

# La validation de contenu d'un outil d'observation des fonctions exécutives selon la méthode Delphi

## Content validation of an executive function observation tool using the Delphi method

## Validação de conteúdo de uma ferramenta de observação das funções executivas utilizando o método Delphi

Stéphanie Duval , Jennifer Couttet and Noémie Montminy

Volume 46, Number 2, 2023

URI: <https://id.erudit.org/iderudit/1111097ar>

DOI: <https://doi.org/10.7202/1111097ar>

[See table of contents](#)

Publisher(s)

ADMEE-Canada

ISSN

0823-3993 (print)

2368-2000 (digital)

[Explore this journal](#)

Cite this article

Duval, S., Couttet, J. & Montminy, N. (2023). La validation de contenu d'un outil d'observation des fonctions exécutives selon la méthode Delphi. *Mesure et évaluation en éducation*, 46(2), 1–40. <https://doi.org/10.7202/1111097ar>

Article abstract

*There are few measures to globally examine the executive functions (EFs) of preschool children in educational contexts, which nevertheless constitute a predictor of their educational success. Indeed, the components of EFs (e.g., inhibition) are usually measured by standardized tasks and reported measures, leading to a discrepancy between the results. We also observe a predominance for the use of tasks measuring a single component of EFs. Thus, the creation of a systematic observation instrument seems necessary, as well as the validation of its content, in order to have a representative and global portrait of the child's EF in the classroom. In this study, the content of a newly developed observation grid on EFs was validated with a panel of experts using the Delphi method. This article, which presents the application of the validation procedure and the resulting outcomes, discusses the effects and limits of the Delphi method, which is increasingly used in educational studies.*

© Stéphanie Duval, Jennifer Couttet et Noémie Montminy, 2024



This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

<https://apropos.erudit.org/en/users/policy-on-use/>

érudit

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

<https://www.erudit.org/en/>

## **La validation de contenu d'un outil d'observation des fonctions exécutives selon la méthode Delphi**

### **Content validation of an executive function observation tool using the Delphi method**

### **Validação de conteúdo de uma ferramenta de observação das funções executivas utilizando o método Delphi**

**Stéphanie Duval**

ID ORCID: 0000-0002-3347-2640

*Université Laval*

**Jennifer Couttet**

*Université Laval*

**Noémie Montminy**

*Université Laval*

MOTS CLÉS: Delphi, fonctions exécutives, outil d'observation, validité de contenu

*Peu de mesures permettent d'examiner globalement les fonctions exécutives (FE) chez l'enfant d'âge préscolaire en contextes éducatifs. Elles constituent pourtant un prédicteur de réussite éducative. Les composantes des FE (p. ex. l'inhibition) sont habituellement mesurées par des tâches standardisées et des mesures rapportées, amenant une divergence entre les résultats. On observe aussi une prédominance de l'utilisation de tâches mesurant une seule composante des FE. Ainsi, la création d'un instrument d'observation systématique semble nécessaire, de même que la validation*



*du contenu de celui-ci, pour pouvoir brosser un portrait représentatif et global des FE de l'enfant en classe. Dans cette étude, le contenu d'une grille d'observation nouvellement élaborée sur les FE a été validé auprès d'un panel d'experts à l'aide de la méthode Delphi. Cet article, qui présente notamment l'application de la procédure de validation et les résultats qui en découlent, discute des retombées et des limites de la méthode Delphi, qui est de plus en plus exploitée dans les études en éducation.*

**KEY WORDS:** content validity, Delphi, executive functions, observation tool

*There are few measures to globally examine the executive functions (EFs) of preschool children in educational contexts, which nevertheless constitute a predictor of their educational success. Indeed, the components of EFs (e.g., inhibition) are usually measured by standardized tasks and reported measures, leading to a discrepancy between the results. We also observe a predominance for the use of tasks measuring a single component of EFs. Thus, the creation of a systematic observation instrument seems necessary, as well as the validation of its content, in order to have a representative and global portrait of the child's EF in the classroom. In this study, the content of a newly developed observation grid on EFs was validated with a panel of experts using the Delphi method. This article, which presents the application of the validation procedure and the resulting outcomes, discusses the effects and limits of the Delphi method, which is increasingly used in educational studies.*

**PALAVRAS-CHAVE:** Delphi, ferramenta de observação, funções executivas, validade de conteúdo

*Poucas medidas permitem examinar globalmente as funções executivas (FE) em crianças em idade pré-escolar em contextos educativos, embora sejam um preditor de sucesso educacional. As componentes das FE (por exemplo, a inibição) são normalmente medidas por tarefas padronizadas e medidas reportadas, levando a uma divergência entre os resultados. Observa-se também uma predominância do uso de tarefas que medem uma única componente das FE. Assim, a criação de uma ferramenta de observação sistemática parece necessária, bem como a validação do conteúdo desta, para poder traçar um retrato representativo e global das FE da criança em sala de aula. Neste estudo, o conteúdo de uma grelha de observação recém-elaborada sobre as FE foi validado junto a um painel de especialistas usando o método Delphi. Este artigo, que apresenta especialmente a aplicação do procedimento de validação e os resultados decorrentes, discute os impactos e os limites do método Delphi, que é cada vez mais utilizado em estudos na área da educação.*

---

Note des auteurs : La correspondance liée à cet article peut être adressée à [stephanie.duval@fse.ulaval.ca](mailto:stephanie.duval@fse.ulaval.ca).

## Introduction

Les fonctions exécutives (FE) représentent un important prédicteur du développement, des apprentissages et de la réussite éducative de l'enfant d'âge préscolaire (3-6 ans), (Ribner et al., 2017). Plus précisément, des travaux ont montré que le niveau de FE mesuré à l'éducation préscolaire permettait de prédire les habiletés de l'enfant en mathématiques et en lecture (Mulcahy et al., 2021), en plus d'être relié à son développement langagier (Foy & Mann, 2012) et à son fonctionnement cognitif en général (Blair & Razza, 2007).

L'âge préscolaire représente un moment clé pour reconnaître et soutenir les habiletés de FE qui sont en pleine progression (Carlson, 2005; Duval et al., 2018), en raison de la maturation accélérée des aires cérébrales qui les supportent (Kovács & Mehler, 2009). Compte tenu de leur rôle central dans la réussite éducative de l'enfant et de l'accroissement rapide des composantes qui y sont liées (p. ex. l'inhibition) à l'âge préscolaire, il apparaît essentiel de se questionner sur les outils qui mesurent les FE à ce moment charnière du développement.

### *Les outils de mesure des FE à l'âge préscolaire*

Selon une récente revue systématique des écrits ayant pour objectif de répertorier les outils disponibles pour mesurer les FE et leurs composantes chez des enfants d'âge préscolaire (Silva et al., 2022), la mesure rapportée *Behavior Rating Inventory of Executive Function (BRIEF-P)* (Gioia et al., 2002) et les tâches standardisées, par exemple les échelles de Wechsler (2003), représentent les instruments les plus souvent utilisés par les chercheurs.

L'utilisation de ces outils présente par ailleurs des limites. Par exemple, l'utilisation d'une mesure rapportée (p. ex. un questionnaire) remplie par le parent de l'enfant ou par le personnel enseignant peut amener des biais de perception. De plus, des facteurs comme le type de langage utilisé par la personne évaluatrice, sa relation plus ou moins proximale avec l'enfant (Denckla, 2002), de même que ses représentations et ses attentes à l'égard de ce dernier peuvent influencer les résultats (Silva et al., 2022).

De surcroît, des limites sont observées quant à l'utilisation de tâches standardisées, qui sont pourtant encore considérées comme la méthode d'évaluation de référence (Zelazo et al., 2016). Ces tâches sont le plus souvent réalisées en laboratoire entre un expérimentateur et un enfant, sur un ordinateur ou sur une tablette électronique. Par exemple, il peut s'agir d'un test de mémoire où l'on demande à l'enfant de répéter à l'envers une série de chiffres (*Backward Digit Span*) (Davis & Pratt, 1995). Lors de ces tâches, il est possible d'attribuer un score de performance à l'égard des FE de l'enfant, mais les comportements observés sont difficilement transférables et généralisables aux situations vécues au quotidien, comme en contextes éducatifs de la petite enfance.

Ainsi, plusieurs chercheurs et chercheuses (Anderson & Reidy, 2012; McClelland & Cameron, 2012) sont d'avis que le contexte dans lequel l'enfant est placé lors de tâches standardisées, c'est-à-dire dans un environnement exempt d'interférences et en présence d'une personne souvent inconnue, n'est pas représentatif des situations réelles qu'il vit en contexte naturel. À titre illustratif, un enfant qui démontre une bonne capacité à inhiber dans une tâche de laboratoire peut avoir de la difficulté à inhiber un comportement similaire en contexte éducatif, en présence de plusieurs autres enfants et de distractions (McClelland & Cameron, 2012). Pour ces raisons, Anderson et Reidy (2012) soutiennent que les mesures standardisées ont une faible validité écologique, d'où l'importance de créer un instrument qui mesure les FE de l'enfant dans des situations naturelles.

De plus, les chercheurs ont montré la prédominance de l'utilisation de tâches individuelles qui évaluent une seule composante des FE de manière isolée, l'habileté d'inhibition étant la plus souvent mesurée (Silva et al., 2022). D'ailleurs, il n'existe que très peu d'outils de mesure globale des FE à l'âge préscolaire (Blair et al., 2005), ce qui oblige les chercheurs à utiliser plusieurs outils ou plusieurs tâches pour obtenir un portrait global des FE des jeunes enfants.

Enfin, la comparaison des données avec les outils existants (mesures rapportées et tâches standardisées axées sur la performance) montre une divergence entre les résultats obtenus (Toplak et al., 2013; Young et al., 2017), ce qui amène à penser qu'ils ne mesurent pas les mêmes construits. De même, bien qu'ils présentent des avantages (p. ex. la validité éprouvée), les outils existants ne suffisent pas à rendre compte des habiletés de l'enfant dans un contexte naturel, comme en services de garde éducatifs à l'enfance

ou en classe d'éducation préscolaire. Sans les discréditer, il semble essentiel de se questionner sur les instruments qui mesurent les composantes des FE de façon écologique, afin de pouvoir brosser un portrait des habiletés que l'enfant peut déployer dans des situations naturelles, tel qu'en contextes éducatifs de la petite enfance.

Des chercheurs insistent sur l'importance de miser sur la création d'un instrument pour examiner les FE en contexte écologique (Cadavid-Ruiz & del Río, 2018 ; Escolano-Pérez et al., 2017 ; Nieto et al., 2016 ; Willoughby et al., 2016), afin de mieux les reconnaître et de comprendre comment elles se déploient et comment elles peuvent être soutenues en contextes éducatifs. L'observation directe représente une piste prometteuse pour examiner les FE, car elle permet d'étudier les comportements naturels et spontanés de l'enfant en misant sur les processus qui sont déployés dans la situation observée (et non sur les résultats) (Bakeman & Quera, 2012).

### ***L'observation et la mesure écologique des FE***

Selon Doebel (2020), le fait de comprendre de quelle manière les habiletés de FE se développent et peuvent être soutenues nécessite d'abord de les reconnaître, compte tenu du fait qu'elles sont toujours manifestées pour répondre à un objectif précis, que ce soit en laboratoire ou lors de situations naturelles. En effet, les objectifs que l'enfant souhaite atteindre activent un contenu mental qui façonne la manière dont il mobilise ses habiletés par rapport à une situation particulière, par exemple dans le cadre du jeu, en présence d'autres enfants et d'interférences (distractions sonores et/ou visuelles, etc.). Ainsi, il paraît indispensable d'observer les habiletés de FE en contexte naturel (Isquith et al., 2004), dans des situations qui permettent à l'adulte d'examiner les manifestations comportementales des enfants (Duval et al., 2018). Le fait de reconnaître globalement les habiletés de l'enfant à l'égard des FE dès la fréquentation d'un contexte éducatif de la petite enfance, comme une classe d'éducation préscolaire, permettrait de mieux caractériser sa trajectoire développementale et d'identifier des moyens de le soutenir.

Pour ce faire, Cadavid-Ruiz et del Río (2018) proposent de miser sur la création de nouveaux outils, comme une grille d'observation directe, pour mesurer les FE de l'enfant de façon écologique. En effet, il existe peu de données sur la manière dont les processus mentaux associés aux FE peuvent être observés en contextes éducatifs de la petite enfance (Duval et al., 2023). Le développement d'une telle mesure s'avère complexe, puisqu'il faut réaliser plusieurs étapes afin que la grille d'observation puisse fournir des informations valides et fiables sur un sujet d'étude.

Selon Lynn (1986), chercheuse éminente dans le champ de la validité de contenu (Haidari & Uzun, 2019), deux phases sont nécessaires pour développer une nouvelle mesure : a) la phase de développement, qui permet de conceptualiser les attributs du concept à documenter en fonction des connaissances disponibles sur le sujet, puis b) la phase d'appréciation-quantification, qui consiste à réaliser une validation du contenu grâce à la consultation d'un comité d'experts. Cette étape est cruciale puisqu'elle permet de s'assurer qu'un outil recueille bel et bien une information adéquate et complète de l'attribut ou du phénomène à documenter (Corbière & Fraccaroli, 2014; Streiner et al., 2014).

Afin de mener la deuxième phase proposée par Lynn (1986), de plus en plus de chercheurs suggèrent de miser sur la méthode Delphi, une procédure de recherche utilisée pour parvenir à une opinion ou à une décision de groupe en interrogeant un panel d'experts (Sablatzky, 2022). Autrement dit, cette méthode réunit un panel d'experts autour d'un objet commun, afin d'obtenir un consensus au sujet de celui-ci. Ce processus repose sur l'idée que les réponses collectives apportent davantage de richesse à une question ou à un objet d'étude en comparaison aux réponses individuelles (Sablatzky, 2022). Développée dans les années 1950 et 1960 par la RAND Corporation, la méthode Delphi a été utilisée dans de nombreux domaines, notamment en économie et en santé (Ciptono et al., 2019; Parrish & Sandera, 2022). Toutefois, peu d'études en éducation ont utilisé cette méthode de recherche, encore moins dans les travaux en petite enfance.

Est-il possible de penser que la validation de contenu d'une grille d'observation des FE nouvellement développée puisse constituer un instrument complémentaire aux mesures existantes destinées aux enfants d'âge préscolaire? Considérant l'importance de reconnaître comment et quand les FE de l'enfant sont naturellement mobilisées en contexte écologique, cette étude propose d'étudier cette question, en validant le contenu d'une grille d'observation directe permettant d'évaluer les habiletés de FE des enfants en contextes éducatifs de la petite enfance.

## Cadre théorique

### *Les composantes des fonctions exécutives-*

Les FE peuvent être définies comme des processus cognitifs d'ordre supérieur, découlant de façon importante d'une activité dans le cortex préfrontal (Barros & Hazin, 2013). Les FE, qui s'avèrent essentielles pour l'exécution et le suivi de plusieurs comportements axés sur des objectifs (Barros et al., 2016; Silva et al., 2022), sont généralement regroupées en trois principales composantes : l'inhibition, la mémoire de travail [MdT] et la flexibilité cognitive (Miyake et al., 2000; Wiebe et al., 2011).

L'inhibition renvoie au contrôle intentionnel des émotions (p. ex. régulation des émotions), des comportements (p. ex. autocontrôle, minutie) et des pensées (p. ex. attention, concentration, pensées intrusives). Grâce à elle, l'enfant peut mettre de côté les distractions et se centrer sur la situation en cours (Diamond, 2013). La MdT représente la capacité de l'enfant à stocker, à mettre à jour et à récupérer des informations dans sa mémoire afin de les utiliser de manière appropriée, selon un objectif précis (Kazemi & Mohammadi, 2019; Willoughby et al., 2016). Elle lui permet ainsi de raisonner, d'analyser et de combiner différentes informations, en plus de faire des rappels, de comparer différentes alternatives et d'élaborer des plans (Kazemi & Mohammadi, 2019; Silva et al., 2022). Enfin, la flexibilité cognitive permet à l'enfant de changer de perspective (Diamond, 2013), d'alterner son attention d'une tâche à une autre (Fleer et al., 2017; Willoughby et al., 2016) et de modifier une réponse en fonction du contexte (Davidson et al., 2006).

À ces trois principales composantes s'ajoutent des habiletés plus complexes, telles que la planification (Diamond, 2013; Nigg, 2017). La capacité à planifier, qui est fortement encouragée pendant l'éducation préscolaire (Ministère de l'Éducation, 2021), peut être décrite comme la manière de prédire et d'évaluer ses comportements ainsi que de constituer et de coordonner une séquence d'actions afin d'atteindre un but précis (Bjorklund & Causey, 2018).

### *La validité de contenu d'un instrument de mesure*

La validité de contenu désigne le degré auquel l'outil et ses composantes sont représentatifs du construit mesuré (Zamanzadeh et al., 2015). Pour ce faire, Rusticus (2014) soulève le fait que tous les éléments de l'instrument (manuel de cotation, items, grille de cotation) susceptibles d'avoir un effet sur les résultats et les interprétations doivent être soumis à une validation du contenu, le plus souvent auprès d'un comité d'experts.



Trois aspects sont à considérer afin d'assurer la validité de contenu d'un instrument : 1) la définition du construit à l'étude, 2) la représentation de ce dernier et 3) sa pertinence (Laveault & Grégoire, 2014; Sireci, 1998). Le premier aspect, soit la définition, fait référence aux définitions conceptuelles et opérationnelles du construit, ce qui permet notamment de préciser les items à mesurer. La représentation et la pertinence font référence au degré selon lequel l'instrument (dans son ensemble) mesure le construit (tel que défini dans le premier aspect) et à la pertinence des items par rapport au contenu à l'étude. La validité de contenu est compromise dans la mesure où : 1) l'instrument n'examine pas tous les items du construit ciblé, 2) il mesure les items du construit de manière disproportionnée ou 3) il mesure des items qui se trouvent en dehors du construit ciblé (Haynes et al., 1995; Laveault & Grégoire, 2014; Rubio et al., 2003).

Il est essentiel de tenir compte de ces trois aspects dans le développement de l'outil de mesure, toujours dans l'idée d'assurer la validité de son contenu. Après la phase de développement de l'outil, Lynn (1986) suggère la phase d'appréciation-quantification, également recommandée par plusieurs chercheurs (Yusoff, 2019). Différentes méthodes tiennent compte de cette phase, dont la consultation d'un panel d'experts, contribuant à une version satisfaisante et consensuelle d'un outil d'évaluation (Colquitt et al., 2019; Polit et al., 2007). La méthode Delphi peut être appliquée dans cette phase de conception d'un nouvel instrument.

### ***La méthode Delphi pour mesurer et améliorer la validité de contenu***

La méthode Delphi permet de consulter, de manière individuelle, simultanée et sous le couvert de l'anonymat, un panel d'experts au sujet d'un même objet, concept ou problème (Ekionea et al., 2011; Parrish & Sandera, 2022). Elle vise principalement à : a) atteindre systématiquement un consensus entre les experts du panel, b) prendre des décisions sur les composantes d'un instrument à partir des données recueillies (Hohmann et al., 2018), c) apporter des modifications pour bonifier un instrument, ce qui a pour effet d'améliorer la validité de contenu d'un outil (Rusticus, 2014).

Initialement, la conception conventionnelle de la méthode Delphi comportait au moins quatre cycles (ou tours) d'évaluation de la part d'un panel d'experts. Cependant, l'essence d'une bonne enquête Delphi réside dans un processus itératif et dans un *feedback* contrôlé pour générer un consensus. Ainsi, dans les travaux récents, les critères de clôture de la plupart des études qui utilisent la méthode Delphi incluent «[...] le consensus

obtenu après un cycle préfixé», c'est-à-dire généralement deux tours d'évaluation (Nasa et al., 2021). Pour leur part, pour améliorer le consensus, Holey et al. (2007) recommandent un certain nombre de tours puisque les participants peuvent encore changer d'avis au troisième tour. Ils ont observé qu'en trois tours, il y avait généralement une augmentation du pourcentage d'accords entre les experts (Holey et al., 2007).

Quelle que soit la décision des chercheurs sur le nombre de tours qu'ils souhaitent réaliser dans leur étude, il faut suivre une séquence prédéterminée à l'intérieur de ceux-ci (Berquez et al. 2011):

- 1) D'abord, les experts sont invités à se prononcer sur une première version de l'instrument de mesure; leurs commentaires sont ensuite analysés de manière qualitative et quantitative (Ekionea et al., 2011).
- 2) À la suite des premières analyses, des modifications tenant compte des commentaires des experts sont apportées à l'outil de mesure, en considérant les divergences dans leurs évaluations.
- 3) Un rapport présentant toutes les modifications effectuées à l'outil de mesure est ensuite transmis au panel d'experts, en plus d'une nouvelle version de l'instrument.
- 4) Les experts sont alors invités à prendre connaissance du rapport et à évaluer la nouvelle version de l'outil, selon la même procédure qu'au premier tour d'évaluation (et ainsi de suite).

En conclusion, la méthode Delphi ne prescrit pas un nombre de tours à réaliser. Elle ne dicte pas non plus le nombre de participants à recruter, ni le degré de consensus à atteindre. Les chercheurs doivent donc adapter la méthode Delphi à leur étude particulière et identifier les avantages et les inconvénients potentiels des choix établis (Keeney et al., 2006).

### ***L'objectif de la recherche***

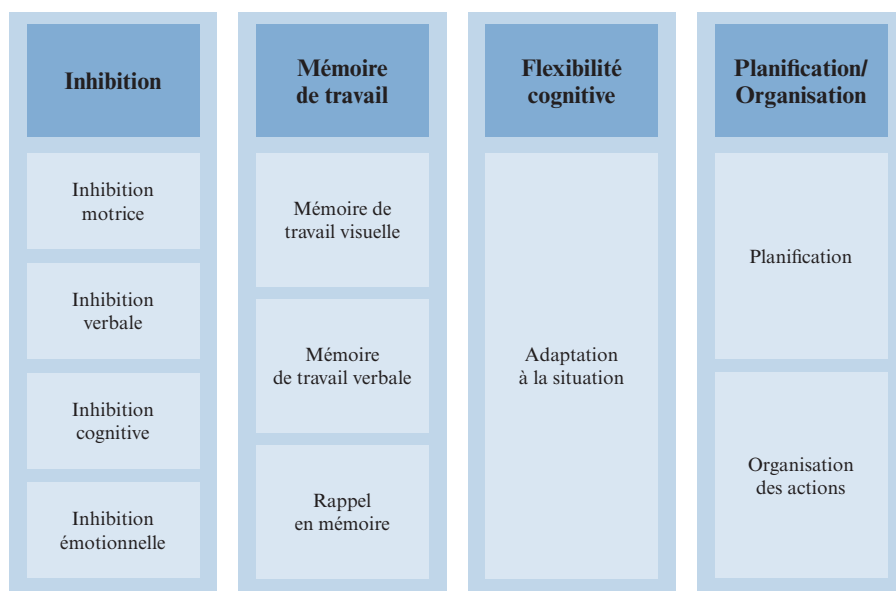
L'objectif de cette recherche, qui est la première étape d'un projet de validation plus large (Duval et al., 2020-2023), vise à évaluer et à améliorer la validité de contenu d'un outil d'observation nouvellement développé, à l'aide de la méthode Delphi. Ainsi, le présent article poursuit un double objectif : 1) présenter la validation d'un outil de recherche et 2) aborder l'utilisation de la méthode Delphi.

## Méthodologie

Conformément aux phases suggérées par Lynn (1986) dans le processus de validation de contenu, la phase de développement implique d'abord d'identifier et de circonscrire les attributs à évaluer (étape 1), ce qui implique de s'attarder à la définition et à la conceptualisation des items que l'outil dit évaluer (Désormeaux-Moreau et al., 2020). Pour ce faire, il est nécessaire de procéder à une recension des écrits scientifiques et des modèles sous-jacents aux outils qui existent déjà dans le domaine (Désormeaux-Moreau et al., 2020). Cette première étape a permis de développer un outil d'observation des FE chez l'enfant âgé de 3 à 6 ans en contextes éducatifs de la petite enfance (Duval & Montminy, 2021). Tel que le montre la figure 1, cet outil comporte quatre composantes et 10 sous-composantes, qui ont été définies en cohérence avec le cadre théorique présenté plus haut.

Figure 1

### *Composantes et sous-composantes de la grille d'observation*



Après la phase de développement vient celle d'appréciation-quantification qui sert à évaluer la pertinence de l'outil et qui se fait en consultant des experts. Ces deux phases doivent s'inscrire dans une séquence qui comporte un aspect itératif afin que les données issues de la phase d'appréciation-quantification alimentent et améliorent le développement de l'outil (Désormeaux-Moreau et al., 2020).

Cet article présente brièvement la phase de développement en décrivant les composantes et les sous-composantes de l'outil, mais aborde surtout la phase d'appréciation-quantification, en explicitant la méthode Delphi qui a été utilisée. L'application de cette méthode, qui a contribué à préciser les informations sur la définition, sur la représentativité et sur la pertinence des items composant la grille d'observation des FE, a permis d'améliorer l'instrument de mesure nouvellement développé à l'aide des commentaires d'un panel d'experts. Les sections suivantes décrivent plus précisément l'application de la méthode Delphi, en commençant par la présentation des participants, puis en exposant le protocole utilisé et les analyses réalisées.

### ***Les participants***

Un expert peut être défini comme une « personne informée », un « spécialiste dans le domaine » ou « quelqu'un qui a des connaissances sur un sujet spécifique » (Goodman, 1987 ; Keeney et al., 2001). Il est essentiel de choisir soigneusement les personnes qui composent le panel d'experts, puisque leurs perspectives et leurs opinions influencent largement les résultats de l'étude (Powell et al., 2021). Pour faire cela, il importe d'abord d'établir les critères de sélection du panel, conformément à d'autres travaux qui ont utilisé la méthode Delphi (Bahadir & Tuncer, 2020 ; Ekionea et al., 2011).

Dans ce projet de recherche, trois critères ont servi à sélectionner les experts, qui ont par ailleurs librement consenti à participer à l'étude :

- 1) Ils devaient avoir publié dans les champs du développement cognitif et des FE en petite enfance ;
- 2) Ils ne devaient pas avoir collaboré avec les chercheuses de la présente étude ;
- 3) Ils ne devaient pas avoir travaillé sur le développement ou la validation d'une grille d'observation des FE auparavant

Il faut ensuite définir le nombre d'experts requis avant d'entreprendre leur recrutement. Selon Lynn (1986), le nombre d'experts nécessaires pour valider le contenu d'un instrument ne devrait pas être inférieur à trois, et le fait de constituer un panel d'au moins cinq experts évite l'apparition de problèmes d'accord fortuit (Lynn, 1986). De façon similaire, d'autres chercheurs (Hoyt et al., 2020) considèrent qu'un panel comptant entre trois et dix experts est idéal. Pour cette étude, un panel de cinq experts a été formé, afin de recueillir des avis différents et complémentaires sur toutes les composantes des FE. Puis, il a fallu recruter les experts et leur transmettre les informations sur le projet. Ils ont été conviés au début de l'automne 2019, via une lettre électronique comportant la présentation du projet, ses objectifs et la nature de l'implication demandée.

Les experts intéressés par le projet ont rempli un formulaire de consentement approuvé au préalable par le comité d'éthique de l'université d'attache de la chercheuse principale. Puis, une rencontre virtuelle a eu lieu à l'hiver 2020, notamment pour définir le rôle de chaque personne, clarifier la démarche et les échéanciers et présenter le canal de communication.

### ***L'outil d'observation à valider***

L'instrument de mesure à valider, titré *Grille d'observation des FE des enfants âgés entre 3 et 6 ans en contextes éducatifs de la petite enfance* (Duval & Montminy, 2021), était initialement composé de quatre composantes, subdivisées en 10 sous-composantes (voir la figure 1 présentée plus haut), puis en 27 items décrivant de manière plus fine ces composantes (voir le tableau 2 présenté plus loin). Le panel a été invité à commenter l'ensemble du *Manuel de l'observatrice ou de l'observateur* (Duval & Montminy, 2021) à l'intérieur duquel s'inscrivait l'outil. Ce manuel comporte six sections :

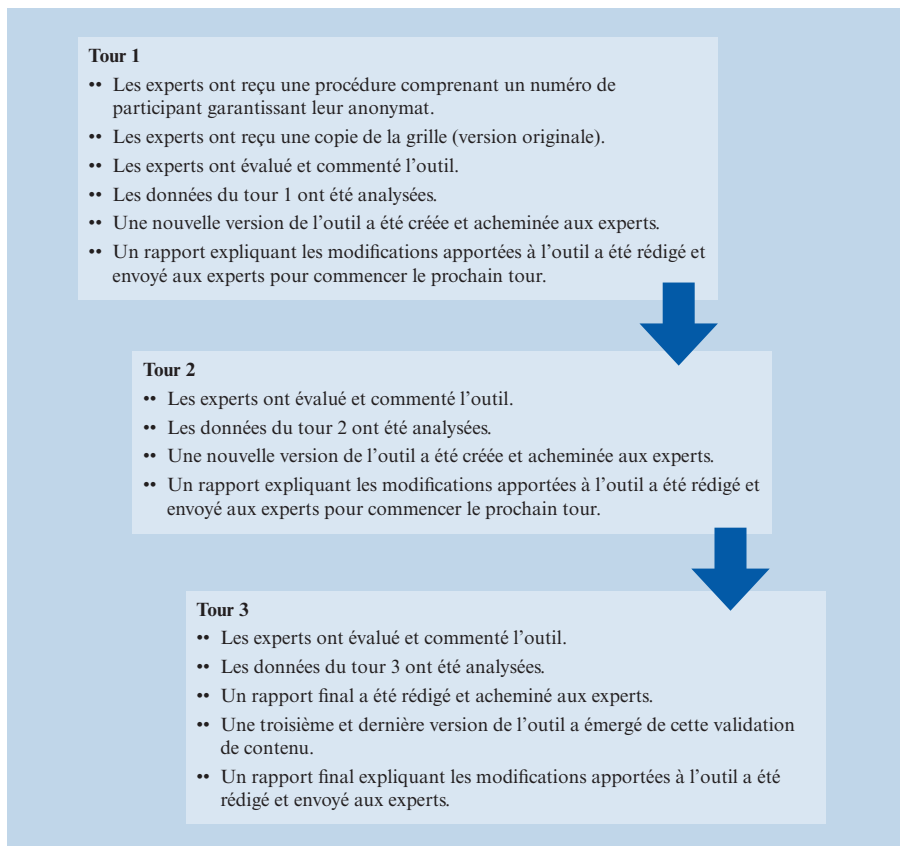
- 1) L'introduction
- 2) Le construit théorique des FE
  - Les composantes des FE : l'inhibition, la mémoire de travail, la flexibilité cognitive, la planification
- 3) Les mesures existantes (mesures administrées, mesures rapportées)
  - Les avantages et les limites
  - La pertinence de concevoir une mesure d'observation
- 4) La définition, la description et les items à observer (avec des exemples pour chaque item ; n = 27)
- 5) La procédure de cotation
- 6) La feuille de cotation.

Les experts ont d'abord analysé la pertinence théorique et la structure globale de l'instrument de mesure. Puis, après avoir examiné et hiérarchisé tous leurs commentaires, l'équipe de recherche a effectué les modifications requises. Elle a ensuite invité les experts à évaluer plus finement la section intitulée « La définition, la description et les items à observer » (le cœur de l'outil), selon la démarche explicitée ici-bas.

### ***L'application de la méthode Delphi pour valider l'outil d'observation***

Holey et al. (2007) suggèrent de mener trois tours d'évaluation afin d'augmenter le pourcentage d'accord entre les experts et, ainsi, améliorer le consensus. Cette manière de faire, appliquée par plusieurs autres chercheurs (Berquez et al. 2011 ; Holey et al., 2007 ; Parrish & Sandera, 2020), est celle qui a été retenue dans cette étude (voir la figure 2).

Figure 2  
*Les tours de la méthode Delphi*



Comme l'illustre la figure 2, trois tours ont été réalisés. Chaque tour de la méthode Delphi était composé de différentes étapes soit a) l'évaluation de l'outil par les experts, b) l'analyse qualitative des commentaires et l'analyse quantitative des scores accordés aux items de l'outil, c) la modification de l'outil, puis d) la rédaction d'un rapport expliquant aux experts les modifications apportées. Les paragraphes qui suivent explicitent l'application de ces étapes qui permet de dégager un consensus entre les experts.

### ***L'évaluation de l'outil d'observation des FE***

Pour chaque tour, les experts ont été invités à réaliser une double tâche d'évaluation : 1) analyser la pertinence théorique globale et la structure de l'outil, en commentant le manuel de cotation et 2) évaluer la pertinence des composantes/sous-composantes/marqueurs comportementaux en attribuant un score sur une échelle en cinq points allant d'atypique (1 = ne représente pas du tout le concept) à typique (5 = représente parfaitement le concept). Cette échelle partagée et utilisée par les experts se présente comme suit :

1 = en désaccord - l'indicateur ne représente pas du tout le concept

2 = plutôt en désaccord

3 = ni en accord ni en désaccord

4 = plutôt en accord

5 = en accord - l'indicateur représente tout à fait le concept.

Ils ont ensuite commenté chacune des sections de l'outil et remis une copie cotée et commentée de l'outil d'observation. Pour préserver leur anonymat tel que le requiert la méthode Delphi (Parrish & Sandera, 2020), une auxiliaire de recherche a anonymisé les copies reçues, en attribuant un numéro à chaque expert (expert 1, expert 2, etc.). Elle a ensuite transmis les grilles commentées et anonymisées à une autre auxiliaire qui a analysé les données anonymisées recueillies et collaboré à la production d'un rapport.

### ***L'analyse par l'équipe de recherche***

La validation de contenu, un processus tant qualitatif que quantitatif (Rusticus, 2014), vise à minimiser la variance d'erreur potentielle (sous-représentation, surreprésentation, mauvaise représentation) associée à l'instrument et à augmenter la probabilité de recueillir des preuves de la validité dans des études ultérieures (Haynes et al., 1995 ; Rubio et al., 2003). À cette fin, l'équipe de recherche a réalisé des analyses qualitatives et quantitatives à chacun des tours. Elle a tout d'abord effectué une analyse de contenu pour chaque section du manuel de l'observateur, à l'aide du

logiciel MaxQDA. Ce logiciel d'analyse qualitative sert à regrouper des ensembles de données selon un processus de tri, de manière à hiérarchiser les énoncés par ordre de priorité, à les structurer et à les analyser selon une approche analytique spécifique (Creswell & Poth, 2017). Ainsi, l'analyse des commentaires a permis de coder les données de manière systématique selon des thèmes («codes»), de noter des idées et de faire des associations.

Puis, afin d'examiner le niveau d'accord entre les experts (en tenant compte de leurs commentaires et des scores attribués aux items), des analyses descriptives ont été réalisées pour chaque item<sup>1</sup> de l'outil, à l'aide du logiciel Excel. Selon Waltz et al. (2017), il y a accord lorsque la valeur est égale ou supérieure à 0,80. Ainsi, les items ayant obtenu un accord de 0,80 et plus ont été conservés sans modification. L'utilisation d'un coefficient de kappa est conforme à d'autres travaux ayant utilisé la méthode Delphi (Holey et al., 2007). À chaque tour, des modifications tenant compte des divergences entre les données ont été apportées à l'outil, et ce, tant au niveau de sa structure que des items qui le composent. Ainsi, en conformité avec la méthode Delphi, toutes les décisions quant à l'outil de mesure ont été prises en tenant compte des divergences entre les experts (Ekionea et al., 2011). Après le troisième tour, seuls les items ayant un accord interjuges satisfaisant (égal ou supérieur à 0,80) ont été conservés dans la version validée de la grille.

Enfin, à chaque tour, l'équipe a produit un rapport présentant l'ensemble des modifications apportées à la grille, qui ont été faites grâce aux commentaires des experts. De façon plus précise, les rapports (n = 3) présentaient les modifications apportées : a) au manuel de l'observateur (procédure de cotation) et b) aux composantes/sous-composantes/indicateurs de la grille d'observation. Ainsi, les experts étaient invités à considérer les évaluations des pairs et à prendre connaissance des changements effectués par l'équipe de recherche (rapport des modifications et nouvelle grille d'observation) à chaque tour, afin de s'investir dans un nouveau cycle d'évaluation. Pour mieux comprendre les changements effectués, la section suivante présente les résultats des analyses (qualitatives et quantitatives), au regard de la méthode Delphi mise en place qui a mené à une grille validée et à un rapport final.

---

1. Dans le présent article, les sous-composantes de la grille d'observation et leurs marqueurs comportementaux sont nommés items.



## Résultats

Initialement, l'outil d'observation présentait quatre composantes des FE, 10 sous-composantes et 27 items (voir le tableau 1). À la fin des tours d'évaluation, la structure de l'outil d'observation a été modifiée, passant de 27 à 22 items. La structure finale de l'outil est présentée en conclusion de cette section. Focalisant sur la phase d'appréciation-quantification du processus de validation de contenu (Lynn, 1986), la présentation des résultats s'oriente sur les évaluations des experts, en dégageant les données découlant des analyses qualitatives et quantitatives menées à chacun des tours.

### *Le tour 1*

#### *Les analyses qualitatives*

À chaque tour, les commentaires des experts ont été classés selon leur valence «plutôt positive», «neutre» et «plutôt négative». Conformément à la méthode Delphi (Ekionea et al., 2011), seules les sections pour lesquelles la majorité des experts ( $n = 3$  et plus) était en désaccord ont été modifiées et révisées. L'équipe a par la suite procédé à une analyse des codes émergents pour s'assurer de bien saisir les questionnements et les demandes de modification des experts.

#### *La pertinence théorique globale et la structure de l'outil*

Les experts ont d'abord considéré la pertinence théorique de l'instrument de mesure, de même que sa structure globale. Rappelons que six sections composent l'instrument de mesure, tel qu'exposé dans la méthodologie. Tous les commentaires des experts ont d'abord été examinés et hiérarchisés, puis les modifications nécessaires ont été apportées. Le tableau 1 montre les ajustements qui ont été faits (tour 1) à la suite de cette analyse des commentaires.

En ce qui concerne le construit théorique des FE, des nuances ont été apportées à l'outil d'observation, et des précisions et des références ont été ajoutées. Puis, pour donner suite aux commentaires recensés, des nuances ont été apportées à la section «Mesures administrées et mesures rapportées», afin de mieux comprendre l'usage de cet instrument (grille d'observation) en contexte éducatif. Enfin, des précisions ont été ajoutées à la procédure de cotation, à la demande des experts qui ont formulé des commentaires en ce sens.

Tableau 1  
*Éléments constitutifs du premier rapport présenté aux experts*

Section concernée	Actions réalisées
Le construit théorique des FE	<ul style="list-style-type: none"> <li>• Nuances apportées, ajout de précisions sur le construit théorique des FE</li> <li>• Ajout de références scientifiques</li> </ul>
Les mesures administrées et les mesures rapportées	<ul style="list-style-type: none"> <li>• Nuances apportées concernant l'importance de développer une mesure écologique des FE en contexte éducatif</li> <li>• Clarification concernant les mesures administrées et les mesures rapportées: la grille d'observation des FE est complémentaire aux outils existants (tâches standardisées). Il n'est pas question de discréditer la pertinence de ces derniers.</li> </ul>
La procédure de cotation	<ul style="list-style-type: none"> <li>• Ajout de précisions sur la procédure de cotation</li> </ul>

### *L'analyse des commentaires sur les items de la grille*

Lors du premier tour d'évaluation, le panel d'experts a formulé 58 commentaires sur les items qui composent la grille d'observation ( $n = 27$ ). Les modifications sur l'instrument de mesure ont été réalisées à partir de ces commentaires. Rappelons qu'une analyse de contenu a été menée afin de classer les commentaires selon leur valence «plutôt positive», «neutre» et «plutôt négative». À la fin du tour 1, la majorité des experts a mentionné que les items 7 et 8 (reliés à l'*inhibition cognitive*) se ressemblaient. Par exemple, l'expert 3 a affirmé: «Je trouve que ces deux indicateurs renvoient à la même chose, la résistance à la distraction». Puisque ces deux items (7 et 8) ne sont pas mutuellement exclusifs et exhaustifs, aspects pourtant essentiels à considérer dans un outil d'observation (Bakeman & Quera, 2012; Heyman et al., 2014), seul l'item 8 a été conservé dans le but de se rapprocher de la définition d'inhibition cognitive (*interference control*) proposée, entre autres, par Nigg (2017) et par Diamond (2013).

Puis, la majorité des experts s'est interrogée sur la présence de l'item 12 (*MdT visuelle*) en contexte éducatif et sur la réelle possibilité de le mesurer de manière fidèle. À cet effet, l'expert 1 a commenté: «Je me demande s'il sera aisé pour l'observateur de distinguer entre le premier indicateur (visuel) et le troisième (spatial)?», et l'expert 4 a indiqué: «Pour moi c'est du rappel. L'enfant utilise des indices comme stratégie de rappel». Après avoir analysé l'ensemble des commentaires formulés par les experts, il a été convenu de jumeler l'item 12 à l'item précédent (tous deux rattachés à la *MdT visuelle*).

Toujours en lien avec la MdT, la plupart des experts se sont interrogés sur l'item 15, particulièrement au regard de son sens en contextes éducatifs de la petite enfance et de la possibilité de le mesurer de manière fidèle. Ce commentaire de l'expert 4 en témoigne : « Même commentaire que plus haut. Pour moi, ce sont des stratégies de rappel en mémoire ». Cet item a donc été jumelé à l'item 14 (*MdT verbale*). De surcroît, les items 17 et 18 (*rappel en mémoire*) ont été modifiés pour tenir compte des commentaires des experts, qui ont émis un doute sur la possibilité de les mesurer en contexte naturel (p. ex. pendant une période de jeu dans une classe). Ainsi, l'expert 3 a indiqué : « C'est un peu la même chose que l'indicateur 1 ». L'expert 2, pour sa part, s'est posé la question suivante :

Est-ce que cette section est en lien avec le *episodic buffer* du modèle de MdT ? Parce que si cette section rappel en mémoire est censée capturer le concept de *episodic buffer*, je ne suis pas sûr que ce soit le genre de chose que l'on puisse observer sur le plan comportemental... Les deux indicateurs me semblent associés à la mémoire à long terme et pas à la MdT.

Considérant que ces deux items ne répondaient pas au critère d'exclusivité mutuelle, seul le premier a été conservé.

Finalement, en ce qui a trait à l'item 23 (*planification*), la majorité des experts a remis en question sa pertinence, comme l'indique l'expert 1 : « Je ne suis pas certain de vraiment comprendre cet indicateur. À quoi ressemblerait un comportement insuffisant ? Par exemple, l'enfant n'arrive pas à choisir entre les différentes activités et donc ne fait rien du tout ou bien essaie de tout faire en même temps ? » L'expert 3 a formulé un commentaire similaire : « J'ai du mal à trouver une pertinence à dissocier *planification* et *organisation* des actions. J'aurai tendance à mettre les indicateurs sous une même catégorie (composante). » Conséquemment, l'item 23 a été supprimé.

### ***Les analyses quantitatives : Examen des niveaux d'accord sur les items de la grille***

Tel que le présente le tableau 2, les niveaux d'accord entre les experts variaient entre 0,48 et 1 à la suite de la première évaluation. Plus précisément, neuf items (sur 27) présentaient un niveau d'accord en deçà de 0,80. Rappelons que seuls les items ayant obtenu une médiane de trois ou moins, suivant l'échelle 1) en désaccord, 2) Plutôt en désaccord et 3) Ni en accord, ni en désaccord, pouvaient être révisés selon la méthode utilisée. De fait, parmi les items au niveau d'accord en deçà de 0,80, six items ont fait l'objet d'une modification conformément à la méthode Delphi

(Berquez et al., 2011). Puis, toujours en en adéquation avec la méthode utilisée, deux items ont été supprimés (items 7 et 23), tandis que d'autres ont été regroupés (items 17 + 18; items 11 + 12) puisqu'ils n'étaient pas mutuellement exclusifs selon le panel des experts.

Tableau 2  
*Les niveaux d'accord sur les items de la grille d'observation – Tour 1*

Item	Définition	Niveau d'accord
<b>Inhibition motrice</b>		
1	L'enfant peut arrêter (par lui-même) une action qu'il ne juge pas pertinente pour la tâche en cours.	0,92
2	L'enfant peut mettre en place des actions réfléchies en prenant son temps (contraire d'impulsivité).	0,72
3	L'enfant est en mesure d'arrêter une action lorsqu'un pair ou un adulte lui demande de le faire.	0,88
<b>Inhibition verbale</b>		
4	L'enfant est en mesure d'arrêter une action verbale (p. ex. chanter à voix haute) qu'il ne juge pas pertinente pour la tâche en cours.	0,96
5	Même s'il a une idée à énoncer, l'enfant arrive à attendre son tour de parole (p. ex. ne pas parler par-dessus les autres) lors d'une situation.	1
6	L'enfant est en mesure d'arrêter une action verbale (p. ex. chanter à voix haute) lorsqu'un pair ou un adulte lui demande de le faire.	0,92
<b>Inhibition cognitive</b>		
7	L'enfant parvient à rester centré sur l'autre lorsqu'un pair ou un adulte lui parle (attention envers les autres).	0,72
8	Malgré les distractions (p.ex. bruits, personnes), l'enfant ne semble pas dérangé/perturbé (contraire de <i>overstimulated</i> ).	0,92
<b>Inhibition émotionnelle</b>		
9	L'enfant reste facilement calme et disposé.	0,80
10	L'enfant parvient à retrouver son état émotionnel (régulation des émotions) rapidement.	0,88
<b>Mémoire de travail visuelle</b>		
11	L'enfant peut conserver en mémoire quelques informations visuelles en lien avec la situation.	0,96
12	L'enfant utilise délibérément les médiateurs visuels de son environnement pour parvenir à ses fins.	0,64
13	L'enfant parvient à conserver quelques caractéristiques spatiales en mémoire.	1

Tableau 2 (suite)  
*Les niveaux d'accord sur les items de la grille d'observation – Tour 1*

<b>Item</b>	<b>Définition</b>	<b>Niveau d'accord</b>
<b>Mémoire de travail verbale</b>		
14	L'enfant peut conserver en mémoire quelques informations verbales en lien avec la situation.	1
15	L'enfant utilise délibérément des indices d'ordre langagier pour arriver à ses fins.	0,64
16	L'enfant parvient à communiquer des informations de manière logique.	0,72
<b>Rappel en mémoire</b>		
17	L'enfant utilise délibérément des connaissances antérieures au profit de la situation en cours.	0,60
18	L'enfant fait des liens entre des éléments de sa vie au profit de la situation qu'il est en train de vivre.	0,48
<b>Adaptation à une nouvelle situation</b>		
19	L'enfant parvient à changer d'objectifs, de tâche ou de situation.	0,96
20	L'enfant s'ajuste aux rebondissements d'une situation.	0,96
21	L'enfant accepte de changer ses actions/stratégies par lui-même ou à la suite d'une demande (pair/adulte).	0,92
<b>Planification</b>		
22	L'enfant arrive à déterminer un but/objectif réaliste.	0,76
23	L'enfant parvient à faire un choix (p.ex. activité, matériel) malgré les différentes situations qui s'offrent à lui.	0,56
24	L'enfant discute, avec un pair ou avec l'adulte, des étapes à réaliser pour arriver à l'objectif donné, et ce, avant et/ou pendant l'activité prévue.	0,96
<b>Organisation des actions</b>		
25	L'enfant coordonne une séquence d'actions complexes pour arriver à ses fins.	0,88
26	L'enfant utilise le matériel de façon cohérente avec son objectif en cours.	0,96
27	L'enfant parvient à anticiper des événements futurs.	0,96

Après cette première analyse des commentaires (analyses qualitatives) et des scores attribués aux items (analyses quantitatives), l'équipe a rédigé un rapport et conçu une nouvelle version de la grille, qui présentait alors quatre composantes, 10 sous-composantes et 22 items (voir tableau 4 plus loin). Les documents ont été renvoyés aux experts pour une seconde évaluation (tour 2). La même procédure a été suivie pour le tour 3. La prochaine section présente les résultats obtenus pour ces deux derniers tours d'évaluation.

### ***Les tours 2 et 3***

#### ***Les analyses qualitatives***

##### *La pertinence théorique globale et la structure de l'outil*

Tel que le démontre le tableau 3, lors du deuxième tour, les commentaires des experts concernaient les items ( $n = 22$ ) de la grille plutôt que la pertinence théorique du construit. De plus, aucune modification n'a été effectuée lors du troisième et dernier tour, considérant la valence « plutôt positive » et « neutre » des commentaires des experts.

Tableau 3

#### ***Éléments constitutifs du deuxième rapport présenté aux experts***

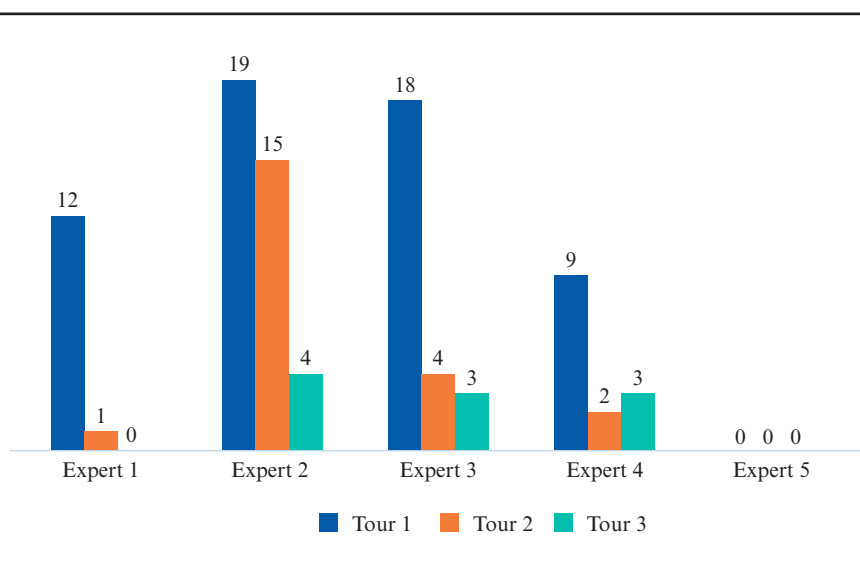
<b>Section concernée</b>	<b>Actions réalisées</b>
Tour 2 L'inhibition	<ul style="list-style-type: none"> <li>• Ajout des sous-types d'inhibition dans les items de la grille</li> <li>• Ajout de références scientifiques</li> </ul>
Exemples liés aux items	<ul style="list-style-type: none"> <li>• Modification de certains exemples liés aux items</li> </ul>
Tour 3 Aucune	<ul style="list-style-type: none"> <li>• Aucune modification</li> </ul>

Lors de la seconde ronde d'évaluation, les experts ont formulé des commentaires sur les items de la grille, en suggérant, notamment, d'y ajouter les sous-types d'inhibition comme l'indique le tableau 3. De plus, ils ont demandé des précisions concernant les exemples liés aux items de la grille. À titre illustratif, l'exemple du premier item de la sous-échelle inhibition motrice a été précisé afin de correspondre davantage à l'item et à sa description. Il en a été de même pour l'exemple concernant le deuxième item de la sous-échelle inhibition émotionnelle.

### *L'analyse des commentaires sur les items de la grille*

Outre l'augmentation du niveau d'accord observé entre les tours 2 et 3, une diminution importante du nombre de commentaires de la part des experts a aussi été relevée (T1 : n = 58 ; T2 : n = 22 ; T3 : n = 0), tel que l'illustre la figure 3.

Figure 3  
*Diminution des commentaires en fonction des tours*



Effectivement, 58 commentaires ont été formulés au premier tour contre seulement 10 à la dernière ronde d'évaluation. De plus, lors du troisième tour, deux experts n'ont formulé aucun commentaire et les trois autres n'en ont émis que quelques-uns (trois ou quatre). Entre les tours 2 et 3, nous sommes passées de trois commentaires (n = 3 experts) à un commentaire (n = 1 expert) pour un item relié à l'inhibition motrice.

À la fin du tour 3, l'expert 4 a réitéré son désaccord pour cet item (inhibition motrice) en mentionnant : « Je suis toujours en désaccord, car ce n'est pas parce que tu inhibes que tu agiras avec minutie et contrôle ». Par ailleurs, puisque la majorité des experts était en accord avec ledit item, le commentaire de l'expert 4 n'a pas été pris en considération, comme le suggère la méthode Delphi (Belquez et al., 2011).

### ***Les analyses quantitatives : examen des niveaux d'accord sur les items de la grille***

À la fin du tour 2, aucun item n'a été modifié en raison de l'augmentation du niveau d'accord. Le tableau 4 montre que le niveau d'accord varie de 0,72 à 1 au tour 2; deux items (5 - inhibition verbale, 21 - organisation des actions) ont fait l'unanimité à cette seconde ronde d'évaluation. De plus, comme 20 items sur 22 ont un niveau d'accord de plus de 0,88, ils font consensus selon la méthode Delphi (Belquez et al., 2011).

Tableau 4  
*Niveaux d'accord sur les items de la grille d'observation – tours 2 et 3*

<b>Item</b>	<b>Définition</b>	<b>Niveau d'accord T2</b>	<b>Niveau d'accord T3</b>
1	L'enfant peut arrêter (par lui-même) une action qu'il ne juge pas pertinente pour la tâche en cours.	0,96	1
2	L'enfant peut mettre en place des actions réfléchies en prenant son temps (contraire d'impulsivité).	0,72	0,80
3	L'enfant est en mesure d'arrêter une action lorsqu'un pair ou un adulte lui demande de le faire.	0,92	0,92
4	L'enfant est en mesure d'arrêter une action verbale (p. ex. chanter à voix haute) qu'il ne juge pas pertinente pour la tâche en cours.	0,96	1
5	Même s'il a une idée à énoncer, l'enfant arrive à attendre son tour de parole (p. ex. ne pas parler par-dessus les autres) lors d'une situation.	1	1
6	L'enfant est en mesure d'arrêter une action verbale (p. ex. chanter à voix haute) lorsqu'un pair ou un adulte lui demande de le faire.	0,96	0,96
7	Malgré les distractions (p. ex. bruits, personnes), l'enfant ne semble pas dérangé/perturbé (contraire de <i>overstimulated</i> ).	0,96	0,96
8	L'enfant reste facilement calme et disposé.	0,96	0,88
9	L'enfant parvient à retrouver son état émotionnel (régulation des émotions) rapidement.	0,96	1
10	L'enfant peut conserver en mémoire quelques informations visuelles en lien avec la situation et/ou utilise délibérément les médiateurs visuels de l'environnement pour parvenir à ses fins.	0,92	0,96



Tableau 4 (suite)  
*Niveaux d'accord sur les items de la grille d'observation – tours 2 et 3*

Item	Définition	Niveau d'accord T2	Niveau d'accord T3
11	L'enfant parvient à conserver quelques caractéristiques spatiales en mémoire.	0,96	0,96
12	L'enfant peut conserver en mémoire quelques informations verbales en lien avec la situation et/ou utilise délibérément des indices d'ordre langagier pour arriver à ses fins.	0,96	0,92
13	L'enfant parvient à communiquer des informations de manière logique.	0,88	0,92
14	L'enfant utilise délibérément des connaissances antérieures au profit de la situation en cours.	0,76	0,84
15	L'enfant parvient à changer d'objectifs, de tâche ou de situation.	0,92	0,96
16	L'enfant s'ajuste aux rebondissements d'une situation.	0,96	0,96
17	L'enfant accepte de changer ses actions/stratégies par lui-même ou à la suite d'une demande (pair ou adulte).	0,88	0,84
18	L'enfant arrive à déterminer un but/objectif réaliste.	0,92	0,92
19	L'enfant discute, avec un pair ou l'adulte, des étapes à réaliser pour arriver à l'objectif donné, et ce, avant et/ou pendant l'activité prévue.	0,96	0,92
20	L'enfant coordonne une séquence d'actions complexes pour arriver à ses fins.	0,88	0,92
21	L'enfant utilise le matériel de façon cohérente avec son objectif en cours.	1	1
22	L'enfant parvient à anticiper des événements futurs.	0,96	0,96

De façon similaire, aucun item n'a fait l'objet de modification à la fin du tour 3. En effet, tous les items ont obtenu un niveau d'accord supérieur à 0,80 indiquant des scores attribués de 4 et 5 (4 = Plutôt en accord ; 5 = En accord) à chacun des items. Comme le présente le tableau 4, le niveau d'accord varie entre 0,80 et 1 lors de ce dernier tour d'évaluation, ce qui montre le consensus des experts pour l'ensemble des items et leurs

marqueurs comportementaux. Plus précisément, cinq items sur 22 font l'unanimité (cote de 5 = En accord), tandis que quatre items obtiennent un niveau d'accord variant entre 0,80 et 0,88.

Pour conclure, au fil des tours d'évaluation, les commentaires des experts ont diminué et les niveaux d'accord ont augmenté. Ainsi, le niveau d'accord variait de 0,48 à 1 (neuf items sous 0,80) au tour 1, de 0,72 à 1 (deux items sous 0,80) au tour 2 et de 0,80 à 1 au tour 3. La prochaine section discute des résultats de l'étude de validation du contenu qui a permis d'obtenir un consensus entre les experts et propose une version finale de l'outil d'observation. Le tableau 5 présente la structure de l'outil validé.

Tableau 5  
*Structure finale de la grille d'observation des FE chez l'enfant de 3 à 6 ans en contextes éducatifs de la petite enfance*

Sous-échelles	Faible (1-2)	Moyen (3-4-5)	Élevé (6-7)
<b>1.1. Inhibition motrice</b>			
• Arrête une action par lui-même.	Il y a peu ou pas d'indices selon lesquels l'enfant arrive à contrôler ses comportements non verbaux et à mettre en place des actions réfléchies.	Les comportements non verbaux de l'enfant sont parfois contrôlés et mis en place de manière réfléchie, de façon autonome ou à la suite de la demande d'un pair.	Les comportements non verbaux de l'enfant sont souvent contrôlés et mis en place de manière réfléchie, de façon autonome ou à la suite de la demande d'un pair ou d'un adulte.
• Met en place des actions réfléchies (contraire d'impulsivité).			
• Arrête une action à la suite de la demande d'un pair/adulte.			
<b>1.2. Inhibition verbale</b>			
• Arrête une action verbale par lui-même.	Il y a peu ou pas d'indices selon lesquels l'enfant arrive à contrôler ses comportements verbaux.	Les comportements verbaux de l'enfant sont parfois contrôlés de manière autonome ou à la suite de la demande d'un pair ou d'un adulte.	Les comportements verbaux de l'enfant sont souvent contrôlés, de manière autonome ou à la suite de la demande d'un pair ou d'un adulte.
• Met en place des actions verbales réfléchies (contraire d'impulsivité).			
• Arrête une action à la suite de la demande d'un pair/adulte			

Tableau 5 (suite)  
*Structure finale de la grille d'observation des FE chez l'enfant de 3 à 6 ans  
 en contextes éducatifs de la petite enfance*

Sous-échelles	Faible (1-2)	Moyen (3-4-5)	Élevé (6-7)
<b>1.3. Inhibition cognitive</b>			
• N'est pas dérangé/perturbé.	Il y a peu d'indices selon lesquels l'enfant parvient à rester centré malgré les distractions. Il ne sait pas résister à l'interférence entre plusieurs informations.	L'enfant est parfois capable de rester centré malgré les distractions et de résister à l'interférence entre plusieurs informations.	L'enfant est souvent capable de rester centré malgré les distractions et de résister à l'interférence entre plusieurs informations.
<b>1.4. Inhibition émotionnelle</b>			
• Déploie des états émotionnels en synchronie avec son environnement.	Il y a peu d'indices voulant que l'enfant déploie des affects en synchronie avec son environnement et	L'enfant parvient parfois à déployer des affects en synchronie avec son environnement et à réguler ses émotions.	L'enfant est souvent capable de déployer des affects en synchronie avec son environnement et à réguler ses émotions.
• Retrouve son état émotionnel (régulation des émotions).	régule ses émotions.		
<b>2.1. Mémoire de travail visuelle</b>			
• Conserve en mémoire quelques informations visuelles et/ou utilise délibérément des médiateurs visuels.	Il y a peu d'indices selon lesquels l'enfant parvient à retenir des informations visuelles pertinentes.	L'enfant arrive parfois à retenir des informations visuelles pertinentes.	L'enfant arrive souvent à retenir des informations visuelles pertinentes.
• Conserve quelques caractéristiques spatiales en mémoire.			

Tableau 5 (suite)  
*Structure finale de la grille d'observation des FE chez l'enfant de 3 à 6 ans  
 en contextes éducatifs de la petite enfance*

Sous-échelles	Faible (1-2)	Moyen (3-4-5)	Élevé (6-7)
<b>2.2. Mémoire de travail verbale</b>			
<ul style="list-style-type: none"> <li>• Conserve en mémoire quelques informations verbales et/ou utilise délibérément des indices d'ordre langagier.</li> <li>• Communique des informations de manière à être compris.</li> </ul>	Il y a peu d'indices selon lesquels l'enfant parvient à retenir et à communiquer des informations verbales.	L'enfant arrive parfois à retenir et à communiquer des informations verbales.	L'enfant arrive souvent à retenir et à communiquer des informations verbales.
<b>2.3. Rappel en mémoire</b>			
<ul style="list-style-type: none"> <li>• Utilise délibérément des connaissances antérieures.</li> </ul>	Il y a peu d'indices selon lesquels l'enfant parvient à utiliser des connaissances antérieures au profit de la situation.	L'enfant arrive parfois à utiliser des connaissances antérieures au profit de la situation.	L'enfant arrive souvent à utiliser des connaissances antérieures au profit de la situation.
<b>3.1. Adaptation à une nouvelle situation</b>			
<ul style="list-style-type: none"> <li>• Parvient à changer d'objectif, de tâche ou de situation.</li> <li>• Ne montre pas de résistance devant un changement.</li> <li>• Accepte de changer ses actions/stratégies.</li> </ul>	Il y a peu d'indices indiquant que l'enfant parvient à s'ajuster à une nouvelle situation et aux rebondissements. De plus, peu d'indices sont observés quant aux changements d'actions/stratégies.	L'enfant parvient parfois à s'ajuster à une nouvelle situation et aux rebondissements. De plus, quelques indices sont observés quant aux changements d'actions/stratégies.	L'enfant est souvent en mesure de s'ajuster à une nouvelle situation et aux rebondissements. De plus, plusieurs indices sont observés quant aux changements d'actions/stratégies.

Tableau 5 (suite)  
**Structure finale de la grille d'observation des FE chez l'enfant de 3 à 6 ans en contextes éducatifs de la petite enfance**

Sous-échelles	Faible (1-2)	Moyen (3-4-5)	Élevé (6-7)
<b>4.1. Planification</b>			
<ul style="list-style-type: none"> <li>• Détermine un but/objectif réaliste.</li> <li>• Discute des étapes à réaliser pour arriver à l'objectif.</li> </ul>	<p>Il y a peu d'indices indiquant que l'enfant arrive à déterminer un but/objectif réaliste et à discuter des étapes à réaliser pour arriver à l'objectif donné.</p>	<p>L'enfant parvient parfois à déterminer un but/objectif réaliste et à discuter des étapes à réaliser pour arriver à l'objectif donné.</p>	<p>L'enfant est souvent en mesure de déterminer un but/objectif réaliste et de discuter des étapes à réaliser pour arriver à l'objectif donné.</p>
<b>4.2. Organisation des actions</b>			
<ul style="list-style-type: none"> <li>• Coordonne une séquence d'actions complexes.</li> <li>• Utilise le matériel de façon cohérente.</li> <li>• Anticipe les événements futurs.</li> </ul>	<p>Il y a peu d'indices indiquant que l'enfant peut coordonner une séquence d'actions complexes et utiliser le matériel de façon cohérente. L'enfant ne semble pas anticiper les événements futurs.</p>	<p>L'enfant parvient parfois à coordonner une séquence d'actions complexes, à utiliser le matériel de façon cohérente et à anticiper les événements futurs.</p>	<p>L'enfant est souvent en mesure de coordonner une séquence d'actions complexes, d'utiliser le matériel de façon cohérente et d'anticiper les événements futurs.</p>

Source: Duval et Montminy (2021)

## Discussion

L'importance de la validation du contenu d'un instrument psychométrique en fait une étape fondamentale de son élaboration (Zamanzadeh et al., 2015). Sans cette étape, il est impossible d'établir la validité et la fiabilité d'un instrument nouvellement développé. Dans le cadre de la présente étude, l'équipe a pu valider le contenu d'une grille d'observation des FE des enfants âgés entre 3 et 6 ans en contextes éducatifs de la petite enfance, en mettant en place les deux phases proposées par Lynn (1986), soit celle de développement et celle d'appréciation-quantification. Pendant cette deuxième phase, elle a appliqué la méthode Delphi pour prendre des décisions quant à la structure de l'instrument d'observation qui a été développé (Hohmann et al., 2018). Grâce à cette *méthode*, la phase

d'appréciation-quantification a pu atteindre systématiquement un consensus entre les experts d'un panel reconnu pour leurs travaux dans le champ à l'étude. Les prochaines sections discutent du processus de validation de cette recherche à partir de ces deux phases, en approfondissant la seconde et en présentant notamment les avantages et les limites de la méthode Delphi.

### ***La phase de développement***

#### ***Identifier et circonscrire les attributs à évaluer***

Découlant d'abord de travaux en neuropsychologie et en psychologie, le concept de FE est encore peu abordé dans les travaux en éducation. Il s'agit ainsi d'un champ en émergence, particulièrement en ce qui concerne l'évaluation des habiletés qui y sont liées (Zelazo et al., 2016). Pour créer un outil d'observation des FE chez les jeunes enfants, il a d'abord été essentiel de dresser l'état des connaissances sur l'objet à l'étude. L'équipe a ensuite développé un instrument d'observation en s'appuyant sur les dernières connaissances issues de la recherche. Cette étape majeure ne doit pas être prise à la légère (Arborio & Fournier, 2005), car elle permet de structurer la création de l'instrument. Dans le cadre de cette étude, l'état des connaissances sur les FE des jeunes enfants a permis de rendre compte de la pertinence de l'objet d'étude sur le plan théorique, en plus de concevoir une structure réfléchie et cohérente des items qui composent la grille d'observation. Les commentaires du panel d'experts en témoignent : ils ont positivement évalué la pertinence théorique globale et la structure de l'outil.

### ***La phase d'appréciation-quantification***

#### ***La méthode Delphi pour évaluer l'outil d'observation développé***

Selon Keeney et al. (2006), les équipes de recherche doivent adapter la méthode Delphi à leur étude, en fonction de leurs objectifs, en prenant soin d'équilibrer les avantages et les inconvénients potentiels de la démarche. Bien que la méthode Delphi ne soit pas appliquée de la même manière par tous, il est important de prendre certains postulats en considération afin d'atteindre un consensus entre les experts. Cette étude a permis, de façon novatrice, de valider le contenu d'un nouvel outil d'observation sur les FE des jeunes enfants. Il importe toutefois de préciser les limites de l'application de la méthode, telle que mise en place dans cette recherche.

### ***Le recrutement des participants***

Une critique courante concernant le recrutement des experts est l'existence potentielle de biais, de jugements ou d'opinions (Berquez et al., 2011) en fonction de leur champ d'expertise. Il est donc important de définir des critères de sélection avant leur recrutement et de constituer un échantillon hétérogène pour s'assurer d'obtenir un large éventail de commentaires (Ekionea et al., 2011). Dans le cadre de cette étude, les participants sont des spécialistes du développement cognitif et des FE de l'enfant, mais leurs travaux ne s'inscrivent pas autour des mêmes composantes. À titre illustratif, un participant a publié au sujet de la MdT tandis qu'un autre concentre plutôt ses travaux sur les habiletés d'inhibition.

Le nombre d'experts à convoquer est également un élément fondamental à prendre en compte. Ici, le nombre de participants a été limité à cinq, ce qui peut paraître peu comparé à d'autres études (p. ex. Leigh (2020) qui a convoqué 21 experts). Par contre, le fait de recruter un panel d'experts moins important a permis d'assurer la rétention des participants, et ce, aux trois tours d'évaluation. En effet, tous les experts recrutés ont pris part à l'entièreté de la démarche, comparativement à d'autres études qui ont montré de l'attrition dans leur panel d'évaluation. Par exemple, Jones et al. (2021), qui ont recruté 12 participants au tour 1 de leur démarche, ont terminé leur étude avec sept experts.

Le processus choisi pour la présente étude a favorisé une stabilité dans le panel d'experts, qui constitue un critère important pour assurer la fiabilité et la validité des résultats obtenus avec la méthode Delphi (Green, 2014). Enfin, Keeney et al. (2006) estiment qu'il n'existe pas de nombre idéal pour constituer le panel d'experts selon cette méthode.

### ***La rigueur de la démarche utilisée***

Il a été possible d'atteindre un consensus entre les experts, soit la dernière étape du protocole, grâce aux trois tours d'évaluation qui ont été réalisés avec rigueur. Avant d'enclencher le processus d'évaluation, il a fallu prendre des décisions concernant le nombre de tours, leur durée (c.-à-d., le temps alloué pour procéder à l'évaluation de l'outil) et la manière de tenir compte des commentaires des experts.

### ***Le nombre de tours***

Rappelons que certains auteurs suggèrent de réaliser trois tours d'évaluation (Berquez et al., 2011 ; Holey et al., 2007), puisque les participants peuvent changer d'avis sur leur niveau d'accord jusqu'à la toute fin du

protocole. Dans cette étude, le troisième et dernier tour d'évaluation a mené à l'atteinte d'un consensus, d'un point de vue tant qualitatif que quantitatif. Conformément aux travaux de Holey et al. (2007), la présente étude suggère, tant par la réduction du nombre de commentaires que par l'augmentation du niveau d'accord, une diminution de la divergence et un renforcement du consensus entre les experts.

### *La durée de la démarche*

La durée de la démarche peut constituer une limite dans la méthode Delphi, entre autres en ce qui concerne le temps et l'énergie à déployer. En effet, le panel d'experts peut ressentir de la fatigue au troisième tour d'évaluation, ce qui peut amener les participants à « forcer le consensus » (Hsu & Sandford, 2007 ; Nworie, 2011). Pour pallier cette limite, une stratégie consiste à diminuer le nombre de tours lorsque les participants manifestent de la fatigue (Keeney et al., 2006), ce qui, selon McKenna (1994), se produit notamment avec des experts très occupés, qui ont beaucoup d'engagements en même temps. Ainsi, en cohérence avec d'autres études (Ekionea et al., 2011 ; Sekayi & Kennedy, 2017), trois tours ont été effectués et ont suffi pour atteindre le consensus. Les analyses qualitatives et quantitatives en témoignent d'ailleurs, comme évoqué précédemment.

Aussi, il est recommandé de valoriser l'apport des experts et de leur montrer qu'ils sont des partenaires dans la démarche, en sollicitant leur collaboration sur un sujet qui les intéresse (Berquez et al., 2011). Dans le cadre de cette étude, les travaux des experts portaient sur au moins une composante des FE, ce qui témoigne de leur intérêt envers le sujet et de leur possible apport à la recherche. Il est effectivement essentiel que les experts sachent que leurs réponses influencent réellement le protocole de recherche (Berquez et al., 2011), ici, la création d'une grille d'observation sur les FE.

### *La manière de tenir compte des commentaires des experts*

Dans le protocole de la recherche, il importe de réfléchir à la manière de tenir compte des commentaires des experts et à la façon dont les ajustements et les modifications sur l'instrument de mesure sont communiqués après chaque tour ( $n = 3$ ). Grâce à la méthode Delphi, l'équipe a pu prendre en considération les commentaires des experts à chaque tour et, par conséquent, ajuster la grille de manière rigoureuse. Il faut cependant relever certaines limites quant à la formulation des commentaires des experts.



D'abord, certains ont énoncé le même commentaire aux différents tours d'évaluation, mais il n'a jamais été pris en compte puisque le consensus sur l'item en question était atteint, dans le respect de la méthode Delphi. Ainsi, un expert a pu émettre un commentaire sur un item (aspect qualitatif), tout en étant en accord avec sa présence dans l'outil en attribuant une cote de 4 ou 5 sur l'échelle de Likert (aspect quantitatif). Lorsque l'accord a été atteint pour l'item en question, tous les commentaires qui y étaient associés n'ont pas été considérés (du moins, aucune modification à la grille n'a été apportée). Cette façon de procéder, conforme à la méthode Delphi (Berquez et al., 2011), peut donner l'impression aux experts que leurs commentaires ne sont pas pris en compte, ce qui peut leur occasionner une certaine frustration. Par conséquent, puisqu'ils peuvent modifier leurs niveaux d'accord jusqu'au dernier tour, il est possible de penser que cette «non-considération» de leurs commentaires influence le niveau d'accord pour certains items tout au long de la démarche.

D'ailleurs, dans le cadre de cette étude, quatre items ont connu une baisse d'accord entre le premier tour d'évaluation et le troisième (items 8, 12, 17 et 19). À titre illustratif, l'item 17 présentait un niveau d'accord de 0,88 au tour 1, qui a baissé à 0,84 au tour 3. Bien qu'il importe de reconnaître cette limite, il demeure que les taux d'accord sont restés très élevés, c'est-à-dire au-dessus de 0,80. Ainsi, selon la méthode Delphi, l'équipe de chercheurs développant l'outil n'y a apporté aucune modification, car un niveau de consensus entre les participants était atteint.

Dans le même ordre d'idées, selon Hasson et al. (2000), une autre limite à relever dans la méthode Delphi est le fait que les experts ne puissent pas discuter ouvertement des commentaires qu'ils ont émis sur les items à évaluer. De même, des chercheurs estiment que cette méthode ne permet pas aux participants de préciser leur point de vue (Hasson et al., 2000; Walker & Selfe, 1996;). Bien que les experts puissent commenter les items de l'outil par écrit en formulant des remarques dans le document (aspect qualitatif), certains aimeraient pouvoir les expliquer de vive voix à l'équipe de recherche, de manière à expliciter leurs points de vue (Hohmann et al., 2018). Toutefois, comme le protocole, qui est rigoureux et systématique, permet aux experts de s'exprimer librement sous le couvert de l'anonymat, il va de soi qu'ils ne peuvent pas entrer directement en contact avec l'équipe de recherche.

Il est donc essentiel, en tout début de parcours, d'expliquer la procédure de la méthode Delphi aux participants, en rappelant que l'équipe tente de trouver un consensus sur un concept, une question ou un problème. L'existence d'un consensus ne signifie pas que les personnes ont trouvé la bonne réponse, opinion ou jugement, mais qu'une majorité d'entre elles sont en accord sur un même construit (Hasson et al., 2000). Pour conclure, malgré quelques limites rattachées à la procédure utilisée, le nombre de travaux en éducation qui font appel à la méthode Delphi est assez limité, mais il est en augmentation. Si quelques travaux récents l'ont utilisée (Puertas-Aguilar et al., 2021), peu ont porté sur les contextes éducatifs de la petite enfance. Étant donné que la démarche a été définie selon des critères de validité rigoureux, il est possible de penser que d'autres équipes de recherche pourront s'inspirer de la présente étude, notamment dans la validation du contenu d'un nouvel instrument de mesure.

## Conclusion

Cette étude a permis de valider une grille d'observation des FE chez les enfants âgés entre 3 et 6 ans en contextes éducatifs de la petite enfance. Initialement, 27 items composaient l'outil d'observation. Après la phase d'appréciation-quantification, 22 items ont été retenus dans la structure de l'instrument. Bien que des limites concernant l'application de la méthode Delphi aient été relevées dans la validation de l'outil d'observation, cette étude s'avère novatrice. Ayant réussi à obtenir un consensus à l'égard des composantes, des sous-composantes et des items qui composent l'instrument d'observation des FE chez les jeunes enfants (outil qui n'existait pas jusqu'alors), cette étude se révèle l'une des rares à avoir appliqué la méthode Delphi dans le champ de la petite enfance, une période pourtant critique pour les apprentissages et le développement des enfants.

En effet, à notre connaissance, une seule étude dans le champ de la petite enfance a utilisé cette méthode en misant sur l'évaluation de certaines habiletés développementales inhérentes à la préparation scolaire (traduction libre de *school readiness*) des jeunes enfants (Skelton & Leclair, 2019). Or, il serait souhaitable de s'intéresser davantage à la méthode Delphi en éducation à la petite enfance, car elle donnerait l'occasion à différents professionnels de développer une vision commune d'un concept ou d'un problème donné (ici, les FE), en plus de prendre des décisions éclairées et de faire des suggestions pertinentes pour soutenir les apprentissages et

le développement des enfants, voire leur réussite éducative. Plus précisément, la méthode Delphi pourrait favoriser le développement ou l'évaluation d'outils efficaces (p. ex. s'assurer de la validité de contenu d'un outil élaboré avant de le proposer aux intervenants). Elle pourrait également informer le milieu de la pratique (p. ex. le personnel enseignant) sur différents enjeux et, ainsi, améliorer les connaissances des praticiens sur le développement de l'enfant (Skelton & Leclair, 2019).

Enfin, pour reconnaître les manifestations comportementales des FE en contextes éducatifs de la petite enfance et les soutenir au quotidien, il semble pertinent d'appliquer la méthode Delphi auprès de plusieurs types d'experts, c'est-à-dire auprès des professionnels du milieu de la pratique (intervenants, éducateurs, enseignants). En effet, après avoir sollicité un panel d'experts dans le milieu de la recherche, il serait pertinent de sonder les experts dans le milieu de la pratique, afin d'obtenir une validité éprouvée dans la consolidation de l'outil d'observation. Plus précisément, cela donnerait l'occasion d'envisager plusieurs perspectives au sujet d'un même objet (les FE), de contribuer à l'actualisation des connaissances dans un même champ et de réfléchir à leur transférabilité et à leur crédibilité (Nworie, 2011 ; Leigh, 2020). En somme, cela favoriserait le développement de liens entre la recherche et la pratique, en prenant en compte des expertises diversifiées et complémentaires (chercheurs universitaires, praticiens), qui auront, dans tous les cas, un impact sur la mise en place de l'outil dans les milieux de pratique (Green, 2014).

Révision linguistique : Marie-Claire Legaré

Mise en page : Emmanuel Gagnon

Résumé en portugais : Eusébio André Machado

Réception : 14 avril 2023

Version finale : 01 décembre 2023

Acceptation : 15 janvier 2024

## LISTE DE RÉFÉRENCES

- Anderson, P. J. & Reidy, N. (2012). Assessing Executive Function in Preschoolers. *Neuropsychology Review*, 22, 345-360. <http://dx.doi.org/10.1007/s11065-012-9220-3>.
- Arborio, A. -M. & Fournier, P. (2005). L'enquête et ses méthodes: l'observation directe, *Population*, 55(4/5). doi: 10.2307/1534705.
- Bahadir, F. & Tuncer, M. (2020). Determining the Standards of Teaching and Learning Process as a Component of Curriculum. *International Journal of Progressive Education*, 16(3), 34-52.
- Bakeman, R. & Quera, V. (2011). Sequential Analysis and Observational Methods for the Behavioural Sciences. Cambridge University Press.
- Barros, P. & Hazin, I. (2013). Avaliação das funções executivas na infância: Revisão dos conceitos e instrumentos [Assessment of executive functions in childhood: Review of concepts and instruments]. *Psicologia Em Pesquisa*, 7(1), 13-22. <https://doi.org/10.5327/Z19821247201300010003>
- Barros, P., Metta, L., Peralba, C., Vilar, C., Guerra, A., Paula, A., Argollo, N. & Hazin, I. (2016). Perfil desenvolvimental das funções executivas utilizando o NEPSY-II em crianças de 5 a 8 anos [Developmental profile of executive functions using the NEPSY-II in 5- to 8-year-old children]. *Revista de Neuropsicologia Latinoamericana*, 8(2), 1-15. <https://doi.org/10.5579/rnl.2016.0295>
- Berquez, A. E., Cook, F. M., Millard, S. K. & Jarvis, E. (2011). The Stammering Information Programme: a Delphi study. *Journal of Fluency Disorders*, 36(3), 206-221. doi: 10.1016/j.jfludis.2011.02.007.
- Bjorklund, D. F. & Causey, K. (2018). Children's thinking: Cognitive development and individual differences. SAGE Publications.
- Blair, C., Granger, D. & Razza, R. P. (2005). Cortisol reactivity is positively related to executive function in preschool children attending Head Start. *Child Development*, 76, 554-567.
- Blair, C. & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, 78(2), 647-663. <https://doi-org.acces.bibl.ulaval.ca/10.1111/j.1467-8624.2007.01019.x>
- Cadavid-Ruiz, N. & del Río, P. (2018). An Observational Analysis of Executive Performance in School Children. *Psychology in Russia: State of the Art*, 11(3). doi: 10.11621/PIR.2018.0314
- Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental neuropsychology*, 28(2), 595-616. [https://doi.org/10.1207/s15326942dn2802\\_3](https://doi.org/10.1207/s15326942dn2802_3)
- Ciptono, A., Setiyono, S., Nurhidayati, F. & Vikaliana, R. (2019). Fuzzy Delphi method in education: A mapping. *Journal of Physics Conference Series*, 1360(1). doi:10.1088/1742-6596/1360/1/012029
- Colquitt, J. A., Sabey, T. B., Rodell, J. B. & Hill, E. T. (2019). Content validation guidelines: Evaluation criteria for definitional correspondence and definitional distinctiveness. *Journal of Applied Psychology*, 104(10), 1243. <https://doi.org/10.1037/apl0000406>

- Corbière, M. & Fraccaroli, F. (2014) La conception, la validation, la traduction et l'adaptation transculturelle d'outils de mesure : des exemples en santé mentale et travail. Dans M. Corbière & N. Larivière (dir.), *Méthodes qualitatives, quantitatives et mixtes : dans la recherche en sciences humaines, sociales et de la santé* (p. 577-624). Presses de l'Université du Québec.
- Creswell, J. & Poth, C. (2017) *Qualitative Inquiry and Research Design: Choosing among Five Approaches*. SAGE Publications.
- Davidson, M. C., Amso, D., Anderson, L. C. & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037-2078. <https://doi-org.acces.bibl.ulaval.ca/10.1016/j.neuropsychologia.2006.02.006>
- Davis, H. L. & Pratt, C. (1995). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology*, *47*(1), 25-1. <https://doi-org.acces.bibl.ulaval.ca/10.1080/00049539508258765>
- Denckla, M. (2002). The behavior rating inventory of executive function: Commentary. *Child Neuropsychology*, *8*(4), 304-306. <https://doi.org/10.1076/chin.8.4.304.13512>
- Désormeaux-Moreau, M., Naud-Lepage, G. & Drolet, M. -J. (2020). Assurer la validité de contenu lors du développement d'un outil d'évaluation : Soutenir l'ergothérapeute pas à pas. *Revue Francophone de Recherche en Ergothérapie*, *6*(2), 75-97. Doi:10.13096/rfre.v6n2.178
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135-168. <https://doi-org.acces.bibl.ulaval.ca/10.1146/annurev-psych-113011-143750>
- Doebel, S. (2020). Rethinking Executive Function and Its Development. *Perspectives on psychological science: a journal of the Association for Psychological Science*, *15*(4), 942-956. <https://doi.org/10.1177/1745691620904771>
- Duval, S., Bouchard, C., Matte-Gagné, C. & Brault-Foisy, L.-M. (2020-2023). *Validation d'une grille d'observation des FE chez les enfants âgés de trois à cinq ans en contextes éducatifs*. Projet subventionné par le CRSH, Développement Savoir. Gouvernement du Canada.
- Duval, S. & Montminy (2021). Grille d'observation des fonctions exécutives en contextes éducatifs de la petite enfance. Document inédit de recherche. Université Laval, Québec.
- Duval, S., Montminy, N., Brault-Foisy, L.-M., Arapi, E. & Vézina, S.-A. (2023). Examining the relation between adult scaffolding of make-believe play and children's executive functions: an observational study conducted in a natural educational setting. *Early Child Development and Care*, *193*(8), 1022-1040. <https://doi.org/10.1080/03004430.2023.2216891>
- Duval, S., Montminy, N. & Gaudette-Leblanc, A. (2018). Perspectives théoriques à l'égard des fonctions exécutives en contexte éducatif préscolaire, *Neuroéducation*, *5*(2), 93-108.
- Ekionea, B. J. P., Bernard, P. & Plaisent, M. (2011). Consensus par la méthode Delphi sur les concepts clés des capacités organisationnelles spécifiques de la gestion des connaissances. *Recherches qualitatives*, *29*(3), 168-192. <https://doi.org/10.7202/1085878ar>
- Escolano-Pérez, E., Herrero-Nivela, M. L., Blanco-Villaseñor, A. & Anguera, M. T. (2017). Systematic Observation: Relevance of This Approach in Preschool Executive Function Assessment and Association with Later Academic Skill. *Frontiers in Psychology*, *8*. <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2017.02031>

- Fleer, M., Veresov, N. & Walker, S. (2017). Re-conceptualizing executive functions as social activity in children's playworlds. *Learning, Culture and Social Interaction*, 14, 1–11. <https://doi.org/10.1016/j.lcsi.2017.04.003>
- Foy, J. & Mann, V. (2012). Executive function and early reading skills. *Reading and Writing*, 26. <https://doi.org/10.1007/s11145-012-9376-5>
- Gioia, G. A., Espy, K. A. & Isquith, P. K. (2002). Behavior Rating Inventory of Executive Function, Preschool Version (BRIEF-P). *Odessa: Psychological Assessment Resources*.
- Goodman, C. M. (1987). The Delphi technique: a critique. *Journal of Advanced Nursing*, 12, 729–734. <http://dx.doi.org/10.1111/j.1365-2648.1987.tb01376.x>
- Green, R. A. (2014). The Delphi Technique in Educational Research. *SAGE Open*, 4(2). <https://doi-org.acces.bibl.ulaval.ca/10.1177/2158244014529773>
- Haidari, S. M. & Uzun, N. B. (2019). Content validity and confirmatory factor analysis of Cooperative Learning Attitude Scale (CLAS) for the EFL students. *International Journal of Social Sciences and Education Research*, 5(4), 418-433.
- Hasson, F., Keeney, S. & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of advanced nursing*, 32(4), 1008–1015.
- Haynes, S. N., Richard, D. C. S. & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7, 238-247.
- Heyman, R. E., Lorber, M. F., Eddy, J. M. & West, T. V. (2014). Behavioral observation and coding. Dans H. T. Reis & C. M. Judd (dir.), *Handbook of research methods in social and personality psychology* (p. 345–372). Cambridge University Press.
- Hohmann, E., Brand, J. C., Rossi, M. J. & Lubowitz, J. H. (2018). Expert Opinion Is Necessary: Delphi Panel Methodology Facilitates a Scientific Approach to Consensus *Arthroscopy: the journal of arthroscopic & related surgery: official publication of the Arthroscopy Association of North America and the International Arthroscopy Association*, 34(2), 349–351. <https://doi.org/10.1016/j.arthro.2017.11.022>
- Holey, E. A., Feeley, J. L., Dixon, J. & Whittaker, V. J. (2007). An exploration of the use of simple statistics to measure consensus and stability in Delphi studies. *BMC Medical Research Methodology*, 7(1), 52. <https://doi.org/10.1186/1471-2288-7-52>
- Hoyt, C. R., Fernandez, J. D., Varughese, T. E., Grandgeorge, E., Manis, H. E., O'Connor, K. E., Abel, R. A. & King, A. A. (2020). The Infant Toddler Activity Card Sort: A Caregiver Report Measure of Children's Occupational Engagement in Family Activities and Routines. *Occupational Therapy Journal of Research*, 40(1), 3641. <https://doi.org/10.1177/1539449219852030>
- Hsu, C. & Sandford, B. (2007). The Delphi Technique: Making Sense of Consensus. Practical Assessment, *Research & Evaluation*, 12, 1-8. <http://pareonline.net/pdf/v12n10.pdf>
- Isquith, P., Gioia, G. & Espy, K. (2004). Executive Function in Preschool Children: Examination Through Everyday Behavior. *Developmental neuropsychology*, 26, 403-422. [https://doi.org/10.1207/s15326942dn2601\\_3](https://doi.org/10.1207/s15326942dn2601_3)
- Jones, J., Belcher, G. & Elliott, K. (2021). Identifying Technical Competencies for Architecture and Construction Education Using the Delphi Method. *Career and Technical Education Research*, 46(1), 3-15. <https://doi.org/10.5328/cter46.1.3>

- Kazemi, A. S. & Mohammadi, Z. (2019). The Effect of Working Memory Training on Executive Function of Children with Attention Deficit/Hyperactivity Disorder. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, 10(3), 134-141. <https://lumenpublishing.com/journals/index.php/brain/article/view/2175>
- Keeney, S., Hasson, F. & McKenna, H. P. (2001). A critical review of the Delphi technique as a research methodology for nursing. *International journal of nursing studies*, 38(2), 195–200. [https://doi.org/10.1016/s0020-7489\(00\)00044-4](https://doi.org/10.1016/s0020-7489(00)00044-4)
- Keeney, S., Hasson, F. & McKenna, H. (2006). Consulting the oracle: ten lessons from using the Delphi technique in nursing research. *Journal of advanced nursing*, 53(2), 205–212. <https://doi.org/10.1111/j.1365-2648.2006.03716.x>
- Kovács, Á. M. & Mehler, J. (2009). Flexible Learning of Multiple Speech Structures in Bilingual Infants. *Science*, 325(5940), 611–612. <https://doi.org/10.1126/science.1173947>
- Laveault, D. & Grégoire, J. (2014). Introduction aux théories des tests en psychologie et en sciences de l'éducation (3<sup>e</sup> éd.). De Boeck Université.
- Leigh, H. (2020). Signature Pedagogies for Art Therapy Education: A Delphi Study. *Art Therapy*, 38, 18. <https://doi.org/10.1080/07421656.2020.1728180>
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35, 382-385.
- McClelland, M. M. & Cameron, C. E. (2012). Self-regulation in early childhood: Improving conceptual clarity and developing ecologically valid measures. *Child Development Perspectives*, 6(2), 136–142. <https://doi.org/10.1111/j.1750-8606.2011.00191.x>
- McKenna, H. P. (1994). The essential elements of a practitioners' nursing model: a survey of psychiatric nurse managers. *Journal of Advanced Nursing*, 19, 870-877. <https://doi.org/10.1111/j.1365-2648.1994.tb01163.x>
- Ministère de l'Éducation du Québec. (2021). Programme de formation de l'école québécoise. Programme-cycle de l'éducation préscolaire. Gouvernement du Québec.
- Miyake, A., Friedman, N., Emerson, M., Witzki, A., Howerter, A. & Wager, T. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100. <https://doi.org/10.1006/cogp.1999.0734>
- Mulcahy, C., Day Hess, C. A., Clements, D. H., Ernst, J. R., Pan, S. E., Mazzocco, M. M. & Sarama, J. (2021). Supporting Young Children's Development of Executive Function Through Early Mathematics. *Policy Insights from the Behavioral and Brain Sciences*, 8(2), 192–199.
- Nasa, P, Jain, R. & Juneja, D. (2021). Delphi methodology in healthcare research: How to decide its appropriateness. *World Journal of Methodology*, 11(4), 116-129. doi: 10.5662/wjm.v11.i4.116
- Nieto, M., Ros, L., Medina, G., Ricarte, J. J. & Latorre, J. M. (2016). Assessing executive functions in preschoolers using shape school task. *Frontiers in Psychology*, 7(1489). <https://doi-org.acces.bibl.ulaval.ca/10.3389/fpsyg.2016.01489>
- Nigg, J. T. (2017). Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of child psychology and psychiatry, and allied disciplines*, 58(4), 361–383. <https://doi.org/10.1111/jcpp.12675>
- Nworie, J. (2011). Using the Delphi Technique in Educational Technology Research. *TechTrends*, 55, 24–30. <https://doi.org/10.1007/s11528-011-0524-6>



- Parrish, A. H. & Sadera, W. A. (2020). Teaching Competencies for Student-Centered, One-to-One Learning Environments: A Delphi Study. *Journal of Educational Computing Research*, 57(8), 1910–1934. <https://doi.org/10.1177/0735633118816651>
- Polit, D. F., Beck, C. T. & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity appraisal and recommendations. *Research in Nursing & Health*, 30(4), 459-467. <https://doi.org/10.1002/nur.20199>
- Powell, A., Bennett, N. & Lamm, K. (2021). Expanding 4-H alumni life perspectives: An experiential and programmatic evaluation. *Journal of Agricultural Education*, 62(2), 185-200. <http://doi.org/10.5032/jae.2021.02185>
- Puertas-Aguilar, M.-Á., Álvarez-Otero, J. & de Lázaro-Torres, M.-L. (2021). The Challenge of Teacher Training in the 2030 Agenda Framework Using Geotechnologies. *Education Sciences*, 11(8), 381. <http://dx.doi.org/10.3390/educsci11080381>
- Ribner, A. D., Willoughby, M. T. & Blair, C. B. (2017). Executive Function Buffers the Association between Early Math and Later Academic Skills. *Frontiers in Psychology*, 8(869). <https://doi.org/10.3389/fpsyg.2017.00869>
- Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S. & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27(2), 94-104. <https://doi.org/10.1093/swr/27.2.94>
- Rusticus, S. (2014). Content Validity. Dans A. C. Michalos (dir.), *Encyclopedia of Quality of Life and Well-Being Research*. Springer, Dordrecht. [https://doi-org.acces.bibl.ulaval.ca/10.1007/978-94-007-0753-5\\_553](https://doi-org.acces.bibl.ulaval.ca/10.1007/978-94-007-0753-5_553)
- Sablitzky, T. (2022). The Delphi Method. *Hypothesis: Research Journal for Health Information Professionals*, 34(1). <https://doi.org/10.18060/26224>
- Sekayi, D. & Kennedy, A. (2017). Qualitative Delphi Method: A Four Round Process with a Worked Example. *Qualitative Report*, 22, 2755-2763. <https://doi.org/10.46743/2160-3715/2017.2974>
- Silva, C., Sousa-Gomes, V., Fávero, M., Oliveira-Lopes, S., Merendeiro, C. S., Oliveira, J. & Moreira, D. (2022). Assessment of preschool-age executive functions: A systematic review. *Clinical Psychology & Psychotherapy*, 29(4), 1374-1391. <https://doi.org/10.1002/cpp.2718>
- Sireci, S. G. (1998). Gathering and analyzing content validity data. *Educational Assessment*, 5(4), 299-321. [https://doi.org/10.1207/s15326977ea0504\\_2](https://doi.org/10.1207/s15326977ea0504_2)
- Skelton, H. & Leclair, L. (2019). The Early Development Instrument – Creation of a Fine Motor/Visual Motor Index. *Journal of Occupational Therapy, Schools, & Early Intervention*, 12(3), 284-297. <https://doi.org/10.1080/19411243.2019.1590753>
- Streiner, D. L. (2014) Measurement. Dans K. Bassil & D. Zabkiewicz (dir.), *Health Research Methods: A Canadian Perspective* (p. 119-142). Oxford University Press.
- Toplak, M. E., West, R. F. & Stanovich, K. E. (2013). Practitioner review: do performance-based measures and ratings of executive function assess the same construct? *Journal of child psychology and psychiatry, and allied disciplines*, 54(2), 131–143. <https://doi.org/10.1111/jcpp.12001>
- Walker, A. M. & Selfe, J. (1996). The Delphi method: A useful tool for the allied health researcher. *British Journal of Therapy and Rehabilitation*, 3. <https://doi.org/10.12968/bjtr.1996.3.12.14731>
- Waltz, C. F., Strickland, O. & Lenz, E. R. (2017). Chapter 6. Validity of Measures. *Measurement in Nursing and Health Research* (5<sup>th</sup> ed.). Springer Publishing Company



- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children* (4<sup>th</sup> ed.). PsychCorp.
- Wiebe, S., Sheffield, T., Nelson, J., Clark, C., Chevalier, N. & Espy, K. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, 108(3), 436–452. <https://doi.org/10.1016/j.jecp.2010.08.008>
- Willoughby, M. T., Blair, C. B. & Family Life Project Investigators. (2016). Measuring executive function in early childhood: A case for formative measurement. *Psychological Assessment*, 28(3), 319–330. <https://doi.org/10.1037/pas0000152>
- Young, A. R., Gurm, M. K. & O'Donnell, K. A. (2017). Assessing executive functions in young children. Dans M. J. Hoskyn, G. Iarocci & A. R. Young (dir.), *Executive functions in children's everyday lives: A handbook for professionals in applied psychology* (p. 21–37). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199980864.003.0003>
- Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal*, 11(2), 49-54. <https://doi.org/10.21315/eimj2019.11.2.6>
- Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H. & Nikanfar, A. R. (2015). Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. *Journal of caring sciences*, 4(2), 165–178. <https://doi.org/10.15171/jcs.2015.017>
- Zelazo, P. D., Blair, C. B. & Willoughby, M. T. (2016). *Executive Function: Implications for Education* (NCER 2017-2000) National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. <http://ies.ed.gov/>.