évaluation représentait la différence entre le score au jugement d’auto-évaluation sur 100 et la performance sur 100. Ce score a été calculé pour les deux questionnaires à choix multiples (QCM 1 et QCM 2) ainsi que pour l’exercice d’application. Un score positif signifiait que les apprenants se sur-évaluaient, un score négatif signifiait que les apprenants se sous-évaluaient (Baars et al., 2013, 2014).

**L’AMPLITUDE DU BIAIS D’AUTO-ÉVALUATION.** L’amplitude du biais a également été mesurée comme dans l’expérience précédente en calculant la racine carrée de l’écart entre le jugement d’auto-évaluation (aux questionnaires et à l’exercice d’application) et la performance élevée au carré (aux questionnaires et à l’exercice d’application). Une note proche de 0 indiquait que l’auto-évaluation est exacte, une note proche de 100 reflétait un écart maximum entre la performance et le jugement d’auto-évaluation (Baars et al., 2013, 2014).

#### *Le temps de complétion des cartes conceptuelles*

La mesure du temps de complétion débutait au moment où les apprenants voyaient la carte s’afficher à l’écran et se terminait lorsque ces derniers cliquaient sur un bouton pour valider la fin de la tâche de complétion. Cette mesure était automatiquement et systématiquement enregistrée par le dispositif informatique.

#### 6.4.4 *Procédure*

Cette expérience s’est déroulée en deux grandes phases à savoir la phase d’entraînement et la phase test. Ces deux phases se déroulaient sur ordinateur et ont été conçues et développées par un maître de conférences en informatique, enseignant du C2I.

Ces phases n'ont pas toutes deux été réalisées par l'ensemble des participants. La phase d'entraînement à l'utilisation des cartes conceptuelles était spécifique au groupe 'entraîné' et aucun des participants des deux autres groupes ne l'a réalisée. Les participants 'entraînés' réalisaient dans un premier temps l'entraînement à distance sachant qu'en moyenne les cartes conceptuelles leur étaient proposées au rythme d'une par semaine, après que la leçon correspondante ait été traitée en classe. Toutefois, les cartes étaient ensuite à leur disposition et les apprenants pouvaient aussi bien réaliser l'entraînement au fur et à mesure de la mise à disposition des cartes qu'à la fin du semestre en réalisant toutes les cartes dans un temps restreint.

La phase d'entraînement était précédée d'une présentation des cartes conceptuelles (définition, exemple et utilité pour l'apprentissage) ainsi que d'une vidéo expliquant à l'apprenant comment utiliser le dispositif de complétion de cartes conceptuelles. Lors de la phase d'entraînement, les apprenants devaient compléter des cartes conceptuelles en lien avec les leçons du C2I abordées lors des douze séances d'enseignement.

La phase test débutait par un formulaire de consentement et un questionnaire démographique. Dans un second temps, les connaissances initiales des participants ainsi que l'auto-évaluations de leur performance globale étaient mesurées grâce au QCM 1. Dans un troisième temps, les participants étaient familiarisés à l'outil de complétion de cartes conceptuelles. L'étape de familiarisation présentait une vidéo explicative de l'outil de complétion de cartes (différente pour les participants entraînés et peu entraînés) ainsi qu'un exercice pratique (identique pour tous les participants) permettant aux participants de manipuler cet outil. La vidéo présentée aux participants entraînés ne récapitulait que les modifications apportées à l'outil de complétion des cartes entre la phase d'entraînement et la phase test, à savoir le retrait des aides à la complétion et la possibilité de quitter la carte sans avoir terminé la complétion. Pour les participants peu entraînés et ceux du groupe contrôle, les cartes conceptuelles leur été présentées (définition, exemple et utilité pour l'apprentissage), suivi de la vidéo expliquant à l'apprenant comment utiliser le dispositif de complétion de cartes conceptuelles (vidéo donnée aux apprenants entraînés avant le début de l'entraînement). L'ensemble des participants réalisaient ensuite un exercice pratique qui consistait à compléter une carte conceptuelle. Cet exercice de familiarisation portait sur des concepts sans liens avec ceux de la séquence d'apprentissage. Les participants visionnaient par la suite la vidéo d'apprentissage sur l'utilisation des styles. Les participants des groupes 'entraîné' et 'peu entraîné' complétaient la carte conceptuelle portant sur les styles après avoir visionné la vidéo puis estimaient l'effort mental et la difficulté ressentis pour comprendre la vidéo et compléter les cartes. Les participants du groupe contrôle évaluaient seulement l'effort

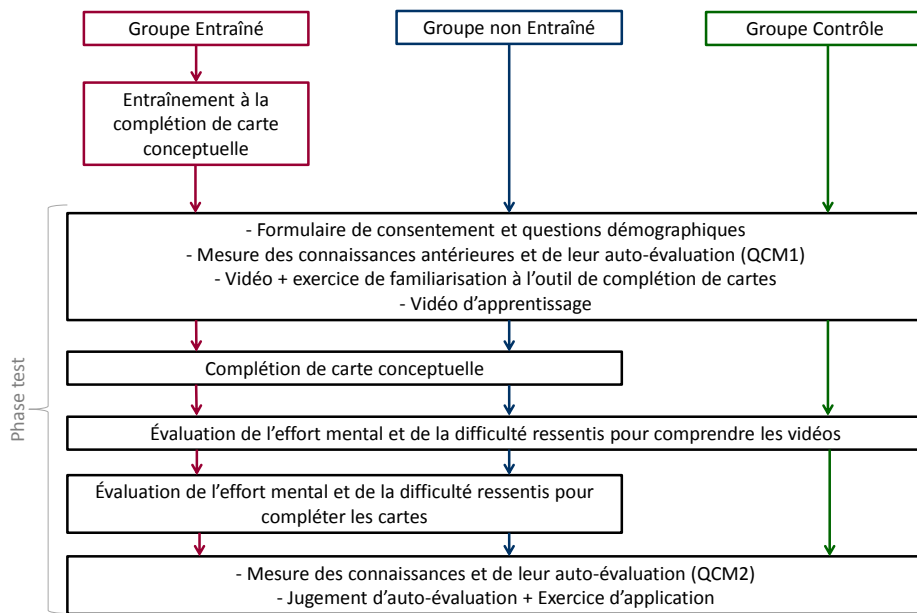


FIGURE 15. – Synthèse de la procédure de l'étude 3

mental et la difficulté ressentis pour comprendre la vidéo. Finalement, tous les participants ré-évaluaient et répondaient de nouveau au questionnaire (QCM 2) puis réalisaient l'exercice d'application, précédé d'une mesure de l'auto-évaluation.

Une schématisation du déroulement de cette expérience est représentée dans la Figure 15.

## 6.5 RÉSULTATS

L'objectif principal de cette dernière étude est de vérifier qu'un entraînement à l'utilisation des cartes conceptuelles permet de réduire la charge associée à l'utilisation des cartes conceptuelles et d'améliorer l'apprentissage et l'exactitude de l'auto-évaluation. L'analyse des données consistera donc, dans un premier temps, à réaliser des Anovas à un facteur visant à comparer les trois groupes de participants sur les variables dépendantes de cette étude. Les variables dépendantes analysées étaient : les trois mesures de la performance (i.e., les performances au QCM 1, au QCM 2 et à l'exercice d'application) ; les mesures du jugement d'auto-évaluation, de la direction du biais ainsi que l'amplitude du biais d'auto-évaluation associés aux trois mesures de performances citées ci-dessus ; les mesures de l'effort mental et de la difficulté ressentis pour comprendre les vidéos. Un t-test à mesures indépendantes sera employé pour les mesures de l'effort mental et de la difficulté associés à l'utilisation des cartes car seuls les groupes 'entraînés' et 'peu entraînés' sont concernés par ces mesures. En complé-

ments de ces analyses, des anovas mixtes  $3 \times 2$  ont été conduites sur les performances aux QCM 1 et QCM 2, les jugements d'auto-évaluation, la direction et l'amplitude des biais d'auto-évaluation aux QCM 1 et QCM 2. La variable indépendante en intra-sujet était le moment de réalisation du QCM (modalités : QCM 1 ; QCM 2). Le groupe expérimental constituait la variable indépendante en inter-sujet (modalités : 'entraîné', 'peu entraîné', 'contrôle'). Ces analyses visaient à détecter une évolution des performances d'apprentissage, du jugement d'auto-évaluation et de son exactitude selon le groupe expérimental. Le temps de complétion a finalement été comparé entre nos deux groupes de participants à l'aide d'un t-test à mesures indépendantes. Les mesures concernant la qualité de la carte n'ont pas été rapportées car l'intégralité des participants ont correctement placé les six concepts manquants.

Le niveau de connaissances initiales des participants des trois groupes expérimentaux a été contrôlé à l'aide d'une Anova à un facteur conduite sur la performance au QCM 1. Les résultats n'ont pas révélé de différence significative selon le groupe expérimental,  $F(2, 41) = 2.21, p = .1$ . Les participants du groupe entraîné ont un niveau de connaissances initiales comparable ( $M = 57.86, SD = 17.62$ ) à celui des participants 'peu entraînés' ( $M = 49.23, SD = 14.98$ ) ainsi qu'à celui des participants du groupe 'contrôle' ( $M = 46.47, SD = 13.67$ ).

### 6.5.1 *Statistiques descriptives*

Les statistiques descriptives, présentées dans le Tableau 8, se basent sur les moyennes et écarts-types de chaque variable dépendante. Ces moyennes indiquent que :

- Globalement, le jugement d'auto-évaluation au QCM 1 est moyen ( $M = 41$ ) et relativement équivalent pour les trois groupes de participants. Les participants du groupe 'peu entraînés' ( $M = 42.2$ ) ont émis un jugement proche de celui des participants du groupe 'entraîné' ( $M = 42.1$ ) et supérieur à celui des participants du groupe 'contrôle' ( $M = 39.1$ ). La moyenne globale du jugement d'auto-évaluation augmente pour le QCM 2 ( $M = 59.3$ ) mais reste semblable pour les trois groupes de participants. À nouveau, les participants du groupe 'peu entraîné' ( $M = 59.5$ ) ont émis un jugement proche de celui des participants du groupe 'entraîné' ( $M = 59.4$ ) et supérieur à celui des participants du groupe 'contrôle' ( $M = 58.9$ ). Le jugement émis à l'exercice d'application est globalement plus élevé ( $M = 64,11$ ) que ceux émis aux questionnaires (QCM 1 et QCM 2). Les participants du groupe 'contrôle' ( $M = 72$ ) ont émis un jugement d'auto-évaluation plus élevé que les participants des groupes 'peu entraîné' ( $M = 66.9$ ) et 'entraîné' ( $M = 61.6$ ).

- La moyenne totale de la direction du biais d'auto-évaluation au QCM 1 indique que l'ensemble des participants tendaient à légèrement sousestimer leurs performances ( $M = -9.9$ ). Les participants entraînés à l'utilisation des cartes ( $M = -15.7$ ) se sousestimaient davantage que les participants des groupes 'contrôle' ( $M = -7.4$ ) et 'peu entraîné' ( $M = -7$ ). La moyenne totale de la direction du biais au QCM 2, diminue comparativement à celle du QCM 1 ( $M = -12.8$ ). Les participants 'entraîné' ( $M = -17.7$ ) se sousestimaient à nouveau davantage, que les participants des groupes 'peu entraîné' ( $M = -11.3$ ) et 'contrôle' ( $M = -9.9$ ). En revanche en lien avec l'exercice d'application, les participants du groupe 'entraîné' étaient globalement exacts ( $M = 1.1$ ), les participants du groupe 'peu entraîné' tendaient à se sousestimer légèrement ( $M = -12.1$ ) tandis que les participants du groupe 'contrôle' tendaient à se sur-évaluer ( $M = 19.1$ ).
- L'amplitude moyenne des biais d'auto-évaluation au QCM 1 était plutôt basse ( $M = 24.85$ ). Les participants des groupes 'entraîné' ( $M = 22.9$ ) et 'contrôle' ( $M = 21.3$ ) tendaient à être plus exacts que les participants du groupe 'peu entraîné' ( $M = 27$ ). L'amplitude moyenne au QCM 2 ( $M = 19.07$ ) est plus basse que l'amplitude moyenne au QCM 1, ce qui indique que les participants tendaient à être plus exacts. Ici, les participants du groupe 'peu entraîné' sont les plus exacts ( $M = 17.5$ ) comparativement aux participants des groupes 'entraîné' ( $M = 20.6$ ) ou 'contrôle' ( $M = 19.1$ ). Enfin, l'amplitude moyenne du biais à l'exercice d'application ( $M = 32.7$ ) est supérieure à l'amplitude des biais des questionnaires 1 et 2. Les apprenants des groupes 'entraîné' ( $M = 42.6$ ) et 'contrôle' ( $M = 39.6$ ) ont obtenus des biais d'amplitude élevée comparativement aux participants du groupe 'peu entraîné' ( $M = 22$ ).
- Les performances au QCM 1 indique que les participants ont un niveau global de connaissances initiales moyen ( $M = 50.9$ ). Les participants du groupe 'entraîné' ( $M = 57.9$ ) ont obtenus des performances plus élevées que les participants des groupes 'peu entraîné' ( $M = 49.2$ ) 'contrôle' ( $M = 46.5$ ). L'ensemble des participants a progressé au QCM 2 ( $M = 72$ ). À nouveau, les participants du groupe 'entraîné' ( $M = 77.1$ ) ont obtenu des performances plus élevées que les participants des groupes 'peu entraîné' ( $M = 70.7$ ) et 'contrôle' ( $M = 68.8$ ). Enfin, concernant les performances à l'exercice d'application, les tendances diffèrent puisque les participants du groupe 'peu entraîné' ont obtenus des performances supérieures ( $M = 79$ ) à celles des participants des groupes 'entraîné' ( $M = 60.5$ ) et 'contrôle' ( $M = 52.6$ ).



- L'effort mental ressenti globalement pour comprendre la vidéo est moyen ( $M = 4$ ). Les tendances indiquent que les participants du groupe 'peu entraîné' ( $M = 3.62$ ) ont ressenti un effort mental un peu plus faible pour comprendre les vidéos comparativement aux participants des groupes 'entraîné' ( $M = 4.36$ ) et 'contrôle' ( $M = 4.53$ ).
- Les tendances sur la difficulté ressentie pour comprendre la vidéo sont similaires à celles obtenues sur l'effort mental. La moyenne reste globalement plutôt basse ( $M = 3.37$ ) et les participants du groupe 'entraîné' tendaient à trouver la vidéos un peu plus difficile ( $M = 3.64$ ) que les participants des groupes 'peu entraîné' ( $M = 3.08$ ) et 'contrôle' ( $M = 3.12$ ).
- L'effort mental moyen ressenti pour utiliser les cartes est relativement faible ( $M = 3.59$ ) et équivalent pour les participants des groupes 'entraîné' ( $M = 3.79$ ) et 'peu entraîné' ( $M = 3.38$ ).
- De façon similaire, la difficulté ressentie pour utiliser les cartes est globalement faible ( $M = 3.7$ ). Les participants du groupe 'entraîné' ( $M = 3.57$ ) ont ressenti une difficulté comparable à celle des participants du groupe 'peu entraîné' ( $M = 3.85$ ) pour utiliser les cartes.

### 6.5.2 *Statistiques inférentielles : effet des cartes et de l'entraînement à l'utilisation des cartes sur les jugements d'auto-évaluation et leurs exactitudes*

#### *Le jugement d'auto-évaluation au QCM 1*

Une Anova à un facteur a été conduite sur le jugement d'auto-évaluation au QCM 1 avec le groupe expérimental en variable indépendante. Les résultats n'ont révélé aucun effet significatif du groupe,  $F(2, 41) = 1, p = .9, \eta_p^2 = .01$ . Le jugement d'auto-évaluation des participants du groupe 'peu entraîné' ( $M = 42.2, SD = 25.46$ ) est similaire à celui des participants du groupe 'entraîné' ( $M = 42.1, SD = 21.64$ ) ainsi qu'à celui des participants du groupe 'contrôle' ( $M = 39.06, SD = 22.23$ ).

#### *La direction du biais d'auto-évaluation au QCM 1*

Une Anova à un facteur, conduite sur la direction du biais d'auto-évaluation au QCM 1 n'a révélé aucun effet significatif du groupe expérimental,  $F(2, 41) = .46, p = .6, \eta_p^2 = .02$ . Les participants du groupe 'entraîné' ( $M = -15.7, SD = 22.69$ ) ne se sont pas significativement plus sous-évalués que les participants des groupes 'peu entraîné' ( $M = -7, SD = 33.15$ ) ou 'contrôle' ( $M = -7.4, SD = 26.15$ ).

TABLEAU 8. – Ensemble des moyennes (et écart-types) pour les variables dépendantes de cette étude. Le symbole \* indique les variables exprimées en pourcentage.

	Min ; Max	Entrainé (N = 17)	Peu entrainé (N = 13)	Contrôle (N = 17)	Total (N = 47)
<b>• Indicateurs de l'auto-évaluation et de son exactitude</b>					
Jugement d'auto-évaluation au QCM 1*	0 ; 100	42.1 (21.6)	42.2 (25.5)	39.1 (22.2)	41 (22.6)
Jugement d'auto-évaluation au QCM 2*	0 ; 100	59.4 (17.7)	59.5 (21.2)	58.9 (18.5)	59.3 (18.8)
Jugement d'auto-évaluation à l'exercice d'application*	0 ; 100	61.6 (17.6)	66.9 (26.1)	72 (15.7)	64.11 (22.12)
Direction du biais d'auto-évaluation au QCM 1	-100 ; 100	-15.7 (22.7)	-7 (33.2)	-7.4 (26.2)	-9.9 (27.1)
Direction du biais d'auto-évaluation au QCM 2	-100 ; 100	-17.7 (22.8)	-11.3 (18.8)	-9.9 (22.9)	-12.8 (21.5)
Direction du biais d'auto-évaluation à l'exercice d'application	-100 ; 100	1.1 (50.4)	-12.1 (29)	19.5 (45.2)	-5.3 (41.3)
Amplitude du biais d'auto-évaluation au QCM 1*	0 ; 100	22.9 (14.8)	27 (19)	21.3 (16.1)	24.85 (16.8)
Amplitude du biais d'auto-évaluation au QCM 2*	0 ; 100	20.6 (20)	17.5 (12.7)	19.1 (15.6)	19.07 (16.6)
Amplitude du biais d'auto-évaluation à l'exercice d'application*	0 ; 100	42.6 (24.3)	22 (21.8)	39.6 (28)	32.7 (25)
<b>• Indicateurs de la performance</b>					
Performances au QCM 1*	0 ; 100	57,9 (17,6)	49,2(15)	46,5 (13,7)	50,9 (15,8)
Performances au QCM 2*	0 ; 100	77,1 (12)	70,7 (11,2)	68,8 (16,2)	72 (13,7)
Performances à l'exercice d'application*	0 ; 100	60,5 (47,1)	79 (25,5)	52,6 (46,18)	69,4 (38,7)
<b>• Indicateurs de la charge mentale</b>					
Effort mental associé à la compréhension de la vidéo	1 ; 9	4,36 (1,82)	3,62 (1,94)	4,53 (1,97)	4 (1,9)
Difficulté associée à la compréhension de la vidéo	1 ; 9	3,64 (1,65)	3,08 (1,44)	3,12 (1,73)	3,37 (1,5)
Effort mental associé à l'utilisation de la carte	1 ; 9	3,79 (1,48)	3,38 (1,56)	X	3,59 (1,5)
Difficulté associée à l'utilisation de la carte	1 ; 9	3,57 (1,28)	3,85 (1,63)	X	3,70 (1,4)

### *L'amplitude du biais d'auto-évaluation QCM 1*

L'anova à un facteur conduite sur l'amplitude du biais d'auto-évaluation au QCM 1 n'a révélé aucun effet significatif du groupe expérimental,  $F(2, 41) = .45, p = .6, \eta_p^2 = .02$ . Les participants du groupe 'entraîné' (M = 22.86, SD = 14.77) ont eu un biais initial d'une amplitude comparable à celle des participants des groupes 'peu entraîné' (M = 27, SD = 19.02) et 'contrôle' (M = 21.29, SD = 16.14).

### *Le jugement d'auto-évaluation au QCM 2*

Une Anova à un facteur conduite sur le jugement d'auto-évaluation au QCM 2 n'a pas révélé d'effet significatif du groupe expérimental,  $F(2, 41) = .004, p = 1, \eta_p^2 = 0$ . Les participants du groupe 'peu entraîné' (M = 59.46, SD = 21.72) ont émis un jugement comparable à celui des participants des groupes 'contrôle' (M = 58.94, SD = 18.46) et 'entraîné' (M = 59.43, SD = 17.73).

### *La direction du biais d'auto-évaluation au QCM 2*

Une Anova à un facteur conduite sur la direction du biais d'auto-évaluation au QCM 2 n'a révélé aucun effet significatif du groupe expérimental,  $F(2, 41) = .54, p = .6, \eta_p^2 = .03$ . Les participants du groupe 'entraîné' (M = -17.71, SD = 22.76) ne se sont pas significativement plus sous-évalués que les participants des groupes 'peu entraîné', (M = -11.31, SD = 18.76) et 'contrôle' (M = -9.88, SD = 22.9).

### *L'amplitude du biais d'auto-évaluation au QCM 2*

L'anova réalisée sur l'amplitude du biais d'auto-évaluation au QCM 2 n'a révélé aucun effet significatif du groupe expérimental,  $F(2, 41) = .12, p = .9, \eta_p^2 = .01$ . Contrairement à ce qui était attendu, les participants des groupes 'contrôle' (M = 19.06, SD = 15.56), 'entraîné' (M = 20.57, SD = 20.01) et 'peu entraîné' (M = 17.46, SD = 12.66) se sont évalués avec une exactitude similaire.

### *Le jugement d'auto-évaluation à l'exercice d'application*

Une Anova à un facteur, conduite sur le jugement d'auto-évaluation à l'exercice d'application, n'a révélé aucun effet significatif du groupe expérimental,  $F(2, 44) = .85, p = .43, \eta_p^2 = .04$ . Le jugement d'auto-évaluation à l'exercice d'application des participants du groupe 'peu entraîné' (M = 66.85, SD = 26.61) était similaire celui des participants du groupe 'contrôle' (M = 72, SD = 15.73) ainsi qu'à celui des participants du groupe 'entraîné' (M = 63.06, SD = 17.96).

### *La direction du biais d'auto-évaluation à l'exercice d'application*

Une Anova à un facteur, conduite sur la direction du biais d'auto-évaluation à l'exercice d'application, n'a pas révélé d'effet significatif du groupe expérimental,  $F(2, 44) = 2.35, p = .11, \eta_p^2 = .1$ . Le biais moyen des participants du groupe 'entraîné' ( $M = -4, SD = 47.29$ ), 'peu entraîné' ( $M = -12.13, SD = 29.04$ ) et 'contrôle' ( $M = 19.45, SD = 45.24$ ) à l'exercice d'application était comparable.

### *L'amplitude du biais d'auto-évaluation à l'exercice d'application*

Une Anova à un facteur conduite sur l'amplitude du biais d'auto-évaluation à l'exercice d'application, n'a pas révélé de différence significative entre nos trois groupes expérimentaux,  $F(2, 44) = 2.41, p = .1, \eta_p^2 = .1$ . Contrairement à nos hypothèse, les participants des groupes 'peu entraîné' ( $M = 22.03, SD = 21.84$ ), 'contrôle', ( $M = 39.57, SD = 28.02$ ) et 'entraîné' ( $M = 40, SD = 23.52$ ) ont des biais d'amplitude comparable à l'exercice d'application.

### *6.5.3 Statistiques inférentielles : effet des cartes et de l'entraînement à l'utilisation des cartes sur les performances d'apprentissage*

L'anova à un facteur conduite sur les performances au QCM 1 a été présentée en introduction de la partie résultats. Cette analyse n'avait pas détecté de différences significatives entre nos trois groupes expérimentaux.

#### *Performances au QCM 2*

En contradiction avec notre hypothèse, l'anova à un facteur conduite sur les performances au QCM 2 n'a révélé aucun effet significatif du groupe expérimental,  $F(2, 41) = 1.52, p = .23, \eta_p^2 = .07$ . Les participants des groupes 'peu entraîné' ( $M = 70.77, SD = 11.15$ ), 'contrôle' ( $M = 68.82, SD = 16.16$ ) et 'entraîné' ( $M = 77.14, SD = 12.04$ ) ont obtenu des performances comparables au QCM 2.

#### *Performances à l'exercice d'application*

À nouveau, l'anova conduite sur les performances à l'exercice d'application n'a pas validée notre hypothèse. Aucun effet significatif du groupe expérimental n'a été détecté,  $F(2, 44) = 1.55, p = .22, \eta_p^2 = .07$ . Les participants des groupes 'contrôle' ( $M = 52.55, SD = 46.18$ ), 'peu entraîné' ( $M = 78.97, SD = 25.51$ ) et 'entraîné' ( $M = 67.06, SD = 44.9$ ) ont des performances comparables.

#### 6.5.4 *Statistiques inférentielles : effet des cartes et de l'entraînement à l'utilisation des cartes sur l'effort mental et la difficulté ressentis*

##### *pour la compréhension de la vidéo*

Une anova à un facteur conduite sur l'effort mental ressenti pour comprendre la vidéo n'a pas révélé d'effet significatif du groupe expérimental  $F(2, 41) = .9, p = .41, \eta_p^2 = .04$ . Les participants des groupes 'contrôle' (M = 4.53, SD = 1.97), 'peu entraîné' (M = 3.62, SD = 1.94) et 'entraîné' (M = 4.36, SD = 1.82) ont ressenti un effort semblable pour comprendre la vidéo d'apprentissage.

L'anova à un facteur, conduite sur la difficulté ressentie pour comprendre les vidéos, n'a pas révélé d'effet significatif du groupe expérimental  $F(2, 41) = .54, p = .59, \eta_p^2 = .03$ . Les participants des groupes 'contrôle' (M = 3.12, SD = 1.73), 'peu entraîné' (M = 3.08, SD = 1.44) et 'entraîné' (M = 3.64, SD = 1.65) ont tous trouvé la vidéo d'apprentissage aussi peu difficile.

##### *pour la complétion de la carte*

Les t-tests conduits sur l'effort mental,  $t(25) = -.69, p = .5, d = .27$ , et la difficulté,  $t(25) = .49, p = .6, d = .19$ , ressentis pour compléter les cartes n'ont pas révélé d'effet significatif du groupe expérimental. Contrairement à notre hypothèse, les participants du groupe 'entraîné' ont ressenti un effort mental (3.79, SD = 1.48) et une difficulté (3.57, SD = 1.28) comparable aux participants des groupes 'peu entraîné' (effort mental : M = 3.38, SD = 1.56 ; difficulté : M = 3.85, SD = 1.63).

#### 6.5.5 *Statistiques inférentielles : évolution de la performance, du jugement d'auto-évaluation et de son exactitude en fonction du groupe expérimental*

Les anovas mixtes conduites sur les performances, les jugements d'auto-évaluation, les directions et amplitudes des biais aux QCM 1 et QCM 2, sont présentées ci-après. Pour ces anovas, le groupe expérimental constituait la variable indépendante en inter-sujet, et le moment de réalisation du QCM, variable indépendante en intrasujet.

##### *Le jugement d'auto-évaluation*

L'anova conduite sur le jugement d'auto-évaluation n'a pas révélé d'interaction significative entre le moment de réalisation du QCM et le groupe expérimental,  $F(2, 41) = .07, p = .9, \eta_p^2 = .003$ . De même, aucun effet principal du groupe n'a été détecté,  $F(2, 41) = .05, p = .95, \eta_p^2 = .003$ .

En revanche, un effet du moment,  $F(1, 41) = 27.69, p < .001, \eta_p^2 = .4$ , indiquait que l'ensemble des participants ont émis un jugement de prédiction de réussite moins élevé pour le QCM 1 ( $M = 40.98, SD = 22.55$ ) que pour le QCM 2 ( $M = 59.25, SD = 18.80$ ).

#### *La direction du biais d'auto-évaluation*

L'anova conduite sur la direction du biais d'auto-évaluation n'a pas détecté d'interaction significative entre nos deux variables,  $F(2, 41) = .02, p = 1, \eta_p^2 = .001$ . De même aucun effet principal du moment,  $F(1, 41) = .38, p = .5, \eta_p^2 = .01$ , ou du groupe expérimental n'a été révélé,  $F(2, 41) = .81, p = .45, \eta_p^2 = .04$ .

#### *L'amplitude du biais d'auto-évaluation*

L'anova réalisée sur l'amplitude du biais d'auto-évaluation n'a pas révélé d'interaction significative,  $F(2, 41) = .45, p = .6, \eta_p^2 = .02$ , ni un effet principal du moment,  $F(1, 41) = 1.75, p = .2, \eta_p^2 = .04$ , ou du groupe expérimental,  $F(2, 41) = .13, p = .9, \eta_p^2 = .01$ .

#### *La performance*

L'anova mixte conduite sur la performance obtenue aux deux QCM n'a pas détectée d'interaction significative entre nos deux variables,  $F(2, 41) = .10, p = .9, \eta_p^2 = .01$ . En revanche, elle a révélé un effet significatif du moment de la mesure,  $F(1, 41) = 52.1, p < .001, \eta_p^2 = .6$  indiquant que l'ensemble des apprenants ont obtenu des performances plus élevées au QCM 2 ( $M = 72.05, SD = 13.74$ ) qu'au QCM 1 ( $M = 50.91, SD = 15.82$ ). L'anova a également détecté un effet significatif du groupe,  $F(2, 41) = 3.33, p = .05, \eta_p^2 = .1$ . Les tests post-hoc de Tukey révèlent une différence significative entre les participants des groupes 'entraîné' et 'contrôle',  $p = .04$ , suggérant que les participants du groupe 'entraîné' ( $M = 67.5$ ) ont obtenu sur l'ensemble des questionnaires des performances plus élevées que les participants du groupe 'contrôle' ( $M = 54.65$ ).

### 6.5.6 Mesures complémentaires

#### *Le temps de complétion*

Le temps de complétion des cartes conceptuelles a également été vérifié. Un t-test à mesures indépendantes a été conduit,  $t(14.9) = -.98, p = .3, d = .37$  et s'est révélé non significatif. Les participants du groupe 'entraîné' ( $M = 1509$  [2.50 minutes],  $SD = 1245$ ) ont mis autant de temps

que les participants ‘peu entraînés’ ( $M = 1171$  [1.95 minutes],  $SD = 328$ ) pour compléter les cartes.

## 6.6 DISCUSSION

### 6.6.1 *Rappel des hypothèses et synthèse des résultats*

L’objectif de cette étude était de déterminer si un entraînement à l’utilisation de cartes conceptuelles pouvait influencer leur efficacité sur l’exactitude de l’auto-évaluation et l’apprentissage. Nous supposions que les individus entraînés à l’utilisation de cartes ressentiraient moins de charge lors de l’utilisation d’une carte dans un contexte nouveau d’apprentissage. Cette diminution de la charge mentale devait faciliter le traitement du contenu des cartes pour non seulement, améliorer l’intégration des connaissances, mais également rendre saillants des indices plus pertinents pour l’auto-évaluation. En conséquence, l’exactitude de l’auto-évaluation ainsi que les performances d’apprentissage devaient s’améliorer. Pour tester cette hypothèse, la charge mentale ressentie pour utiliser les cartes, de même que les performances et l’exactitude de l’auto-évaluation de participants entraînés à utiliser des cartes conceptuelles ont été comparées à celles d’apprenants peu entraînés. Les analyses des résultats de cette étude n’ont pas permis de confirmer cette hypothèse. Aucune différence significative n’a émergé au niveau de la charge mentale ressentie pour utiliser les cartes entre les deux groupes de participants concernés. De même, les participants du groupe ‘entraîné’ à l’utilisation des cartes conceptuelles ne se sont pas révélés plus exacts dans leur auto-évaluation et n’ont pas obtenu de performances plus élevées que les participants des groupes ‘peu entraîné’ et ‘contrôle’. Ainsi, l’entraînement à l’utilisation des cartes conceptuelles n’a pas permis de diminuer la charge mentale associée à l’utilisation de ces cartes, et aucun effet significatif des cartes n’a été observé sur l’apprentissage et l’auto-évaluation. L’analyse des résultats a, néanmoins, permis d’observer les effets de l’apprentissage à travers l’amélioration des performances entre le QCM 1 et le QCM 2. La progression d’apprentissage s’est faite de façon équivalente pour les trois groupes. Aucune progression n’a été constatée sur l’exactitude des biais d’auto-évaluation.

Plusieurs facteurs peuvent expliquer les résultats obtenus dans cette étude. En premier lieu, la faible taille de l’échantillon a sûrement pu impacter la puissance statistique des tests réalisés. L’activité de complétion de cartes conceptuelles, moins complexe que l’activité de construction, ainsi que le contenu de l’entraînement proposé, différent de celui délivré dans de précédentes études, ont également pu limiter l’influence de l’entraînement. Ces trois points sont développés ci-après.

La faible taille de l'échantillon dans cette étude a pu occasionner une perte de la puissance statistique des analyses réalisées. En effet, comme le rappelle [Yergeau \(2009\)](#) 'la taille de l'échantillon a un effet direct sur la puissance statistique d'une étude', et 'cet élément [i.e. la taille de l'effectif] est celui qui a le plus d'impact sur le calcul de la puissance statistique.' (p.202). Lorsqu'un test statistique a une faible puissance, cela signifie qu'il est peu probable que l'hypothèse nulle soit rejetée, même quand elle est fausse. En d'autres termes, cela signifie qu'il y a peu de chance qu'un outil d'aide soit considéré comme efficace bien qu'il puisse l'être. Dans notre étude, la faible puissance des tests statistiques devrait, ainsi, nous conduire à ne pas nous prononcer sur l'efficacité réelle de l'entraînement à l'utilisation des cartes conceptuelles. Par ailleurs, les écarts-types sont relativement importants pour l'ensemble des groupes. Cette large amplitude des écarts-types indique l'existence d'une forte variabilité inter-individuelle. Malgré le fait que la puissance statistique de notre étude soit faible et que les résultats obtenus ne nous permettent pas de valider nos hypothèses, les statistiques descriptives montrent des tendances intéressantes. Ainsi, les participants du groupe 'peu entraîné' sont ceux qui voient l'amplitude du biais d'auto-évaluation diminuer le plus entre le QCM 1 ( $M = 27$ ) et le QCM 2 ( $M = 17.5$ ). Ce sont également eux qui ont obtenu une amplitude de biais la plus faible ( $M = 22$ ) et la performance la plus élevée ( $M = 79$ ) à l'exercice d'application. Il est, par ailleurs, intéressant de constater que les participants peu entraînés de cette étude ont une amplitude moyenne des biais inférieure à celles des participants expérimentés, qui ont utilisé une carte ( $M = 33.98$ ), dans l'étude précédente (étude 2). De même, la performance moyenne des participants peu entraînés de cette étude est supérieure à celle des participants expérimentés de l'étude précédente ( $M = 55.72$ ).

Par ailleurs, le temps de complétion des cartes indique que, tous les participants, qu'ils aient été entraînés ou non à compléter les cartes, ont complété relativement rapidement la carte proposée (en moyenne moins de 2 minutes et 30 secondes). De plus, l'intégralité des participants a correctement placé les six concepts manquants. La durée de l'entraînement ne semble, par conséquent, pas avoir eu d'effet sur l'utilisation de la carte. L'activité de complétion de cartes, ne nécessiterait pas un entraînement long. Cette explication est soutenue par les résultats de l'étude de [Van Loon et al. \(2014\)](#) qui ont proposé une courte familiarisation à la complétion de diagrammes, mais pas d'entraînement spécifique. Ils ont néanmoins observé que les apprenants qui avaient complété un diagramme étaient plus exacts dans l'évaluation qu'ils faisaient de leurs performances comparativement à des apprenants qui n'avaient pas eu de diagramme.

Finalement, l'entraînement proposé dans cette étude diffère de celui présenté dans l'étude de [Redford et al. \(2012\)](#) sur deux aspects. D'une



part, dans l'étude de Redford et al. (2012), les participants étaient non seulement entraînés à construire des cartes mais ils étaient également familiarisés avec la mesure de la performance finale. Lors des huit séances, les participants étaient, en effet, entraînés à répondre aux questions inférentielles de mesure de la compréhension. D'autre part, les participants étaient également entraînés à utiliser les cartes en association avec une tâche de compréhension de texte. Ainsi, la tâche associée aux cartes pendant l'entraînement était la même que celle proposée lors de la phase test. Au contraire, dans notre étude, les cartes proposées lors de l'entraînement étaient librement utilisées par les apprenants. Les étudiants pouvaient notamment les étudier en accompagnement d'un support de cours, ou les utiliser pour tester leurs connaissances. De plus, les cartes proposées en entraînement l'étaient après que le cours ait eu lieu, en tant qu'outils de révision. Par conséquent, aucun exercice pratique de mise en application des connaissances acquises ne suivait l'utilisation des cartes.

Nous avons vu que :

- Un entraînement long à la complétion de cartes conceptuelles n'a pas permis de diminuer la charge mentale associée à leur utilisation.
- Un entraînement long à la complétion de cartes conceptuelles n'a pas permis d'améliorer l'exactitude de l'auto-évaluation ou les performances d'apprentissage.
- La faible taille de notre échantillon ne nous permet pas de trancher quant à l'efficacité ou l'inefficacité de la durée de l'entraînement sur l'amélioration de l'exactitude de l'auto-évaluation et l'apprentissage.

---

## DISCUSSION ET PERSPECTIVES

---

Ce dernier chapitre offre une synthèse des résultats obtenus dans ce travail, afin d'en extraire des éléments qui feront progresser la compréhension du processus d'auto-évaluation, ainsi que la façon dont il est possible d'en améliorer l'exactitude.

Ce travail a été réalisé en collaboration avec les chercheurs de l'Institut de Recherche en Informatique de Toulouse et s'intégrait dans un projet appliqué, financé par l'Agence Nationale de la Recherche. Théoriquement, il se basait initialement sur le constat que l'auto-évaluation est un processus clé dans l'apprentissage autorégulé (e.g., [Nelson & Narens, 1990](#) ; [Winne & Hadwin, 1998](#)), qui est néanmoins difficile, et souvent mal réalisé par les apprenants (e.g., [Maki, 1998](#)). Plusieurs outils d'aide à l'auto-évaluation ont été testés afin d'aider l'apprenant dans la réalisation de cette activité métacognitive (e.g., [Griffin et al., 2008](#) ; [Rawson et al., 2000](#) ; [Thiede et al., 2005](#)). La plupart de ces outils ont, néanmoins, été testés dans des conditions très contrôlées et/ou sur des tâches d'apprentissage classiques, telle que la lecture de textes linéaires. Le développement des technologies conduit pourtant à une diversification des supports d'apprentissage et rend les processus métacognitifs d'autant plus importants ([Baccino et al., 2008](#)). L'objectif de notre recherche était, par conséquent, de tester divers outils d'aide à l'auto-évaluation en mettant l'accent sur la validité externe des études. Nos trois études se sont attachées à intégrer pleinement les dispositifs d'aide à l'auto-évaluation à la formation des apprenants. Nous nous sommes également intéressés aux conditions à réunir afin de proposer des outils d'aide à l'auto-évaluation efficaces. Des variables telles que le type de connaissance à évaluer, le mode de présentation de ces outils, le niveau d'expertise de l'apprenant, et la durée de l'entraînement à l'utilisation de l'outil d'aide à l'auto-évaluation, ont été intégrées. Les trois études conduites ont permis de tester trois dispositifs d'aide à l'auto-évaluation sur 446 élèves possédant différents niveaux d'expertise. Les contenus d'apprentissage et des outils de mesures des performances ont été conçus en collaboration avec des enseignants en informatique, ou étaient extraits directement de la base de données qui permet l'obtention de la certification du C2I. Enfin, les passations ont été réalisées dans

des situations les plus naturelles possible, à savoir, directement dans les salles de classes ou les salles informatiques habituellement utilisées par les élèves. Notre volonté de réaliser des études écologiques, toutefois, s'est confrontée à d'inévitables contraintes de terrain qui ont généré des limites à ce travail.

Dans une première partie de cette discussion, nous récapitulons les principaux résultats des trois études présentées dans ce travail de thèse. Nous proposons, dans une deuxième partie, des pistes explicatives à l'absence générale d'effet des outils d'aide à l'auto-évaluation. Nous discutons, finalement, des pistes de travail qu'il serait pertinent d'explorer afin d'améliorer la compréhension du processus d'auto-évaluation dans des contextes écologiques variés.

## 7.1 SYNTHÈSE DES PRINCIPAUX RÉSULTATS

L'objectif de la première étude était d'utiliser conjointement différents outils d'amélioration de l'exactitude de l'auto-évaluation. Les résultats n'ont pas révélé d'effet bénéfique du dispositif d'aide à l'auto-évaluation. Nous avons pu constater, néanmoins, que les participants s'auto-évaluaient avec plus d'exactitude sur des connaissances déclaratives que procédurales.

La seconde étude testait l'effet des cartes conceptuelles en tant qu'outil d'aide à l'auto-évaluation. L'objectif était de déterminer les conditions de présentation des cartes conceptuelles permettant aux apprenants d'améliorer l'exactitude de leur auto-évaluation et leurs performances d'apprentissage, en fonction de leur niveau d'expertise. Cette étude n'a pas répliqué les résultats obtenus dans de précédents travaux rapportés dans la littérature. La consultation de cartes s'est globalement révélée plus efficace que la construction pour l'amélioration de l'exactitude de l'auto-évaluation et des performances d'apprentissage. Les apprenants qui n'ont pas eu de cartes ont obtenu de meilleures performances que les apprenants aidés de cartes. Par ailleurs, l'expertise a eu un impact positif significatif sur l'apprentissage et l'auto-évaluation.

La troisième étude s'est attachée à tester l'hypothèse selon laquelle le manque d'entraînement à l'utilisation des cartes ait pu avoir un impact sur l'absence d'efficacité des cartes. En d'autres termes, une possibilité d'explication de cette absence d'effet considérerait que les apprenants, peu entraînés à l'utilisation des cartes, n'ont pas su les utiliser correctement. Les résultats observés n'ont pas montré de différence significative entre les apprenants peu entraînés et les apprenants entraînés en termes d'exactitude de l'auto-évaluation ou de performances d'apprentissage.

Il est important de retenir de ces trois études que les outils et méthodes déployés pour améliorer l'exactitude absolue de l'auto-évaluation ont conduit à une absence globale d'effets. Seuls les facteurs propres aux par-

participants ou à la tâche semblaient affecter l'exactitude de l'auto-évaluation mesurée. Dans la suite de cette discussion, nous revenons, dans un premier temps, sur les effets qu'a pu avoir le niveau d'expertise des apprenants sur les résultats de notre première étude. Nous proposons, dans un second temps, des explications aux résultats des Études 2 et 3 en relation avec le type de connaissances évaluées et le choix de la mesure de l'exactitude de l'auto-évaluation.

## 7.2 PISTES D'EXPLICATION DES RÉSULTATS

### 7.2.1 *Expertise et efficacité des outils d'aide à l'auto-évaluation*

L'effet de l'expertise inversée suggère que l'efficacité d'outils d'aide à l'apprentissage varie en fonction du niveau d'expertise des apprenants (Kalyuga et al., 2003). Les experts peuvent, par là même, devenir moins performants que des novices lorsque des outils d'aide à l'apprentissage viennent supporter l'apprentissage. L'explication sous-jacente est que les experts disposent de davantage de connaissances antérieures que les novices et sont, par conséquent, capables de faire davantage d'inférences entre différentes informations. Dès lors que des informations (utiles aux novices) sont présentées, les experts doivent traiter une information potentiellement redondante avec leur représentation cognitive existante. Cette redondance dans les informations occasionne une surcharge mentale qui perturbe le traitement de nouvelles informations pertinentes. Un principe similaire peut s'appliquer sur l'exactitude de l'auto-évaluation. Les experts sont généralement plus exacts que les novices dans l'évaluation qu'ils font de leurs connaissances (e.g., Chi & Ceci, 1978 ; Kostons et al., 2010). Cet avantage serait lié à l'utilisation d'indices plus pertinents pour l'auto-évaluation et/ou probablement au moindre coût cognitif nécessaire à la mise en œuvre du processus d'auto-évaluation (Griffin et al., 2009). Le manque d'efficacité des outils d'aide à l'auto-évaluation pourrait, par conséquent, être engendré par une surcharge mentale : face à des informations non pertinentes à traiter, les experts ne disposent plus des ressources nécessaires à la mise en place d'un processus efficace d'évaluation des performances.

### 7.2.2 *Considérer le type de connaissances et le type de tâches évaluées*

Différentes études scientifiques ont attesté de l'efficacité des cartes conceptuelles sur l'exactitude de l'auto-évaluation (Redford et al., 2012 ; Thiede et al., 2010). Les cartes étaient toutefois associées à des tâches classiques de compréhension de textes linéaires qui présentaient des connaissances de nature déclarative. Dans les Études 2 et 3 de ce travail, les cartes

conceptuelles ont été associées à un matériel multimédia et à une tâche de mesure des performances impliquant des connaissances procédurales. De plus, les cartes proposées aux apprenants synthétisaient des informations déclaratives, tandis que l'exercice de mesure de la compréhension requérait l'utilisation de connaissances procédurales. Les connaissances procédurales et déclaratives ont, cependant, des caractéristiques sensiblement différentes. L'acquisition de connaissances procédurales, par exemple, débute par l'acquisition de connaissances déclaratives, mais nécessite des étapes supplémentaires de compilation et d'automatisation de la compétence (e.g., [J. R. Anderson & Fincham, 1994](#)). Également, l'évaluation de l'acquisition de ces deux types de connaissances ne se fait pas de la même façon. Bien que nous ayons mesuré les connaissances procédurales à l'aide d'un questionnaire (cf. Étude 1), celles-ci sont plus souvent mesurées à l'aide d'exercices d'applications, alors que les connaissances déclaratives sont généralement mesurées à l'aide de questions. Il est aussi possible que les indices pertinents pour l'auto-évaluation des performances associées à des connaissances procédurales ne soient pas les mêmes que les indices nécessaires à l'auto-évaluation de performances basées sur des connaissances déclaratives. En ce sens, l'Étude 1 a mis en avant que les participants s'auto-évaluaient avec plus d'exactitude sur les connaissances déclaratives que sur les connaissances procédurales. Dès lors, les outils d'aide à l'auto-évaluation devraient également nécessiter des adaptations. Ces variations dans le matériel et le type de connaissances évaluées, comparativement aux précédentes études, constitue une piste potentielle pour expliquer l'absence globale d'efficacité des cartes conceptuelles sur l'exactitude de l'auto-évaluation et l'apprentissage.

### *7.2.3 Considérer le type d'exactitude de l'auto-évaluation mesurée*

Initialement, le choix de la mesure de l'exactitude de l'auto-évaluation s'est fait en fonction des objectifs de l'étude et des contraintes propres à la situation d'apprentissage. L'exactitude absolue semblait plus adaptée à l'étude de l'auto-évaluation en situation naturelle. De plus, l'exactitude relative ne pouvait pas être mesurée sur la base d'une mesure unique du jugement de performances comme l'exigeait la mesure du jugement d'auto-évaluation à la tâche principale des Études 2 et 3.

De nombreux auteurs, toutefois, ont mis en évidence des résultats différents sur l'exactitude absolue et relative (e.g., [Griffin et al., 2009](#) ; [Koriat, 1997](#) ; [Mengelkamp & Bannert, 2010](#)). [Koriat \(1997\)](#) a ainsi montré que la pratique répétée d'une même liste de mots améliorait l'exactitude relative de l'auto-évaluation mais détériorait l'exactitude absolue. Dans cette même perspective, [Mengelkamp et Bannert \(2010\)](#) ont observé des résultats différents sur l'exactitude relative et absolue de la mesure du jugement

de confiance (jugement rétrospectif). Griffin et al. (2009) ont également mis en exergue que certains facteurs tel que le niveau d'expertise exerçait une influence sur l'exactitude absolue mais pas sur l'exactitude relative. Par ailleurs, les effets des outils d'aide à l'auto-évaluation ont souvent été mis en avant sur l'exactitude relative de l'auto-évaluation (e.g. M. Anderson & Thiede, 2008 ; Redford et al., 2012 ; Thiede et al., 2005, 2010). En ce qui concerne l'exactitude absolue, les effets des outils d'aide sont non seulement moins étudiés, mais semblent également être plus mitigés. Baars et al. (2013) ont observé les effets des outils (dans ce cas-ci, des exemples à compléter) sur la direction du biais d'auto-évaluation mais pas sur l'amplitude du biais. De plus, leurs résultats indiquent qu'alors que les apprenants se surestimaient au préalable, ils se sous-estimaient davantage sous l'effet des exemples à compléter. Les apprenants, finalement, n'étaient pas plus exacts. Ces résultats attestent que les deux indicateurs ne mesurent pas la même dimension de l'exactitude de l'auto-évaluation. Maki et al. (2005) ajoutent, qu'en conséquence, l'exactitude relative et absolue de l'auto-évaluation ne seraient pas affectées par les mêmes facteurs. Les auteurs ont montré que l'exactitude absolue était influencée par des facteurs relatifs aux apprenants (i.e., habiletés verbales) et à la tâche (difficulté), tandis que l'exactitude relative ne l'était pas. Nos résultats confirment cette influence en ce que nous avons principalement obtenu un effet du type de connaissances évaluées (Étude 1) et du niveau d'expertise des apprenants (Étude 2) sur l'exactitude absolue de l'auto-évaluation. Par ailleurs, nos résultats, associés à ceux de Baars et al. (2013), laissent penser que les facteurs 'environnementaux', tels que les outils d'aide à l'auto-évaluation (Schraw, 1998), pourraient davantage influencer l'exactitude relative que l'exactitude absolue de l'auto-évaluation.

Finalement, dans la continuité des explications d'ordre méthodologiques, la faible taille de nos échantillons et des effets suggèrent que la puissance statistique était faible et ne permettait pas de tirer des conclusions fermes quant à l'absence d'efficacité de nos outils. Les écarts-types, quant à eux, témoignent d'une forte variabilité inter-individuelles. Les facteurs motivationnels, non contrôlés dans nos études, peuvent constituer une variable responsable de la forte variabilité inter-individuelle. Ce point est abordé ci-après dans les pistes à explorer.

### 7.3 PISTES DE TRAVAIL POUR DE FUTURES ÉTUDES

#### 7.3.1 *Considération des facteurs motivationnels*

Parmi les différents modèles de l'apprentissage autorégulé, les modèles de Pintrich et De Groot (1990) et Zimmerman (1990) intègrent des composantes motivationnelles. Ces facteurs jouent un rôle essentiel dans l'appren-

tissage autorégulé et leur influence a été mis en évidence dans plusieurs études. Ainsi, le sentiment d'auto-efficacité<sup>1</sup> des apprenants influence l'évaluation qu'ils font de leurs performances dans une tâche spécifique (Koriat, 2007). Lorsque des participants s'estimaient compétents dans un domaine, par exemple, ils estimaient également avoir de meilleures performances à une tâche spécifique de ce domaine. Au contraire, lorsque des participants pensaient réaliser un test dans un domaine où ils sont moins compétents, leur jugement de confiance était plus faible (Ehrlinger & Dunning, 2003). Les performances étaient, toutefois, similaires pour les deux groupes de participants. Dans une autre étude, Noury, Dupeyrat, Escribe, et Huet (2008) ont mis en évidence que le sentiment d'auto-efficacité, relatif à une tâche de résolution de problèmes de statistiques dans un environnement numérique, prédisait le jugement de confiance et son exactitude.

Dupeyrat, Escribe, Huet, et Régner (2011) ont, par ailleurs, comparé la direction du biais d'auto-évaluation selon les buts d'accomplissements<sup>2</sup> adoptés par les apprenants. Les résultats ont montré que les apprenants qui se surestimaient avaient des buts de performance<sup>3</sup>. Les apprenants possédant des buts de maîtrise<sup>4</sup> n'étaient, quant à eux, pas plus exacts que les autres apprenants. Il est à noter, néanmoins, que cette étude s'intéressait à l'auto-évaluation de la compétence globale en mathématique et non sur la prédiction de performances futures. Par ailleurs, d'autres travaux ont également montré que les buts d'accomplissement avaient des répercussions sur les comportements d'apprentissage. Plus précisément, les apprenants avec des buts de performances tendent à utiliser moins de processus d'autorégulation (e.g., Ames & Archer, 1988 ; Wolters, Yu, & Pintrich, 1996).

L'ensemble de ces études met en évidence l'impact des variables motivationnelles sur le jugement d'auto-évaluation. Dans le travail présent, bien que la motivation n'ait pas été mesurée et que des efforts aient été faits pour intégrer les dispositifs à la formation des apprenants, le manque de motivation pourrait justifier que les individus aient peu utilisé les outils d'aide à l'auto-évaluation proposés. Cette piste semble plausible car le

---

1. Le sentiment d'auto-efficacité fait référence à un jugement émis par un individu sur ses propres capacités à réaliser une tâche dans une situation donnée (Bouffard-Bouchard & Pinard, 1988).

2. Les buts d'accomplissements sont 'des constructions cognitives internes, conscientes et accessibles, précisant ce que les individus tentent d'accomplir et les raisons qui les poussent à agir' (Cury, 2004).

3. Les buts de performance renvoient à la volonté qu'a l'apprenant de réussir la tâche pour mettre en avant ses capacités et réussir mieux que les autres (Darnon & Butera, 2005).

4. Les buts de maîtrise renvoient au désir qu'a l'apprenant de réaliser la tâche pour acquérir de nouvelles connaissances et maîtriser la tâche (Darnon & Butera, 2005).

contenu du C2I n'était pas directement lié à la formation des participants de la première étude. En effet, un contenu d'apprentissage relatif au C2I 2MI était initialement prévu pour ces participants. L'absence de matériel disponible au moment de l'étude nous a conduit, à utiliser un autre support. Dans la seconde étude, les novices n'étaient pas non plus directement concernés par le contenu proposé.

### 7.3.2 *Quelle application de l'hypothèse de l'utilisation d'indices de Koriat pour l'évaluation des connaissances procédurales ?*

L'hypothèse d'utilisation d'indices de [Koriat \(1997\)](#) s'applique à des tâches d'apprentissage simples. Rappelons que le principe sous-jacent de cette hypothèse est de considérer le jugement d'auto-évaluation d'autant plus exact que les apprenants utilisent des indices pertinents (des indices pertinents sont des indices qui prédisent les performances finales). Dans une tâche d'apprentissage de paires de mots, un indice pertinent peut être, par exemple, la difficulté des paires d'items à apprendre. Cette hypothèse a été adaptée à des tâches plus complexes au travers de l'hypothèse du modèle de situation ([Griffin et al., 2009](#)). Cette dernière hypothèse est, cette fois, adaptée à des tâches de compréhension de textes. Dans le cas d'une tâche de compréhension d'un texte, la mesure de la performance consiste à évaluer le degré d'élaboration du modèle de situation. Les indices pertinents pour le jugement des performances futures sont donc ceux extraits du modèle de situation. Dans le cas d'une tâche procédurale, les performances sont généralement mesurées à l'aide d'un exercice d'application. Toutefois, l'acquisition de connaissances procédurales nécessite préalablement l'intégration de connaissances déclaratives (e.g., [J. R. Anderson & Fincham, 1994](#)). C'est pourquoi les outils proposés dans nos études traitaient de connaissances déclaratives. Toutefois, nos résultats laissent penser qu'inciter les apprenants à utiliser des indices issus de l'apprentissage profond de connaissances déclaratives ne suffit pas à améliorer l'exactitude de l'auto-évaluation à une tâche procédurale. Il serait ainsi nécessaire d'identifier plus précisément les indices pertinents pour l'évaluation des connaissances procédurales afin de proposer des outils plus adaptés. Une piste potentielle dans la prolongation de ce travail serait d'approfondir la question de l'évaluation des connaissances procédurales. Il serait pertinent, par exemple, d'aider les apprenants à utiliser des indices relatifs aux différentes étapes nécessaires à la résolution d'un problème procédural. En effet, la mesure des performances consiste notamment à évaluer le degré de réussite de plusieurs étapes successives lors de l'évaluation de connaissances procédurales.



## 7.4 CONCLUSION

En conclusion de ce travail, il apparaît qu'aider les apprenants à s'auto-évaluer avec exactitude est une tâche complexe. Comme nous l'avons vu tout au long de ce travail de thèse, les outils d'aide à l'auto-évaluation ne conduisent pas systématiquement à une amélioration de l'exactitude de l'auto-évaluation. La conception d'outils d'aide à l'auto-évaluation doit, ainsi, considérer un ensemble de facteurs tels que le niveau de connaissances des participants, ou encore le type de connaissances évaluées. De futurs travaux portant sur la question de l'exactitude de l'auto-évaluation méritent d'être conduits afin de concevoir des outils d'aide à l'auto-évaluation efficaces dans de multiples contextes. Il semble ainsi important de diversifier les tâches d'apprentissages proposées aux apprenants et de les étudier dans un environnement le plus écologique possible. Ultimement, l'intérêt des outils d'aide à l'auto-évaluation est qu'ils soient utilisés sur un panel de tâches le plus large possible et dans des situations naturelles.

---

## REFERENCES

---

- Ackerman, R., & Lauterman, T. (2012). Taking reading comprehension exams on screen or on paper? a metacognitive analysis of learning texts under time pressure. *Computers in human behavior*, 28(5), 1816–1828. doi: 10.1016/j.chb.2012.04.023 13
- Adesope, O. O., & Nesbit, J. C. (2013). Animated and static concept maps enhance learning from spoken narration. *Learning and Instruction*, 27, 1–10. doi: 10.1016/j.learninstruc.2013.02.002 82
- Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy : Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive science*, 26(2), 147–179. doi: 10.1016/S0364-0213(02)00061-7 43
- Allen, R. M., & Casbergue, R. M. (1997). Evolution of novice through expert teachers' recall : Implications for effective reflection on practice. *Teaching and Teacher Education*, 13(7), 741–755. doi: 10.1016/S0742-051X(97)00018-8 87
- Álvarez-Montero, F. J., Sáenz-Pérez, F., & Vaquero-Sánchez, A. (2015). Using datalog to provide just-in-time feedback during the construction of concept maps. *Expert Systems with Applications*, 42(3), 1362–1375. doi: 10.1016/j.eswa.2014.08.055 126, 128
- Amadiou, F., Mariné, C., & Laimay, C. (2011). The attention-guiding effect and cognitive load in the comprehension of animations. *Computers in Human Behavior*, 27(1), 36–40. doi: 10.1016/j.chb.2010.05.009 99
- Amadiou, F., & Salmerón, L. (2014). Concept maps for comprehension and navigation of hypertexts. In R. Hanewald & D. Ifenthaler (Eds.), *Digital knowledge maps in education* (pp. 41–59). Springer, New York. 45, 82
- Amadiou, F., Tricot, A., & Mariné, C. (2010). Interaction between prior knowledge and concept-map structure on hypertext comprehension, coherence of reading orders and disorientation. *Interacting with computers*, 22(2), 88–97. doi: 10.1016/j.intcom.2009.07.001 44, 87
- Amadiou, F., Van Gog, T., Paas, F., Tricot, A., & Mariné, C. (2009). Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning. *Learning and Instruction*, 19(5), 376–386. doi: 10.1016/j.learninstruc.2009.02.005 99
- Amer, A. A. (1994). The effect of knowledge-map and underlining training on the reading comprehension of scientific texts. *English for Specific*

- Purposes*, 13(1), 35–45. doi: 10.1016/0889-4906(94)90023-X 44, 82
- Ames, C., & Archer, J. (1988). Achievement goals in the classroom : Students' learning strategies and motivation processes. *Journal of educational psychology*, 80(3), 260–267. doi: 10.1037/0022-0663.80.3.260 154
- Anderson, J. R., & Fincham, J. M. (1994). Acquisition of procedural skills from examples. *Journal of experimental psychology : learning, memory, and cognition*, 20(6), 1322–1340. doi: 10.1037/0278-7393.20.6.1322 152, 155
- Anderson, M., & Thiede, K. W. (2008). Why do delayed summaries improve metacomprehension accuracy ? *Acta psychologica*, 128(1), 110–118. doi: 10.1016/j.actpsy.2007.10.006 2, 25, 40, 51, 56, 57, 91, 153
- Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems : Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology*, 95(4), 774–783. doi: 10.1037/0022-0663.95.4.774 43, 57, 60
- Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of educational psychology*, 51(5), 267–272. doi: 10.1037/h0046669 27, 36, 45
- Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia ? *Journal of educational psychology*, 96(3), 523–535. doi: 10.1037/0022-0663.96.3.523 2
- Azevedo, R., Moos, D. C., Greene, J. A., Winters, F. I., & Cromley, J. G. (2008). Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia ? *Educational Technology Research and Development*, 56(1), 45–72. doi: 10.1007/s11423-007-9067-0 48, 49, 58
- Baars, M., Gog, T., Bruin, A., & Paas, F. (2014). Effects of problem solving after worked example study on primary school children's monitoring accuracy. *Applied Cognitive Psychology*, 28(3), 382–391. doi: 10.1002/acp.3008 13, 56, 65, 98, 99, 135
- Baars, M., Visser, S., Gog, T. v., Bruin, A. d., & Paas, F. (2013). Completion of partially worked-out examples as a generation strategy for improving monitoring accuracy. *Contemporary Educational Psychology*, 38(4), 395–406. doi: 10.1016/j.cedpsych.2013.09.001 2, 8, 12, 13, 26, 35, 36, 40, 56, 57, 58, 60, 65, 98, 99, 120, 135, 153
- Baccino, T., Salmeron, L., & Cañas, J. (2008). La lecture des hypertextes. In A. Chevalier & A. Tricot (Eds.), *Ergonomie des documents électroniques* (pp. 9–34). Paris : Presse Universitaire de France. 2,

- Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science*, 40(1), 193–211. doi: 10.1007/s11251-011-9167-4 48, 49, 50, 58, 78
- Bartholome, T., & Bromme, R. (2009). Coherence formation when learning from text and pictures : What kind of support for whom ? *Journal of Educational Psychology*, 101(2), 282–293. doi: 10.1037/a0014312 13
- Berardi-Coletta, B., Buyer, L. S., Dominowski, R. L., & Rellinger, E. R. (1995). Metacognition and problem solving : A process-oriented approach. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 21(1), 205–223. doi: 10.1037/0278-7393.21.1.205 57, 78
- Berry, D. C., & Broadbent, D. E. (1987). Explanation and verbalization in a computer-assisted search task. *The Quarterly Journal of Experimental Psychology*, 39(4), 585–609. doi: 10.1080/14640748708401804 78
- Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies : Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and instruction*, 13(2), 221–252. doi: 10.1207/s1532690xci1302\_3 43, 57, 66, 78
- Boekaerts, M., & Rozendaal, J. S. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction*, 20(5), 372–382. doi: 10.1016/j.learninstruc.2009.03.002 13
- Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education : A critical analysis of findings. *Higher education*, 18(5), 529–549. doi: 10.1007/BF00138746 6
- Bouffard, T., Vezeau, C., Roy, M., & Lengelé, A. (2011). Stability of biases in self-evaluation and relations to well-being among elementary school children. *International Journal of Educational Research*, 50(4), 221–229. doi: 10.1016/j.ijer.2011.08.003 6
- Bouffard-Bouchard, T., & Pinard, A. (1988). Sentiment d'auto-efficacité et exercice des processus d'autorégulation chez des étudiants de niveau collégial\*. *International Journal of Psychology*, 23(1-6), 409–431. 7, 9, 154
- BouJaoude, S., & Attieh, M. (2008). The effect of using concept maps as study tools on achievement in chemistry. *Eurasia Journal of Mathematics, Science & Technology Education*, 4(3), 233–246. 82
- Butcher, K. R. (2006). Learning from text with diagrams : Promoting mental model development and inference generation. *Journal of Educational Psychology*, 98(1), 182–197. doi: 10.1037/0022-0663

- Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, 62(2), 233–246. doi: 10.1111/j.2044-8279.1992.tb01017.x 219
- Chang, K.-E., Sung, Y.-T., & Chen, I.-D. (2002). The effect of concept mapping to enhance text comprehension and summarization. *The Journal of Experimental Education*, 71(1), 5–23. doi: 10.1080/00220970209602054 122, 124, 126, 129
- Chang, K.-E., Sung, Y.-T., & Chen, S. (2001). Learning through computer-based concept mapping with scaffolding aid. *Journal of Computer Assisted Learning*, 17(1), 21–33. doi: 10.1111/j.1365-2729.2001.00156.x 46, 47, 82, 83, 126, 128, 129
- Chi, M. T. (2000). Self-explaining : the dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), (pp. 161–238). Mahwah, NJ : Lawrence Erlbaum Associates. 42, 43
- Chi, M. T. (2006). Two approaches to the study of experts' characteristics. In P. F. K.A. Ericsson N. Charness & R. Hoffman (Eds.), (p. 21-30). Cambridge University Press. 29, 78, 128
- Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations : How students study and use examples in learning to solve problems. *Cognitive science*, 13(2), 145–182. doi: 10.1207/s15516709cog1302\_1 43, 57, 58, 81
- Chi, M. T., & Ceci, S. J. (1978). Knowledge structures and memory development. In R. Siegler (Ed.), *Children's thinking : What develops ?* (pp. 73–96). Hillsdale, NJ : Erlbaum. 1, 29, 30, 32, 120, 151
- Chi, M. T., Glaser, R., & Rees, E. (1981). *Expertise in problem solving*. (Rapport technique N° TR-5). University of Pittsburgh : Learning Research and Development Center. 87
- Chi, M. T., Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive science*, 18(3), 439–477. doi: 10.1207/s15516709cog1803\_3 43
- Chularut, P., & DeBacker, T. K. (2004). The influence of concept mapping on achievement, self-regulation, and self-efficacy in students of english as a second language. *Contemporary Educational Psychology*, 29(3), 248–263. doi: 10.1016/j.cedpsych.2003.09.001 45, 46
- Cimolino, L., Kay, J., & Miller, A. (2003). Incremental student modelling and reflection by verified concept-mapping. In *Supplementary proceedings of the aied2003 : learner modelling for reflection workshop* (pp. 219–227). 128
- Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational psychology review*, 3(3), 149–210. doi: 10.1007/BF01320076 45, 83
- Collins, A., & Brown, J. S. (1988). The computer as a tool for learning

- through reflection. In H. Mandl & L. Alan (Eds.), *Learning issues for intelligent tutoring systems* (p. 1-18). New York : Springer US. 128
- Conati, C., & Vanlehn, K. (2000). Toward computer-based support of meta-cognitive skills : A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education (IJAIED)*, 11, 389–415. 43, 57
- Cury, F. (2004). Evolution conceptuelle de la théorie des buts d’accomplissement dans le domaine du sport. *L’année psychologique*, 104(2), 295–329. doi: 10.3406/psy.2004.29669 154
- Dancey, C. P., & Reidy, J. (2007). *Learning how to learn*. De Boeck Supérieur. 76
- Darnon, B., & Butera, F. (2005). Buts d’accomplissement, stratégies d’étude, et motivation intrinsèque : présentation d’un domaine de recherche et validation française de l’échelle d’elliot et mcgregor (2001). *L’année psychologique*, 105(1), 105–131. doi: 10.3406/psy.2005.3821 154
- de Bruin, A. B., & van Gog, T. (2012). Improving self-monitoring and self-regulation : From cognitive psychology to the classroom. *Learning and Instruction*, 22(4), 245–252. doi: 10.1016/j.learninstruc.2012.01.003 6
- Dinsmore, D. L., & Parkinson, M. M. (2013). What are confidence judgments made of ? students’ explanations for their confidence ratings and what that means for calibration. *Learning and Instruction*, 24, 4–14. doi: 10.1016/j.learninstruc.2012.06.001 9, 11
- Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension a brief history and how to improve its accuracy. *Current Directions in Psychological Science*, 16(4), 228–232. doi: 10.1111/j.1467-8721.2007.00509.x 8, 12, 14, 53, 86
- Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic–extrinsic distinction of judgments of learning (jols) : Effects of relatedness and serial position on jols. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 27(5), 1180–1191. doi: 10.1037/0278-7393.27.5.1180 23
- Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (jol) and the delayed-jol effect. *Memory & Cognition*, 20(4), 374–380. doi: 10.3758/BF03210921 1, 5, 14, 50
- Dunlosky, J., & Nelson, T. O. (1997). Similarity between the cue for judgments of learning (jol) and the cue for test is not the primary determinant of jol accuracy. *Journal of Memory and Language*, 36(1), 34–49. doi: 10.1006/jmla.1996.2476 18
- Dunlosky, J., & Rawson, K. A. (2005). Why does rereading improve metacomprehension accuracy ? evaluating the levels-of-disruption

- hypothesis for the rereading effect. *Discourse Processes*, 40(1), 37–55. doi: 10.1207/s15326950dp4001\_2 18, 19, 20, 25, 37, 38
- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement : Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, 22(4), 271–280. doi: 10.1016/j.learninstruc.2011.08.003 1, 5, 6, 60
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58. doi: 10.1177/1529100612453266 85
- Dunlosky, J., Rawson, K. A., & Middleton, E. L. (2005). What constrains the accuracy of metacomprehension judgments ? testing the transfer-appropriate-monitoring and accessibility hypotheses. *Journal of Memory and Language*, 52(4), 551–565. doi: 10.1016/j.jml.2005.01.011 10, 16, 17, 26, 51
- Dupeyrat, C., Escribe, C., Huet, N., & Régner, I. (2011). Positive biases in self-assessment of mathematics competence, achievement goals, and mathematics performance. *International Journal of Educational Research*, 50(4), 241–250. doi: 10.1016/j.ijer.2011.08.005 154
- Ehrlich, M.-F., & Cahour, B. (1991). Contrôle métacognitif de la compréhension : cohésion d'un texte expositif et auto-évaluation de la compréhension. *Bulletin de psychologie*, 44, 147–155. 27
- Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of personality and social psychology*, 84(1), 5-17. doi: 10.1037/0022-3514.84.1.5 154
- Fiorella, L., Vogel-Walcutt, J. J., & Fiore, S. (2012). Differential impact of two types of metacognitive prompting provided during simulation-based training. *Computers in Human Behavior*, 28(2), 696–702. doi: 10.1016/j.chb.2011.11.017 49, 50
- Flavell, J. H. (1977). *Cognitive development*. Prentice-Hall Publishing : Englewood Cliffs, NJ. 1
- Flavell, J. H. (1979). Metacognition and cognitive monitoring : A new area of cognitive–developmental inquiry. *American psychologist*, 34(10), 906-911. doi: 10.1037/0003-066X.34.10.906 8, 29
- Flavell, J. H., Friedrichs, A. G., & Hoyt, J. D. (1970). Developmental changes in memorization processes. *Cognitive psychology*, 1(4), 324–340. doi: 10.1016/0010-0285(70)90019-8 1
- Gama, C. A. (2004). *Integrating metacognition instruction in interactive learning environments* (Thèse de doctorat non publiée). University of Sussex. 49, 58, 65
- García, T., Rodríguez, C., González-Castro, P., González-Pienda, J. A., &

- Torrance, M. (2015). Elementary students' metacognitive processes and post-performance calibration on mathematical problem-solving tasks. *Metacognition and Learning*, 1–32. doi: 10.1007/s11409-015-9139-1 53
- Glaser, R., Chi, M. T., & Farr, M. J. (1988). *The nature of expertise*. Lawrence Erlbaum Associates Hillsdale, NJ. 31
- Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 11(4), 702–718. doi: 10.1037/0278-7393.11.1-4.702 14, 35
- Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition*, 15(1), 84–93. doi: 10.3758/BF03197714 1, 10, 14, 30, 35, 87
- Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing : Failure in the self-assessment of comprehension. *Memory & Cognition*, 10(6), 597–602. doi: 10.3758/BF03202442 9
- Gobert, J. D., & Clement, J. J. (1999). Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. *Journal of research in science teaching*, 36(1), 39–53. doi: 10.1002/(sici)1098-2736(199901)36:1b 84
- Gramzow, R. H., Elliot, A. J., Asher, E., & McGregor, H. A. (2003). Self-evaluation bias and academic performance : Some ways and some reasons why. *Journal of Research in Personality*, 37(2), 41–61. doi: 10.1016/S0092-6566(02)00535-4 6
- Griffin, T. D., Jee, B. D., & Wiley, J. (2009). The effects of domain knowledge on metacomprehension accuracy. *Memory & Cognition*, 37(7), 1001–1013. doi: 10.3758/MC.37.7.1001 12, 23, 24, 29, 30, 31, 35, 36, 54, 56, 81, 87, 151, 152, 153, 155
- Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation : Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition*, 36(1), 93–103. doi: 10.3758/MC.36.1.93 8, 26, 37, 39, 40, 43, 44, 47, 57, 66, 79, 87, 120, 149
- Gurlitt, J., & Renkl, A. (2008). Are high-coherent concept maps better for prior knowledge activation ? differential effects of concept mapping tasks on high school vs. university students. *Journal of Computer Assisted Learning*, 24(5), 407–419. doi: 10.1111/j.1365-2729.2008.00277.x 82, 85, 86, 88, 89, 90, 92, 122
- Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts : the effects of incentives, reflection, and explanatory style. *Metacognition and Learning*, 3(2), 101–121. doi: 10.1007/s11409-008-9021-5 12, 13, 49



- Hart, J. (1967). Memory and the memory-monitoring process. *Journal of verbal learning and verbal behavior*, 6(5), 685–691. doi: 10.1016/S0022-5371(67)80072-0 16
- Hausmann, R. G., & Chi, M. H. (2002). Can a computer interface support self-explaining. *Cognitive Technology*, 7(1), 4–14. 43, 48, 57
- Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of memory and language*, 32(6), 717–742. doi: 10.1006/jmla.1993.1036 37
- Hilbert, T. S., & Renkl, A. (2008). Concept mapping as a follow-up strategy to learning from texts : what characterizes good and poor mappers ? *Instructional Science*, 36(1), 53–73. doi: 10.1007/s11251-007-9022-9 85, 122, 123, 127
- Hilbert, T. S., & Renkl, A. (2009). Learning how to use a computer-based concept-mapping tool : Self-explaining examples helps. *Computers in Human Behavior*, 25(2), 267–274. doi: 10.1016/j.chb.2008.12.006 46, 47
- Hines, J. C., Touron, D. R., & Hertzog, C. (2009). Metacognitive influences on study time allocation in an associative recognition task : An analysis of adult age differences. *Psychology and Aging*, 24(2), 462–475. doi: 10.1037/a0014417 2, 35
- Hobus, P., Schmidt, H., Boshuizen, H., & Patel, V. (1987). Contextual factors in the activation of first diagnostic hypotheses : expert-novice differences. *Medical Education*, 21(6), 471–476. doi: 10.1111/j.1365-2923.1987.tb01405.x 79
- Huet, N. (2004). *Contributions à l'analyse des processus métamnémoniques : connaissances métamnémoniques et expérience professionnelle* (Habilitation à diriger des recherches en Psychologie). Université du Mirail, Toulouse 2. 5, 31, 54
- Huet, N., & Mariné, C. (2009). Expertise and self-regulation processes in a professional task. *Applied Cognitive Psychology*, 23(7), 1027–1041. doi: 10.1002/acp.1540 2, 35, 89, 90
- Huet, N., Mariné, C., & Escribe, C. (1994). Auto-évaluation des compétences en résolution de problèmes chez des adultes peu qualifiés. *International Journal of Psychology*, 29(3), 273–289. doi: 10.1080/00207599408246545 8
- Ifenthaler, D. (2012). Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Journal of Educational Technology & Society*, 15(1), 38–52. 48, 49, 51
- Izaute, M., Larochelle, S., Morency, J., & Tiberghien, G. (1996). La validité du sentiment de savoir au rappel et à la reconnaissance. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 50(2), 163–181. doi: 10.1037/1196-1961.50.2.163 10

- Kalyuga, S. (2009). Knowledge elaboration : A cognitive load perspective. *Learning and Instruction*, 19(5), 402–410. doi: 10.1016/j.learninstruc.2009.02.003 27
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational psychologist*, 38(1), 23–31. doi: 10.1207/S15326985EP3801\_4 2, 79, 81, 89, 92, 151, 220
- Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities : An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of applied psychology*, 74(4), 657–690. doi: 10.1037/0021-9010.74.4.657 48, 85, 87, 92
- Kauffman, D. F., Zhao, R., & Yang, Y.-S. (2011). Effects of online note taking formats and self-monitoring prompts on learning from online text : Using technology to enhance self-regulated learning. *Contemporary Educational Psychology*, 36(4), 313–322. doi: 10.1016/j.cedpsych.2011.04.001 6
- Kelemen, W. L., Frost, P. J., & Weaver, C. A. (2000). Individual differences in metacognition : Evidence against a general metacognitive ability. *Memory & Cognition*, 28(1), 92–107. doi: 10.3758/BF03211579 54
- Kintsch, W. (1988). The role of knowledge in discourse comprehension : a construction-integration model. *Psychological review*, 95(2), 163–182. doi: 10.1037/0033-295X.95.2.163 23, 27, 54
- Kintsch, W. (1998). *Comprehension : A paradigm for cognition*. New York : Cambridge university press. 35
- Kitchin, R. M. (1994). Cognitive maps : What are they and why study them ? *Journal of environmental psychology*, 14(1), 1–19. doi: 10.1016/S0272-4944(05)80194-X 44
- Koriat, A. (1993). How do we know that we know ? the accessibility model of the feeling of knowing. *Psychological review*, 100(4), 609–639. doi: 10.1037/0033-295X.100.4.609 16, 17, 20, 26
- Koriat, A. (1995). Dissociating knowing and the feeling of knowing : Further evidence for the accessibility model. *Journal of Experimental Psychology : General*, 124(3), 311–333. doi: 10.1037/0096-3445.124.3.311 16, 17, 20
- Koriat, A. (1997). Monitoring one's own knowledge during study : A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology : General*, 126(4), 349–370. doi: 10.1037/0096-3445.126.4.349 1, 8, 15, 20, 21, 22, 25, 26, 28, 35, 36, 37, 54, 56, 83, 152, 155
- Koriat, A. (2007). Metacognition and consciousness. In P. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The cambridge handbook of consciousness* (p. 289-326). Cambridge : Cambridge University Press. 12, 60, 154

- Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions : a comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 32(5), 1133–1145. doi: 10.1037/0278-7393.32.5.1133 36
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology : Human learning and memory*, 6(2), 107–118. doi: 10.1037/0278-7393.6.2.107 14, 184
- Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves : judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology : General*, 131(2), 147–162. doi: 10.1037/0096-3445.131.2.147 11, 14, 60
- Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, 14(2), 219–224. doi: 10.3758/BF03194055 2, 35
- Kostons, D., van Gog, T., & Paas, F. (2010). Self-assessment and task selection in learner-controlled instruction : Differences between effective and ineffective learners. *Computers & Education*, 54(4), 932–940. doi: 10.1016/j.compedu.2009.09.025 6, 30, 54, 87, 151
- Lambiotte, J. G., & Dansereau, D. F. (1992). Effects of knowledge maps and prior knowledge on recall of science lecture content. *The Journal of Experimental Education*, 60(3), 189–201. doi: 10.1080/00220973.1992.9943875 86, 88, 90, 92
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive science*, 11(1), 65–100. doi: 10.1016/S0364-0213(87)80026-5 85
- Leplat, J. (2005). Les automatismes dans l'activité : pour une réhabilitation et un bon usage. *Activités. Revue électronique*, 2(2), 43–67. 28, 128
- Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities : The state of the art. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty : Heuristics and biases* (p. 306-334). Cambridge, UK : Cambridge University Press. 11, 183, 184
- Lin, L.-M., Zabrocky, K. M., & Moore, D. (2002). Effects of text difficulty and adults' age on relative calibration of comprehension. *American Journal of Psychology*, 115(2), 187–198. 38
- Lin, X., & Lehman, J. D. (1999). Supporting learning of variable control in a computer-based biology environment : Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching*, 36(7), 837–858. 49, 58, 65, 81
- Maillard, A., Motàk, L., Sakdavong, J.-C., Dupeyrat, C., & Huet, N. (2013). How and when presenting a concept map for learning and an accurate self-evaluation ? In *Proceedings of the 5th international*

- conference on computer supported education. 94, 96
- Maki, R. H. (1998). Predicting performance on text : Delayed versus immediate predictions and tests. *Memory & Cognition*, 26(5), 959–964. doi: 10.3758/BF03201176 1, 35, 51, 90, 120, 149
- Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 10(4), 663–679. doi: 10.1037/0278-7393.10.4.663 8
- Maki, R. H., & Serra, M. (1992). The basis of test predictions for text material. *Journal of experimental psychology : Learning, memory, and cognition*, 18(1), 116–126. doi: 10.1037/0278-7393.18.1.116 54, 87
- Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology*, 97(4), 723–731. doi: 10.1037/0022-0663.97.4.723 12, 153
- Mariné, C., & Huet, N. (1998). Techniques d'évaluation de la métacognition. i les mesures indépendantes de l'exécution de tâches. ii les mesures dépendantes de l'exécution de tâches. *L'année psychologique*, 98(4), 711–742. 7, 9
- Mayer, R. E. (1979). Twenty years of research on advance organizers : Assimilation theory is still the best predictor of results. *Instructional Science*, 8(2), 133–167. doi: 10.1007/BF00117008 45, 88
- Mayer, R. E. (2001). *Multimedia learning*. New York : Cambridge university press. 90
- Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and instruction*, 12(1), 107–119. doi: 10.1016/S0959-4752(01)00018-4 92
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, 38(1), 43–52. doi: 10.1207/S15326985EP3801\_6 85
- Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words ? extensions of a dual-coding theory of multimedia learning. *Journal of educational psychology*, 86(3), 389–401. doi: 10.1037/0022-0663.86.3.389 90
- Mazzoni, G., & Nelson, T. O. (1995). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 21(5), 1263–1274. doi: 10.1037/0278-7393.21.5.1263 16, 25, 40
- McCagg, E. C., & Dansereau, D. F. (1991). A convergent paradigm for examining knowledge mapping as a learning strategy. *The Journal of Educational Research*, 84(6), 317–324. doi: 10.1080/00220671.1991.9941812 126, 129

- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and instruction*, *14*(1), 1–43. [24](#), [87](#)
- Meijer, J., Veenman, M. V., & van Hout-Wolters, B. (2012). Multi-domain, multi-method measures of metacognitive activity : what is all the fuss about metacognition. . . indeed? *Research Papers in Education*, *27*(5), 597–627. doi: 10.1080/02671522.2010.550011 [54](#)
- Mengelkamp, C., & Bannert, M. (2009). Judgements about knowledge : Searching for factors that influence their validity. *Electronic Journal of Research in Educational Psychology*, *17*(7), 163–190. [9](#), [10](#)
- Mengelkamp, C., & Bannert, M. (2010). Accuracy of confidence judgments : Stability and generality in the learning process and predictive validity for learning outcome. *Memory & cognition*, *38*(4), 441–451. doi: 10.3758/MC.38.4.441 [9](#), [12](#), [152](#)
- Metcalf, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*(1), 174–179. doi: 10.3758/PBR.15.1.174 [2](#), [35](#)
- Mihalca, L. (2014). Why prompting metacognition in computer-based learning environments. *Cognitie, Creier, Comportament*, *18*(4), 299–314. [49](#), [58](#)
- Miller, T. M., & Geraci, L. (2011). Unskilled but aware : reinterpreting overconfidence in low-performing students. *Journal of experimental psychology : learning, memory, and cognition*, *37*(2), 502–506. doi: 10.1037/a0021802 [35](#), [49](#), [58](#)
- Minguela, M., Solé, I., & Pieschl, S. (2015). Flexible self-regulated reading as a cue for deep comprehension : evidence from online and offline measures. *Reading and Writing*, *28*(5), 721–744. doi: 10.1007/s11145-015-9547-2 [37](#)
- Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning : The role of modality and contiguity. *Journal of educational psychology*, *91*(2), 358–368. doi: 10.1037/0022-0663.91.2.358 [90](#)
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of verbal learning and verbal behavior*, *16*(5), 519–533. doi: 10.1016/S0022-5371(77)80016-9 [18](#)
- Mulligan, N. W. (2002). The effects of generation on conceptual implicit memory. *Journal of Memory and Language*, *47*(2), 327–342. doi: 10.1016/S0749-596X(02)00006-2 [39](#)
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological bulletin*, *95*(1), 109–133. doi: 10.1037/0033-2909.95.1.109 [13](#), [14](#), [60](#)

- Nelson, T. O. (1996). Consciousness and metacognition. *American psychologist*, 51(2), 102–116. doi: 10.1037/0003-066X.51.2.102  
27
- Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (jols) are extremely accurate at predicting subsequent recall : The "delayed-jol effect". *Psychological Science*, 2(4), 267–270. doi: doi:10.1111/j.1467-9280.1991.tb00147.x 2, 8, 11, 14, 36, 50, 56, 81
- Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science*, 5(4), 207–213. doi: 10.1111/j.1467-9280.1994.tb00502.x 7
- Nelson, T. O., & Narens, L. (1990). Metamemory : A theoretical framework and new findings. *The psychology of learning and motivation*, 26, 125-141. doi: 10.1016/S0079-7421(08)60053-5 1, 5, 6, 7, 9, 11, 35, 57, 149, 179
- Nelson-Le Gall, S., Kratzer, L., Jones, E., & DeCooke, P. (1990). Children's self-assessment of performance and task-related help seeking. *Journal of Experimental Child Psychology*, 49(2), 245–263. doi: 10.1016/0022-0965(90)90057-F 6
- Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps : A meta-analysis. *Review of educational research*, 76(3), 413–448. doi: 10.3102/00346543076003413 44, 83, 86, 88, 91, 122
- Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *The Journal of Experimental Educational*, 7–28. 7, 12
- Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning*, 1(2), 159–179. doi: 10.1007/s10409-006-9595-6 12
- Nietfeld, J. L., & Schraw, G. (2002). The effect of knowledge and strategy training on monitoring accuracy. *The Journal of Educational Research*, 95(3), 131–142. doi: 10.1080/00220670209596583 87
- Noël, B. (1991). *La métacognition*. Paris : De Boeck Université. 1
- Noury, F., Dupeyrat, C., Escribe, C., & Huet, N. (2008). Influence of self-efficacy and perception of competence on learners' confidence judgments during problem solving in a web-based learning environment.. 154
- Novak, J. D. (1984). *Learning how to learn*. New York : Cambridge University Press. 44
- Novak, J. D. (2010). Learning, creating, and using knowledge : Concept maps as facilitative tools in schools and corporations. *Journal of*

- e-learning and Knowledge Society*, 6(3), 21–30. [83](#)
- Novak, J. D., & Cañas, A. J. (2008). *The theory underlying concept maps and how to construct them* (Rapport technique). Florida : Florida Institute for Human and Machine Cognition. [44](#), [82](#), [83](#), [120](#), [121](#)
- O'donnell, A. M., Dansereau, D. F., & Hall, R. H. (2002). Knowledge maps as scaffolds for cognitive processing. *Educational Psychology Review*, 14(1), 71–86. doi: 10.1023/A:1013132527007 [44](#), [46](#), [47](#), [48](#), [82](#)
- Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics : A cognitive-load approach. *Journal of educational psychology*, 84(4), 429-434. doi: 10.1037/0022-0663.84.4.429 [27](#), [99](#), [129](#)
- Paivio, A. (1986). *Mental representations : A dual coding approach*. Oxford, England : Oxford University Press. [45](#)
- Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving : A path analysis. *Journal of educational psychology*, 86(2), 193-203. doi: 10.1037/0022-0663.86.2.193 [11](#)
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology*, 82(1), 33–40. doi: 10.1037/0022-0663.82.1.33 [1](#), [153](#), [180](#)
- Ponce, H. R., & Mayer, R. E. (2014). Qualitatively different cognitive processing during online reading primed by different study activities. *Computers in Human Behavior*, 30, 121–130. doi: 10.1016/j.chb.2013.07.054 [45](#)
- Pressley, M. (1995). *Advanced educational psychology*. New York : Harper-Collins. [87](#)
- Pressley, M., & Ghatala, E. S. (1990). Self-regulated learning : Monitoring learning from text. *Educational Psychologist*, 25(1), 19–33. doi: 10.1207/s15326985ep2501\_3 [1](#), [8](#), [9](#)
- Raney, G. E. (1993). Monitoring changes in cognitive load during reading : an event-related brain potential and reaction time analysis. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 19(1), 51–69. doi: 10.1037/0278-7393.19.1.51 [39](#)
- Rawson, K. A., & Dunlosky, J. (2007). Improving students' self-evaluation of learning for key concepts in textbook materials. *European Journal of Cognitive Psychology*, 19(4-5), 559–579. doi: 10.1080/09541440701326022 [10](#), [11](#)
- Rawson, K. A., Dunlosky, J., & Thiede, K. W. (2000). The rereading effect : Metacomprehension accuracy improves across reading trials. *Memory & Cognition*, 28(6), 1004–1010. doi: 10.3758/BF03209348 [23](#), [24](#), [37](#), [149](#)

- Redford, J. S., Thiede, K. W., Wiley, J., & Griffin, T. D. (2012). Concept mapping improves metacomprehension accuracy among 7th graders. *Learning and Instruction*, 22(4), 262–270. doi: 10.1016/j.learninstruc.2011.10.007 25, 35, 36, 44, 46, 47, 82, 83, 84, 86, 89, 90, 91, 92, 120, 121, 122, 123, 127, 147, 148, 151, 153
- Renkl, A. (1997). Learning from worked-out examples : A study on individual differences. *Cognitive science*, 21(1), 1–29. doi: 10.1207/s15516709cog2101\_1 43, 48, 58
- Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgments of learning on metacognitive accuracy : a meta-analytic review. *Psychological bulletin*, 137(1), 131–148. doi: 10.1037/a0021705 50, 51
- Rouet, J.-F. (2005). La conception des ressources multimédias pour l'apprentissage : apports des recherches en psychologie du langage. *Revue française de Pédagogie*, 79–87. doi: http://www.jstor.org/stable/41202068 90, 91
- Sakdavong, J.-C., Amadiou, F., Huet, N., & Dupeyrat, C. (2011). Help design and help use in a computer based learning environment. In *International conference on computer supported education*. 94, 96
- Salomon, G. (1984). "Television is" easy" and print is" tough" : The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of educational psychology*, 76(4), 647–658. doi: 10.1037/0022-0663.76.4.647 85
- Santhanam, E., Leach, C., & Dawson, C. (1998). Concept mapping : How should it be introduced, and is there evidence for long term benefit ? *Higher Education*, 35(3), 317–328. doi: 10.1023/A:1003028902215 126
- Scevak, J. J., Moore, P. J., & Kirby, J. R. (1993). Training students to use maps to increase text recall. *Contemporary Educational Psychology*, 18(4), 401–413. doi: 10.1006/ceps.1993.1030 127
- Schneider, W. (1996). The effects of expertise and iq on children's memory : When knowledge is, and when it is not enough. *International Journal of Behavioral Development*, 19(4), 773–796. doi: 10.1080/016502596385578 54, 87
- Schnotz, W. (2002). Commentary : Towards an integrated view of learning from text and visual displays. *Educational psychology review*, 14(1), 101–120. doi: 10.1023/A:1013136727916 45
- Schraw, G. (1994). The effect of metacognitive knowledge on local and global monitoring. *Contemporary educational psychology*, 19(2), 143–154. doi: 10.1006/ceps.1994.1013 10
- Schraw, G. (1995). Measures of feeling-of-knowing accuracy : A new look at an old problem. *Applied Cognitive Psychology*, 9(4), 321–332. doi: 10.1002/acp.2350090405 13, 183



- Schraw, G. (1998). On the development of adult metacognition. In M. C. Smith & T. Pourchot (Eds.), *Adult learning and development* (pp. 89–106). Mahwah, NJ : Lawrence Erlbaum Associates. [2](#), [9](#), [30](#), [31](#), [32](#), [153](#)
- Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, *4*(1), 33–45. doi: 10.1007/s11409-008-9031-3 [9](#), [60](#)
- Schraw, G., Dunkle, M. E., Bendixen, L. D., & Roedel, T. D. (1995). Does a general monitoring skill exist ? *Journal of Educational Psychology*, *87*(3), 433–444. doi: 10.1037/0022-0663.87.3.433 [31](#), [54](#)
- Schraw, G., Kuch, F., & Gutierrez, A. P. (2013). Measure for measure : Calibrating ten commonly used calibration scores. *Learning and Instruction*, *24*, 48–57. doi: 10.1016/j.learninstruc.2012.08.007 [13](#)
- Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational psychology review*, *7*(4), 351–371. doi: 10.1007/BF02212307 [5](#)
- Schraw, G., & Nietfeld, J. (1998). A further test of the general monitoring skill hypothesis. *Journal of Educational Psychology*, *90*(2), 236–248. doi: 10.1037/0022-0663.90.2.236 [31](#), [54](#)
- Schraw, G., Potenza, M. T., & Nebelsick-Gullet, L. (1993). Constraints on the calibration of performance. *Contemporary Educational Psychology*, *18*(4), 455–463. doi: 10.1006/ceps.1993.1034 [13](#), [49](#), [58](#)
- Schraw, G., & Robinson, D. H. (2011). *Assessment of higher order thinking skills*. Charlotte : NC : Information Age Publishers. [55](#)
- Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, *78*(1), 153–189. doi: 10.3102/0034654307313795 [128](#)
- Slamecka, N. J., & Graf, P. (1978). The generation effect : Delineation of a phenomenon. *Journal of experimental Psychology : Human learning and Memory*, *4*(6), 592–604. doi: 10.1037/0278-7393.4.6.592 [39](#)
- Stull, A. T., & Mayer, R. E. (2007). Learning by doing versus learning by viewing : Three experimental comparisons of learner-generated versus author-provided graphic organizers. *Journal of educational psychology*, *99*(4), 808–820. doi: 10.1037/0022-0663.99.4.808 [2](#), [44](#), [46](#), [47](#), [48](#), [85](#), [86](#), [90](#), [92](#), [120](#), [122](#), [123](#), [129](#)
- Sweller, J. (1988). Cognitive load during problem solving : Effects on learning. *Cognitive science*, *12*(2), 257–285. doi: 10.1207/s15516709cog1202\_4 [48](#), [85](#)
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York : Springer. [219](#), [220](#)
- Sweller, J., Van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, *10*(3), 251–296. doi: 10.1023/A:1022193728205 [2](#), [90](#), [217](#), [218](#), [219](#)

- Tabbers, H. K., Martens, R. L., & Merriënboer, J. J. (2004). Multimedia instructions and cognitive load theory : Effects of modality and cueing. *British Journal of Educational Psychology*, 74(1), 71–81. doi: 10.1348/000709904322848824 [90](#), [91](#)
- Tabbers, H. K., Martens, R. L., & Van Merriënboer, J. J. (2001). The modality effect in multimedia instructions. In *Proceedings of the 23rd annual conference of the cognitive science society* (pp. 1024–1029). [90](#)
- Taconnat, L. (2005). Apport du paradigme de production d'informations dans l'étude des dissociations entre les mesures de mémoire. *L'année psychologique*, 105(3), 521–537. doi: 10.3406/psy.2005.29708 [39](#)
- Taconnat, L., & Isingrini, M. (2004). Cognitive operations in the generation effect on a recall test : role of aging and divided attention. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 30(4), 827–837. doi: 10.1037/0278-7393.30.4.827 [39](#)
- Thiede, K. W., & Anderson, M. (2003). Summarizing can improve meta-comprehension accuracy. *Contemporary Educational Psychology*, 28(2), 129–160. doi: 10.1016/S0361-476X(02)00011-5 [8](#), [25](#), [40](#), [41](#), [51](#)
- Thiede, K. W., Anderson, M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of educational psychology*, 95(1), 66-73. doi: 10.1037/0022-0663.95.1.66 [25](#), [40](#), [41](#), [51](#), [56](#), [57](#)
- Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study : An analysis of selection of items for study and self-paced study time. *Journal of experimental psychology : Learning, Memory, and Cognition*, 25(4), 1024–1037. doi: 10.1037/0278-7393.25.4.1024 [2](#), [35](#)
- Thiede, K. W., Dunlosky, J., Griffin, T. D., & Wiley, J. (2005). Understanding the delayed-keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 31(6), 1267–1280. doi: 10.1037/0278-7393.31.6.1267 [2](#), [8](#), [14](#), [25](#), [40](#), [41](#), [42](#), [47](#), [51](#), [56](#), [57](#), [91](#), [149](#), [153](#)
- Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes*, 47(4), 331–362. doi: 10.1080/01638530902959927 [23](#), [25](#), [40](#), [44](#), [46](#), [56](#), [57](#), [82](#), [83](#), [89](#), [90](#), [91](#), [92](#), [120](#), [123](#), [128](#), [151](#), [153](#)
- Thiede, K. W., Griffin, T. D., Wiley, J., & Redford, J. (2009). Metacognitive monitoring during and after reading. In D. J. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Handbook of metacognition in education* (pp. 85–106). New York : Routledge. [90](#)

- Thillmann, H., Künsting, J., Wirth, J., & Leutner, D. (2009). Is it merely a question of ‘what’ to prompt or also ‘when’ to prompt? the role of point of presentation time of prompts in self-regulated learning. *Zeitschrift für Pädagogische Psychologie*, *23*(2), 105–115. doi: 10.1024/1010-0652.23.2.105 51
- Thuair, F. (2013). *Les processus d'évaluation et de régulation de l'apprentissage : psychopathologie cognitive de la schizophrénie et études préliminaires auprès de participants sains* (Thèse de Doctorat en Psychologie). Université Blaise Pascal, Clermont-Ferrand II. 7
- Tsai, C., Lin, S. S., & Yuan, S.-M. (2001). Students' use of web-based concept map testing and strategies for learning. *Journal of Computer Assisted Learning*, *17*(1), 72–84. doi: 10.1111/j.1365-2729.2001.00160.x 83, 85
- Van den Boom, G., Paas, F., Van Merriënboer, J. J., & Van Gog, T. (2004). Reflection prompts and tutor feedback in a web-based learning environment : Effects on students' self-regulated learning competence. *Computers in Human Behavior*, *20*(4), 551–567. doi: 10.1016/j.chb.2003.10.001 49
- Van Gog, T., Kester, L., & Paas, F. (2011). Effects of concurrent monitoring on cognitive load and performance as a function of task complexity. *Applied cognitive psychology*, *25*(4), 584–587. doi: 10.1002/acp.1726 28, 87
- Vanides, J., Yue, Y., Tomita, M., & Ruiz-Primo, M. A. (2005). Using concept maps in the science classroom. *Science Scope*, *28*(8), 27 - 31. 83
- Van Loon, M. H., de Bruin, A. B., van Gog, T., Van Merriënboer, J. J., & Dunlosky, J. (2014). Can students evaluate their understanding of cause-and-effect relations? the effects of diagram completion on monitoring accuracy. *Acta Psychologica*, *151*(0), 143 - 154. doi: http://dx.doi.org/10.1016/j.actpsy.2014.06.007 84, 91, 92, 120, 122, 124, 129, 147
- Van Merriënboer, J., Schuurman, J., De Croock, M., & Paas, F. (2002). Redirecting learners' attention during training : Effects on cognitive load, transfer test performance and training efficiency. *Learning and Instruction*, *12*(1), 11–37. doi: 10.1016/S0959-4752(01)00020-2 27
- Van Merriënboer, J. J. (1990). Strategies for programming instruction in high school : Program completion vs. program generation. *Journal of educational computing research*, *6*(3), 265–285. doi: 10.2190/4NK5-17L7-TWQV-1EHL 27
- Van Merriënboer, J. J., & Sluijsmans, D. M. (2009). Toward a synthesis of cognitive load theory, four-component instructional design, and self-directed learning. *Educational Psychology Review*, *21*(1), 55–66.

- doi: 10.1007/s10648-008-9092-5 218
- Van Overschelde, J. P., & Nelson, T. O. (2006). Delayed judgments of learning cause both a decrease in absolute accuracy (calibration) and an increase in relative accuracy (resolution). *Memory & cognition*, 34(7), 1527–1538. doi: 10.3758/BF03195916 51, 56
- Veenman, M., & Elshout, J. J. (1999). Changes in the relation between cognitive and metacognitive skills during the acquisition of expertise. *European journal of psychology of education*, 14(4), 509-523. doi: 10.1007/BF03172976 32, 55
- Veenman, M., Prins, F., & Elshout, J. (2002). Initial inductive learning in a complex computer simulated environment : the role of metacognitive skills and intellectual ability. *Computers in Human Behavior*, 18(3), 327–341. doi: 10.1016/S0747-5632(01)00038-3 27
- Veenman, M. V., Elshout, J. J., & Meijer, J. (1997). The generality vs domain-specificity of metacognitive skills in novice learning across domains. *Learning and instruction*, 7(2), 187-209. doi: 10.1016/S0959-4752(96)00025-4 32, 54, 55
- Veenman, M. V. J. (2011). Learning to self-monitor and self-regulate. In R. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (p. 197-218). New York : Routledge. 5
- Veenman, M. V. J., Van Hout-Wolters, B. H., & Afflerbach, P. (2006). Metacognition and learning : Conceptual and methodological considerations. *Metacognition and learning*, 1(1), 3–14. doi: 10.1007/s11409-006-6893-0 27
- Vidal, P., Huet, N., Miravette, S., & Marquié, D. (2015). *Processus global de réingénierie pour l'acquisition et le développement de compétences* (Rapport technique N° ANR-10-CORD-011). Toulouse : Université Toulouse III. 53, 61
- Vörös, Z., Rouet, J.-F., & Pléh, C. (2011). Effect of high-level content organizers on hypertext learning. *Computers in Human Behavior*, 27(5), 2047–2055. doi: 10.1016/j.chb.2011.04.005 82, 83
- Weaver, C. A., & Bryant, D. S. (1995). Monitoring of comprehension : The role of text difficulty in metamemory for narrative and expository text. *Memory & Cognition*, 23(1), 12–22. doi: 10.3758/BF03210553 38, 78
- Weaver III, C. A., & Kelemen, W. L. (2003). Processing similarity does not improve metamemory : evidence against transfer-appropriate monitoring. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 29(6), 1058–1065. doi: 10.1037/0278-7393.29.6.1058 18, 19, 25, 26
- Weinerth, K., Koenig, V., Brunner, M., & Martin, R. (2014). Concept maps : A useful and usable tool for computer-based knowledge assessment ? a literature review with a focus on usability. *Computers*

- & *Education*, 78, 201–209. doi: 10.1016/j.compedu.2014.06.002  
126, 127
- Wigent, C. A. (2013). High school readers : A profile of above average readers and readers with learning disabilities reading expository text. *Learning and Individual Differences*, 25, 134–140. doi: 10.1016/j.lindif.2013.03.011 37
- Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources : Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, 91(2), 301–311. doi: 10.1037/0022-0663.91.2.301 42
- Winne, P. H. (2004). Students' calibration of knowledge and learning processes : Implications for designing powerful software learning environments. *International Journal of Educational Research*, 41(6), 466–488. doi: 10.1016/j.ijer.2005.08.012 13
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Hillsdale, NJ : Erlbaum. 1, 149, 180, 181
- Wolters, C. A., Yu, S. L., & Pintrich, P. R. (1996). The relation between goal orientation and students' motivational beliefs and self-regulated learning. *Learning and individual differences*, 8(3), 211–238. doi: 10.1016/S1041-6080(96)90015-1 154
- Yates, J. F. (1990). *Judgment and decision making*. Englewood Cliffs, NJ : Prentice-Hall. 49
- Yergeau, É. (2009). Étude sur la puissance statistique des devis de recherche en éducation. *Revue des sciences de l'éducation*, 35(2), 199–221. doi: 10.7202/038735ar 147
- Zaromb, F. M., Karpicke, J. D., & Roediger III, H. L. (2010). Comprehension as a basis for metacognitive judgments : effects of effort after meaning on recall and metacognition. *Journal of experimental psychology : learning, memory, and cognition*, 36(2), 552–557. doi: 10.1037/a0018277 23
- Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of educational psychology*, 81(3), 329–339. doi: 10.1037/0022-0663.81.3.329 1, 2, 180
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement : An overview. *Educational psychologist*, 25(1), 3–17. doi: 10.1207/s15326985ep2501\_2 6, 153, 180
- Zimmerman, B. J. (2000). Attaining self-regulation. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press, San Diego, CA. 58

---

## ANNEXES

---

ANNEXE A :Les principaux modèles de l'apprentissage auto-régulé

ANNEXE B :Précisions sur les méthodes de calcul de l'exactitude de l'auto-évaluation

ANNEXE C :Intégralité du matériel de l'expérience 1

ANNEXE D :Résultats complémentaires de l'expérience 2

ANNEXE E :Compléments d'informations sur le matériel utilisé dans l'expérience 3

ANNEXE F :La théorie de la charge cognitive : compléments d'informations



# A

---

## L'AUTO-ÉVALUATION DANS LES PRINCIPAUX MODÈLES DE L'APPRENTISSAGE AUTORÉGULÉ

---

Le modèle de la métamémoire de (Nelson & Narens, 1990) pose les fondations des modèles de l'apprentissage autorégulé en définissant les principes fondamentaux telle que la relation qui unit l'auto-évaluation à l'autorégulation. Dans ce modèle, les auteurs explicitent trois principes : 'les processus cognitifs sont divisés en au moins deux niveaux spécifiques inter-reliés' (p.125). Ici, les deux niveaux sont le niveau 'objet' et le niveau 'méta' ; 'le niveau méta contient un modèle dynamique (e.g., une simulation mentale) du niveau objet' (p.126) ; 'il y a deux relations dominante, appelée 'contrôle' et 'évaluation', qui sont définies en terme de direction du flux d'information entre le niveau méta et le niveau objet. En clair, l'apprenant va traiter les informations du niveau cognitif (= niveau objet) et 'informer' le niveau métacognitif. La représentation qu'a le niveau 'méta' de la situation au niveau 'objet' va donc être modifiée (ou non) selon l'information reçue. Grâce à cette information, le niveau métacognitif va en retour modifier (ou non) le niveau cognitif. Trois types de modifications sont possibles, (1) l'initiation d'une action ; (2) la poursuite d'une action déjà engagée ; (3) l'arrêt d'une action déjà commencée. Cette régulation au niveau 'objet' va modifier l'état cognitif du sujet qui va renvoyer de nouvelles informations au niveau métacognitif, etc. Cette boucle prend fin lorsque l'apprenant décide d'arrêter l'épisode d'apprentissage, soit pour une raison environnementale (e.g., contrainte temporelle), soit parce qu'il estime que son objectif d'apprentissage est atteint.

Ce modèle met en exergue l'importance du processus d'auto-évaluation. En effet, si le niveau méta au travers du processus d'auto-évaluation informe l'apprenant qu'une information est parfaitement intégrée en mémoire, celui-ci pourra décider de stopper le processus d'apprentissage. Si cette l'auto-évaluation est inexacte et incongruente avec l'état réel des connaissances de l'apprenant, alors celui-ci pourra soit continuer d'apprendre inutilement (dans le cas où il se sous-évalue) ou arrêter trop précocément (dans le cas où il se sur-évalue). Le modèle de (Nelson & Narens, 1990) permet de comprendre la relation entre auto-évaluation et autorégulation dans une tâche de mémorisation. D'autres modèles s'en



sont inspirés pour expliquer l'apprentissage autorégulé, seuls les trois principaux, à savoir ceux de [Pintrich et De Groot \(1990\)](#) ; [Winne et Hadwin \(1998\)](#) ; [Zimmerman \(1990\)](#) seront présentés ici, de façon globale et avec l'objectif de discuter la place consacrée au processus d'auto-évaluation.

Ces modèles s'opposent avant tout sur l'orientation qui les gouverne respectivement puisque les modèles de [Pintrich et De Groot \(1990\)](#) et de [Zimmerman \(1990\)](#) ont une orientation motivationnelle tandis que le modèle de ([Winne & Hadwin, 1998](#)) est davantage axé sur les répercussions métacognitives de l'apprentissage autorégulé. En revanche les trois modèles s'accordent à dire que l'apprentissage autorégulé est un processus cyclique et itératif qui se compose à minima d'une phase de définition de la tâche et/ou de 'planification' et d'une phase de 'performance', une troisième (ou quatrième) phase d'auto-évaluation peut venir clôturer le cycle.

Lors de la phase de représentation de la tâche (qui est confondue avec celle de planification chez [Zimmerman \(1990\)](#)) les apprenants vont activer des connaissances antérieures et des buts motivationnels, mais également définir leur objectif d'apprentissage, le 'standard' à atteindre. Cette représentation détermine la planification de l'action sur les plans environnemental (e.g., temporel), cognitif et métacognitif (i.e. stratégies à utiliser).

Ensuite, les apprenants vont mettre en place les stratégies définies dans l'étape précédente (i.e. traiter le contenu à apprendre dans le but de l'intégrer). D'après [Zimmerman \(1989\)](#), les apprenants doivent également pendant cette phase s'auto-observer pour que, lors de la phase suivante, ils puissent s'auto-évaluer sur la base de ces informations.

Enfin, pour [Pintrich et De Groot \(1990\)](#) et [Zimmerman \(1990\)](#), les apprenants évaluent la phase d'apprentissage afin de déterminer l'efficacité des stratégies cognitives et métacognitives mises en place précédemment ainsi que le niveau d'atteinte de leur objectif. Le résultat de cette évaluation permettra aux apprenants de réguler leurs comportements pour entamer ou non un nouveau cycle d'apprentissage.

Dans ces deux modèles, la place de l'auto-évaluation pourrait être considérée comme une limite. En effet, pour [Pintrich et De Groot \(1990\)](#) et [Zimmerman \(1990\)](#), l'auto-évaluation ne se fait pas de manière continue mais ponctuellement lors d'une phase bien définies. Même si dans le modèle de [Pintrich et De Groot \(1990\)](#), l'auto-évaluation apparaît à deux reprises (i.e. après la planification et après la performance), ce processus n'est pas omniprésent comme le considèrent [Winne et Hadwin \(1998\)](#). La [Table 9](#) reprend les différentes phases de l'apprentissage autorégulé selon les trois modèles présentés.

Ainsi, [Winne et Hadwin \(1998\)](#) considèrent le processus d'auto-évaluation comme permanent, quelle que soit la phase de l'apprentissage

autorégulé. Il ne leur semble donc pas nécessaire de dédier une phase entière à ce processus. En revanche, les auteurs proposent une étape finale et optionnelle au cycle de l'apprentissage autorégulé qui serait un 'bilan métacognitif'. Suite aux trois premières phases mais aussi grâce au processus sous-jacent d'auto-évaluation, les apprenants peuvent extraire des informations sur l'efficacité du cycle d'apprentissage et acquérir de nouvelles connaissances métacognitives. En intégrant cette étape, [Winne et Hadwin \(1998\)](#) expliquent, du fait de l'acquisition de connaissances métacognitives supplémentaires, que les experts d'un domaine développent de meilleures habiletés métacognitives, notamment en terme d'exactitude de l'auto-évaluation. Les auteurs pensent d'ailleurs que les habiletés métacognitives peuvent s'automatiser avec le développement de l'expertise.

TABLEAU 9. – Les modèles de l'apprentissage autorégulé.

	Modèle de Pintrich	Modèle de Zimmerman	Modèle de Winne et Hadwin
Définition de la tâche	o	o	o
Planification	o		o
Performance	o	o	o
Auto-évaluation	o	o	x
Adaptation de la métacognition	x	x	o



# B

---

## PRÉCISIONS SUR LES MÉTHODES DE CALCUL DE L'EXACTITUDE DE L'AUTO-ÉVALUATION

---

### B.1 LE GAMMA DE GOODMAN-KRUSKAL

À partir du Tableau 10, le coefficient se calcule de la façon suivante (Schraw, 1995) :

$$\text{Gamma} = (adb) / (ad + bc)$$

Par exemple : un participant a lu un texte et doit répondre à six questions traitant du texte afin de mesurer sa compréhension des concepts du texte. Avant chaque question, il a dû estimer à quel point il pensait pouvoir répondre correctement à chaque question. La réponse peut-être binaire ou sur une échelle type Lickert en quelques points.

Schraw (1995) a fait de nombreuses critiques sur le Gamma de Goodman-Kruskal tel que le fait qu'il nécessite plusieurs observations par individus. Pour que le Gamma soit 'fiable', il faut que toutes les valeurs du tableau soient différentes de 0 et que 80 % des données contiennent au moins cinq observations. Par ailleurs, le Gamma peut être égal à 1 (i.e. indiquer que l'exactitude relative de l'auto-évaluation est élevée) tandis que l'auto-évaluation n'est pas exacte. Par exemple, lorsque  $a = 1$  ;  $b = 10$  ;  $c = 0$  ;  $d = 1$ .

TABLEAU 10. – Tableau permettant le calcul du coefficient Gamma de Goodman-Kruskal

	réponse jugée bonne	Réponse jugée fausse
Réponse juste	A = auto-évaluation exacte	B = biais d'auto-évaluation
Réponse fausse	C = biais d'auto-évaluation	D = auto-évaluation exacte

### B.2 LES COURBES DE CALIBRATION

Le calcul des courbes de calibration nécessite plusieurs étapes et la procédure décrite ci-après est issue de l'article de Lichtenstein et al. (1982).

Pour calculer les courbes, il faut : (1) collecter des valeurs de jugement d'évaluation pour plusieurs items, (2) regrouper les jugements similaires en différentes classes. Par exemple, si le jugement a été recueilli sur une échelle de 0 à 100, les évaluations de 0 à 9 sont catégorisées ensemble, puis celles de 10 à 19, etc. (3) Pour chaque 'classe' de jugement, il faut calculer la proportion d'items corrects puis (4) rapporter la réponse moyenne sur les abscisses correspondante à la proportion qui est correcte sur les ordonnées (Lichtenstein et al., 1982). Si la moyenne des biais (différence entre auto-évaluation et performance) se situe au-dessus de la droite ( $y = x$ ), les individus se sur-estiment, si la moyenne des biais est au-dessous de la droite, les individus se sous-estiment (Koriat et al., 1980).



---

## INTÉGRALITÉ DU MATÉRIEL DE L'EXPÉRIENCE

1

---

### c.1 QUESTIONNAIRE DE MESURE DE CONNAISSANCES ANTÉRIEURES

1. Quelles sont les propriétés que l'on peut affecter à un style de paragraphe ?
  - L'en-tête de page
  - Les tabulations
  - L'alignement horizontal
  - Les retraits
  - Les marges
  - Aucune des réponses précédentes
  - Je ne sais pas
2. Dans un traitement de texte, afin de réaliser une table des matières à deux niveaux, que doit-on mettre en place ?
  - Des styles pour les niveaux de titre 1 et 2
  - Un signet pour chacun des titres
  - Un index des titres
  - Un tableau avec tous les titres des deux premiers niveaux hiérarchiques
  - Aucune des réponses précédentes
  - Je ne sais pas
3. Dans un traitement de texte, si on souhaite automatiser l'indexation des légendes, il est nécessaire d'utiliser ...
  - Un signet pour chacun des titres
  - Un style
  - Deux styles
  - Une table des matières
  - Aucune des réponses précédentes
  - Je ne sais pas
- ?
4. Si on souhaite faire apparaître le nom du chapitre courant en tête de chaque page, il est nécessaire d'utiliser ...

- Un style
  - Une formule
  - Une formule et un style
  - Un signet
  - Aucune des réponses précédentes
  - Je ne sais pas
5. La modification d'un style de paragraphe sert notamment à . . .
- Modifier l'apparence de tous les paragraphes du document
  - Modifier l'apparence de tous les titres du document
  - Modifier l'apparence des paragraphes auxquels le style était appliqué
  - Modifier l'ordre d'apparition des paragraphes concernés
  - Aucune des réponses précédentes
  - Je ne sais pas
6. Dans un traitement de texte, comment force-t-on le passage à la ligne sans changer de paragraphe ?
- On insère des espaces en fin de ligne
  - On insère un saut de ligne
  - On diminue le retrait droit de la ligne
  - On insère un caractère de tabulation
  - On appuie sur la touche Entrée
  - Aucune des réponses précédentes
  - Je ne sais pas
7. Quelles sont les caractéristiques d'un paragraphe justifié ?
- Le texte est centré
  - Le texte est aligné à droite ou à gauche
  - Le texte est aligné à gauche et à droite
  - Les marges de droites et de gauche ont la même valeur
  - Les retraits droits et gauche (avant/après) ont la même valeur
  - Aucune des réponses précédentes
  - Je ne sais pas
8. Quelle est la propriété propre aux caractères
- La justification
  - Le soulignement
  - Le centrage
  - L'alignement
  - Aucune des réponses précédentes
  - Je ne sais pas
9. Dans un traitement de texte, on souhaite diminuer la largeur d'un paragraphe. Quelle est la meilleure méthode pour cela ?
- Placer des taquets de tabulations à droite et à gauche du paragraphe

- Modifier les marges du paragraphe
  - Modifier les retraits du paragraphe
  - Centrer le paragraphe
  - Diminuer la taille des caractères
  - Aucune des réponses précédentes
  - Je ne sais pas
10. Pour appliquer la même mise en forme à tous les titres d'un document, quelle est la méthode recommandée ?
- On utilise la commande 'Rechercher/Remplacer'
  - On utilise le pinceau
  - On utilise la table des matières
  - On utilise un style
  - On utilise la commande 'Copier' puis 'Coller'
  - Aucune des réponses précédentes
  - Je ne sais pas
11. Dans un document, pour modifier la mise en forme de tous les titres de style 'Titre 2', quelle est la méthode la plus efficace ?
- Il faut créer un nouveau style basé sur le style 'Titre 2'
  - Il faut modifier le style 'Titre 2' et l'appliquer de nouveau aux titres de même niveau
  - Il faut utiliser pour chaque titre la commande 'Copier' puis 'Coller'
  - Il faut utiliser pour chaque titre le pinceau pour reproduire la mise en forme
  - Il faut seulement modifier le style 'Titre 2'
  - Aucune des réponses précédentes
  - Je ne sais pas
12. Que faut-il faire pour automatiser la mise en forme d'un document avec un logiciel de traitement de texte ?
- Il faut insérer une table des matières
  - Il faut utiliser les styles
  - Il faut convertir le document au format RTF
  - Il faut mettre le texte en forme ligne par ligne à l'aide du pinceau
  - Aucune des réponses précédentes
  - Je ne sais pas

## c.2 QUESTIONNAIRE SIMILAIRE

1. Quelle(s) propriété(s) n'a aucun sens pour formater un caractère dans un traitement de texte ?
- Le retrait
  - La casse



- La police
  - La couleur
  - Aucune des réponses précédentes
  - Je ne sais pas
2. Dans un traitement de texte, quelle est la meilleure façon d’avoir un saut de ligne au début de chaque paragraphe d’un groupe de paragraphes
- Appliquer un format de paragraphe ‘double interligne’ en ayant le curseur dans le groupe de paragraphes
  - Appliquer un format de paragraphe ‘saut de ligne avant’ en sélectionnant le groupe de paragraphes
  - Appliquer un double interligne en sélectionnant le groupe de paragraphes
  - Insérer un saut de ligne avant chacun des paragraphes
  - Aucune des réponses précédentes
  - Je ne sais pas
3. Dans un traitement de texte, les styles de paragraphes ne peuvent pas être associés à :
- Un paragraphe
  - Plusieurs paragraphes
  - Une partie de paragraphe
  - Plusieurs paragraphes disjoints (qui ne suivent pas dans le texte)
  - Aucune des réponses précédentes
  - Je ne sais pas
4. Dans un traitement de texte, si on applique un style de paragraphes à différents paragraphes, puis on modifie ce style en le passant en alignement à gauche
- Tous les paragraphes auxquels on appliquera ce style par la suite seront alignés à gauche
  - Tous les paragraphes auxquels on appliquera ce style seront alignés à gauche
  - Tous les paragraphes qui n’avaient pas d’alignement avant l’application seront alignés à gauche
  - Tous les paragraphes sans mise en forme seront alignés à gauche
  - Aucune des réponses précédentes
  - Je ne sais pas
5. Dans un traitement de texte, afin d’automatiser la génération d’une table des matières, il est nécessaire de
- Mettre en forme les titres avec une indentation correcte
  - Appliquer des styles hiérarchisés aux titres
  - Créer sa propre hiérarchie de styles de titres
  - Appliquer les styles Titre1, Titre2, etc à la table des matières

- Aucune des réponses précédentes
  - Je ne sais pas
6. Si on souhaite mettre en forme une table des matières générée automatiquement, il faut
- Mettre en forme la table des matières qui a été générée
  - Ce n'est pas possible
  - Modifier les styles de titre
  - Modifier des styles spécifiques
  - Aucune des réponses précédentes
  - Je ne sais pas

### c.3 QUESTIONS DE GÉNÉRALISATION

1. A quel type de référence fait appel la notation suivante : \$B\$3 ?
- La référence monétaire
  - La référence absolue
  - La référence relative
  - La référence symbolique
  - Aucune des réponses précédentes
  - Je ne sais pas
2. Dans un tableur, de quoi est formé un classeur ?
- De sections
  - De cellules
  - De pages
  - De feuilles
  - Aucune des réponses précédentes
  - Je ne sais pas
3. Dans un tableur, pour écrire une formule dont le calcul utilise d'autres cellules, que faut-il ?
- Il faut que la formule contienne les valeurs des cellules sur lequel porte le calcul
  - Il faut que la formule contienne les références des cellules sur lequel porte le calcul
  - Il faut que les données sur lequel porte le calcul soient alignées
  - Il faut insérer la calculatrice de Windows dans la feuille de calcul
  - Aucune des réponses précédentes
  - Je ne sais pas
4. Dans un tableur, combien de cellules désigne la référence A1 :A3 ;D3 ?
- 3
  - 4

- 6
  - 12
  - Aucune des réponses précédentes
  - Je ne sais pas
5. Dans un tableur, que doit-on faire pour qu'un nombre s'affiche avec le format monétaire \$ ?
- On ne saisit '\$' qu'une fois et les autres cellules seront transformées automatiquement
  - On ne peut pas le faire car \$ sert aux références absolues : il faut mettre le symbole \$ dans la colonne précédente
  - On est obligé de saisir '\$' à côté de chaque nombre
  - On sélectionne les cellules et on change le format de nombre
  - Aucune des réponses précédentes
  - Je ne sais pas
6. Dans un tableur, si on copie la formule '= A\$1 \* B4' de A4 en B5, quelle formule obtient-on en B5 ?
- = A\$2 \* C5
  - = B2 \* C5
  - = B\$1 \* C5
  - = A\$1 \* C4
  - Aucune des réponses précédentes
  - Je ne sais pas
7. Dans un tableur, comment peut-on référencer la plage contenant les cellules B3, C3 et D3 ?
- B3 ;D3
  - B3-D3
  - B3 :D3
  - B3 :E3-D3
  - Aucune des réponses précédentes
  - Je ne sais pas

#### c.4 RÉSULTATS COMPLÉMENTAIRES

##### c.4.1 *Statistiques descriptives selon le type de connaissance*

L'ensemble des valeurs présentées ici sont synthétisées dans le Tableau 11 pour ce qui est des connaissances procédurales et dans le Tableau 12 pour ce qui est des connaissances déclaratives.

Sur le plan descriptif, il apparaît que les performances étaient plus élevées au questionnaire de généralisation ( $M = 85.03$  ;  $SD = 21.47$ ) que pour les deux premiers ( $M = 74.66$  ;  $SD = 12.18$  pour le questionnaire de mesure des connaissances antérieures et  $M = 72.86$  ;  $SD = 25.28$  pour le

TABLEAU 11. – Ensemble des moyennes (et écarts-types) de la performance, du jugement d'auto-évaluation et de la direction du biais pour les connaissances procédurales selon la condition expérimentale.

	groupe expérimental	groupe contrôle	total	
performance	connaissances antérieures	70.19 (14.25)	73.6 (12.76)	71.9 (13.5)
	questions similaires	61.73 (20.05)	62.62 (19.14)	62.17 (19.4)
	questions de généralisation	59.23 (32.23)	59.2 (26.4)	59.21 (29.15)
	total	63.71 (17.02)	65.14 (11.55)	64.43 (14.41)
jugement d'auto-évaluation	connaissances antérieures	71.51 (15.25)	70.03 (14.34)	70.77 (14.67)
	questions similaires	67.19 (20.08)	72.1 (18.44)	69.65 (19.23)
	questions de généralisation	57.99 (29.01)	66.74 (17.12)	62.36 (23.98)
	total	65.56 (14.51)	69.62 (10.46)	67.59 (12.68)
direction du biais d'auto-évaluation	connaissances antérieures	1.54 (18.29)	-3.37 (19.3)	-91 (18.77)
	questions similaires	5.47 (27.10)	9.47 (22.10)	7.47 (24.55)
	questions de généralisation	-1.24 (30.61)	7.54 (26.3)	3.15 (28.58)
	total	1.92 (17.78)	4.55 (15.1)	3.24 (16.36)

TABLEAU 12. – Ensemble des moyennes (et écarts-types) de la performance, du jugement d'auto-évaluation et de la direction du biais pour les connaissances déclaratives selon la condition expérimentale

	groupe expérimental	groupe contrôle	total	
Performance	connaissances antérieures	73.92 (12.82)	75.4 (11.74)	74.66 (12.18)
	questions similaires	75.43 (25.6)	70.3 (25.84)	72.86 (25.58)
	questions de généralisation	77.04 (26.4)	93.02 (10.58)	85.03 (21.47)
jugement d'auto-évaluation	total	75.46 (15.41)	79.58 (11.97)	77.52 (13.81)
	connaissances antérieures	69.88 (13.77)	69.97 (12.03)	69.92 (12.79)
	questions similaires	64.18 (18.82)	67.88 (16.46)	66.03 (17.59)
	questions de généralisation	66.93 (26.67)	78.60 (16.12)	72.77 (22.59)
	total	66.99 (15.62)	72.15 (7.75)	69.57 (12.47)
direction du biais d'auto-évaluation	connaissances antérieures	-3.99 (16.48)	-5.38 (15.76)	-4.69 (15.97)
	questions similaires	-11.24 (27.06)	-2.43 (28.08)	-6.84 (27.64)
	questions de généralisation	-10.12 (17.59)	-14.42 (18.11)	-12.27 (17.8)
	total	-8.45 (14.73)	-7.41 (12.85)	-7.93 (13.69)

questionnaire similaire) uniquement pour les connaissances déclaratives. Cette tendance s'inverse avec les connaissances procédurales puisque les performances sont moins bonnes au questionnaire de généralisation ( $M = 59.21$  ;  $SD = 29.15$ ) que pour les deux précédents ( $M = 71.9$  ;  $SD = 13.5$  pour le questionnaire de mesure des connaissances antérieures et  $M = 62.17$  ;  $SD = 19.4$  pour le questionnaire similaire). Quel que soit le type de connaissances, les participants du groupe contrôle ont des performances sensiblement meilleures ( $M = 79.58$  ;  $SD = 11.97$  pour les connaissances déclaratives et  $M = 65.14$  ;  $SD = 11.55$  pour les connaissances procédurales) que les participants du groupe expérimental ( $M = 75.46$  ;  $SD = 15.41$  pour les connaissances déclaratives et  $M = 63.71$  ;  $SD = 17.02$  pour les connaissances procédurales).

Pour ce qui est du jugement d'auto-évaluation, les participants ont évalués plus fortement leurs performances aux questions déclaratives lors du questionnaire de généralisation ( $M = 72.77$  ;  $SD = 22.59$ ) que lors des deux questionnaires précédents ( $M = 69.92$  ;  $SD = 12.79$  pour le questionnaire de mesure des connaissances antérieures et  $M = 66.03$  ;  $SD = 17.59$  pour le questionnaire similaire). Par ailleurs, les participants du groupe contrôle ( $M = 72.15$  ;  $SD = 7.75$ ) ont eu tendance à estimer légèrement plus fortement leurs performances que les participants du groupe expérimental ( $M = 66.99$  ;  $SD = 15.62$ ). Au contraire, lorsque les participants ont estimé leurs performances aux questions procédurales, leurs estimations étaient plus fortes au questionnaire de mesure des connaissances antérieures ( $M = 70.77$  ;  $SD = 14.67$ ) qu'au questionnaire similaire ( $M = 69.65$  ;  $SD = 19.23$ ) et plus encore qu'au questionnaire de généralisation ( $M = 62.36$  ;  $SD = 23.98$ ). À nouveau, les participants du groupe contrôle ( $M = 69.62$  ;  $SD = 10.46$ ) ont estimé mieux réussir que les participants du groupe expérimental ( $M = 65.56$  ;  $SD = 14.51$ ).

Lors de l'évaluation des connaissances déclaratives, les participants se sont davantage sous-évalués lors du questionnaire de généralisation ( $M = -7.93$  ;  $SD = 13.69$ ) que lors des précédents questionnaires ( $M = -4.69$  ;  $SD = 15.97$  pour le questionnaire des connaissances antérieures et  $M = -6.84$  ;  $SD = 27.64$  pour le questionnaire similaire). Lors de l'évaluation des connaissances procédurales, les participants tendaient à ne pas avoir de biais lors du questionnaire de mesure des connaissances antérieures ( $M = .91$  ;  $SD = 18.77$ ) puis ils se sont davantage sur-estimés sur le questionnaire similaire ( $M = 7.47$  ;  $SD = 24.55$ ) que sur le questionnaire de généralisation ( $M = 3.15$  ;  $SD = 28.58$ ). Les participants du groupe contrôle ( $M = -7.41$  ;  $SD = 12.85$ ) ont eu tendance à se sous-estimer autant que les participants du groupe expérimental ( $M = -8.45$  ;  $SD = 14.73$ ) sur les questions déclaratives ; tandis qu'ils ont eu tendance à se sur-estimer très sensiblement ( $M = 4.55$  ;  $SD = 15.1$ ) sur les questions procédurales

quand les participants du groupe expérimental étaient plus exacts ( $M = 1.92$ ;  $SD = 17.78$ ).

c.4.2 *Effet du type de connaissances, du moment de la mesure et de la condition expérimentale sur la performance, le jugement d'auto-évaluation et la direction du biais d'auto-évaluation*

Des Anova  $3 \times 2 \times 2$  ont été conduites sur la performance, le jugement d'auto-évaluation et la direction du biais d'auto-évaluation avec pour variables indépendantes le moment de la mesure (avant le dispositif versus après sur les questions similaires et sur les questions de généralisation); la 'condition expérimentale' (dispositif d'aide à l'auto-évaluation versus dispositif contrôle) et le type de connaissances (déclaratives versus procédurales).

*Résultats sur la performance*

L'Anova à trois facteurs conduite sur la performance n'a pas révélé d'interaction entre nos trois variables indépendantes,  $F(2, 92) = 2.69$ ,  $p = .1$ . Aucune interaction entre le moment et la condition expérimentale n'a été révélé,  $F(2, 92) = 1.47$ ,  $p = .23$ ; entre la condition expérimentale et le type de connaissance,  $F(1, 46) = .46$ ,  $p = .5$  n'a été détectée.

En revanche, une interaction entre le moment de la mesure et le type de connaissances,  $F(2, 92) = 10.75$ ,  $p < .001$ ,  $\eta_p^2 = .19$  a été détectée. Trois t-tests à mesures appariés ont été conduits, le seuil de significativité a donc été divisé par trois et est devenu égal à  $.017$ . Les analyses ne révèlent aucune différence significative entre les performances obtenues aux questions déclaratives et procédurales pour les questionnaires de mesure des connaissances antérieures,  $t(47) = 1.16$ ,  $p = .25$  tandis qu'une différence significative est apparue pour les performances au questionnaire similaire,  $t(47) = 2.78$ ,  $p = .008$  et les performances au questionnaire de généralisation,  $t(47) = 6.17$ ,  $p < .001$ . Les données suggèrent que les participants réussissent davantage les questions déclaratives au questionnaire similaire ( $M = 72.86$ ;  $SD = 25.58$ ) et au questionnaire de généralisation ( $M = 85.03$ ;  $SD = 21.47$ ) que les questions procédurales ( $M = 62.17$ ;  $SD = 19.4$  pour le questionnaire similaire et  $M = 59.21$ ;  $SD = 29.15$  pour le questionnaire de généralisation).

Par ailleurs, aucun effet principal du moment,  $F(2, 92) = 2.14$ ,  $p > .05$  ou de la condition expérimentale,  $F(1, 46) = .6$ ,  $p > .05$  n'a été révélé. Seul un effet principal du type de connaissance a été révélé,  $F(1, 46) = 43.92$ ,  $p < .001$ ,  $\eta_p^2 = .49$  suggérant que les participants réussissent globalement mieux les questions déclaratives ( $M = 77.52$ ,  $SD = 13.81$ ) que procédurales ( $M = 64.43$ ,  $SD = 14.41$ ).

### *Résultats sur le jugement d'auto-évaluation*

Comme précédemment, aucune interaction entre nos trois variables,  $F(2, 92) = .25, p = .78$  n'a été révélé par l'Anova, de même qu'aucune interaction entre le moment et condition expérimentale,  $F(2, 92) = 1.47, p = .24$  ou entre condition expérimentale et le type de connaissance,  $F(1, 46) = .25, p = .62$  n'a été révélé.

Aucun effet principal du moment de la mesure,  $F(2, 92) = .46, p = .63$  ou de la condition expérimentale n'a été détecté,  $F(1, 46) = 1.8, p = .19$  et l'effet du type de connaissance était marginalement significatif,  $F(1, 46) = 3.29, p = .08, \eta_p^2 = .07$ .

Seule une interaction entre le moment de la mesure et le type de connaissance a été détectée,  $F(2, 92) = 12.55, p < .001, \eta_p^2 = .21$ . L'analyse des t-tests a révélé une absence de différence selon le type de connaissance pour le questionnaire de mesure des connaissances antérieures,  $t(47) = -.58, p = .56$  et le questionnaire similaire,  $t(47) = -1.59, p = .12$  tandis qu'une différence est apparue pour le questionnaire de généralisation,  $t(47) = 4.7, p < .001$  suggérant que les participants ont évalué leurs performances aux questions déclaratives ( $M = 72.77$ ;  $SD = 22.59$ ) plus fortement qu'aux questions procédurales ( $M = 62.36$ ;  $SD = 23.98$ ).

### *Résultats sur la direction du biais d'auto-évaluation*

L'analyse de l'anova à trois facteurs n'a pas révélé d'interaction entre nos trois variables,  $F(2, 92) = 1.52, p = .22$  ni entre le moment de la mesure et le type de connaissance,  $F(2, 92) = 2.53, p = .09$ , entre condition expérimentale et le type de connaissance,  $F(1, 46) = .15, p = .7$ , ou entre le moment de la mesure et condition expérimentale,  $F(2, 92) = 1.17, p = .32$ .

De même, aucun effet du moment,  $F(2, 92) = 1.24, p = .29$  ou de la condition expérimentale n'a été détecté,  $F(1, 46) = .22, p = .64$ . Un effet principal du type de connaissance a été détecté,  $F(1, 46) = 29.71, p < .001, \eta_p^2 = .39$  suggérant que les participants se sont davantage sous-estimés sur les questions déclaratives ( $M = -7.93$ ;  $SD = 13.69$ ) et ont eu tendance à se sur-estimer sur les questions procédurales, ( $M = 3.24$ ;  $SD = 16.36$ ).

#### *c.4.3 Distinction selon le sous domaine de compétence*

Les analyses considérant le sous domaine des questions sont maintenant présentées. À nouveau, des Anovas  $2 \times 3 \times 2$  ont été conduites avec en variables indépendantes le moment de la mesure (avant le dispositif versus après sur des questions similaires); la 'condition expérimentale' (dispositif d'aide à l'auto-évaluation versus dispositif contrôle) et le sous domaine



de compétence (fonctionnalité de base du traitement de texte ; feuilles de styles ; automatisation des styles). L'ensemble des moyennes et écarts-types pour les trois variables dépendantes présentées sont présentés dans le Tableau 13 pour le domaine du traitement de texte, dans le Tableau 14 pour le domaine des styles et dans le Tableau 15 pour le domaine de l'automatisation.

#### c.4.4 *Statistiques descriptives selon le sous-domaine de compétence*

Les performances des participants aux trois sous domaines étaient particulièrement hautes puisque la performance moyenne la plus basse était de 59% de réussite pour les questions relatives à l'automatisation.

Plus précisément, les performances aux questions relatives au traitement de texte étaient plus élevées lors du questionnaire similaire ( $M = 74.94$  ;  $SD = 24.11$ ) que lors de la mesure des connaissances antérieures ( $M = 68.38$  ;  $SD = 16.04$ ). De plus, sur ce domaine, les participants du groupe expérimental ont été sensiblement meilleurs ( $M = 72.68$  ;  $SD = 13.04$ ) que les participants du groupe contrôle ( $M = 70.65$  ;  $SD = 17.53$ ). Relativement au sous domaine des styles, les participants du groupe contrôle ( $M = 79.65$  ;  $SD = 15.19$ ) et du groupe expérimental ( $M = 81.17$  ;  $SD = 19.44$ ) ont obtenu des performances similaires. En revanche, les participants ont globalement eu de meilleures performances au questionnaire de mesure des connaissances antérieures ( $M = 86.39$  ;  $SD = 11$ ) qu'au questionnaire similaire ( $M = 74.43$  ;  $SD = 30.07$ ). Enfin, pour ce qui est du domaine de l'automatisation, les participants du groupe expérimental ( $M = 57.1$  ;  $SD = 16.18$ ) ont été à peine moins performants que les participants du groupe contrôle ( $M = 61.15$  ;  $SD = 13.75$ ). Globalement, les participants ont été plus performants sur le questionnaire des connaissances antérieures ( $M = 65.07$  ;  $SD = 18.29$ ) que sur le questionnaire similaire ( $M = 53.18$  ;  $SD = 21.9$ ).

Concernant le jugement d'auto-évaluation des questions sur le traitement de texte, les participants ont émis un jugement plus haut sur leurs performances au questionnaire de connaissances antérieures ( $M = 70.84$  ;  $SD = 16.01$ ) que sur leurs performances au questionnaire similaire ( $M = 63.69$  ;  $SD = 20.5$ ). Les participants du groupe contrôle ( $M = 67.02$  ;  $SD = 17.15$ ) et expérimental ( $M = 67.51$  ;  $SD = 14.69$ ) ont estimé leurs performances de façon égale. Sur les questions relatives au domaine des styles, le pattern est comparable principalement selon le moment de la mesure. En effet, le jugement des participants était globalement haut ( $M = 71.53$  ;  $SD = 14.7$ ) et les participants ont prédit avoir de meilleures performances lors du questionnaire de mesure des connaissances antérieures ( $M = 77.07$  ;  $SD = 14.15$ ) que lors du questionnaire similaire, ( $M = 65.98$  ;  $SD = 21.36$ ). Par ailleurs, les participants du groupe contrôle ( $M = 73.73$  ;

SD = 12.56) ont émis un jugement moyen sensiblement supérieur à celui des participants du groupe expérimental (M = 69.32 ; SD = 16.55). Sur le domaine de l'automatisation, les participants ont à l'inverse émis un jugement moyen plus haut sur les questions similaires (M = 73.84 ; SD = 16.89) que sur les questions de mesure des connaissances antérieures (M = 63.13 ; SD = 16.08). Le groupe expérimental (M = 67.75 ; SD = 15.76) et contrôle (M = 69.22 ; SD = 12.38) avaient des moyennes comparables, bien que le jugement moyen du groupe contrôle soit légèrement plus élevé.

Pour les questions relatives au traitement de texte, la direction du biais de l'ensemble des participants était légèrement positive au questionnaire de mesure des connaissances antérieures (M = 2.77 ; SD = 20.53), les apprenants se sur-évaluaient à peine. Cette moyenne a baissé lors du questionnaire similaire (M = -11.26 ; SD = 29.61), les participants ont eu davantage tendance à sous-estimer leurs performances. Sur ce domaine, les participants du groupe expérimental (M = -5.01 ; SD = 18.26) se sous-évaluaient sensiblement plus que les participants du groupe contrôle (M = -3.48 ; SD = 22.3).

Concernant le domaine des styles, la direction moyenne du biais indique de façon globale que les participants tendaient à se sous-estimer ((M = -8.84 ; SD = 18.29). Les participants du groupe expérimental (M = -11.8 ; SD = 21) tendaient généralement à se sous-évaluer davantage que les participants du groupe contrôle (M = -5.87 ; SD = 14.98). En revanche, le biais moyen était similaire pour les questions de mesure des connaissances antérieures (M = -9.22 ; SD = 16) et les questions similaires (M = -8.45 ; SD = 28.54). Enfin, sur le domaine de l'automatisation, les participants ont eu tendance à sur-évaluer leurs performances ((M = 9.36 ; SD = 19.45). Les participants du groupe expérimental (M = 10.65 ; SD = 22.6) se sur-estimer un peu plus que les participants du groupe contrôle (M = 8.07 ; SD = 16.09) et les participants se sur-estimer beaucoup plus lors du questionnaire similaire (M = 20.66 ; SD = 27.26) que lors du questionnaire de mesure des connaissances antérieures (M = -1.94 ; SD = 21.52).

#### c.4.5 *Effet du sous-domaine de compétence, du moment de la mesure et de la condition expérimentale sur la performance, le jugement d'auto-évaluation et la direction du biais d'auto-évaluation*

##### *Effet sur l'amplitude des biais*

Une Anova à trois facteurs, n'a pas révélé d'interaction entre la condition expérimentale, le moment de la mesure et le sous domaine de compétence,  $F(2, 92) = .28, p = .8$ . Aucune interaction n'a été révélé entre le sous domaine et le moment de la mesure,  $F(2, 92) = 1.26, p = .3$ , entre le moment et la condition expérimentale,  $F(2, 92) = 2.16, p = .15$  ou entre

TABLEAU 13. – Ensemble des moyennes (et écarts-types) de la performance, du jugement d'auto-évaluation et de la direction du biais pour les connaissances sur le domaine du traitement de texte selon la condition expérimentale

		groupe expérimental	groupe contrôle	total
performance	connaissances antérieures	69.02 (18.32)	67.75 (13.77)	68.38 (16.04)
	questions similaires	76.35 (20.24)	73.54 (27.83)	74.94 (24.11)
	total	72.68 (13.04)	70.65 (17.53)	71.66 (15.32)
jugement d'auto-évaluation	connaissances antérieures	72.01 (15.32)	69.68 (16.92)	70.84 (16.01)
	questions similaires	63 (17.69)	64.38 (23.35)	63.69 (20.50)
	total	67.51 (14.69)	67.02 (17.15)	67.27 (15.79)
direction du biais d'auto-évaluation	connaissances antérieures	3.32 (20.77)	2.22 (20.73)	2.77 (20.53)
	questions similaires	-13.35 (25.53)	-9.17 (33.63)	-11.26 (29.61)
	total	-5.01 (18.26)	-3.48 (22.3)	-4.25 (20.18)

TABLEAU 14. – Ensemble des moyennes (et écarts-types) de la performance, du jugement d'auto-évaluation et de la direction du biais pour les connaissances sur le domaine des styles selon la condition expérimentale

	groupe expérimental	groupe contrôle	total	
performance	connaissances antérieures	85.29 (12.76)	87.48 (9.05)	86.39 (11)
	questions similaires	77.04 (31.79)	71.82 (28.68)	74.43 (30.07)
	total	81.17 (19.44)	79.65 (15.19)	80.41 (17.28)
jugement d'auto-évaluation	connaissances antérieures	76.7 (15.59)	77.44 (12.87)	77.07 (14.15)
	questions similaires	61.94 (22.56)	70.02 (19.73)	65.98 (21.36)
	total	69.32 (16.55)	73.73 (12.56)	71.53 (14.7)
direction du biais d'auto-évaluation	connaissances antérieures	-8.5 (17.17)	-9.95 (15.1)	-9.22 (16)
	questions similaires	-15.1 (31.15)	-1.8 (24.54)	-8.45 (28.54)
	total	-11.8 (21)	-5.87 (14.98)	-8.84 (18.29)

TABLEAU 15. – Ensemble des moyennes (et écarts-types) de la performance, du jugement d'auto-évaluation et de la direction du biais pour les connaissances sur le domaine de l'automatisation selon la condition expérimentale

		groupe expérimental	groupe contrôle	total
performance	connaissances antérieures	61.86 (16.43)	68.28 (19.8)	65.07 (18.29)
	questions similaires	52.34 (22.73)	54.02 (21.49)	53.18 (21.9)
	total	57.1 (16.18)	61.15 (13.75)	59.12 (14.99)
jugement d'auto-évaluation	connaissances antérieures	63.38 (18.21)	62.89 (14.01)	63.13 (16.08)
	questions similaires	72.13 (19.62)	75.56 (13.85)	73.84 (16.89)
	total	67.75 (15.76)	69.22 (12.38)	68.49 (14.03)
direction du biais d'auto-évaluation	connaissances antérieures	1.52 (22.39)	-5.4 (20.51)	-1.94 (21.52)
	questions similaires	19.79 (29.59)	21.54 (25.33)	20.66 (27.26)
	total	10.65 (22.6)	8.07 (16.09)	9.36 (19.45)

le sous domaine et la condition expérimentale,  $F(2, 92) = .75, p = .5$ . Seul un effet principal du sous domaine,  $F(2, 92) = 7.03, p = .001$  ainsi qu'un effet du moment,  $F(2, 92) = 17.23, p < .001$  ont été révélés.

Les comparaisons par paires à l'aide de t-tests à mesures appariées réalisées sur le sous domaine ont révélés que les apprenants s'auto-évaluent avec plus d'exactitude sur le sous domaine des feuilles de styles ( $M = 27.58, SD = 11.01$ ) que sur les fonctionnalités de base du traitement de texte, ( $M = 33.62, SD = 11.03$ ),  $t(47) = 3.4, p = .001$ , ou que sur l'automatisation des styles ( $M = 35.48, SD = 10.43$ ),  $t(47) = -3.47, p = .001$ . En revanche, aucune différence n'est apparue entre le sous domaine des fonctionnalités de base du traitement de texte et celui de l'automatisation des styles,  $t(47) = -.75, p = .46$ .

#### *Effet sur les performances*

L'anova à trois facteurs réalisée n'a pas détecté d'interaction entre nos trois variables indépendantes,  $F(2, 92) = .14, p = .87$ . Aucune interaction entre le moment de la mesure et la condition expérimentale,  $F(1, 46) = .84, p = .37$ , entre le sous-domaine et la condition expérimentale,  $F(2, 92) = .83, p = .44$  n'a été détectée.

Un effet d'interaction entre le moment de la mesure et le sous domaine de compétence a été détecté,  $F(2, 92) = 7.5, p = .001, \eta_p^2 = .14$ . Les trois t-tests réalisés ont détecté une différence significative entre le questionnaire de mesure de connaissances antérieures et le questionnaire similaire pour le sous domaine des 'styles',  $t(47) = 2.8, p = .007$  et de 'l'automatisation',  $t(47) = 3.05, p = .004$  mais pas pour celui du traitement de texte,  $t(47) = -1.67, p = .1$ . Ces résultats montrent que les participants ont eu de meilleures performance au questionnaire de mesures des connaissances antérieures sur le domaine des styles ( $M = 86,39$ ;  $SD = 11$ ) et de l'automatisation ( $M = 65.07, SD = 18.29$ ) qu'au questionnaire similaire ( $M = 74.43, SD = 30.07$  pour les styles;  $M = 53.18, SD = 21.9$  pour l'automatisation). Dans le sous domaine du traitement de texte, les participants ont obtenus des performances comparables au questionnaire de mesure de connaissances antérieures ( $M = 68.38, SD = 16.04$ ) et au questionnaire similaire ( $M = 74.94, SD = 24.11$ ).

En parallèle, un effet principal du moment de la mesure,  $F(1, 46) = 5.33, p = .03, \eta_p^2 = .10$  et du sous-domaine de compétence,  $F(2, 92) = 33.22, p < .001, \eta_p^2 = .42$  ont été détectés tandis qu'aucun effet de la condition expérimentale n'a été détecté,  $F(1, 46) = .002, p = .96$ .

#### *Effet sur le jugement d'auto-évaluation*

L'anova n'a détecté aucune interaction triple entre nos trois variables indépendantes,  $F(2, 92) = .2, p = .8$ .

Aucune interaction n'a été détectée entre la condition expérimentale et le sous-domaine de compétence,  $F(2, 92) = .95, p = .39$  ni entre la condition expérimentale et le moment de la mesure,  $F(1, 46) = 1.49, p = .23$ .

Une interaction a été détectée entre le moment de la mesure et le sous domaine de compétence,  $F(2, 92) = 25.88, p < .001$ . Les t-tests ont révélé une différence significative entre le premier questionnaire et le second sur le domaine du traitement de texte,  $t(47) = 2.6, p = .011$  suggérant que le jugement est plus élevé au premier questionnaire ( $M = 70.84, SD = 16.01$ ) qu'au second ( $M = 63.69, SD = 20.5$ ). Cette différence a été à nouveau détectée sur le domaine des styles,  $t(47) = 3.6, p = .001$ , suggérant comme précédemment que le jugement des performances était plus élevé au premier questionnaire ( $M = 77.07, SD = 14.15$ ) qu'au second ( $M = 65.98, SD = 21.36$ ). Concernant le domaine de l'automatisation, le jugement émis sur le premier questionnaire était plus faible ( $M = 63.13, SD = 16.08$ ) que sur le second ( $M = 73.84, SD = 16.89$ ),  $t(47) = -4.29, p < .001$ .

Aucun effet principal du moment de la mesure,  $F(1, 46) = 1.5, p = .23$  et aucun effet principal de la condition expérimentale n'ont été détectés,  $F(1, 46) = .23, p = .64$ . Un effet principal du sous domaine de compétence a été révélé,  $F(2, 92) = 3.02, p = .05$  suggérant que les participants s'auto-évaluent plus fortement sur le domaine des styles ( $M = 71.53, SD = 14.7$ ) que sur le domaine du traitement de texte ( $M = 67.27, SD = 15.79$ ) ou de l'automatisation ( $M = 68.49, SD = 14.03$ ).

#### *Effet sur la direction du biais d'auto-évaluation*

L'anova à trois facteurs n'a détecté aucune interaction triple entre nos trois variables indépendantes,  $F(2, 92) = .35, p = .7$ .

Une interaction a été détectée entre la variable moment de la mesure et le sous domaine de compétence,  $F(2, 92) = 20.86, p < .001$ . Les t-tests ont révélé une différence significative entre les deux questionnaires sur le domaine du traitement de texte,  $t(47) = 3.12, p = .003$  et celui de l'automatisation,  $t(47) = -5.22, p < .001$ . Aucune différence significative n'a été détecté sur le domaine des styles,  $t(47) = -.19, p = .85$ . Les données suggèrent que, sur le traitement de texte, les participants se sur-estimaient très légèrement au premier questionnaire ( $M = 2.78, SD = 20.53$ ) et se sous-évaluaient davantage au second ( $M = -11.26, SD = 29.61$ ). Sur le domaine de l'automatisation, les apprenants se sous-estiment très peu ( $M = -1.94, SD = 21.52$ ) tandis qu'ils se sur-estiment plus largement au second questionnaire ( $M = 20.66, SD = 27.26$ ).

Une interaction marginalement significative a été détectée entre le moment de la mesure et la condition expérimentale,  $F(1, 46) = 3.04, p =$

.09,  $\eta_p^2 = .06$ . Les t-tests n'ont révélé aucune différence significative pour le groupe expérimental entre les deux moments de la mesure,  $t(23) = .47, p = .64$ . La direction du biais d'auto-évaluation au questionnaire de mesure des connaissances antérieures ( $M = -1.22, SD = 13.98$ ) était comparable à celle du questionnaire similaire ( $M = -2.89, SD = 22.38$ ). Chez le groupe contrôle, une différence marginalement significative a été détectée,  $t(23) = -1.88, p = .07$ . Les participants tendaient à se sous-évaluer au questionnaire de mesure des connaissances antérieures ( $M = -4.38, SD = 14.82$ ) et tendaient à se sur-évaluer lors du questionnaire similaire ( $M = 3.52, SD = 20.52$ ).

En revanche, aucune interaction n'a été détectée entre la variable domaine de compétence et la condition expérimentale,  $F(2, 92) = 1.08, p = .3$ .

L'analyse des effets principaux ont révélé un effet du domaine de compétence,  $F(2, 92) = 21.47, p < .001, \eta_p^2 = .32$  suggérant que les participants se sont sur-évalués sur le domaine de l'automatisation ( $M = 9.36, SD = 19.45$ ) tandis qu'ils se sous-évaluaient sur le domaine du traitement de texte ( $M = -4.25, SD = 20.18$ ) et davantage encore sur le domaine des styles ( $M = -8.84, SD = 18.29$ ). Aucun effet du moment,  $F(1, 46) = 1.29, p = .26$  ou de la condition expérimentale,  $F(1, 46) = .13, p = .7$  n'ont été détectés.





# D

---

## RÉSULTATS COMPLÉMENTAIRES DE L'EXPÉRIENCE 2

---

### D.1 LE MODE ET LE MOMENT D'UTILISATION DES CARTES CONCEPTUELLES

Afin de détecter une interaction entre le mode et le moment d'utilisation des cartes, des Anova à deux facteurs ont été réalisées pour chaque variable dépendante. Le groupe contrôle n'a pas été considéré dans ces analyses car l'objectif était uniquement de comparer les conditions de construction et la consultation selon le moment d'utilisation des cartes. Les Anovas ont été complétées par des t-tests lorsqu'une interaction était significative.

#### d.1.1 *Le jugement d'auto-évaluation*

La première Anova réalisée sur le jugement d'auto-évaluation a mis en évidence une interaction entre le mode et le moment d'utilisation des cartes,  $F(1, 219) = 5.52, p = .02, \eta_p^2 = .03$ . Aucun effet principal du mode  $F(1, 219) = 1.74, p = .2$  et du moment  $F(1, 219) = .57, p = .5$  n'a été révélé.

Lorsque les participants construisent les cartes simultanément à l'apprentissage ( $M = 59.73, SD = 24$ ), leur jugement est significativement plus bas, comparativement à une construction différée ( $M = 70.58, SD = 25.81$ ),  $t(92) = 2.11, p = .04$

Lorsque les participants consultent les cartes, il n'existe aucune différence entre la condition simultanée ( $M = 63.33, SD = 27.44$ ) et différée ( $M = 57.78, SD = 25.01$ ),  $t(127) = -1.2, p = .23$ .

#### d.1.2 *La direction du biais d'auto-évaluation*

L'Anova réalisée sur la direction du biais d'auto-évaluation n'a révélé aucune interaction entre le mode et le moment d'utilisation des cartes,  $F(1, 219) = .01, p = .93$ .

Les participants qui ont construit les cartes simultanément à l'apprentissage ( $M = 36.33, SD = 39.8$ ) ont un biais similaire aux participants qui ont construit les cartes après l'apprentissage ( $M = 36.65, SD = 43.05$ ). Les participants qui ont consulté les cartes après l'apprentissage ( $M = 31.01, SD = 38.84$ ) ont un biais similaire aux participants qui ont consulté les cartes simultanément à l'apprentissage ( $M = 29.7, SD = 39.57$ ).

Aucun effet principal du mode,  $F(1, 219) = 1.27, p = .26$  ou du moment n'a émergé,  $F(1, 219) = .02, p = .9$ .

#### d.1.3 *L'amplitude du biais d'auto-évaluation*

L'Anova réalisée sur l'amplitude du biais ne confirme pas notre hypothèse d'interaction entre le mode et le moment d'utilisation des cartes,  $F(1, 219) = .22, p = .64$ .

Les participants qui ont construit les cartes simultanément à l'apprentissage ( $M = 46.42, SD = 27.06$ ) ont eu un biais d'une amplitude équivalente à celui des participants qui ont construit des participants qui ont construit les cartes après l'apprentissage ( $M = 44.74, SD = 34.35$ ). De même les participants qui ont consulté les cartes après l'apprentissage ( $M = 42.65, SD = 25.24$ ) ont eu un biais d'amplitude comparable aux participants qui ont consulté les cartes pendant l'apprentissage ( $M = 40.69, SD = 27.94$ ).

Par ailleurs, aucun effet principal du mode,  $F(1, 219) = 1.03, p = .31$  ou du moment n'est apparu,  $F(1, 219) = .001, p = 1$ .

#### d.1.4 *La performance à l'exercice d'application*

À nouveau, notre hypothèse d'interaction n'a pas été vérifiée ici,  $F(1, 219) = 2.5, p = .12$ .

La construction simultanée ( $M = 23.4, SD = 35.75$ ) ne permet pas aux apprenants d'obtenir de meilleurs résultats que la construction différée ( $M = 33.93, SD = 43.6$ ); de même que la consultation différée ( $M = 26.77, SD = 39.3$ ) ne permet pas d'obtenir de meilleures performances que la consultation simultanée ( $M = 33.64, SD = 42.83$ ).

À nouveau, aucun effet principal du mode,  $F(1, 219) = .08, p = .78$ . ou du moment n'a émergé,  $F(1, 219) = .11, p = .74$ .

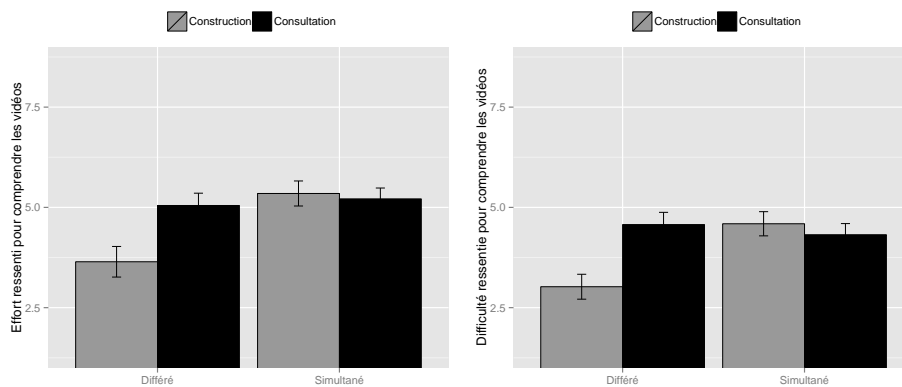


FIGURE 16. – Effort mental et difficulté ressentie pour comprendre les vidéos en fonction du moment et du mode d'utilisation des cartes.

#### D.1.5 L'effort mental et la difficulté associés à la compréhension des vidéos.

L'Anova réalisée sur l'effort mental associé à la compréhension des vidéos a révélé une interaction entre le mode et le moment d'utilisation des cartes,  $F(1, 219) = 5.86, p = .02, \eta_p^2 = .03$  (cf. Figure 16).

L'analyse des t-tests a révélé une différence significative entre les participants qui construisent les cartes après l'apprentissage et ceux qui construisent les cartes pendant l'apprentissage,  $t(92) = -3.48, p = .001$ . Toutefois, la différence n'est pas apparue dans le sens attendu puisque les participants qui ont construit les cartes après l'apprentissage ont ressenti moins d'effort mental pour comprendre les vidéos ( $M = 3.64, SD = 2.56$ ) que les participants qui ont construit les cartes pendant l'apprentissage ( $M = 5.35, SD = 2.18$ ).

Dans le cas de la consultation, aucune différence n'est apparue,  $t(127) = -.40, p = .69$ . Les participants qui ont consulté les cartes après l'apprentissage ( $M = 5.05, SD = 2.43$ ) ont ressenti autant d'effort pour comprendre les vidéos que ceux qui les ont consultées pendant l'apprentissage ( $M = 5.21, SD = 2.2$ ).

Similairement, l'Anova réalisée sur la difficulté à comprendre les vidéos a également révélé un effet d'interaction entre le mode et le moment d'utilisation des cartes,  $F(1, 219) = 8.98, p = .003, \eta_p^2 = .04$ .

Les résultats relatifs à la difficulté ressentie pour comprendre les vidéos sont similaires à ceux obtenus pour l'effort mental. Une différence significative a émergé entre les participants qui ont construit les cartes après l'apprentissage et ceux qui les ont construites pendant l'apprentissage,  $t(92) = -3.62, p < .001$ . Les participants qui ont construit les cartes après

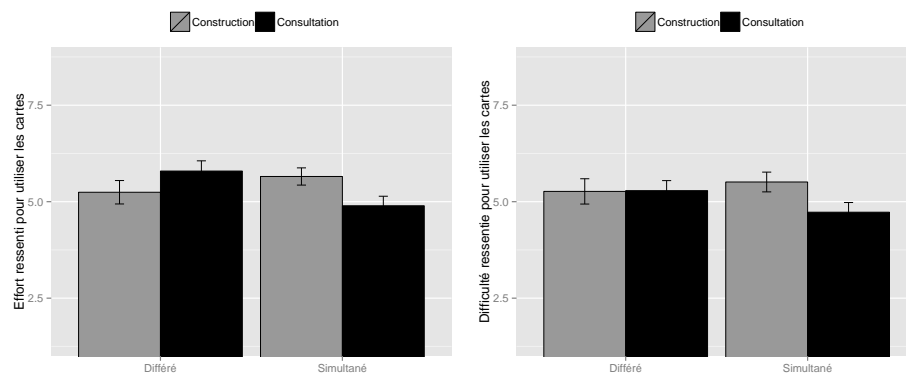


FIGURE 17. – Effort mental et difficulté ressentie pour utiliser les cartes conceptuelles en fonction du moment et du mode d'utilisation des cartes.

l'apprentissage ont ressenti moins de difficulté à comprendre les vidéos ( $M = 3.02, SD = 2.08$ ) que les participants qui ont construit les cartes pendant l'apprentissage ( $M = 4.59, SD = 2.11$ ).

#### D.1.6 L'effort mental et la difficulté associés à l'utilisation des cartes.

L'Anova réalisée sur l'effort mental pour utiliser les cartes a mis en évidence une interaction entre le mode et le moment d'utilisation des cartes,  $F(1, 219) = 6.09, p = .01, \eta_p^2 = .03$ , (cf. Figure 17).

L'analyse des t-tests a mis en évidence que dans le cas de la construction et contrairement à nos hypothèses, l'utilisation différée de la carte ( $M = 5.24, SD = 2.04$ ) ne demandait pas plus d'effort mental que l'utilisation simultanée ( $M = 5.65, SD = 1.56$ ),  $t(92) = -1.1, p = .28$ .

Dans le cas de la consultation, les tendances sont à nouveau contraire à nos hypothèses puisque la consultation différée ( $M = 5.79, SD = 2.1$ ) a nécessité un effort significativement plus important que la consultation simultanée ( $M = 4.89, SD = 2.01$ ),  $t(127) = 2.49, p = .01$ .

Enfin, l'Anova réalisée sur la difficulté ressentie pour utiliser les cartes n'a pas mis en évidence d'interaction entre le mode et le moment d'utilisation des cartes,  $F(1, 219) = 2.12, p = .15$ . Les tendances sont les mêmes que celles observées sur l'effort mental.

#### D.2 L'EFFET DU MOMENT DE PRÉSENTATION DES CARTES SELON LE NIVEAU D'EXPERTISE DE L'APPRENANT

Des anovas  $2 \times 2$  ont été conduites entre le moment de présentation des cartes (simultanée vs. différée) et le niveau d'expertise des apprenants

(expérimentés versus novices). Ces analyses ont été réalisées sur les huit variables dépendantes que sont : le jugement d'auto-évaluation, la direction du biais d'auto-évaluation, l'amplitude du biais d'auto-évaluation, la performance à l'exercice d'application, l'effort mental et la difficulté ressentis pour comprendre les vidéos, l'effort mental et la difficulté ressentis pour utiliser les cartes.

#### d.2.1 *Le jugement d'auto-évaluation*

L'anova conduite sur le jugement d'auto-évaluation n'a révélé aucun effet d'interaction,  $F(1, 214) = 1.94, p = .17$ , et aucun effet du moment,  $F(1, 214) = .23, p = .63$ . Seul un effet principal de l'expertise indiquant que les experts s'auto-évaluaient plus fortement ( $M = 70,25$  ;  $SD = 22.43$ ) que les novices ( $M = 55.1, SD = 27.27$ ) s'est révélé significatif,  $F(1, 214) = 20.2, p < .001, \eta_p^2 = .09$ .

#### d.2.2 *La direction du biais d'auto-évaluation*

L'anova à deux facteurs conduite sur la direction du biais d'auto-évaluation n'a pas détecté d'interaction entre le moment de présentation des cartes conceptuelles et le niveau d'expertise des participants,  $F(1, 214) = .15, p = .7$ . De même, aucun effet du moment n'a été révélé,  $F(1, 214) = .14, p = .7$ . Seul un effet principal de l'expertise,  $F(1, 214) = 48.9, p < .001, \eta_p^2 = .2$  indiquait que les participants expérimentés ( $M = 14.43$  ;  $SD = 40.64$ ) se sur-évaluaient moins fortement que les participants novices ( $M = 49.09, SD = 31.59$ ).

#### d.2.3 *L'amplitude du biais d'auto-évaluation*

L'anova conduite entre le moment de présentation des cartes et le niveau d'expertise des apprenants sur l'amplitude du biais d'auto-évaluation à l'exercice d'application n'a pas détecté d'effet d'interaction,  $F(1, 214) = .009, p = .9$  ou d'effet principal du moment,  $F(1, 214) = .08, p = .8$ . En revanche, un effet principal de l'expertise,  $F(1, 214) = 22.2, p < .001, \eta_p^2 = .09$  indique que les participants expérimentés ( $M = 33.98$  ;  $SD = 26.44$ ) s'auto-évaluaient avec plus d'exactitude que les participants novices ( $M = 51.36, SD = 27.71$ ).

#### d.2.4 *La performance à l'exercice d'application*

Aucun effet d'interaction,  $F(1, 214) = .41, p = .5$  ou du moment de présentation des cartes,  $F(1, 214) = .002, p = 1$  n'ont été détectés par

l'anova conduite sur la performance à l'exercice d'application. Un effet principal de l'expertise,  $F(1, 214) = 128.3, p < .001, \eta_p^2 = .38$  indiquait que les participants expérimentés ( $M = 55.72; SD = 41.2$ ) avaient des performances plus élevées que les participants novices ( $M = 6.01, SD = 20.58$ ) à l'exercice d'application.

#### D.2.5 *L'effort mental et la difficulté ressentis pour comprendre les vidéos*

##### *L'effort mental*

L'anova conduite sur l'effort mental ressenti pour comprendre les vidéos d'apprentissage a détecté une interaction significative entre nos deux variables indépendantes,  $F(1, 214) = 8.24, p = .01, \eta_p^2 = .04$ . L'analyse des effets simples, réalisée à l'aide de t-tests à mesures indépendantes, a révélé que les participants expérimentés qui ont utilisé les cartes après l'apprentissage ont ressenti un effort moins important ( $M = 3.59, SD = 2.33$ ) comparativement aux participants expérimentés qui ont utilisé les cartes pendant l'apprentissage, ( $M = 5.35, SD = 1.84$ ),  $t(104) = -4.31, p < .001$ . Au contraire, chez les novices, il n'existe pas de différence entre les participants qui ont utilisé les cartes pendant l'apprentissage, et ceux qui les ont utilisées après,  $t(110) = .09, p = .93$ .

##### *La difficulté*

L'anova conduite sur la difficulté ressentie pour comprendre les vidéos d'apprentissage a détecté une interaction significative entre nos deux variables indépendantes,  $F(1, 214) = 4.02, p = .05, \eta_p^2 = .02$

L'analyse des effets simples, réalisée à l'aide de t-tests à mesures indépendantes, a révélé que les participants expérimentés qui ont utilisé les cartes après l'apprentissage ont ressenti moins de difficulté ( $M = 3.21, SD = 2.15$ ) comparativement aux participants expérimentés qui ont utilisé les cartes pendant l'apprentissage, ( $M = 4.29, SD = 1.73$ ),  $t(104) = -2.86, p = .005$ . Au contraire, chez les novices, il n'existe pas de différence dans la difficulté ressentie pour comprendre les vidéos entre les participants qui ont utilisé les cartes pendant l'apprentissage, et ceux qui les ont utilisées après,  $t(110) = .27, p = .79$ .

#### D.2.6 *L'effort mental et la difficulté ressentis pour utiliser les cartes*

L'anova conduite sur l'effort mental ressenti pour utiliser les cartes n'a pas détecté d'effet d'interaction entre le moment de présentation des cartes et le niveau d'expertise des participants,  $F(1, 214) = 1.2, p = .27$ .

Par ailleurs, aucun effet du moment,  $F(1, 214) = 1.32, p = .25$  ou de l'expertise,  $F(1, 214) = .002, p = .97$ , n'ont été détectés.

Comme pour l'effort mental, l'anova conduite sur la difficulté ressentie pour utiliser les cartes n'a pas détecté d'interaction entre nos deux variables indépendantes,  $F(1, 214) = .06, p = .8$ ; ni d'effet de l'expertise,  $F(1, 214) = .06, p = .82$ , ou du moment,  $F(1, 214) = .37, p = .54$ .





# E

---

## COMPLÉMENTS D'INFORMATIONS SUR LE MATÉRIEL UTILISÉ DANS L'EXPÉRIENCE 3

---

### E.1 QUESTIONNAIRE DE MESURE DE CONNAISSANCES DÉCLARATIVES (QCM 1 ET 2)

1. Dans un traitement de texte, si on souhaite automatiser l'indexation des légendes, il est nécessaire d'utiliser ...
  - Une table des matières
  - Un style
  - Un signet pour chacun des titres
  - Un format de paragraphe
2. Si on souhaite faire apparaître le nom du chapitre courant en tête de chaque page, il est nécessaire d'utiliser ...
  - Un style
  - Un formule
  - Une formule associée à un style
  - Un signet
3. La modification d'un style de paragraphe sert notamment à ...
  - Modifier l'apparence de tous les paragraphes du document
  - Modifier l'apparence de tous les titres du document
  - Modifier l'apparence des paragraphes auxquels le style était appliqué
  - Modifier l'ordre d'apparition des paragraphes concernés
4. Dans un document, pour modifier la mise en forme des paragraphes où le style 'Titre 2' est appliqué, quelle est la méthode la plus efficace ?
  - Utiliser le pinceau sur chaque paragraphe pour reproduire la mise en forme
  - Modifier le style 'Titre 2' et l'appliquer de nouveau aux paragraphes concernés
  - Simplement modifier le style 'Titre 2'
  - Utiliser pour chaque titre la commande 'Copier/Coller'

5. Quelles sont les caractéristiques d'un paragraphe justifié ?
  - Le texte est centré
  - Le texte est aligné à gauche et à droite
  - Les marges de droites et de gauche ont la même valeur
  - Les retraits droits et gauche (avant/après) ont la même valeur
6. Quelle est la propriété propre aux caractères ?
  - La justification
  - Le soulignement
  - Le centrage
  - L'alignement
7. Dans un traitement de texte, on souhaite diminuer la largeur d'un paragraphe. Quelle est la meilleure méthode pour cela ?
  - Modifier les marges du paragraphe
  - Modifier les retraits associés au paragraphe
  - Centrer le paragraphe
  - Diminuer la taille des caractères
8. Pour appliquer la même mise en forme à tous les titres d'un document, quelle est la méthode recommandée ?
  - On utilise la commande 'Rechercher/Remplacer'
  - On utilise le pinceau
  - On applique un style de titre
  - On utilise la commande 'Copier' puis 'Coller'
9. Que faut-il faire pour automatiser la mise en forme d'un document avec un logiciel de traitement de texte ?
  - Il faut utiliser les styles
  - Il faut insérer une table des matières
  - Il faut convertir le document au format RTF
  - Il faut mettre le texte en forme ligne par ligne à l'aide du pinceau
10. L'interligne, l'alignement et le retrait sont des exemples de
  - style de paragraphe
  - format de paragraphe
  - style de caractères
  - format de caractère

## E.2 THÈMES DES CARTES CONCEPTUELLES PRÉSENTÉES LORS DE LA PHASE D'ENTRAÎNEMENT

Les cartes conceptuelles proposées lors de l'entraînement, dont un exemple est présenté dans la Figure 18, avaient pour thème : Le protocole TCP/IP ; Le droit numérique ; Les licences informatiques ; Le téléchargement ; L'identité numérique ; La recherche d'information ; La veille

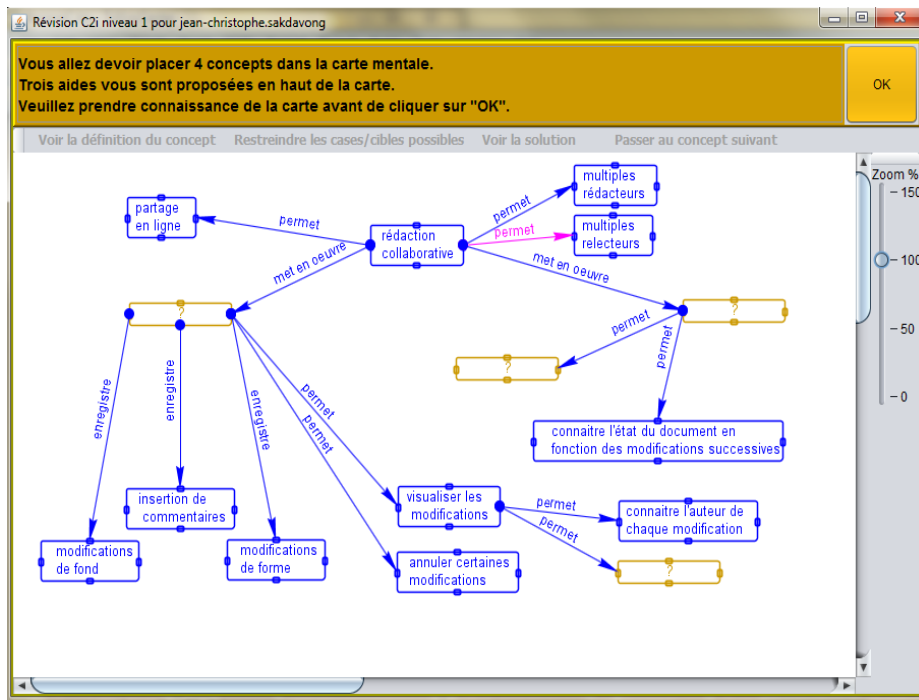


FIGURE 18. – Exemple de carte conceptuelle présentée lors de la phase d'entraînement

informationnelle ; L'exploitation de résultats d'une recherche d'information ; Les mails et forums ; Les moyens de communication ; La rédaction collaborative ; L'archivage et la compression de fichiers.

### E.3 EXEMPLE DE CARTE CONCEPTUELLE PRÉSENTÉE LORS DE LA PHASE D'ENTRAÎNEMENT

La Figure 18 présente un exemple de carte conceptuelle présentée lors de la phase d'entraînement de l'étude 3.



---

## LA THÉORIE DE LA CHARGE COGNITIVE : COMPLÉMENTS D'INFORMATION

---

La théorie de la charge cognitive défend l'idée que les performances d'apprentissage peuvent être améliorées ou détériorées en fonction de la bonne ou mauvaise présentation des informations à apprendre. Des informations mal présentées déclenchent des activités cognitives supplémentaires qui utilisent des ressources en mémoire de travail. Compte tenu de leur caractère limité, ces ressources ne seront alors plus disponibles pour le traitement des informations pertinentes. Des informations bien présentées, au contraire, peuvent simplement inciter l'apprenant à investir l'intégralité des ressources disponibles dans l'apprentissage (Sweller et al., 1998).

Dans cette Annexe, les notions de base de la théorie de la charge cognitive sont d'abord présentées afin d'expliquer comment les ressources cognitives se répartissent lors du traitement d'information. Nous présentons à la suite certains effets associés à ces notions.

### F.1 NOTIONS ÉLÉMENTAIRES SUR LA THÉORIE DE LA CHARGE COGNITIVE

L'apprentissage nécessite en premier lieu que les informations soient traitées en mémoire de travail. Les ressources en mémoire de travail peuvent toutefois être mobilisées pour d'autres activités que la seule intégration des connaissances. Sweller et al. (1998) ont distingué trois types de charges distinctes qui peuvent occuper les ressources en mémoire de travail lors du traitement de l'information.

- La charge intrinsèque renvoie au traitement du matériel à apprendre. Un contenu à apprendre est constitué d'informations plus ou moins nombreuses et plus ou moins reliées entre elles. Cette constitution intrinsèque du contenu d'apprentissage nécessite plus ou moins de ressources pour être traitée et représente la charge intrinsèque. En d'autres termes, elle renvoie au niveau de difficulté du matériel. Par exemple, un contenu avec peu d'information (une liste de courses de dix articles) représentera une charge intrinsèque moins importante

qu'un contenu avec beaucoup d'informations (apprendre la liste des verbes irréguliers en anglais). De la même façon, un contenu où les éléments sont très interactifs (apprendre la théorie de la charge cognitive) aura une charge intrinsèque plus importante qu'un contenu où les éléments n'ont pas de lien entre eux (apprendre la liste des verbes irréguliers en anglais). Dans le premier cas, les éléments doivent être traités avec un maximum de simultanéité pour pouvoir être intégrés. Dans le second cas, les éléments peuvent être traités séquentiellement. Le niveau de cette charge est fortement dépendant du niveau d'expertise puisque des individus avec un haut niveau de connaissances antérieures dans un domaine disposent de schémas pré-existants, activés en mémoire de travail comme un seul élément ; une personne ne possédant pas de schémas pré-existants devra activer plusieurs éléments.

- La charge superflue est associée à la présentation de l'information et impacte la gestion des éléments à apprendre en mémoire de travail. Une mauvaise présentation de l'information, en termes de qualité et/ou de quantité, peut avoir pour conséquence de créer des activités cognitives supplémentaires non pertinentes à l'intégration des connaissances en mémoire à long terme. Son existence est liée à une mauvaise conception de l'interface ou du matériel d'apprentissage (Sweller et al., 1998 ; J. J. Van Merriënboer & Sluijsmans, 2009).
- La charge essentielle renvoie à l'effort nécessaire au traitement de l'information et aux manipulations cognitives permettant la construction du schéma.

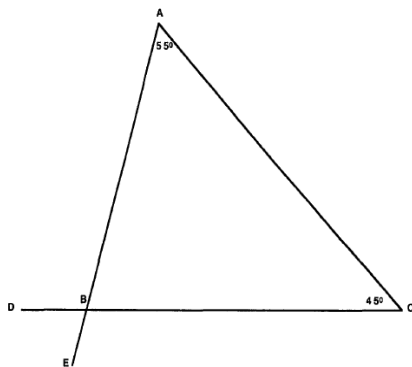
Une tâche d'apprentissage optimale se reflète par une diminution de la charge superflue et une augmentation de la charge essentielle, permettant d'optimiser l'utilisation des ressources à des fins d'acquisition des connaissances. Bien souvent, il est nécessaire que l'attention de l'apprenant soit redirigée des processus non pertinents vers les processus pertinents à l'intégration des schémas.

En lien avec cette répartition des ressources, des recherches ont mis en évidence l'existence de différents effets, liés à la présentation d'informations, qui perturbent l'apprentissage. La compréhension des conditions d'apparition de ces effets a permis de trouver des solutions afin de les éviter. Quatre de ces effets sont présentés ci-après.

## F.2 EFFET DE DISSOCIATION DE L'ATTENTION

L'effet de dissociation de l'attention apparaît dès lors que deux sources d'informations complémentaires, et non compréhensibles l'une sans l'autre, ne sont pas intégrées physiquement l'une avec

1. Condition illustrative de l'effet de dissociation de l'attention



Dans la figure ci-dessus, trouver la valeur de l'angle DBE

Solution:

Angle ABC =  $180^\circ - \text{Angle BAC} - \text{Angle BCA}$  (La somme des angles d'un triangle vaut  $180^\circ$ )

$$= 180^\circ - 55^\circ - 45^\circ$$

$$= 80^\circ$$

Angle DBE = Angle ABC (les angles verticalement opposés sont égaux)

$$= 80^\circ$$

2. Condition sans effet de dissociation de l'attention

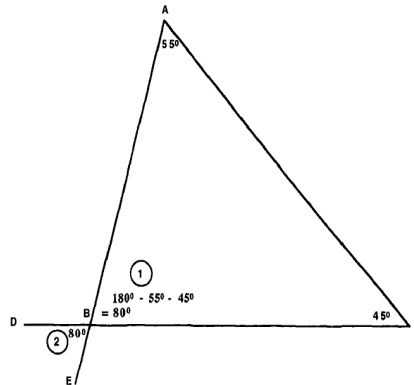


FIGURE 19. – illustration de l'effet de dissociation de l'attention

l'autre (Chandler & Sweller, 1992). En d'autres termes, cet effet s'observe quand les performances d'apprentissage d'un individu sont moins bonnes dans une condition où les informations sont spatialement et/ou temporellement séparées, comparativement à une condition où les informations sont spatialement et temporellement réunies. L'éloignement physique des informations oblige les apprenants à s'engager dans des processus cognitifs coûteux mais non pertinents pour la construction de schéma. L'apprenant doit alors maintenir en mémoire une information et l'associer à l'information complémentaire avant de pouvoir comprendre le contenu à intégrer. Un exemple de matériel occasionnant un effet de dissociation de l'attention est présenté en Figure 19 (issue de l'article de Sweller et al., 1998).

### F.3 L'EFFET DE REDONDANCE

L'effet de redondance apparaît lorsque la présentation d'une seule information est plus efficace que la présentation de deux informations au contenu similaire. C'est le cas notamment lorsqu'un diagramme est accompagné d'un texte descriptif. Les auteurs de la théorie de la charge cognitive considèrent comme redondantes toutes les informations superflues (Sweller, Ayres, & Kalyuga, 2011). L'effet de



redondance s'explique par l'allocation de ressources cognitives au traitement d'informations qui ont déjà été traitées, ou qui n'ont pas à l'être. Les ressources occupées ne sont dès lors plus disponibles pour traiter l'information pertinente.

#### F.4 L'EFFET DE L'EXPERTISE INVERSÉE

Les experts sont définis comme des individus qui ont acquis des compétences et des connaissances spécifiques en quantité et en qualité sur un domaine spécifique. Ces connaissances spécifiques leur permettent d'être plus performants dans la réalisation de certaines tâches en lien avec leur domaine de compétence, comparativement à des individus novices. Il existe toutefois certaines situations où les experts s'avèrent aussi peu, voire moins compétents que les novices. Ainsi, l'effet de l'expertise inversée peut apparaître lorsque certains dispositifs d'apprentissage sont néfastes pour l'apprentissage des apprenants qui possèdent davantage de connaissances antérieures. Ces dispositifs peuvent amener des apprenants disposant d'un niveau de connaissances élevé à obtenir de moins bonnes performances que des apprenants qui disposaient de moins de connaissances antérieures [Kalyuga et al. \(2003\)](#). Ces auteurs proposent que l'apparition de cet effet est dû au traitement d'informations, utiles pour les moins expérimentés, mais inutiles, voire redondantes, pour les plus expérimentés qui possèdent déjà ces connaissances dans leur représentation mentale. Les ressources des plus expérimentés sont dès lors utilisées pour traiter ces informations non pertinentes et pour les rendre cohérentes avec les schémas pré-existants. Par conséquent, ces ressources ne sont plus disponibles pour traiter des informations nouvelles.

#### F.5 L'EFFET DE MODALITÉ

L'effet de modalité apparaît lorsque les performances d'apprentissage d'un apprenant sont moins bonnes lorsque celui-ci traite une information présentée sous une seule modalité (i.e. auditive *ou* visuelle) plutôt que sous deux modalités différentes (i.e. auditive *et* visuelle) ([Sweller et al., 2011](#)). Par exemple, un schéma accompagné d'un commentaire présenté sous forme de texte à lire (même modalité) sera moins efficace qu'un schéma accompagné d'un commentaire audio (modalités différentes). Le postulat sous-jacent à l'explication de cet effet est que la mémoire de travail est composée d'une mémoire de travail auditive et d'une mémoire de travail visuelle. Il serait ainsi possible d'optimiser l'utilisation des capacités

en mémoire de travail en couplant les deux formats. Au contraire, présenter plusieurs informations sous un même format tendrait à surcharger la mémoire visuelle ou auditive. Le fait de présenter des informations sous deux modalités différentes, cependant, ne suffit pas toujours à augmenter la performance d'apprentissage. Tout comme pour l'effet de dissociation de l'attention, il est nécessaire de présenter des informations complémentaires. Des informations identiques, au contraire, provoqueraient un effet de redondance susceptible de détériorer les performances d'apprentissage.



## COLOPHON

Ce document a été mis en page avec  $\text{\LaTeX} 2_{\mathcal{E}}$ , selon le style *Classic Thesis* mis au point par André Miede (© 2006) à partir des recommandations typographiques de Robert Bringhurst.