

VALIDITÉ EN SCIENCE DE LA CONCEPTION

Kai R. Larsen

École de commerce de Leeds, Université du Colorado, Boulder, CO 80309, États-Unis
{Kai.Larsen@colorado.edu}

Roman Lukyanenko

École de commerce McIntire, Université de Virginie, Charlottesville, VA 22908 États-Unis
{romanl@virginia.edu}

Roland M. Mueller

Département d'économie et de commerce, École d'économie et de droit de Berlin, Berlin, Allemagne
{roland.mueller@hwr-berlin.de}

Veda C. Storey

Collège de commerce J. Mack Robinson, Université d'État de Géorgie, Atlanta, GA 30303 États-Unis
{vstorey@gsu.edu}

Jeffrey Parsons

Faculté d'administration des affaires, Université Memorial de Terre-Neuve, St. John's, Terre-Neuve-et-Labrador, Canada {jeffreyp@mun.ca}

Debra VanderMeer

Faculté de commerce, Université internationale de Floride, Miami, FL 33199 États-Unis
{vanderd@fiu.edu}

Dirk S. Hovorka

École de commerce de l'Université de Sydney, Université de Sydney, Sydney, Australie
{dirk.hovorka@sydney.edu.au}

Traduit par : **Wanda Brunel et Lila Belahlou**

Département de Français, Université de Virginie, Charlottesville, VA 22908 États-Unis
{wanda.brunel@univ-lyon2.fr}, {lila.belahlou@ens.fr}

Résumé — Il est nécessaire, pour les chercheurs, de s'assurer que le fruit de leur recherche soit vrai. Cependant, la notion de validité n'est ni bien comprise ni systématiquement faite dans la science de la conception, ce qui résulte en le développement et l'évaluation d'artefacts (modèles, méthodes, exemplifications et théories) pour résoudre des problèmes. Par conséquent, il est difficile de démontrer et de communiquer la validité de ce qui est affirmé au sujet des artefacts. Cet article définit la notion de validité dans la science de la conception et développe un système de validité de la science de la conception ainsi qu'une modélisation de processus pour l'appliquer. Le système consiste en trois types d'affirmations et de validité de degré élevé (critère, causalité et contexte) de même que des sous-types de validité. Le système guide les chercheurs à considérer l'intégration de notion de validité dans leurs projets qui font usage de la science de la conception et contribue au corpus croissant de recherches sur la méthodologie de la science de la conception. Il fournit également un moyen systématique de structurer et de valider les affirmations faites par les projets de science de la conception. Nous appliquons le système sur des exemples de recherches préexistantes, puis nous l'utilisons pour démontrer la validité des affirmations relative au système lui-même.

Abstract — Researchers must ensure that the claims about the knowledge produced by their work are valid. However, validity is neither well-understood nor consistently established in design science, which involves the development and evaluation of artifacts (models, methods, instantiations, and theories) to solve problems. As a result, it is challenging to demonstrate and

communicate the validity of knowledge claims about artifacts. This paper defines validity in design science and derives the Design Science Validity Framework and a process model for applying it. The framework comprises three high-level claim and validity types—criterion, causal, and context—as well as validity subtypes. The framework guides researchers in integrating validity considerations into projects employing design science and contributes to the growing body of research on design science methodology. It also provides a systematic way to articulate and validate the knowledge claims of design science projects. We apply the framework to examples from existing research and then use it to demonstrate the validity of knowledge claims about the framework itself.

Mots-clés : science de la conception, science du design, recherche en science du design, Cadre de Validité de la Science du Design, argument, validité de la recherche, validité de critère, validité causale, validité de contexte, validité caractéristique, validité de l'efficacité, validité externe, validité écologique, évaluation, validation

Keywords: design science, design science research (DSR), Design Science Validity Framework, knowledge claim, research validity, criterion validity, causal validity, context validity, characteristic validity, efficacy validity, external validity, ecological validity, evaluation, validation

Droits de partage : Ce document est partagé sous <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en> par un accord signé et exécuté avec *MIS Quarterly*. Il est partagé selon les règles des National Institutes of Health (NIH) des États-Unis pour la recherche soutenue pour la subvention 3U24AG052175-08S1.

Citer : Larsen, K. R., Lukyanenko, R., Mueller, R., Storey, V. C., Parsons, J., VanderMeer, D., and Hovorka, D. 2025. « Validité en Science de la conception » *MIS Quarterly*, Traduction par : Brunel, W. et Belahlou, L., pp. 1–40. original anglais : bit.ly/4inr9J9

Introduction

Les connaissances scientifiques doivent être crédibles et fiables (Burton-Jones et al. 2021; Creswell et Miller 2000). Une tradition de validité établie facilite une science crédible et fiable car elle fournit un moyen systématique d'évaluer les contributions faites à la recherche (Cook et Campbell 1979; Guba et Lincoln 1994). Chaque discipline scientifique dispose de modèles et de procédures pour évaluer la validité des arguments concernant les phénomènes d'intérêt. L'articulation de ces modèles apporte une clarté d'évaluation, augmente l'efficacité de la recherche, favorise le partage des meilleures pratiques, facilite la science cumulative et contribue à une plus grande confiance du public dans la science.

En *science de la conception* (science du design, design science) – un genre de recherche en systèmes d'information, informatique, ingénierie, médecine, gestion et sciences des

matériaux et biomédicales – les chercheurs avancent des arguments au sujet des caractéristiques ou des performances des artefacts qu'ils inventent. Les chercheurs ont proposé des processus pour mener des recherches en science de la conception, chacun d'entre eux soulignant l'évaluation comme une étape essentielle (Gregor et Hevner 2013; Hevner et al. 2004; Peffers et al. 2007; Prat et al. 2015; Sein et al. 2011). Cependant, les méthodes et cadres existants ne prennent pas en compte la validité des arguments comme un élément essentiel de l'évaluation des artefacts et ne s'accordent pas sur ce que signifie évaluer la validité des arguments sur les artefacts.

La science de la conception présente des modèles récurrents d'arguments. Parmi les exemples, on peut citer: l'artefact surpasse l'état de l'art (Padmanabhan et al. 2022); une version récente de l'artefact fonctionne mieux que la ou les versions précédentes (Sein et al. 2011); le modèle correspond à un système de référence du monde

réel (Gonzalez-Huerta et al. 2017); et l'artefact fonctionne en raison de la présence de caractéristiques de conception particulières (Abbasi et al. 2012). Il existe également un répertoire d'approches d'évaluation courantes en science de la conception (Prat et al. 2015; Venable et al. 2016), bien que celles-ci n'abordent pas la logique qui relie les affirmations à la validation, le processus d'établissement de la validité d'un type particulier d'argument.

Bien que les types de validité soient utilisés dans certains projets de science de la conception, ils sont généralement adaptés d'autres traditions. Par exemple, la validité interne et la validité externe sont adaptées de la psychométrie (Baskerville et al. 2015; vom Brocke, Hevner et al. 2020), tandis que les mesures de validité, telles que la précision et le rappel, sont tirées de la recherche d'informations, des statistiques et d'autres disciplines (Larsen et Becker 2020). Une exception est la validité d'exemplification (Lukyanenko et Parsons 2020), qui a été proposée comme type de validité propre à la science de la conception.

La plupart des validations de la science de la conception ne sont pas étiquetées, ce qui rend difficile de les référencer efficacement et de partager les meilleures pratiques avec la communauté. De plus, certaines procédures d'évaluation, telles que les vérifications d'applicabilité (Rosemann et Vessey 2008), répondent à de multiples affirmations (souvent non formulées), ce qui conduit à se demander *quelles affirmations sont valides*. Malgré les avantages potentiels, aucune tentative systématique n'a été faite pour définir ou étudier la validité de tous les types d'artefacts abordés par les scientifiques du design (par exemple, les modèles, les méthodes, les théories du design ; voir Gregor et Hevner 2013) ou pour fournir un processus de validation des arguments. Par conséquent, la question essentielle de ce qui constitue la validité en science de la conception reste sans réponse.

Une approche systématique de la validité sera bénéfique pour la science de la conception. Premièrement, elle favorisera la qualité des études. Le lien entre les arguments et les preuves

qui les étayent est essentiel à la qualité de la science de la conception. Cela peut être facilité en suivant des procédures d'évaluation conformes à des structures convenues et connues pour étayer les types d'affirmations formulées. Ces procédures exigent généralement une rigueur irréprochable et une argumentation minutieuse (par exemple, des preuves de causalité pour la validité interne, souvent fournies par des expériences faites aléatoirement). L'établissement de la validité rendrait le lien entre les arguments et leur validation plus explicite et plus susceptible d'être examiné et contrôlé. Cette transparence est importante pour susciter la confiance dans les arguments et les résultats rapportés (Burton-Jones et al. 2021).

Deuxièmement, comme des types spécifiques de validité s'appliquent à des types spécifiques d'affirmations (par exemple, la validité interne s'applique aux affirmations de causalité), l'établissement d'une tradition de validité pour la science de la conception conduira à rendre les affirmations de connaissances plus explicites. Cela peut aider les chercheurs à mieux apprécier et communiquer les contributions qu'ils apportent et faciliter le processus de validation. En outre, les affirmations de connaissances explicitées peuvent également être utilisées par d'autres chercheurs voulant étendre les contributions originales à la fois au sein et au-delà de la science de la conception, permettant ainsi aux scientifiques du design de mieux contribuer à une tradition cumulative de recherche sur la conception, l'utilisation et l'impact des artefacts des technologies de l'information.

Troisièmement, une conceptualisation complète de la validité augmentera l'impact de la science de la conception sur le monde réel. En sensibilisant les chercheurs à la nature des arguments et aux types de validité, les artefacts qui ne parviennent pas à remplir leur fonction lorsqu'ils sont exposés à des circonstances inattendues du monde réel sont moins susceptibles de se produire. Par exemple, dans le domaine de l'apprentissage automatique, il est important que les chercheurs sachent quelles caractéristiques des artefacts ont produit les meilleurs résultats. Cependant, ces derniers se concentrent étroitement sur les performances

d'un artefact proposé. Une conception plus inclusive de la validité prendrait en compte l'applicabilité d'un artefact à diverses conditions du monde réel, y compris pour des tâches plus larges et différents groupes d'utilisateurs (Ethayarajh et Jurafsky 2020).

Quatrièmement, l'établissement d'une tradition de validité facilitera la publication d'études en science de la conception. En tant que reflet des normes communautaires et des concepts partagés, une tradition de validité stimule la productivité de la recherche en augmentant la cohérence de la validation (Chan 2014) et en rationalisant les pratiques de validation. Il convient de noter que les domaines prolifiques de l'économétrie et de la psychométrie ont des approches de validité ou de validation distinctes (Taylor 2013), tout comme la recherche qualitative (Creswell et Miller 2000) et les domaines de l'informatique et de l'ingénierie logicielle, tels que la simulation de modèles (par exemple, Pääkkönen et al. 2017) et la décidabilité des problèmes (par exemple, Fan et al. 2018). Dans toutes les disciplines, un niveau plus élevé d'accord sur les normes disciplinaires fondamentales (y compris la validation) est corrélé à une productivité de recherche et à une qualité de publication accrues (Gumpert 2007). La formalisation de conceptualisations disciplinaires partagées est essentielle pour accélérer le progrès scientifique (Académies nationales des sciences 2022).

Enfin, établir la nature de la validité en science de la conception permettra de la positionner par rapport aux traditions établies (par exemple, dans les sciences du comportement) et de mieux faire connaître les contributions distinctes de la science de la conception aux non-initiés. Les modèles courants de validation reçoivent un nom systématique, ce qui aide les spécialistes de la science de la conception et les non-initiés à faire référence à ces approches d'évaluation. Cela contribue à son tour à établir l'identité de la science de la conception.

Cette recherche apporte plusieurs contributions. Nous examinons d'abord comment les concepts de validité ont été utilisés dans la science de la conception. Nous considérons ensuite la nature générale de la validité pour identifier les

fondements établis qui pourraient bénéficier aux scientifiques du design. Pour cela, nous développons les connaissances sur la logique et les relations entre les arguments, les types d'artefacts, les types de validité et le contexte d'évaluation. Ces fondements nous permettent ensuite de définir les types de validité et de les utiliser de manière appropriée pour évaluer affirmations de connaissances en science de la conception.

Pour continuer, nous développons **Le Cadre De Validité De La Science de la conception** (Design Science Validity Framework) pour fournir un moyen systématique d'articuler et de valider les arguments des projets de science de la conception. Le cadre comprend trois types d'arguments et de validité de connaissances de haut niveau (*critère*, *causalité* et *contexte*) ainsi que des sous-types de validité. Cette structure correspond à l'objectif de la science de la conception de développer des artefacts innovants comme solutions aux défis sociétaux, tout en contribuant à des connaissances scientifiques que les praticiens peuvent réutiliser dans divers contextes. Le cadre couvre de façon exhaustive les artefacts de la science de la conception, et comprend les systèmes implémentés (par exemple, les outils et les exemplifications déployées), les contributions abstraites qui encapsulent les plans pour le développement de systèmes (par exemple, les modèles conceptuels ou les méthodes d'apprentissage automatique) et les connaissances théoriques en matière de conception (par exemple, les théories de conception qui prescrivent la conception et l'action pour atteindre des objectifs spécifiques, voir Gregor et Jones 2007).

Enfin, nous évaluons le cadre de validité de la science de la conception à travers une série d'études nourrit de manière répétée par le cadre lui-même. Grâce à deux vérifications d'applicabilité, une analyse approfondie de la recherche sur la validité et une analyse complète de l'évaluation de la science de la conception, nous démontrons l'utilité de notre cadre pour les chercheurs, sa capacité à capturer les pratiques de validation existantes et sa prudence. Nous concluons par une discussion générale sur la nature et l'importance de la validité de la science

de la conception, puis recommandons la manière dont le cadre peut être utilisé et identifions les opportunités de recherche futures.

DÉVELOPPEMENTS MÉTHODOLOGIQUES EN SCIENCE DE LA CONCEPTION

La science de la conception a développé des fondements méthodologiques pour faciliter l'acquisition de connaissances scientifiques

pratiques, crédibles et fiables. Ces efforts ont porté sur la garantie de la rigueur dans le développement des artefacts et sur l'évaluation de l'utilité et de la qualité des artefacts et des connaissances en conception. (Voir le tableau 1 pour les sujets spécifiques abordés). Alors que la rigueur dans le développement des artefacts a reçu beaucoup d'attention, la rigueur dans l'évaluation de l'utilité et de la qualité des artefacts et des connaissances en conception a pris du retard, même si cela constitue un défi majeur.

Tableau 1. Fondements méthodologiques de la science de la conception	
Contribution	Référence
Intégration du développement de systèmes avec une théorie rigoureuse et une évaluation empirique	Nunamaker et al. (1991); March et Smith (1995)
Approches d'évaluation des artefacts et des théories	Hevner et coll. (2004); Venable et coll. (2016); Prat et coll. (2015); Gregor et Jones (2007); Tuunanen et coll. (2024)
Méthodes pour produire et communiquer la science de la conception	Peffer et coll. (2007); Gregor et Hevner (2013); Johannesson et Perjons (2014); Baskerville et al. (2015); Avdiji et Winter (2019); livari (2015); Gregor et coll. (2020); Tuunanen et coll. (2024)
Alignement des théories du noyau avec les artefacts	Arazy et al. (2010); Gregor et Jones (2007); Kuechler et Vaishnavi (2012); de Brocke, Winter et al. (2020)
Transparence	vom Brocke et al. (2021); Burton-Jones et coll. (2021); Lukyanenko et Parsons (2020); Hevner et coll. (2024)
Alignement avec la pratique	Sein et al. (2011); Lukyanenko et Parsons (2020); Hevner et coll. (2024)
Théorisation du design	Mandviwalla (2015); Gregor et Jones (2007); Gregory et Muntermann (2014); Lukyanenko et Parsons (2020); Gregor et al. (2020)

Une étape importante vers l'établissement de la rigueur de la science de la conception consiste à comprendre quelles évaluations existent déjà, comment elles fournissent des connaissances complémentaires et comment elles se chevauchent. Aucun travail antérieur n'a tenté de comprendre la rigueur de l'évaluation de la science de la conception en établissant les types de validités possibles. Pour ce faire, il faut comprendre les types d'arguments, d'artefacts et d'approches de validation.

Nous avons examiné des articles de revues éminentes sur les cadres de la science de la conception, comme résumé dans le tableau 2. La plupart des cadres du tableau 2 traitent principalement de la rigueur du développement

d'artefacts. Par exemple, Sein et al. (2011) se sont concentrés sur la recherche en conception-action (ADR), où les évaluations se déroulent dans un contexte réel et authentique. Venable et al. (2016) ont fait la distinction entre les évaluations formatives et sommatives, en se concentrant sur le moment où les évaluations sont menées. Gregor et Hevner (2013) ont examiné des façons de positionner et de présenter la science de la conception et si les évaluations se déroulent à l'intérieur ou à l'extérieur d'un contexte de développement. Prat et al. (2015) ont organisé les approches d'évaluation en une taxonomie de cinq dimensions de types d'évaluation distincts. Baskerville et al. (2015) ont développé un cadre pour mettre en évidence les différents types de production de connaissances qui peuvent se

Tableau 2. Cadres de la science de la conception et rôle de la validité

Papier	Présentation du cadre	Validité et affirmations	Évaluation
Hevner et al. (2004)	La science de la conception consiste à construire et à évaluer des artefacts et à les connecter à l'environnement et à la base de connaissances de conception.	L'accent est mis sur la validité. Il cite d'autres travaux sur la manière dont l'accumulation de preuves finira par établir la validité des affirmations plus larges de la science de la conception.	Propose deux questions fondamentales qui nécessitent des preuves issues des évaluations : 1) quelle utilité l'artefact fournit-il ? et 2) comment cette utilité est-elle démontrée ? Propose que la contribution découle de l'utilité démontrée par l'évaluation.
Peppers et al. (2007)	Un processus itératif en six étapes pour présenter et évaluer la science de la conception.	Peu d'importance accordée à la validité. Suggère que la science de la conception a dû employer des « arguments ad hoc » (p. 50).	Les chercheurs observent et mesurent dans quelle mesure l'artefact soutient une solution. Cela implique de comparer les objectifs d'une solution aux résultats réels observés à partir de l'utilisation de l'artefact (par exemple, via le temps de réponse, les éléments produits, les enquêtes de satisfaction des utilisateurs et la simulation).
Sein, et al. (2011)	Un modèle en quatre étapes de recherche sur la conception d'actions.	Ne contient aucune mention de validité ou de affirmation de connaissance.	Plaide en faveur du principe d'« évaluation authentique et simultanée » pour souligner une caractéristique clé de la recherche sur la conception de l'action.
Gregor et Hevner (2013)	Un schéma pour la publication en science de la conception, y compris l'évaluation.	Évaluation des artefacts en termes de critères pouvant inclure la validité, l'utilité, la qualité et l'efficacité.	Permet de distinguer l'évaluation à l'intérieur et à l'extérieur de l'environnement de développement. Critères d'utilité liés au transfert des performances en dehors de l'environnement de développement.
Prat et al. (2015)	Une taxonomie des méthodes d'évaluation avec cinq dimensions : objectif, environnement, structure, activité et évaluation.	La validité fait partie de la dimension objectif de l'efficacité et de l'efficacité. La validité est atteinte si l'artefact fonctionne correctement (atteint l'objectif).	Se concentre sur le « quoi » (systèmes d'artefacts) et le « comment » (méthodes utilisées) dans l'évaluation. L'évaluation implique la « relativité » de la supériorité de l'artefact par rapport aux autres solutions. Identifie les techniques d'évaluation typiques et les évaluations secondaires.
Baskerville et al. (2015)	Genres d'enquête. Cadre : conception vs science et nomothétique vs idiographique.	Fournit 18 critères de qualité. Mentionne la validité interne et externe.	Fournit des réflexions sur les genres d'enquête, plutôt que sur l'évaluation elle-même.
Venable et al. (2016)	Les évaluations de la science de la conception existent selon deux dimensions : artificielle et naturaliste et formative et sommative.	La validité est considérée comme découlant de la force de l'évaluation et de la façon dont l'artefact accomplit son objectif.	L'objectif principal de l'évaluation est d'évaluer dans quelle mesure un artefact atteint l'utilité prévue.

produire tout au long d'un projet de science de la conception. Tuunanen et al. (2024) ont proposé une approche pour la gestion de projets complexes à plusieurs étapes en suggérant que ces étapes (« échelons ») produisent différents artefacts qui peuvent être évalués à chaque étape. Il convient de noter que les préoccupations en matière de validité ont augmenté au fil du temps, passant de peu ou pas de considération dans les premiers cadres (par exemple, Hevner et al. 2004;

Peppers et al. 2007) à une prise en compte active dans les études méthodologiques récentes (par exemple, Baskerville et al. 2015; Tuunanen et al. 2024).

Les cadres populaires du tableau 2 constituent les méthodologiques établies en science de la conception. Ils permettent de comprendre l'approche générale de la conduite de la recherche et de l'évaluation des artefacts, mais n'offrent pas

de conseils sur les activités de validation, notamment sur la manière dont les preuves peuvent être fournies pour justifier les arguments.

Certains travaux en science de la conception ont reconnu l'importance d'établir la validité (vom Brocke, Winter et al. 2020) et ont suggéré que les types de validité comportementale existants (par exemple, interne, écologique) pourraient s'appliquer à la science de la conception (Baskerville et al. 2015). Dans la science de la conception utilisant des techniques basées sur l'intelligence artificielle, telles que l'apprentissage automatique et le traitement du langage naturel, il est d'usage de signaler la précision, le rappel et le score F1, ainsi que d'autres mesures de la matrice de confusion (par exemple, Abbasi et Chen 2008; Li et al. 2020). Il ne s'agit pas de types de validité, mais ils sont liés à la validité car ils fournissent des mesures quantifiables pour établir la validité des affirmations. Ils ne s'appliquent pas non plus à tous les artefacts de la science de la conception (par exemple, les théories du design). Pour établir la validité en science de la conception, nous considérons d'abord la validité en science au sens large, que nous synthétisons ensuite avec les préoccupations spécifiques de la science de la conception.¹

VALIDITÉ DE LA RECHERCHE

L'idée fondamentale selon laquelle les arguments doivent être validés remonte à l'Antiquité (Carter 2019). Le terme *validité* trouve son origine dans les sciences sociales quantitatives (par exemple, Cronbach et Meehl 1955) et est devenu plus largement utilisé par la suite, car il a été appliqué à l'évolution des croyances interdisciplinaires sur la création d'évaluations appropriées (par exemple, dans lesquelles les connaissances préalables sont évaluées et les critères d'acceptation sont convenus) pour établir la validité des arguments.

Pour établir la validité en science de la conception, nous avons examiné environ 7 500 sources de la littérature sur la validité de la recherche dans les systèmes d'information (y compris la science de la conception) et dans d'autres domaines de recherche, tels que l'informatique, les sciences sociales, l'ingénierie, la médecine, le droit et les sciences humaines. Cet effort a permis de recueillir 2 418 validités à examiner, ce qui constitue de loin l'effort le plus important pour étudier les validités. Nous rendrons compte de ces efforts plus tard, lorsque nous validerons notre cadre. Nous passons ici brièvement en revue la manière dont la validité est utilisée en science, car cela éclaire nos notions de validité en science de la conception.

Au départ, la validité était définie de manière étroite comme « la proximité entre ce que nous croyons mesurer et ce que nous avons l'intention de mesurer » (Roberts et Priest 2006, p. 41). Cette vision a centré la validité sur les artefacts de mesure, notamment en psychométrie, concernant les tests, les instruments ou les questionnaires administrés aux humains. Au fil du temps, les chercheurs quantitatifs ont développé de nombreux types de validité différentes (par exemple, interne, écologique, discriminante, externe) désormais largement utilisés dans la recherche comportementale sur les systèmes d'information (Boudreau et al. 2001). L'utilisation de ces validités a été reconnue comme une « responsabilité professionnelle » (Shultz et al. 1998, p. 266).

Les chercheurs qualitatifs et interprétativistes ont fait valoir des préoccupations uniques en matière de validité dans leurs travaux, soulignant l'importance du contexte, du cadre et des participants, en plus du rôle des chercheurs dans la création d'un compte rendu naturel, digne de

¹L'annexe A évalue l'état actuel des pratiques de validation.

confiance, confirmable et fiable du processus de recherche (Lincoln et Guba 1985 ; Onwuegbuzie et Leech 2007). Dans cette tradition, la validité est considérée comme le produit d'un consensus social, où ce qui est valide est basé sur une « communauté d'acceptabilité » (Moules et al. 2015, p. 172). Réfléchissant à ces efforts, Creswell et Miller (2000, p. 124) ont noté « un consensus général » selon lequel la validité est fondamentale pour l'acceptation des études – un point de vue largement partagé dans toutes les disciplines (Hoyningen-Huene 2013), y compris les systèmes d'information (Burton-Jones et al. 2021).

En informatique, discipline étroitement liée à des systèmes d'information, les chercheurs de sous-domaines tels que l'apprentissage automatique et l'intelligence artificielle ont adopté un cadre de tâches commun, en se concentrant sur des repères partagés pour valider les nouveaux artefacts par rapport à l'état de l'art tel que reflété dans les classements partagés (Matadamas-Hernández et al. 2012). Ce cadre a catalysé des avancées majeures dans la vision par ordinateur et le traitement du langage naturel. Par exemple, les chercheurs en apprentissage automatique ont soigneusement développé des approches, telles que l'analyse de causalité et les études d'ablation, pour évaluer les artefacts conçus (par exemple, Chowdary et Kanhangad 2022). Ils utilisent ces approches dans le cadre de tâches communes pour tester les arguments et contribuer aux connaissances partagées sur les tâches. Néanmoins, malgré des progrès impressionnants, le cadre de tâches communes peut avoir limité les applications pratiques des modèles résultants et désavantagé les critères d'évaluation tels que « la compacité, l'équité et l'efficacité énergétique » (Ethayarajh et Jurafsky 2020, p. 1).

Chaque type d'enquête, et parfois chaque domaine (par exemple, l'apprentissage automatique), a développé des traditions de validité distinctes. Tout comme les préoccupations en matière de validité qualitative et quantitative diffèrent, on peut s'attendre à ce que la validité de la science de la conception ait son propre objectif. Au-delà du simple emprunt de types de validité à la recherche psychométrique, informatique ou qualitative, il

est possible d'aborder la nature unique de la science de la conception, dans laquelle les chercheurs développent des artefacts pour produire les résultats souhaités et valident les arguments sur ces artefacts. Néanmoins, les recherches sur la validité existantes dans d'autres disciplines ont des implications pour la validité en science de la conception car elles traitent de concepts et d'idées de validité générale. Notre examen de la littérature sur la validité a conduit aux conclusions suivantes qui éclairent la validité de la science de la conception.

Premièrement, les chercheurs conviennent que des recherches scientifiques évaluées et validées de manière systématique sont nécessaires (Cohen et al. 2013). Même les chercheurs qui concluent que le terme « validité » n'est pas adéquat pour leurs recherches reconnaissent les avantages de la diffusion et du partage de modèles de validation efficaces et éprouvés (par exemple, Creswell et Miller 2000 ; Maxwell 1992).

Deuxièmement, les sciences cherchent à faire avancer les connaissances qui constituent les contributions de la recherche (Collier-Reed et Ingerman 2013). Une étude peut comporter de nombreuses connaissances, certaines étant le centre de l'étude (connaissances principales) et d'autres tirées de recherches antérieures (connaissances secondaires). Souvent, seule la première est soumise à validation dans une étude. Dans le contexte de la validité, un **argument** est donc une affirmation sur les phénomènes d'intérêt qui capture la contribution originale de l'étude.

Troisièmement, lors de la validation d'un argument, l'objet de l'affirmation est généralement comparé à une **entité de référence** (reference entity) – un objet abstrait ou concret dont les propriétés ou le comportement peuvent être comparés à l'idée ou à l'objet proposé pour évaluer la qualité de ce dernier. Par exemple, en informatique, la validité fonctionnelle est atteinte lorsque « le modèle imite le comportement d'entrée-sortie du système réel avec un niveau de précision acceptable » (Murray-Smith, 2015, p. 30), le système réel étant l'entité de référence. Les entités de référence apparaissent sous diverses formes, notamment des artefacts, des objets naturels existants (par exemple, un

humain) et des idées mentales. Quelle que soit l'argument de recherche, il existe toujours une entité de référence par rapport à laquelle valider ce dernier. Même pour les innovations ou inventions radicales, il existe des entités de référence par rapport auxquelles celles-ci peuvent être comparées.²

Quatrièmement, pour valider une affirmation de connaissance, les chercheurs s'engagent dans une ou plusieurs évaluations appropriées à cette affirmation. Une **procédure d'évaluation** (validation) est un ensemble de tâches entreprises pour fournir la preuve de la validité de l'argument. Généralement, la procédure consiste à comparer sur certaines dimensions l'entité focale (par exemple, l'artefact développé dans une étude de science de la conception) à une entité de référence, qui peut être matérielle ou abstraite. Ces procédures sont généralement établies par consensus au sein de la discipline (Taylor 2013).

Cinquièmement, la validité est une question de degré, car la pertinence et la qualité de l'entité de référence et l'approche adoptée pour la valider peuvent varier. Ainsi, les comparaisons particulièrement solides permettent de correctement archiver – réalisées par rapport aux meilleures connaissances, à *l'artefact* ou à la façon de faire les choses à **la pointe de la technologie**, ce que l'on appelle le critère. Les chercheurs quantitatifs et qualitatifs (par exemple, Moules et al. 2015 ; Taylor 2013) ont soutenu qu'il était impossible d'atteindre une validité parfaite. Chaque communauté de recherche détermine les normes de ce qui constitue un résultat de comparaison suffisant pour que l'argument correspondant puisse être accepté comme scientifiquement valide.

Enfin, pour systématiser les modèles convenus et largement utilisés d'entités de référence, de procédures d'évaluation et de normes d'acceptation de preuves pour des types spécifiques d'affirmations de connaissances, ces dernières sont étiquetées et organisées en **types de validité** (validity types). Ensuite, garantir une validité particulière (par exemple, la validité convergente) peut être utilisé comme un raccourci pour suggérer que l'affirmation correspondante (que deux mesures sont, en fait, liées) peut être acceptée comme valide sur la base des normes scientifiques existantes. Par exemple, Cook et Campbell (1979) ont examiné les menaces et présenté des approches pour assurer la validité interne dans les sciences sociales, tandis que Lincoln et Guba (1985) ont proposé des étapes pour démontrer la fiabilité et l'authenticité des affirmations dans les études qualitatives. Ces modèles de validité suggèrent des approches spécifiques pour identifier les entités de référence, les comparer et présenter les résultats. Cela facilite le référencement et le partage des pratiques de validation, leur utilisation cohérente et leur amélioration en identifiant les connexions, les chevauchements et les lacunes entre les types de validité, contribuant ainsi à une tradition cumulative.

LE CADRE DE VALIDITÉ DE LA SCIENCE DE LA CONCEPTION

En considérant les fondements de la validité, nous développons maintenant le cadre de validité de la science de la conception.³ Nous commençons par les affirmations de connaissances sur l'artefact focal. L'explication de ces affirmations peut aider les chercheurs à partager et à réutiliser des procédures de validation utiles.

² Prenons l'exemple des grandes inventions de l'histoire humaine, comme le feu, la roue, la dynamite, le clou et l'imprimerie. Toutes ces inventions étaient réalisées à l'aide d'outils manuels ou moins innovants, qui permettaient de les évaluer. Si le génie de nombreuses inventions révolutionnaires réside dans la mesure dans laquelle elles s'écartent des outils existants ou des méthodes de réalisation de tâches (l'entité de référence), lorsqu'aucun outil de référence existant ne peut être identifié ou qu'il est

impossible d'y accéder (par exemple, s'il est exclusif), d'autres outils ou processus pertinents peuvent être pris en compte.

³ Dans ce qui suit, nous désignons le cadre de validité de la science du design simplement par le terme « cadre de validité », à moins que nous ne l'appelions intentionnellement par son nom complet.

Affirmations en science de la conception

La nature des affirmations de la science de la conception est enracinée dans la science de la conception en tant qu'enquête scientifique utilitaire (Hevner et al. 2004). L'un des objectifs clés de la science de la conception est de créer des artefacts qui répondent aux défis du monde réel et de générer des connaissances en conception liées à ces artefacts. Par exemple, Hevner et al. (2004, p. 77) ont défini la science de la conception comme une recherche qui « crée et évalue les artefacts informatiques destinés à résoudre les problèmes organisationnels identifiés. » Depuis lors, la compréhension de la science de la conception a évolué au-delà d'une approche purement organisationnelle, les chercheurs cherchant de plus en plus à relever des défis sociétaux et individuels plus vastes (Weinhardt et al. 2020; Winter et al. 2014). En outre, de nombreux artefacts sont sociotechniques, associant des logiciels, du matériel et des processus à des individus, des groupes et des organisations (Thomas et al. 2022). Ainsi, la science de la conception englobe un large éventail de problèmes sociaux et d'artefacts innovants correspondants, une perspective inclusive que nous adoptons. Ce qui distingue la science de la conception de la pratique est l'objectif de développer des connaissances liées à la construction d'artefacts (Gregor et Jones 2007; Peffers et al. 2018). Nous définissons donc **la science de la conception** (design science) en tant que recherche qui développe de nouveaux artefacts et des connaissances de conception pertinentes pour répondre aux défis et aux opportunités individuels, organisationnels et sociétaux.

De nombreux types d'artefacts sont des contributions à la science de la conception. Les artefacts courants sont des modèles, des méthodes, des exemplifications, des théories de conception et des composants théoriques (par exemple, des constructions, des principes de

conception) (Gregor et Hevner 2013). Les artefacts sont des entités complexes composées de plusieurs parties ou composants interdépendants.⁴ Ils sont également généralement considérés comme des composants de systèmes plus vastes, tels que des organisations ou des systèmes techniques plus larges. Compte tenu de cette diversité, nous définissons un **artefact de la science de la conception** (design science artifact) comme une entité abstraite ou concrète créée en tant que contribution centrale d'un projet de science de la conception pour atteindre les résultats souhaités au niveau individuel, organisationnel ou sociétal.

Les principes fondamentaux selon lesquels la science de la conception apporte des solutions aux problèmes et génère des connaissances de conception prescriptives sont au cœur des affirmations de connaissances de la science de la conception. Plus précisément, pour évaluer la contribution d'un artefact donné, l'artefact doit être supérieur d'une certaine manière (par exemple, plus rapide, plus efficace, moins cher) par rapport aux solutions existantes (Padmanabhan et al. 2022) et l'approche utilisée pour créer l'artefact doit être spécifiée (Hevner et al. 2004). Pour fonctionner en tant que science et permettre à d'autres de bénéficier des connaissances de conception accumulées, les chercheurs doivent communiquer des connaissances sur la manière dont l'artefact a été construit, par exemple en partageant du code dans un référentiel ou en décrivant le fonctionnement de l'artefact au niveau conceptuel (Burton-Jones et al. 2021; Hevner et al. 2024).

De plus, les mécanismes causaux spécifiques qui relient les choix de conception (parties de l'artefact) aux résultats souhaités peuvent être fournis pour approfondir les connaissances en conception découlant du développement de l'artefact (Gregor et Jones 2007; Peffers et al. 2018). Enfin, comme il est peu probable que les praticiens mettent en œuvre leurs solutions dans des contextes identiques au contexte de recherche

⁴ Chaque type d'artefact possède également de nombreux sous-types. Par exemple, les sous-types de *modèle* incluent

les cadres, les taxonomies, les ontologies, les simulations et les modèles mathématiques.

original présenté dans un article, il est utile de spécifier quand et dans quelles conditions et limites un artefact est censé atteindre ses résultats (Hevner et al. 2024). Une tâche essentielle des scientifiques de la conception est d'aider les autres chercheurs et praticiens à réutiliser les artefacts et les connaissances en conception dans de nouveaux contextes (par exemple, Iivari et al. 2021).

Les activités de la science de la conception produisent des types spécifiques d'arguments. Un **argument en science de la conception** (design science knowledge claim) est une proposition concernant un artefact qui affirme sa contribution à la science et à la société à travers une forme ou une fonction particulière. Par exemple, Tiefenbeck et al. (2016) ont développé un compteur de douche et ont affirmé que l'indication précise de l'eau et de l'énergie consommées (la forme de l'artefact) réduisait la consommation d'énergie (la fonction de l'artefact) par rapport aux douches dépourvues de tels indicateurs.⁵ Conformément aux objectifs de la science de la conception, les arguments de la science de la conception suggèrent qu'un artefact construit selon une approche spécifiée apporte un résultat souhaité, que le résultat est causé par les propriétés spécifiques de l'artefact, ou que le résultat est censé se maintenir dans un contexte spécifique ou dans différents contextes. Nous définissons ces types d'arguments respectivement comme *critère*, *causal* et *contexte* et les élaborons ci-dessous.

Énoncé de critère (criterion claims) : dans chaque article de science de la conception que nous avons examiné, les auteurs formulent un ou plusieurs arguments sur les *résultats attendus* de l'artefact, tels que ses avantages ou son utilité pour relever un défi ou saisir une opportunité. Nous définissons un énoncé de *critère* comme un argument sur l'utilité d'un artefact. Il existe

toujours une manière déjà existante ou une alternative pour faire quelque chose, donc les énoncés de critère présentent implicitement ou explicitement une comparaison avec de telles entités ou processus existants (de critères) (Padmanabhan et al. 2022). Par exemple, les auteurs peuvent prétendre avoir développé un moteur de recherche de construction plus efficace (Li et al. 2020) ou une nouvelle théorie de conception de technologie personnalisable (Germonprez et al. 2007). De telles affirmations garantissent ou supposent, respectivement, que les moteurs de recherche conventionnels pourraient mieux fonctionner en extrayant des constructions théoriques à partir d'articles ou que les théories de conception typiques ne traitent pas des technologies personnalisables. Nous désignons ces arguments comme des énoncés de critère même si une affirmation elle-même n'a pas besoin de spécifier explicitement le critère.

Les affirmations relatives aux critères sont particulièrement solides lorsque l'artefact de comparaison est l'état de l'art communément accepté (Padmanabhan et al. 2022). Par exemple, Larsen et Bong (2016) ont comparé les performances de leur algorithme de recherche à un critère du moteur de recherche EBSCO Host, tandis que Li et al. (2020) ont proposé un argument sans doute plus fort en comparant leur algorithme de recherche à la fois à EBSCO Host et à Google Scholar, généralement reconnus comme à la pointe de la recherche universitaire. Dans notre examen des articles sur la science de la conception (rapportés plus tard), nous avons constaté que *tous les articles* contenaient des arguments relatifs aux critères qui indiquaient implicitement ou explicitement des résultats meilleurs que les artefacts ou les processus existants.

Thèses causales : à long terme, une affirmation fondée sur un critère ne suffit pas à garantir le

⁵ Dans le contexte des théories de la conception, Gregor et Jones (2006, p. 327) ont avancé une notion connexe de « propositions ou hypothèses testables » expliquant que ces propositions sur les théories de la conception sont testées « par une instanciation, en construisant un système ou en implémentant une méthode, ou éventuellement dans de rares cas par une logique déductive ». La revendication de

connaissance en matière de conception est un concept plus large : comme nous le montrons dans cet article, les revendications de connaissance s'appliquent à tout artefact et, pour être validées, elles ne nécessitent pas d'instancier leurs composants, même dans le cas de la théorie de la conception.

progrès scientifique. Une fois les affirmations fondées sur un critère établies, les chercheurs peuvent (dans le cadre d'une étude établissant une affirmation fondée sur un critère ou dans des études ultérieures) souhaiter approfondir les connaissances en matière de conception en justifiant les affirmations sur les caractéristiques de conception de l'artefact qui produisent les résultats revendiqués. Nous appelons ces affirmations *les thèses causales*, compte tenu de l'intérêt fondamental de la science pour les causes et les effets (Salmon 1998). Une thèse causale est un argument sur la mesure dans laquelle des parties spécifiques d'un artefact provoquent une utilité particulière.

Bien que fournir le code ou la description d'un artefact contribue à soutenir la valeur d'un artefact et le rend plus accessible à la réplication et à l'extension, l'avancement des thèses causales peut aider à créer de meilleurs artefacts et à développer de nouvelles théories de la technologie. Par exemple, Abbasi et Chen (2008) ont avancé plusieurs thèses causales en affirmant que les ensembles de fonctionnalités de CyberGate étaient plus aptes à représenter les types d'informations que les ensembles de fonctionnalités de base couramment utilisés dans les systèmes antérieurs (notez également la présence d'un énoncé de critère). Ils ont mené des évaluations séparées pour chacun des ensembles de fonctionnalités étendues, à savoir le sujet, l'opinion, le style, le genre et les informations d'interaction, afin d'établir l'influence causale de chacun sur les performances de l'artefact (le critère provenant des performances d'un artefact employant toutes les fonctionnalités).

Les thèses causales ne se limitent pas à l'apprentissage automatique ou même aux artefacts qui servent d'exemples. Dans notre évaluation, nous évaluons les thèses causales sur la valeur des parties de notre propre cadre de validité de la science de la conception. En fait, étant donné que la plupart des types d'artefacts de la science de la conception n'ont pas d'exemples, les méthodes qualitatives jouent un rôle majeur dans l'évaluation de ces artefacts. Ces méthodes, à leur tour, peuvent être fondées sur des connaissances justificatives (par exemple, la théorie du noyau-kernel theory) qui montrent les

mécanismes causaux reliant les caractéristiques des artefacts aux résultats cibles (Gregor et Jones 2007).

Il n'est pas toujours possible de savoir comment ou pourquoi quelque chose fonctionne, en particulier lorsqu'il s'agit de technologies complexes, d'objets innovants et d'« inventions » (Gregor et Hevner 2013). Par conséquent, tous les articles ne font pas de telles affirmations ; dans notre enquête sur les articles de science de la conception, 50 % présentent des thèses causales.

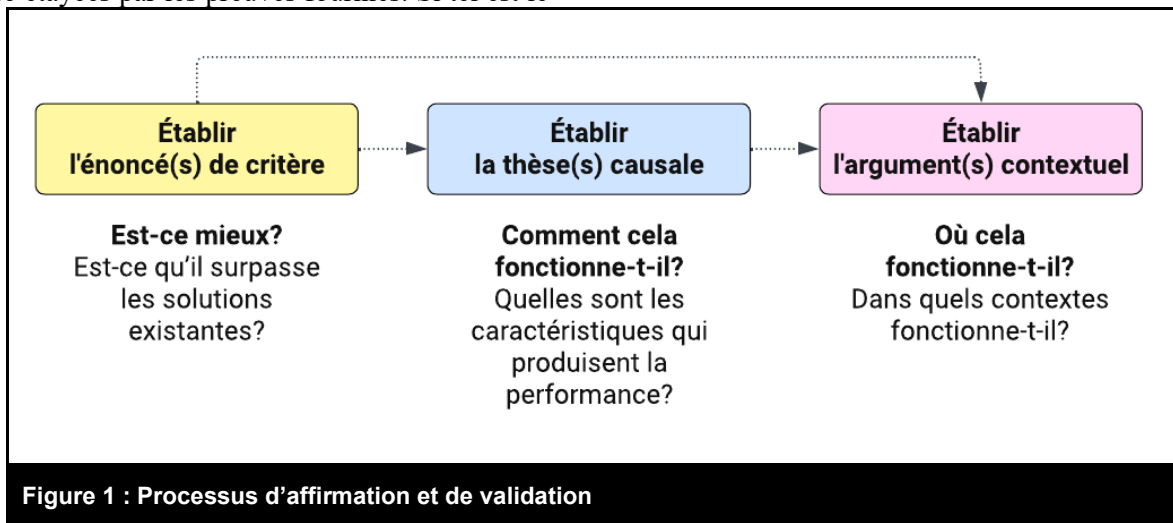
Arguments contextuels (context claims): les artefacts sont toujours créés dans un contexte donné pour répondre à un défi ou à une opportunité donnée. Cependant, le rôle du contexte n'est pas toujours pris en compte dans la science de la conception. Pour aider les praticiens à mettre en œuvre des artefacts, *les arguments contextuels* explicitent les situations ou les conditions dans lesquelles les résultats proposés de l'artefact sont censés se produire. En conséquence, les énoncés de critère et les thèses de causalité sont généralement considérés comme plus rigoureux lorsqu'ils sont évalués dans des contextes multiples ou des contextes qui ressemblent davantage au contexte d'utilisation prévu de l'artefact.

Tous les articles de science de la conception ne font pas d'argument sur le contexte. Une exception est la recherche en design-action (action design research), où le cadre dans lequel l'artefact est mis en œuvre, une organisation réelle (Sein et al. 2011), est le contexte dans lequel toute affirmation concernant l'artefact est censée s'appliquer. Un argument sur le contexte peut être élargi (et donc porter sur le critère ou des thèses causales renforcées) en suggérant que l'artefact fonctionne au-delà de son cadre d'origine, par exemple dans des cadres similaires ou même distincts. En outre, un argument sur le contexte peut aider à délimiter la généralisation des résultats. Par exemple, Lukyanenko et al. (2019) ont explicitement affirmé que leur méthode de collecte de données basée sur les instances est particulièrement efficace pour les tâches ouvertes dans les projets de science citoyenne à grande échelle, mais ont également affirmé qu'elle ne devrait pas apporter

d'avantages pour la collecte fermée dans le cadre du crowdsourcing de microtâches.

Dans notre échantillon d'articles sur la science de la conception, 14 % des articles ont fait des arguments contextuels. Ce chiffre a peut-être été réduit par des auteurs qui n'ont pas explicitement fait de telles affirmations, même s'il était probable que de telles affirmations auraient pu être étayées par les preuves fournies. Si tel est le

cas, cela représente une opportunité pour les chercheurs d'envisager de faire valoir des affirmations contextuelles (afin que d'autres lecteurs de leurs travaux puissent mieux apprécier les autres contextes dans lesquels les solutions qu'ils ont développées pourraient également être utiles).



La figure 1 résume le processus de validation en fonction de l'importance et de la prévalence des arguments. La figure implique que chaque article avance au moins un énoncé de critère. Cela peut être suffisant pour les nouveaux artefacts. Comme le suggèrent Gregor et Hevner (2013, p. 346), les nouveaux artefacts sont « des applications intéressantes où il existe peu de compréhension actuelle du contexte du problème et où aucun artefact efficace n'est disponible comme solution ». Ensuite, en fonction de la nouveauté de l'artefact, un article peut en plus avancer des thèses causales et/ou contextuelles pour améliorer la compréhension des raisons et de quand l'artefact fonctionne. Les thèses causales nécessitent l'établissement simultané ou préalable d'un énoncé de critère ; les arguments sur le contexte nécessitent soit l'établissement d'un argument sur le critère, soit d'une thèse causale.

Les trois types d'arguments peuvent être formulés à l'égard de tous les types d'artefacts.

Comme indiqué précédemment, les artefacts abstraits (par exemple, la théorie de la conception, les modèles conceptuels) peuvent être instanciés et (via l'exemplification) peuvent avoir un impact mesurable sur le monde. En outre, des techniques qualitatives, telles que l'analyse contrefactuelle et les entretiens avec les parties prenantes, peuvent être utilisées pour évaluer les affirmations de critères, de causalité et de contexte pour les entités abstraites. Les types d'arguments sont déterminés par la contribution souhaitée, et non par le type d'artefact.

Artefacts et entités de comparaison de la science de la conception

Voici notre définition de la **validité en science de la conception** (design science validity): mesure dans laquelle les arguments sur un artefact focal sont étayés par des preuves. Conformément aux trois types d'arguments, il existe trois types généraux de validité en science de la conception: **validité de critère** (criterion validity), **validité**

causale (causal validity) et **validité de contexte** (context validity). En étayant leurs affirmations respectives, ces types de validité font progresser la pratique de la science de la conception de trois manières importantes. Premièrement, *la validité de critère* est utilisée pour étayer les affirmations sur les résultats de la conception et de l'utilisation de l'artefact, visant à provoquer le changement souhaité. Deuxièmement, *la validité causale* est utilisée pour étayer les affirmations sur la contribution d'une conception spécifique (caractéristique ou pièce d'artefact) à l'utilité d'un artefact. Troisièmement, *la validité de contexte* aborde la mesure dans laquelle les arguments sont valables pour le contexte prévu ou pour des contextes supplémentaires. Au sein de ces grandes catégories, nous proposons des types de validité plus spécifiques basés principalement sur le type d'entité de référence et, secondairement, sur la nature de la comparaison entre l'artefact focal et l'entité de référence.

La validité de la science de la conception se concentre sur les affirmations concernant un artefact focal. Certains artefacts sont *matériels* et d'autres sont *abstraits* (Gregor et Hevner 2013, p. 341). Par exemple, le compteur de douche de Tiefenbeck et al. (2016) est matériel alors qu'une théorie, qui est un système de concepts et de propositions (Gregor et Jones 2007), est un artefact abstrait.

Les artefacts abstraits peuvent être utilisés pour développer des artefacts matériels tels que des compteurs de douche. Souvent, il existe *deux artefacts distincts mais liés* : la théorie (par exemple, les propositions de conception) et l'artefact matériel généré à partir de celle-ci (par exemple, le compteur de douche). Les artefacts sont des composants de systèmes plus vastes, tels que *les systèmes sociotechniques* de personnes et d'artefacts (Chatterjee et al. 2021; Winter et al.

2014). Pour de nombreux projets, il est important de prendre en compte les systèmes plus vastes dans lesquels les artefacts sont intégrés, car les arguments peuvent être faits en rapport aux impacts au niveau des systèmes. Si un chercheur cherche à concevoir le système sociotechnique lui-même comme un artefact (voir Thomas et al. 2022), des énoncés de critères, de causalité et de contexte peuvent être faits à propos de ce système sociotechnique.⁶ En le même temps, si l'accent est mis sur les propriétés et le comportement de l'artefact (par exemple, un compteur de douche intelligent) intégré dans un système sociotechnique (par exemple, une ville intelligente), le système fournit un cadre de mise en œuvre pour l'artefact, permettant de faire des arguments sur le contexte de l'artefact. Il convient de noter que l'artefact focal et l'entité de référence de comparaison existent au sein de systèmes sociotechniques (généralement différents), générant ainsi des défis en matière de validité de contexte.

Pour dériver les catégories d'entités de référence, nous avons pris en compte les principaux types d'artefacts produits par la science de la conception : constructions (théorique), modèles, méthodes, instanciations et théories de conception, ainsi que les objets naturels existants qui peuvent être utilisés comme critères (vom Brocke, Winter et al. 2020; Gregor et Hevner 2013; Hevner et al. 2004; March et Smith 1995). Le tableau 3 présente la liste des entités de référence auxquelles un artefact focal peut être comparé. Le fait de disposer des entités de référence nous permet d'expliquer les types de validité et de formaliser la procédure de validation spécifique que les chercheurs peuvent utiliser pour étayer leurs affirmations de connaissances.

⁶ Une fois qu'un système de gestion de la relation client (CRM) (l'artefact central) est placé dans le système social et physique plus large d'une organisation, nous pouvons évaluer les attributs du système plus vaste, y compris l'organisation (par exemple, la satisfaction client, l'équité, les ventes, le profit). De plus, en

comparant les organisations dotées d'un CRM à celles qui n'en disposent pas ou à la même organisation avant et après l'introduction du CRM, nous pouvons valider les affirmations concernant l'impact du CRM ou de ses composants sur les résultats d'intérêt.

Tableau 3. Catégories d'entités de référence

Catégorie	Sous-catégorie	Détails
Entités exemplifiées Entités qui produisent de résultats. Lorsqu'elles sont utilisées, elles créent des changements dans le monde, qui peuvent être mesurés et enregistrés. Les résultats de l'entité de référence peuvent être dans le passé, le présent ou le futur. D'autres aspects que les résultats peuvent également être comparés.	Instance de critère Une entité du monde réel qui devient une norme par rapport à laquelle un artefact focal peut être comparé.	La catégorie comprend des artefacts de critère (par exemple, le moteur de recherche Google Scholar), des objets du monde réel (par exemple, un maître d'échecs humain), des processus du monde réel (par exemple, l'embauche d'employés) ou des capteurs rapportant des états du monde réel (par exemple, la température ou les précipitations). La responsabilité du chercheur est de justifier que l'exemple de critère sert de norme raisonnable pour la comparaison. Les exemplifications peuvent opérationnaliser des constructions, des modèles, des méthodes et des théories de conception dans des solutions matérielles (par exemple, des applications, des plateformes, des systèmes d'entreprise). Les exemplifications de critère sont utilisées pour démontrer les performances supérieures de l'artefact focal ou la capacité de l'artefact focal à approximer la sortie, la structure ou les caractéristiques de l'entité de critère.
	Artefact manipulé Un artefact dérivé de l'artefact focal ; généralement développé dans le cadre de la même étude et manipulé pour permettre des inférences à partir de la comparaison de deux artefacts.	Construit en supprimant ou en remplaçant une partie de l'artefact focal.
Entités non exemplifiées Objets de comparaison qui sont des plans pour des artefacts matériels concrets. À moins d'être exemplifiés, ces entités ne produisent pas de résultats matériels. Elles nécessitent généralement une interprétation par un agent tel qu'un expert humain.	Théorie (et ses composants : constructions et principes de conception) Un système de concepts destiné à expliquer, prédire ou guider l'action.	<i>Théories de conception et leurs composants</i> (par exemple, les constructions, les principes de conception) et <i>les théories non liées à la conception</i> (par exemple, les théories de l'explication et de la prédiction), qui peuvent fournir des connaissances pertinentes pour la conception des artefacts focaux.
	Modèle Une représentation conceptuelle d'un aspect de la réalité créée pour améliorer la compréhension ou faciliter l'action.	Un modèle représente un aspect de la réalité ; ignore généralement les aspects qui ne sont pas pertinents pour le modélisateur ; peut introduire des biais intentionnels. Un modèle représente un domaine dans lequel des interventions de conception se produisent, ou il existe d'autres artefacts de conception et est évalué en fonction de son utilité ou de sa correspondance avec la vérité.
	Méthode Une séquence d'étapes autonome et logique utilisée pour accomplir une tâche.	Les méthodes peuvent être utilisées dans de nombreux domaines différents, comme l'analyse des besoins, le contrôle des processus et les listes de contrôle.

Types de validité

Types de validité de critère : Comme indiqué dans la Figure 1, tous les projets de science de la conception doivent faire une affirmation de critère pour montrer que l'artefact conçu offre un certain avantage. Pour ce faire, il faut utiliser un type de validité de critère. Les énoncés de critère (et les thèses causales, comme nous le verrons plus loin) peuvent être comparées de deux manières : par leur efficacité ou par leurs caractéristiques. Les comparaisons d'efficacité prennent en compte la similarité des résultats de l'artefact avec d'autres entités génératrices de résultats, tandis que les comparaisons de caractéristiques évaluent les similarités entre les caractéristiques de l'artefact et ses entités de référence (tableau 3). Par conséquent, la validité de critère comporte deux sous-types : la validité de critère d'efficacité et la validité de critère des caractéristiques.

La **validité de critère d'efficacité** (criterion efficacy validity) soutient les affirmations concernant les artefacts instanciés (tableau 3) lorsque l'artefact focal et l'entité de référence ont des résultats comparables. La validité de critère d'efficacité traite donc des énoncés du critère qui sont étayées par la comparaison de l'efficacité de l'artefact focal à celle d'une entité exemplifiée censée représenter une norme. Ces types de validité soutiennent les affirmations selon lesquelles les résultats de l'artefact (ou les effets de ces résultats sur un système sociotechnique) ont une utilité par rapport aux résultats ou aux effets d'une entité de référence.

On peut distinguer deux sous-types de validité de critère d'efficacité, selon si le temps est important ou non. Dans ce cas, la validité prédictive, avec un enregistrement approprié des prédictions avant que les résultats ne soient générés par l'entité de référence, produit la validité la plus forte pour les affirmations. Cependant, dans la science de la conception, la validité prédictive est plus souvent évaluée par rapport aux données futures qui existent au moment de la création de l'artefact mais qui ne sont pas mises à disposition de l'artefact avant la validation. Si le temps n'est pas un facteur dans les prédictions, si les données futures ne sont pas disponibles ou si une

validation plus faible est suffisante, la validité concurrente peut être évaluée. Lorsqu'une affirmation prédictive n'est pas formulée, la validité concurrente peut être utilisée pour examiner les sorties de l'artefact focal par rapport aux sorties de référence. Souvent, les cas disponibles sont divisés par validation croisée en ensembles « d'entraînement », « de validation » et « de test », où les performances de l'artefact focal par rapport aux valeurs réelles dans les ensembles « de validation » et « de test » sont des types de validité concurrentes. Dans des contextes courants, le terme « validité de critère d'efficacité » peut être supposé de faire référence à la validité concurrente.

La **validité de critère des caractéristiques** (criterion characteristic validity) traite des énoncés de critère qui comparent les caractéristiques de l'artefact focal à celles d'une entité de référence censée représenter une norme. Pour les validités caractéristiques, un agent (le plus souvent un humain) est généralement impliqué dans l'évaluation car : l'artefact de référence ne produit pas de résultats (comme c'est le cas pour les théories, les modèles et les méthodes); le résultat n'est pas directement comparable à l'entité de référence (comme cela peut être le cas lors de la comparaison des résultats de deux IA génératives); l'artefact est exemplifié mais une caractéristique de l'artefact (comme son interface) contient la contribution ; ou l'artefact est comparé aux expériences du répondant avec des artefacts ou des processus pertinents (comme avec l'évaluation de l'utilité perçue). Une validité de critère des caractéristiques est utilisée dans tous ces cas.

Les validités caractéristiques permettent d'évaluer les aspects de l'artefact que les évaluations d'efficacité ne peuvent pas évaluer. Elles sont plus flexibles et permettent de comparer les résultats non identiques ou les résultats qui ne peuvent pas être comparés automatiquement (par exemple, les perceptions de l'utilité d'une ontologie médicale ou d'un site Web). Que l'évaluation soit basée sur la comparaison directe des résultats ou sur les caractéristiques d'un artefact, des types de validité spécifiques sont basés sur l'entité de comparaison (tableau 3). Par exemple, la validité

d'instance de critère (validité d'instance, instance validity) (en abrégé) est la mesure dans laquelle les affirmations concernant un artefact par rapport à une entité de référence exemplifiée sont prises en charge. Abbasi et al. (2018) ont étudié la validité d'instance de critère pour valider une affirmation d'utilité. Les auteurs ont demandé à l'organisation d'estimer les économies réalisées en mettant en œuvre le système par rapport à la poursuite du processus commercial actuel.

De même, la validité théorique (theory validity) aborde la mesure dans laquelle les affirmations concernant un artefact focal par rapport à un artefact théorique sont étayées. Par exemple, Lukyanenko et al. (2019) ont développé une plateforme de science citoyenne avec des caractéristiques de conception qui, selon eux, correspondaient aux principes pertinents d'une théorie de la conception.

La validité du modèle de critère (validité du modèle, model validity) est la mesure dans laquelle un artefact focal est cohérent avec un modèle. Bien que ce type de validité ne soit pas présent dans l'échantillon d'articles sur la science de la conception que nous avons examinés, la prise en compte de la validité du modèle de critère était présente dans notre revue de la littérature. Par exemple, Refsgaard et al. (2006, p. 1596) ont suggéré que lors du développement du code d'un modèle de simulation (artefact focal), il est essentiel d'établir que le « code du modèle est dans un certain sens une représentation fidèle d'un modèle conceptuel » (par exemple, un modèle développé par des experts) du système du monde réel (par exemple, d'un écosystème).

Enfin, la validité de critères de méthode (validité de méthode, method validity) compare une méthode à une autre une méthode existante en tant que séquence autonome et logique d'étapes utilisées pour accomplir une tâche. Les méthodes de critères permettent d'évaluer un artefact focal ou une partie d'un artefact focal par rapport à d'autres entités (naturelles ou artefacts, et leurs parties). Piramuthu et Doss (2017) ont fourni un exemple lorsqu'ils ont évalué leur artefact - un protocole pour l'authentification simultanée de plusieurs étiquettes d'identification par radiofréquence. Ils ont utilisé une preuve

formelle pour valider que le protocole répondait à l'exigence de sécurité la plus stricte (Avoine et al. 2009).

Types de validité causale : comme la validité de critère, la validité causale comporte deux sous-types: la validité causale d'efficacité et la validité causale des caractéristiques.

La **validité causale d'efficacité** (causal efficacy validity) traite des affirmations causales étayées par une évaluation relative à l'efficacité d'une version manipulée de l'artefact qui comporte des parties intentionnellement différentes (par rapport à l'artefact focal). L'artefact manipulé est généralement développé dans le cadre de la même étude par la même équipe de recherche et est manipulé pour permettre des inférences découlant de comparaisons des deux artefacts. On parle parfois d'étude d'ablation, comme l'a introduit Newell (1975) et que l'on retrouve couramment dans l'apprentissage automatique (par exemple, Abbasi et al. 2012; Abbasi et Chen 2008; Etudo et al. 2017).

La **validité causale des caractéristiques** (causal characteristic validity) aborde les affirmations causales qui sont étayées par une comparaison avec les caractéristiques d'un artefact manipulé qui possède des composants intentionnellement différents. Une façon d'obtenir cette différence est de supprimer ou de modifier un composant (une partie). Comme pour la validité d'efficacité causale, l'entité de comparaison est généralement développée dans le contexte de la même étude par la même équipe de recherche et est délibérément manipulée pour permettre des inférences à partir de la comparaison des deux artefacts. Par exemple, Umapathy et al. (2008) ont développé deux versions de leur artefact focal : l'une permettant la sélection de modèles d'intégration avec le support d'actes de parole (artefact focal) et l'autre sans support (artefact de référence). Les chercheurs ont réalisé une expérience dans laquelle les participants se voyaient confier une tâche de conception de processus opérationnel (avec une liste d'éléments de tâche spécifiques) pour produire un modèle de notation et de modèle de processus opérationnel modifié (BPMN) d'une intégration d'entreprise. Les modèles ont

été évalués par des experts et le nombre d'erreurs perçus a été comparé.

Une autre approche repose sur le raisonnement contrefactuel (Collins et al. 2004). La comparaison de l'artefact focal à une version dont un composant a été supprimé peut valider que le composant n'a pas d'impact causal sur les performances de l'artefact et peut être supprimé par parcimonie.

Types de validité de contexte : Le type appuie les affirmations sur le contexte et les conditions (par exemple, laboratoire ou système sociotechnique réel, utilisé par les utilisateurs cibles ou leurs substituts) dans lesquels l'artefact est évalué. L'objectif ici est d'évaluer dans quelle mesure les affirmations relatives aux critères et

aux causes liées à l'artefact sont étayées dans un contexte qui ressemble d'une certaine manière au contexte sociotechnique cible ou qui se généralise à différents contextes. Par conséquent, les procédures d'évaluation de validité de contexte traitent des différences de conditions et de contextes. Les affirmations relatives au contexte peuvent être formulées soit par rapport au contexte d'évaluation de l'artefact par rapport au système sociotechnique cible, soit par rapport à un autre contexte. Il en résulte deux sous-types de validité de contexte : **écologique** (se concentrant sur les attributs du contexte d'évaluation d'origine par rapport à ceux du système sociotechnique cible) et **externe** (se concentrant sur l'évaluation dans des systèmes sociotechniques supplémentaires).

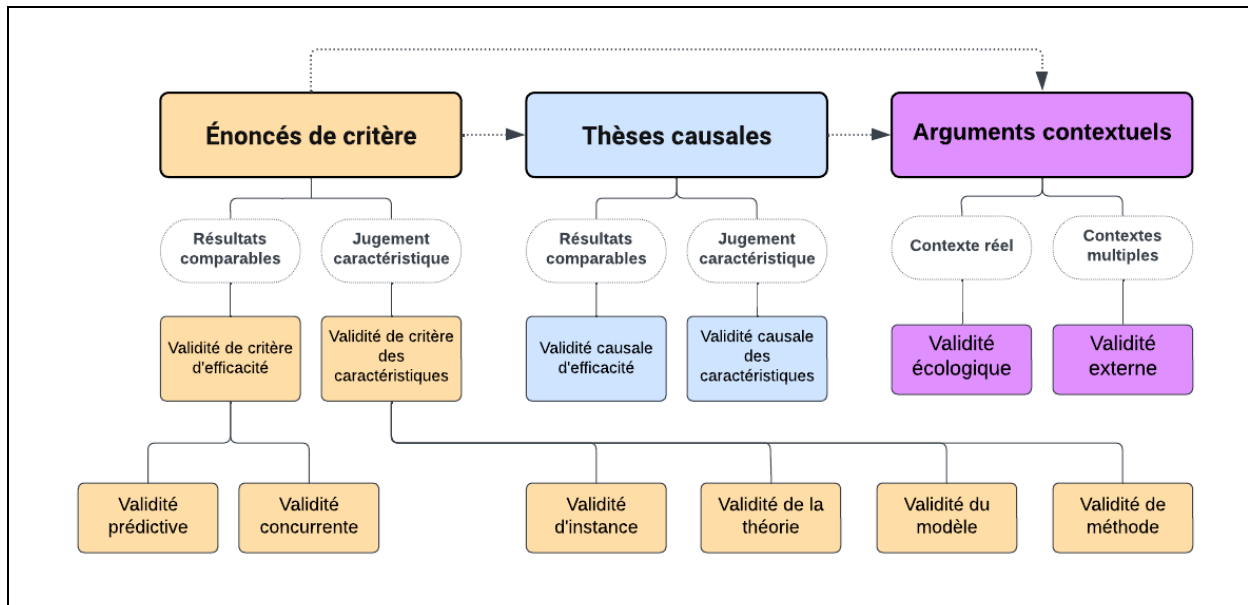


Figure 2. Cadre de validité de la science de la conception

La **validité écologique** (ecologic validity) d'une affirmation augmente avec la similarité entre le système sociotechnique d'évaluation et le système sociotechnique auquel l'artefact est destiné. Certaines approches de la science de la conception, telles que la recherche en design-action, ont tendance à avoir une validité écologique élevée car les évaluations sont effectuées dans des contextes organisationnels naturalistes dans lesquels l'artefact est utilisé. Par exemple, Zaitsev et Mankinen (2022) ont

travaillé avec des participants au Cambodge pour développer une application qui a été utilisée dans un contexte réel.

La **validité externe** (external validity) est définie comme la mesure dans laquelle les affirmations relatives aux critères ou aux connaissances causales sont étayées par une évaluation dans plus qu'un système sociotechnique. Lorsque les systèmes sociotechniques dans lesquels les artefacts focaux sont évalués diffèrent et que

l'artefact focal conserve son utilité, une validité externe plus élevée est obtenue. Par exemple, Zaitsev et Mankinen (2022) ont utilisé une approche de recherche en design-action pour développer une application visant à améliorer la littératie financière par la formation. Le développement initial a eu lieu au Cambodge, mais l'artefact a ensuite été adapté et utilisé au Népal, offrant une validité externe, les auteurs concluant que « la conception originale, déjà flexible et minimaliste, créée selon les principes de conception, a fourni une bonne base pour la localisation » (p. 106).

Le cadre de validité de la science de la conception et son application

La figure 2 illustre le cadre de validité de la science de la conception. Le cadre établit la relation entre les trois types d'affirmations concernant l'artefact focal et les types de validité spécifiques utilisés pour fournir des preuves à l'appui de ces types d'affirmations, via des types de comparaison spécifiques.

Le tableau 4 définit formellement chacun des types de validité.

Tableau 4. Définitions de la validité en science de la conception

Validité en science de la conception	La mesure dans laquelle les affirmations relatives aux connaissances sur un artefact focal sont étayées par des preuves.
Validité de critère	La mesure dans laquelle les affirmations sur un artefact focal sont étayées par une la comparaison d'évaluation à une entité de référence censée représenter une norme.
Validité de critère d'efficacité	La mesure dans laquelle les affirmations sur l'efficacité d'un artefact focal sont étayées par une comparaison de l'évaluation à l'efficacité d'une entité de référence matérielle censée représenter une norme.
Validité prédictive	La mesure dans laquelle les affirmations sur l'efficacité d'un artefact focal sont étayées par l'évaluation de son exactitude par rapport à une efficacité de référence apparue après les données utilisées pour créer l'artefact.
Validité concurrente	La mesure dans laquelle les affirmations sur l'efficacité d'un artefact focal sont étayées par l'évaluation de son efficacité par rapport à une sortie de référence apparue au cours de la même période que les données utilisées pour créer l'artefact.
Validité de critère des caractéristiques	La mesure dans laquelle les affirmations sur les caractéristiques d'un artefact focal sont étayées par l'évaluation de ses caractéristiques par rapport aux caractéristiques abstraites d'une entité de référence censée représenter une norme.
Validité de l'instance	La mesure dans laquelle les affirmations sur les caractéristiques d'un artefact focal sont étayées par l'évaluation de ses caractéristiques par rapport aux caractéristiques abstraites d'une entité de référence instanciée censée représenter une norme.
Validité de la théorie	La mesure dans laquelle les affirmations sur les caractéristiques d'un artefact focal sont étayées par l'évaluation de ses caractéristiques par rapport aux caractéristiques d'une entité de référence théorique censée représenter une norme.
Validité du modèle	La mesure dans laquelle les affirmations sur les caractéristiques d'un artefact focal sont étayées par l'évaluation de ses caractéristiques par rapport aux caractéristiques d'une entité de référence modèle censée représenter une norme.
Validité de méthode	La mesure dans laquelle les affirmations sur les caractéristiques d'un artefact focal sont étayées par l'évaluation de ses caractéristiques par rapport aux caractéristiques d'une entité de référence de méthode censée représenter une norme.
Validité causale	La mesure dans laquelle les affirmations sur l'impact d'une partie d'un artefact focal sont étayées par une évaluation par rapport à un artefact manipulé dépourvu de cette partie.
Validité causale d'efficacité	La mesure dans laquelle les affirmations sur l'impact d'une partie d'un artefact focal sur l'efficacité de cet artefact sont étayées par une évaluation de son efficacité par rapport à l'efficacité d'un artefact manipulé dépourvu de cette partie.

Validité causale des caractéristiques	La mesure dans laquelle les affirmations relatives aux caractéristiques d'une partie d'un artefact focal sont étayées par une évaluation de ses caractéristiques par rapport aux caractéristiques d'un artefact manipulé auquel manque cette partie.
Validité de contexte	La mesure dans laquelle les affirmations relatives aux critères ou aux connaissances causales sont étayées par une évaluation dans un contexte spécifique.
Validité externe	La mesure dans laquelle les affirmations relatives aux critères ou aux connaissances causales sont étayées par une évaluation dans plus d'un système sociotechnique.
Validité écologique	La mesure dans laquelle les affirmations fondées sur des critères ou des causes sont étayées par une évaluation dans un système sociotechnique correspondant aux systèmes sociotechniques du monde réel pour lesquels un artefact est destiné.

En général, différents types de validité sont appropriés, en fonction des types d'affirmations que les chercheurs souhaitent formuler. Comme le montre notre cadre, il existe trois principaux types d'affirmations, correspondant aux trois principaux types de validité. Par conséquent, notre cadre permet aux chercheurs de déterminer les types de validité appropriés en fonction de la nature des arguments concernant un artefact. Une fois que les chercheurs ont décidé d'un type d'affirmation, ils doivent suivre ce type d'affirmation du haut vers le bas jusqu'à ce qu'ils trouvent un type de validité de niveau inférieur approprié pour étayer cette affirmation.

Par exemple, Koornneef et al. (2020) ont proposé une méthode pour améliorer le processus d'identification des informations et des options pertinentes pour résoudre les problèmes de dépannage/maintenance des avions entre les vols. Les auteurs ont affirmé que leur prototype permettait une récupération plus rapide des informations pertinentes pour la résolution des problèmes (une affirmation de critère, voir la figure 2). Ils ont comparé leur prototype à la pratique actuelle dans l'industrie consistant à rechercher des informations pertinentes dans les manuels de maintenance. Comme les méthodes de recherche de prototype et de manuel ont fourni des résultats comparables (c'est-à-dire des informations pertinentes sur les problèmes), il s'agit d'un exemple de validité de critère d'efficacité avec un type spécifique de validité concurrente (car le temps de comparaison est supposé ne jouer aucun rôle). Pour étayer davantage leur affirmation, les auteurs ont réalisé une expérience dans laquelle des stagiaires en maintenance avancée ont été invités à localiser des informations pertinentes en utilisant soit le

prototype, soit la méthode manuelle. Les auteurs ont mesuré le temps de recherche dans les deux cas et ont constaté que le prototype offrait des temps de recherche plus rapides. Si une évaluation confirme l'affirmation, un seul type de validité de nœud feuille peut être nécessaire. Si les auteurs déterminent que leur artefact n'est pas suffisamment nouveau pour justifier une publication dans leur média cible, ils peuvent approfondir la contribution en faisant une affirmation causale. Alors que Koornneef et al. (2020) disposaient déjà d'une évaluation de la validité de l'efficacité, dans un scénario hypothétique, ils auraient pu ajouter une validité d'efficacité causale en évaluant quelles parties de leur artefact contribuaient le plus à la vitesse de récupération des informations pertinentes.

Cet exemple illustre comment le cadre représente les pratiques existantes en science de la conception, ainsi que comment il peut être utilisé rétrospectivement pour analyser une validation existante et mieux comprendre pourquoi certains choix ont été faits. Il peut également suggérer des alternatives (par exemple, évaluer des prototypes en demandant un jugement humain, ce qui aboutit à des types de validité caractéristiques), montrant ainsi la possibilité d'utiliser le cadre pour guider le processus de validation en science de la conception.

Nous démontrons en outre le périmètre d'application du cadre en considérant des exemples de travaux issus de la littérature sur la science de la conception. Nous délimitons cette démonstration en considérant les types d'artefacts proposés par March et Smith (1995) et Gregor et Hevner (2013) : instanciations, théorie de la conception, méthodes et modèles. Pour chacun de

ces types d'artefacts, le tableau 5 présente un exemple d'article et identifie l'artefact, les

affirmations et la validité présentés par les auteurs.

Tableau 5. Validations d'exemples			
Artefact et type d'artefact	Affirmation de connaissances	Type de validité	Évaluation
Meth et al. (2015) proposent un outil d'exploration des exigences, une <i>instanciation</i> .	L'utilisation de l'outil permet aux utilisateurs d'identifier les besoins avec une plus grande précision que sans l'outil.	Validité concurrente	Expérimentez pour évaluer si les utilisateurs identifient avec précision toutes les exigences à l'aide de l'outil.
Ramakrishnan et al. (2023) proposent des <i>principes de conception</i> pour les connaissances communes basées sur une plateforme, une <i>théorie</i> .	Les principes ont une accessibilité, une importance, une nouveauté et une perspicacité, une praticabilité et une orientation, ainsi qu'une efficacité.	Validité de la théorie	Évaluation des principes de conception par un groupe de discussion.
Sedrakyan et al. (2017) proposent une méthode de prototypage rapide incluant la rétroaction (FIRP).	La simulation FIRP améliore la compréhension des aspects comportementaux d'un modèle.	Validité du modèle*	Plan expérimental factoriel évaluant la compréhensibilité de la méthode.
Valecha et al. (2013) proposent un vocabulaire partagé pour la normalisation des messages dans le domaine des interventions d'urgence, un <i>modèle</i> .	Le modèle est complet et correct pour couvrir les besoins de communication des intervenants d'urgence.	Validité du modèle	Groupes de discussion avec des experts en interventions d'urgence.
* Remarque : L'entité de référence peut être différente du type d'artefact, car elle sert à établir la base de référence. Par exemple, les participants peuvent s'appuyer sur leur modèle du monde pour évaluer certains aspects de la méthode proposée.			

ÉVALUATION ET APPLICATION DU CADRE DE VALIDITÉ DE LA SCIENCE DE LA CONCEPTION

Pour évaluer le cadre et démontrer simultanément comment l'utiliser, nous l'avons appliqué à lui-même de façon récurrente. Un cadre (un type de modèle) doit avoir une utilité, ce qui signifie, par exemple, qu'il doit être jugé utile par ses utilisateurs potentiels (Hevner et al. 2004; March et Smith 1995). Un cadre doit également être parcimonieux tout en capturant fidèlement les concepts clés du domaine (*National Academies of Sciences* 2022).

Nous formulons plusieurs arguments sur le cadre de validité, que nous validons ensuite. Nous avançons plusieurs critères, puis détaillons le contexte et la causalité pour illustrer leur utilisation leurs différents types de validité. Tout d'abord, nous affirmons l'utilité du cadre telle que perçue par ses utilisateurs potentiels, les designers scientifiques. Il s'agit d'affirmations sur les critères et le contexte appuyées par des preuves de *validité du modèle* et de *validité écologique*. Deuxièmement, nous affirmons l'utilité du cadre dans la systématisation des *procédures de validation existantes* (nommées et non nommées), comme le prouvent les preuves de

validité du modèle. Troisièmement, nous affirmons l'utilité du cadre dans la capture et l'organisation systématiques des types de validité existants en science de la conception tels que proposés et définis dans toutes les disciplines. Il s'agit d'affirmations sur les critères et le contexte appuyées par des preuves de *validité du modèle* et de *validité externe*. Enfin, nous affirmons que les composants du cadre sont suffisants pour atteindre l'utilité, comme le prouvent les preuves de *validité causales des caractéristiques*.

Affirmation 1: L'énoncé de critère et les arguments sur les contextes avec les utilisateurs cibles

Pour l'affirmation 1, nous avons évalué les affirmations de critère et de contexte en entreprenant deux vérifications d'applicabilité (Rosemann et Vessey 2008), un type courant d'évaluation en science de la conception (par exemple, Li *et al.* 2020; Lukyanenko *et al.* 2019). Le contrôle d'applicabilité est une évaluation composite abordant plusieurs types de validité. Cependant, pour évaluer le cadre de validité, le contrôle d'applicabilité est efficace pour aborder la *validité du modèle*, dans laquelle les participants comparent notre cadre à leur modèle d'un cadre idéal et à d'autres cadres pertinents. Parce qu'il engage les utilisateurs probables du cadre et leur demande de l'utiliser dans le contexte de leur travail, le contrôle d'applicabilité fournit également la preuve de la *validité écologique*. Notre évaluation a eu lieu dans le contexte de scientifiques de la conception qui avaient une vaste expérience de l'évaluation et une exposition aux cadres d'évaluation existants.

Les cadres d'évaluation antérieurs, notamment ceux de Prat *et al.* (2015) et Venable *et al.* (2016), se concentrent respectivement sur les modèles d'évaluation et sur les stratégies d'évaluation contemporaines et *post hoc*. En codant six dimensions des techniques et critères d'évaluation, Prat *et al.* identifient des techniques spécifiques (par exemple, observation, description, expérimentation, analyse dynamique) et des approches de recherche spécifiques (par exemple, qualitatives, expérimentales, de simulation, métriques) qui

sont utilisées dans l'évaluation. Le cadre de Venable *et al.* (2016) se concentre sur le contexte artificiel/naturaliste dans lequel l'évaluation a lieu et sur la manière dont ces évaluations peuvent être formatives (ayant lieu pendant le projet) ou sommatives (ayant lieu à la fin du projet).

Notre cadre de validité fournit le raisonnement scientifique sous-jacent pour valider les affirmations de connaissances par le biais du processus d'évaluation. Nous entendons par là le degré auquel les preuves obtenues grâce à un processus d'évaluation soutiennent les affirmations de connaissances faites par les chercheurs sur l'artefact focal. De plus, notre analyse montre que les évaluations en science de la conception sont toujours des comparaisons : des critères de performance comparés à des artefacts de critères existants (ou à des moyens existants pour atteindre le même résultat), des caractéristiques d'artefacts souhaitées comparées à des artefacts existants, ou des comparaisons au sein et/ou entre des contextes d'application d'artefacts. En articulant les relations entre les validités de critère, causale et contextuelle, notre cadre de validité va au-delà des cadres d'évaluation antérieurs. Nous montrons les processus par lesquels les évaluations sont effectuées et démontrons comment la validité des types d'affirmations de connaissances est établie. En rendant ces comparaisons relationnelles explicites, notre cadre fournit des méta-catégories (par exemple, critère, causal et contexte) pour les évaluations ; organise les validités existantes dans ces méta-catégories ; et identifie les comparaisons spécifiques et la structure d'argumentation *affirmation-évaluation-preuve* des évaluations elles-mêmes.

Lors de la première vérification d'applicabilité, le cadre de validité a été présenté lors d'un atelier régional sur les systèmes d'information auquel ont participé des chercheurs en comportementalisme et des scientifiques de la conception. Nous nous sommes concentrés sur l'évaluation initiale de l'utilité et sur les ajustements à apporter avant d'engager la dernière cohorte d'experts en science de la conception. Quinze participants ont été formés à l'utilisation du cadre de validité et chargés d'organiser quatre définitions de validité et quatre

évaluations de science de la conception dans le cadre, et de répondre à une enquête ouverte. Les résultats ont suggéré que le cadre serait utile aux scientifiques de la conception ainsi qu'à d'autres chercheurs. Par exemple, le cadre a été décrit comme « une perspective stimulante et rafraîchissante ». D'autres commentaires ont souligné que le cadre était assez complexe, en particulier pour les chercheurs ayant une expérience limitée en science de la conception. En réponse, nous avons ajouté un processus de sélection des types de validité à utiliser dans un projet de recherche et avons simplifié le cadre.

Après avoir intégré les commentaires dans le cadre (dont la version finale est présentée dans la figure 2), nous avons effectué le deuxième test d'applicabilité auprès des principaux scientifiques de la conception pour évaluer l'importance du cadre (affirmation de critère), l'accessibilité (affirmation de contexte) et l'adéquation (affirmation de contexte).

Les 11 participants étaient des scientifiques de la conception avec différents niveaux d'expérience universitaire post-doctorale (en moyenne 13 ans). Pour ceux qui avaient un compte Google Scholar (neuf participants), leur nombre moyen de citations était de 4 981. Un examen de leurs articles a révélé que 36% avaient cité Prat *et al.* (2015) ou Venable *et al.* (2016). La vérification de l'applicabilité a été menée en ligne pour inclure des chercheurs de divers endroits, domaines de recherche et sous-communautés de la science de la conception. Ces participants ont utilisé une variété de méthodes, en tenant compte de divers sujets de recherche et types d'analyse. Leurs recherches ont couvert un large éventail de sujets, notamment la modélisation des données et des connaissances, la modélisation analytique et de l'apprentissage automatique, la science des données, le changement des processus d'entreprise, la simulation et la théorie de la conception. Nous avons spécifiquement invité des participants d'un large spectre pour inclure des participants de la science de la conception

dans son ensemble. Les participants étaient des contributeurs réguliers à WITS, DESRIST et d'autres conférences sur la science de la conception. Les étapes de la vérification de l'applicabilité sont décrites ci-dessous.

Préparation: Les participants ont été invités à décrire des exemples de processus de validation qu'ils ont utilisés dans le cadre de recherches en cours ou d'un projet récemment achevé. Ces processus existants, qui font partie des modèles mentaux des experts, ont servi de références aux participants pour évaluer notre cadre.

Dans l'ensemble, les participants ont montré une compréhension cohérente de la nécessité d'évaluer et de valider les artefacts, mais ont démontré une compréhension limitée des types de validité qui correspondaient à leurs évaluations et manquaient de vocabulaire pour décrire leurs validations. Dans de nombreux cas, les types de validité ont été discutés en termes psychométriques (par exemple, validité de construction) ou en termes liés à l'utilité et à l'efficacité (par exemple, preuve de concept, preuve de valeur). La plupart des participants se sont concentrés sur les processus par lesquels les évaluations ont été réalisées (par exemple, expériences sur le terrain, expériences en laboratoire), plutôt que sur le rôle des validations dans l'appui des affirmations concernant leur artefact.

Introduction au cadre de validité de la science de la conception : les participants ont ensuite assisté à une session en ligne d'une heure au cours de laquelle l'équipe d'auteurs a présenté le cadre de validité.⁷ Après l'explication du cadre et l'exemple étendu, les participants ont été invités à utiliser notre processus pour déterminer les types de validité dans leurs projets.

Commentaires: Une fois la tâche terminée, les participants ont rempli un questionnaire contenant six questions ouvertes et une série de questions sur l'utilité des artefacts et leur intention de les utiliser, en appliquant les échelles

⁷ La présentation est disponible dans les documents de transparence à l'adresse <https://osf.io/ca6vg/>.

de type Likert de Venkatesh *et al.* (2003). Les réponses, à la fois qualitatives et quantitatives, ont indiqué que les participants considéraient le cadre de validité comme important, accessible et adapté (Rosemann et Vessey 2008) pour répondre aux besoins identifiés de la communauté et pensaient qu'il serait utile pour leurs propres recherches.

Rosemann et Vessey (2008) définissent l'importance comme une recherche « qui répond aux besoins de la pratique en abordant un problème du monde réel [...] de telle manière qu'elle puisse servir de point de départ pour fournir une solution » (p. 3). Nous considérons le processus d'évaluation de cet aspect comme un exemple de la validité du critère. Dans notre contexte, les scientifiques de la conception sont des praticiens du monde réel qui doivent eux-mêmes valider les affirmations de connaissances sur les artefacts. L'accessibilité est un critère de validité qui « englobe la question de savoir si la recherche est compréhensible, lisible et se concentre sur les résultats plutôt que sur le processus de recherche » (p. 3). Enfin, la pertinence est définie comme la mesure dans laquelle la recherche peut « [répondre] aux besoins de la pratique » (p. 3), ce qui signifie, selon nous, la mesure dans laquelle les chercheurs considèrent le cadre comme approprié au contexte cible. Ces trois évaluations abordent toutes les validités du modèle ainsi que la validité de contexte, étant donné leur évaluation dans un contexte similaire au monde réel.

Les participants ont convenu que le cadre est important pour clarifier et structurer des exigences de plus en plus complexes en matière de méthodes, de critères et de stratégies de validation. Cette importance a été démontrée par des commentaires tels que « il soutient une approche systématique de la validation » et « un cadre comme celui-ci peut donner aux chercheurs une norme acceptée vers laquelle se tourner lorsqu'ils essaient de valider leurs propres artefacts de science de la conception ». D'autres commentaires ont reconnu que le cadre réduit la complexité, relie les affirmations de connaissances aux stratégies de validation et fournit des détails utiles pour élaborer des

affirmations de validité. Tous ces commentaires soutiennent notre affirmation de critère.

Lors de la présentation du cadre, et bien qu'elle ait été brève, son accessibilité a été démontrée par des commentaires tels que: « C'est une bonne chose que j'aie entendu parler de votre cadre de validité au début de mon projet. » Un participant a trouvé le cadre immédiatement accessible, déclarant : « Je peux utiliser le cadre [...] le cadre a permis de valider que mon évaluation est suffisamment bonne, mais j'aurais dû élaborer quelques détails supplémentaires dans mon évaluation pour les évaluateurs. »

Le précurseur final de l'applicabilité, la pertinence, a également été reflété dans les commentaires : « En tant que chercheur et auteur DSR, le cadre m'aide à planifier la validation de mon travail dès le début. En tant que réviseur et éditeur, je pourrai mettre en évidence une compréhension commune de la validité dans DSR. Lorsque j'évalue la valeur d'un article DSR, je peux utiliser le cadre pour identifier les points forts et les potentiels d'amélioration, ou pour inciter les auteurs à partager les aspects de leur travail DSR qu'ils ont exclus de l'article mais qui sont pertinents pour établir sa validité. »

Deux préoccupations ont été soulevées concernant les implications futures du cadre. La première était que le cadre pourrait être utilisé pour entraver la publication de la science de la conception car « les examinateurs peuvent être en désaccord sur les types de validité requis ». Les participants ont convenu qu'il ne s'agissait pas d'un défaut du cadre lui-même. Il est inapproprié d'utiliser le cadre comme argument pour des évaluations supplémentaires. Comme indiqué ci-dessus, des évaluations supplémentaires ne sont justifiées que lorsque les évaluations effectuées ne sont pas conformes aux affirmations de connaissances formulées ou que les contributions de l'article aux affirmations de connaissances existantes sont jugées insuffisantes par l'équipe d'évaluation. Dans ce dernier cas, toute évaluation supplémentaire proposée par les examinateurs doit être adaptée aux affirmations de connaissances supplémentaires.

La deuxième préoccupation concernait le risque d'augmentation des coûts et des délais nécessaires à la production et à la révision des travaux de science de la conception en raison des exigences croissantes en matière de procédures de validité. Nous avons mis à jour notre cadre initial pour mieux expliquer les affirmations de connaissances et le processus d'application du cadre pour se protéger contre une telle utilisation du cadre.

Les participants ont répondu sur une échelle de Likert à 7 points sur l'utilité du cadre ($\mu = 5,64$; $\sigma = 1,04$), une question de vérification d'applicabilité courante en science de la conception (par exemple, Li *et al.* 2020 ; Lukyanenko *et al.* 2019). Ils ont également été interrogés sur leur intention d'utiliser le cadre ($\mu = 6,15$; $\sigma = 0,92$) par rapport à leur processus existant, un autre critère d'affirmation qui, lorsqu'il est évalué dans le cadre cible, fait également office d'affirmation contextuelle. Pour l'utilité, un indicateur aberrant reflétait la deuxième préoccupation quant à savoir si le cadre ralentirait (au moins initialement) les tâches de recherche. Sans cet indicateur, l'utilité a augmenté considérablement (de $\mu = 5,64$ à $\mu = 6,14$; $\sigma = 1,03$).

Le participant qui craignait que le cadre puisse être difficile à utiliser avait également l'intention de l'utiliser, déclarant que « L'existence d'un cadre comme celui-ci peut donner aux chercheurs une norme acceptée à laquelle se référer lorsqu'ils tentent de valider leurs propres artefacts scientifiques de la conception. » Tous les participants ont indiqué qu'ils avaient l'intention d'utiliser le cadre dès qu'il serait disponible. Nous avons donc conclu que les vérifications d'applicabilité ont établi *la validité du modèle* et ajouté une preuve initiale de *validité écologique*.

Affirmation 2 : Le critère d'exhaustivité du cadre

Nous avons validé l'affirmation selon laquelle le cadre de validité permet de représenter l'ensemble des évaluations existantes dans les articles publiés en science de la conception. Étant donné que le cadre est un modèle, il devrait être

capable de représenter ces évaluations, ce qui donne *une validité du modèle*. Nous avons consulté deux sources pour identifier une population d'articles pertinents à partir desquels échantillonner. Tout d'abord, nous avons utilisé les 121 articles de science de la conception identifiés par Prat *et al.* (2015) à partir du panier de huit revues de l'Association for Information Systems (AIS) Senior Scholars (avril 2004 à décembre 2013). Ensuite, nous avons analysé 1 233 articles candidats supplémentaires de l'AIS

Panier de huit articles de 2014 à 2017. Nous avons ensuite codé manuellement ces 1 233 articles en nous basant sur le schéma d'inclusion de Prat *et al.* (2015). Quarante-vingt-six articles ont été codés par le troisième auteur comme étant des articles de science de la conception et vérifiés par le deuxième auteur. Il y avait un accord à 100% sur le fait qu'il s'agissait pour tous d'articles de science de la conception. Les 121 articles originaux de Prat *et al.* (2015), plus nos 86 articles, ont donné un total de 207 articles publiés entre avril 2004 et décembre 2017.

Pour capturer les articles qui suivaient l'approche de la science de la conception mais n'utilisaient pas explicitement l'expression « science de la conception », nous avons créé des ensembles de fonctionnalités basés sur une liste de mots-clés de la science de la conception ainsi que sur des citations d'articles de premier plan sur la science de la conception. Nous avons ensuite formé un modèle d'apprentissage automatique pour distinguer les manuscrits pertinents dans l'ensemble des manuscrits, sur la base du processus décrit par Larsen *et al.* (2019). Nous avons appliqué une combinaison d'apprentissage automatique et d'évaluation manuelle pour étendre les années de couverture de 1994 à 2019 ainsi que les sources, de façon à inclure le panier de huit de l'AIS, ainsi que *les systèmes d'aide à la décision*, *les transactions de l'ACM sur les systèmes d'information* et les actes de la Conférence internationale sur la recherche en science de la conception dans les systèmes d'information et la technologie (DESRIST) et la Conférence internationale sur les systèmes d'information (ICIS). Après avoir sous-échantillonné les quatre dernières sources, nous

Tableau 6. Validité Définitions (les cellules ombrées indiquent « sans objet »)				
Nom	Partie	Sor tir	Arte fact	STS
Validité de la science de la conception				
Validité de critère				
Validité de critère d'efficacité			19	
Validité prédictive			3	
Validité concurrente			4	
Validité de critère des caractéristiques				
Validité d'instance	0	0	5	
Validité de la théorie	0	0	0	
Validité du modèle	0	0	0	
Validité de la méthode	0	0	10	
Validité causale				
Validité de l'efficacité causale	24			
Validité des caractéristiques causales	7	0	1	
Validité de contexte				
Validité externe				7
Validité écologique				1

nous sommes retrouvés avec 527 articles sur la science de la conception.

À partir de notre corpus, nous avons sélectionné au hasard 32 articles empiriques, que les quatrième et sixième auteurs ont codés indépendamment, identifiant 79 descriptions d'évaluation. Les premier et troisième auteurs ont ensuite codé indépendamment chaque description d'évaluation en utilisant le cadre. Les codeurs étaient d'accord dans 76,3% des cas, ce qui a donné un kappa de Cohen de 0,703. Les désaccords concernaient généralement ce que les auteurs originaux avaient l'intention d'affirmer plutôt que la manière d'interpréter le cadre de validité. À partir de cet exercice, nous avons

conclu que les évaluations trouvées dans les articles examinés pouvaient être classées en utilisant les types de validité de notre cadre, démontrant ainsi *la validité du modèle*. Cependant, deux types de validité, *la validité théorique* et *la validité du modèle*, n'ont pas été utilisés dans les articles que nous avons échantillonnés (tableau 6).

Affirmation 3 : Les affirmations de critère et de contexte du pouvoir de représentation

La formalisation de conceptualisations disciplinaires partagées accélère le progrès scientifique (National Academies of Sciences 2022). D'autres disciplines développent et valident également des artefacts, et le cadre de validité doit être capable de représenter les validités utilisées pour évaluer ces artefacts.

L'affirmation 3 est que le cadre de validité permet de représenter l'ensemble des types de validité de la science de la conception, non seulement pour les systèmes d'information, mais aussi pour les sciences du comportement, l'ingénierie et la médecine, lorsque ces disciplines valident les affirmations concernant leurs artefacts. En évaluant l'affirmation du critère selon laquelle le cadre permet de représenter l'ensemble des définitions de validité de la science de la conception, nous établissons la *validité du modèle*. En montrant que notre cadre s'applique à d'autres disciplines, nous établissons la *validité externe*.

La première tâche consistait à identifier les types de validité spécifiques à la science de la conception déjà proposés dans la littérature. Aucun ensemble important de types de validité généraux n'a été trouvé. Nous avons donc construit un tel ensemble de données pour affiner davantage un sous-ensemble de types de validité. Nous avons commencé par identifier et collecter des types de validité et leurs définitions dans divers domaines, notamment les sciences sociales, l'ingénierie, l'informatique, et la médecine. Ces données ont été compilées sur une période de trois ans par le premier auteur et une équipe d'assistants de recherche. Des documents

contenant des ensembles de types de validité, tels que les normes pour les tests éducatifs et psychologiques (par exemple, American Educational Research Association *et al.* 2014) ont constitué les *sources initiales*. La plus grande collection de dénominations de validité, 168, a été fournie par Newton et Shaw (2014). La grande majorité d'entre eux étaient regroupés dans un grand tableau pour illustrer la nature intraitable des types de validité. Pourtant, beaucoup d'entre eux n'apparaissaient dans aucune base de données disponible et ne pouvaient pas être trouvés à l'aide d'un moteur de recherche.

La deuxième tâche consistait à obtenir des définitions pour les types de validité identifiés. Nous nous sommes appuyés principalement sur des livres et des articles scientifiques. Pour chaque source, la section contenant la validité a été examinée plus en détail pour identifier d'autres *types de validité candidats*. Une validité candidate était un concept que l'auteur déclarait être une validité ou un concept non déclaré comme tel mais répertorié avec des types de validité (par exemple, *le réalisme mondain* comme étant étroitement lié à *la validité écologique*). Nous n'avons pas remis en question les affirmations des auteurs. Dans chaque cas, nous avons enregistré de nouveaux types de validité candidats. Au total, 2 418 types de validité candidats ont émergé d'environ 7 500 sources examinées manuellement. Ils ont été classés par nom avant que chaque ensemble de catégories ne soit examiné pour trouver des homonymes (même nom mais signification différente), ce qui a donné 418 types de validité distincts. Cette revue de la littérature a été utilisée pour générer la première version du cadre de validité de la science de la conception.

Dans la troisième tâche, nous avons examiné tous les articles publiés dans le panier de huit revues AIS (Lowry *et al.* 2013) de 1994 à 2017. Au total, 6 083 articles ont été analysés en appliquant 216 requêtes d'expressions récurrentes représentant la validité, la fiabilité et des concepts connexes tels

que la généralisabilité, ce qui a donné lieu à 73 365 phrases. Les phrases ont été classées par nombre de correspondances, avec 9 707 phrases contenant plus de deux correspondances sur les expressions récurrentes examinées manuellement par le premier auteur. Cette analyse n'a donné lieu qu'à 23 types de validité supplémentaires, pour un total de 441 types de validité candidats, ce qui suggère que le processus initial avait été approfondi. Nous avons retiré de l'évaluation ultérieure tout type de validité candidat non couramment utilisé pour lequel nous n'avons pas réussi à localiser cinq définitions provenant de sources différentes. Cent cinquante-huit types de validité candidats ont été supprimés à cette étape.

Au cours de la quatrième tâche, les premier et deuxième auteurs ont lu indépendamment les cinq définitions de chaque type de validité et ont sélectionné une ou deux définitions qui représentaient les aspects généraux exprimés par les autres définitions. Les décisions ont été discutées (et enregistrées) et les désaccords ont été résolus pour obtenir un accord sur une ou deux définitions représentant le type de validité. Étant donné qu'il existait de nombreux cas où des types de validité très similaires, voire identiques, existaient, le calcul des mesures de fiabilité inter-évaluateurs n'était pas approprié. Onze types de validité candidats ont été éliminés à cette étape parce qu'ils ne correspondaient pas à notre définition d'une validité de recherche (par exemple, les validités fondées sur la loi) ou parce qu'une définition claire n'avait pas émergé.⁸

Pour la cinquième tâche, avec un ensemble final de types de validité spécifiés, les premier et deuxième auteurs ont suivi la définition ainsi établie de la validité de la science de la conception et ont codé indépendamment tous les types de validité comme « type de validité de la science du design » ou « autre type de validité ». Les codeurs sont parvenus à un accord de 90,4% avec un kappa de Cohen de 0,79. Les codeurs ont discuté et résolu les désaccords éventuels. Presque tous les désaccords provenaient des types

⁸ La justification de chaque suppression est disponible dans les matériaux fournis par transparence.

de validité d'efficacité, pour lesquels un codeur a utilisé une interprétation plus inclusive. Au total, 79 types de validité de la science de la conception ont été trouvés.

Parmi les types de validité de la science de la conception, nous avons supprimé les types de validité des nœuds internes, qui combinaient plusieurs autres types de validité. Après avoir supprimé ces types de validité, 70 définitions sont restées à catégoriser dans notre cadre. Parmi celles-ci, 23 (33%) provenaient de revues spécialisées en sciences du comportement, 37 (53%) de l'ingénierie et de l'informatique et 10

(14%) de la médecine; 22 (31%) étaient tirées de l'échantillon de revues IS.

Les premier et troisième auteurs ont codé indépendamment toutes les définitions de validité dans le cadre et se sont mis d'accord sur 79,2% des cas (κ de Cohen = 0,731), ce qui incluait des cas d'accord partiel codés comme non-accord (par exemple, validité de critère d'efficacité par rapport à validité prédictive). Les désaccords ont été discutés et résolus. Toutes les définitions correspondent à une catégorie de validité du cadre (colonne de gauche du tableau 7).⁹

Tableau 7. Codage de la Définition de la Validité	
Validité de la science de la conception	Validités dans la littérature
Validité de critère	
Validité de critère d'efficacité	Validité absolue, validité de groupe de critères, validité de critère, validité orientée vers un critère, validité liée à un critère, validité de décision, validité diagnostique, validité discriminante, validité empirique, validité d'ordre inférieur, validité opérationnelle, validité pragmatique, validité procédurale, validité postdictive, validité répliquative, validité rétrospective, validité d'application
Validité prédictive	Validité de critère prédictif, validité prédictive (2)*, validité prédictive liée à des critères
Validité concurrente	Validité concomitante des critères, validité concomitante liée aux critères, validité concomitante, validité transversale, validité relative
Validité de critère des caractéristiques	
Validité d'instance	Validité d'observation, validité physique
Validité de la théorie	Validité étiologique, validité théorique, validité d'exemplification
Validité du modèle	Validité du modèle conceptuel, validité fonctionnelle, validité structurelle, validité sémantique
Validité de la méthode	Validité algorithmique, cohérence

⁹ Remarque: les mesures de validité ont été incluses dans l'analyse, mais comme elles n'utilisent pas d'argument, elles ne sont pas elles-mêmes des validités. Nous avons donc exclu les mesures suivantes de la classification : exactitude, aire sous la courbe, exhaustivité, rejet correct, taux de détection, score F1, retombées, fausse alarme, proportion de fausses découvertes, faux négatifs, taux de faux négatifs,

taux de fausses omissions, faux positifs, taux de faux positifs, succès, taux de succès, coefficient de corrélation de Matthews, échec, taux d'échec, valeur prédictive négative, valeur prédictive positive, précision, rappel, sensibilité, spécificité, vrai négatif, taux de vrai négatif, vrai positif, taux de vrai positif.

Validité causale	
Validité de l'efficacité causale	
Validité des caractéristiques causales	
Validité de contexte	
Validité externe	Validité pragmatique
Validité écologique	Validité comportementale, validité écologique

Remarque : *deux versions légèrement différentes de validité prédictive référant le même type de validité.

Ainsi, l'évaluation a fourni des preuves solides de la validité de l'affirmation 3 en termes d'affirmation de critère et d'affirmation de contexte, renforçant *la validité du modèle* et fournissant *une validité externe*. Elle n'a pas soutenu la nécessité des deux validités causales incluses dans le cadre. Cependant, notre évaluation dans la section 5.2 a établi la nécessité des types de validité causale. Le manque d'attention à la validité causale dans les définitions de types de validité existantes suggère une opportunité de contribuer à la validité de la science de la conception et, par conséquent, à l'évaluation. Cela soutient la nécessité de modèles en science de la conception, comme proposé par Peffers *et al.* (2008).

Affirmation 4 : Thèse causale selon laquelle le cadre est parcimonieux

Le cadre de validité de la science de la conception a été développé de manière itérative, comme c'est souvent le cas dans la science de la conception. Comme nous voulions être aussi exhaustifs que possibles, la version initiale du cadre était plus étendue que celle présentée dans cet article. Par exemple, le cadre initial contenait un type de validité appelé « validité du cahier des charges », un type de validité caractéristique des critères utilisé pour comparer un artefact à un cahier des charges ou aux attentes et à l'expérience d'un utilisateur. Cependant, les cahiers des charges ne sont pas des artefacts de la science de la conception qui capturent les contributions à la

science et à la société, et nous n'avons pas été en mesure de classer clairement la validité du cahier des charges comme un type de validité distinct. Nous avons donc entrepris une évaluation de *la validité causale caractéristique* pour examiner la nécessité de ce type de validité (ce qui confirme que notre cadre, sans ce type de validité, est parcimonieux et ne comporte pas plus de composants que nécessaire).

La validité du cahier des charges a été introduite pour répondre aux arguments relatifs aux exigences explicites (par exemple, la fonction, la facilité d'utilisation, la forme) et aux exigences implicites (par exemple, les besoins, les objectifs ou les expériences avec une classe similaire d'artefacts). *Le cahier des charges* est un artefact commun produit dans le développement de logiciels. Pourtant, il ne figure pas sur les listes communes de types d'artefacts de la science de la conception (par exemple, Gregor et Hevner 2013; March et Smith 1995). Cependant, nous l'avons initialement considéré comme une entité de référence majeure de notre cadre, en faisant l'erreur d'être plus exhaustif et prudent. Au fur et à mesure que notre compréhension du domaine de validité s'est élargie, nous avons commencé à nous demander si *le cahier des charges* était nécessaire. Nous avons choisi de résoudre ce problème en affirmant formellement que *le cahier des charges* est un type de validité nécessaire pour que le cadre soit complet.

Pour valider cette affirmation, nous avons choisi une approche d'évaluation basée sur le raisonnement contrefactuel. Le premier et le troisième auteurs ont codé de manière coopérative deux définitions de validité

précédemment codées comme « validité du cahier des charges » dans le cadre original. Tout d'abord, nous avons examiné la « validité d'application », définie comme la correspondance d'un modèle de simulation à son objectif et son cahier des charges, c'est-à-dire la probabilité que le modèle produise des résultats qui reflètent un artefact externe ou un système sociotechnique. Sur la base de notre meilleure compréhension du cadre, cette instance de validité a été recodée comme une *validité de critère d'efficacité*. La deuxième définition existante était la « validité sémantique », définie en partie sur la base de la pertinence des définitions de catégorie. Nous avons réalisé que l'entité de référence était la compréhension par les évaluateurs sémantiques d'un modèle équivalent ou d'une catégorie de système sociotechnique, suggérant qu'il s'agissait d'une *validité de modèle*. La même logique est devenue plus claire lors de l'examen de sept évaluations de science de la conception initialement classées comme « validités du cahier des charges ». Dans la plupart des cas, l'évaluation initiale s'est concentrée sur des termes tels que « exigence » ou « satisfaction » et « acceptabilité ». Souvent, les évaluations étaient mal décrites par les auteurs originaux et manquaient de clarté en termes d'artefact de référence ou de système sociotechnique réel. Dans cette évaluation de la validité causale, nous nous sommes concentrés sur ce que nous pensions que les chercheurs considéraient comme leur artefact de référence. Pour les sept évaluations, nous avons trouvé cinq cas de *validité du modèle* et deux cas de *validité d'exemplification*.

Ainsi, cette évaluation a montré que le cadre de validité pouvait être rendu plus parcimonieux sans perdre sa capacité de représentation. Ainsi, l'évaluation contrefactuelle de la *validité causale* a démontré que les parties supprimées du cadre n'étaient pas impliquées de manière causale dans la performance du cadre. Les autres types de validité ont tous été construits autour des types d'artefacts communément reconnus et étaient nécessaires pour classer les preuves, suggérant une *validité caractéristique causale* pour l'affirmation initiale de parcimonie pour les parties restantes du cadre. Par conséquent, il est peu probable que des évaluations

supplémentaires de la validité causale permettent de supprimer des parties supplémentaires. Par conséquent, aucune itération supplémentaire n'a été jugée nécessaire.

DISCUSSION ET IMPLICATIONS

Cette recherche apporte une valeur ajoutée à la science de la conception qui pourrait être étendue à d'autres applications. Nos travaux ont des implications à la fois théoriques et pratiques.

Contributions générales

Un aspect essentiel de la science de la conception consiste à valider les connaissances acquises sur les artefacts focaux. Cette recherche explique pourquoi les procédures de validation sont nécessaires et fournit un processus par lequel elles peuvent être identifiées et mises en œuvre à travers différents types de validité.

Nous proposons ce cadre de validité de la science de la conception. Il associe les types de validité aux caractéristiques des arguments, aidant ainsi les auteurs à formuler et à communiquer leurs arguments et les preuves qui les soutiennent. Basé sur la plus grande revue de littérature sur la validité jamais réalisée, ce cadre de validité de la science de la conception est un cadre complet pour la validation des arguments sur les artefacts. Le cadre fournit un vocabulaire standard pour les rapports de validité de la recherche. Il a une validité externe dans la mesure où il représente avec succès toutes les preuves de validité identifiées liées aux artefacts provenant des systèmes d'information, des sciences du comportement, de l'ingénierie, de l'informatique et de la médecine.

Le cadre guide les chercheurs dans l'identification des arguments portant sur les artefacts en considérant les branches de la hiérarchie des types de validité, renforçant ainsi la rigueur et la contribution des projets de science de la conception. En fournissant un processus et une nomenclature pour valider les arguments, le cadre peut être appliqué à tout type de création et d'évaluation d'artefacts. Relier explicitement les types de validité, les processus d'évaluation et les données à l'appui aux arguments devrait être utile

aux chercheurs et aux évaluateurs en science de la conception, ainsi qu'à ceux qui adoptent d'autres approches de recherche. Par conséquent, notre cadre peut aider à forger des liens entre la science de la conception et d'autres types de recherche. Les arguments explicites peuvent également être utilisés pour étendre les contributions aux travaux antérieurs. Cela pourrait être fait, par exemple, en faisant des affirmations causales et contextuelles sur les artefacts ou en ajoutant des critères supplémentaires avec des entités de comparaison plus récentes.

Constats et implications

La construction et l'évaluation du cadre de validité de la science de la conception ont permis de dégager plusieurs enseignements notables qui démontrent sa valeur. Tout d'abord, les scientifiques de la conception ont historiquement utilisé des évaluations *ad hoc* pour étayer des affirmations implicites sur la validité de l'artefact lui-même. Pour centrer et élargir la validité en tant qu'aspect de la rigueur en science de la conception, nous déplaçons ces évaluations vers l'établissement de la validité des affirmations explicites sur les connaissances de l'artefact. Ces affirmations seront étayées par des preuves à des degrés divers, dont la suffisance sera établie par la communauté et potentiellement contestée et modifiée au fil du temps.

Deuxièmement, il y a un manque surprenant de discussion approfondie sur la validité de la science de la conception, même si les évaluations et les types de validité sont bien établis et compris dans d'autres disciplines. La science de la conception publiée utilise une gamme étroite de concepts de validité, se concentrant principalement sur les mesures d'efficacité et les types de validité caractéristiques, ce qui suggère que la validité a été sous-utilisée. Dans notre revue de la littérature, les termes les plus fréquemment utilisés étaient exactitude, précision, rappel, spécificité, vrai positif et faux positif (concepts liés à la matrice de confusion). Il s'agit en fait de mesures plutôt que de types de validité, mais ils sont utilisés pour établir la validité de l'efficacité. Dans l'échantillon de publications analysées, ces mesures de validité

étaient le plus souvent liées à l'évaluation des modèles d'apprentissage automatique, ce qui, bien sûr, ne représente pas le champ d'application de la science de la conception.

Troisièmement, certains types de validité utilisés en science de la conception ont été adoptés par d'autres disciplines. Cependant, cela ne facilite pas une approche évaluative holistique pour la science de la conception. Par exemple, la validité en psychométrie et en économétrie est fortement axée sur la mesure, et certains types sont utiles en science de la conception (par exemple, la validité interne pour étayer les affirmations de causalité dans l'expérimentation). Cependant, les validités psychométriques ne sont pas suffisantes pour étayer l'éventail des arguments dans un projet de science de la conception.

Enfin, comme le confirme notre test d'applicabilité, un défi majeur pour les chercheurs qui créent des artefacts réside dans la nature incertaine de la validation et la tendance à effectuer des activités de validation de manière implicite sans avoir une idée précise de la structure sous-jacente de ces validations. Le cadre de validité fournit aux chercheurs un modèle structuré et un ensemble de types de validité soigneusement expliqués avec des définitions explicites et standardisées. Le processus de validation permet aux chercheurs de recueillir des données qui soutiennent un type de validité explicitement lié à une connaissance sur un artefact conçu.

Conseils pratiques

En résumant les arguments et les conclusions de notre article, nous formulons les recommandations suivantes à l'intention des chercheurs qui souhaitent développer des artefacts contribuant à la science et à la société.

Exprimer explicitement les arguments. Pour atteindre des niveaux élevés d'utilité pratique et de reproductibilité scientifique, les chercheurs doivent expliciter les connaissances relatives à leurs artefacts. Cela peut aider à encadrer les contributions à la recherche et à guider la validation, car la validation dépend des arguments articulés. De plus, les arguments

explicites favorisent l'accumulation des connaissances.

Étant donné que les connaissances de conception peuvent évoluer grâce au raffinement itératif d'un artefact (Tuunanen *et al.* 2024), des arguments peuvent émerger à différentes étapes au cours des itérations. Par exemple, certains arguments peuvent être formulés avant le développement d'un artefact, tandis que d'autres peuvent être formulés après le déploiement d'un artefact dans un certain contexte et l'observation des résultats. Cependant, ces derniers arguments ne doivent pas être considérés comme validés tant qu'ils n'ont pas subi une procédure de validation appropriée, éventuellement lors d'une itération ultérieure de l'artefact. Ce qui est important, c'est qu'un argument soit formellement évalué indépendamment du processus qui l'a généré, ce qui nécessite que les arguments soient énoncés explicitement.

Avancer des arguments à la mesure de la contribution visée. Le cadre de validité de la science de la conception ne doit pas être utilisé pour justifier des validations excessives. Il est souvent impossible/inutile d'énoncer toutes les affirmations concernant l'artefact et toutes les affirmations ne peuvent pas être soumises à une validation (par exemple, en raison de la difficulté à effectuer des comparaisons ou à acquérir des entités de comparaison appropriées). La question de la quantité de preuves nécessaires dépend du contexte et des caractéristiques spécifiques d'un problème.

Les chercheurs doivent formuler au moins une affirmation relative à l'artefact, en s'efforçant de faire des comparaisons avec des artefacts ou des processus de pointe. Cela peut être suffisant si l'artefact est particulièrement nouveau, de sorte que l'on ne sait pas grand-chose de ce qui le rend efficace ou des paramètres supplémentaires (au-delà de l'original) dans lesquels il peut être déployé. Au-delà de cela, les affirmations causales et contextuelles renforcent la contribution de la recherche car elles approfondissent les connaissances en matière de conception et aident les praticiens à appliquer de manière fiable et sûre les résultats de la recherche dans divers contextes.

Assurez-vous que chaque affirmation de connaissance est validée. Si une affirmation originale est formulée à propos de l'artefact, elle doit être validée. Par exemple, si un composant d'un artefact est censé provoquer un résultat spécifique, il convient alors d'établir la validité causale. Bien qu'une seule validation ne prouve pas une affirmation de connaissance, la validation des affirmations augmente la probabilité de produire des connaissances de conception fiables.

Quant à la mesure dans laquelle l'évaluation est nécessaire pour établir la validité d'une affirmation, Galison et D'Agostino (1987) présenter un argument convaincant selon lequel la suffisance des preuves est une question d'accord communautaire. Le nombre d'activités de validation requises n'est généralement pas basé sur une règle spécifique mais sur un consensus concernant l'assemblage d'« arguments convaincants, ceux qui tiendront la route devant un tribunal » (Galison et D'Agostino 1987, p. 227). Les chercheurs et les équipes d'évaluation parviennent à un tel consensus au cours du processus d'évaluation. Ce qui est important, c'est que les procédures de validation entreprises soient appropriées pour valider les connaissances avancées dans un article.

Valider les affirmations tout au long de la conception de l'artefact. De nombreux artefacts émergent grâce à des processus itératifs via l'expérimentation, le bricolage ou l'amélioration progressive. Certains projets impliquent plusieurs étapes et produisent différents artefacts au cours de ces étapes (par exemple, modèle conceptuel, prototype de système, version bêta, système final en production) (Tuunanen *et al.* 2024). Des affirmations de connaissances sur les artefacts résultants peuvent être formulées tout au long de ce processus. La validation formative de ces affirmations peut aider à affiner et à améliorer davantage l'artefact. Les validations « intermédiaires » peuvent être informelles, comme l'utilisation d'un échantillon de commodité d'utilisateurs potentiels ou d'un critère facile à obtenir. Cependant, pour garantir que l'artefact résultant contribue de manière fiable à la science et à la pratique, les affirmations sur la version *finale* ou *publique* de l'artefact doivent

être soumises à une validation sommative rigoureuse.

Assurez-vous d'utiliser des types de validité appropriés. Avec la mise en place du Cadre de validité de la science de la conception, un chercheur peut se référer au cadre lors de la validation de ses arguments. Le cadre organise diverses pratiques de validité dans un système de référence cohérent. Il indique quels types de validité sont appropriés pour chaque type d'argument et suggère les entités de comparaison et les procédures de comparaison raisonnables pour ces types de validité. À mesure que la communauté continue d'appliquer et d'affiner ces types de validité, leur robustesse devrait augmenter, offrant aux chercheurs une base plus solide sur laquelle construire leurs recherches.

Le cadre de validité de la science de la conception offre des possibilités de recherche future. Tout d'abord, le cadre est extensible, ce qui signifie qu'il peut accueillir des sous-types de validité supplémentaires qui pourraient être proposés par la communauté de recherche (par exemple, un affinement supplémentaire de la validité du modèle en fonction des types de modèles). Ensuite, il est possible de mieux suivre les modèles de validation en science de la conception et d'identifier leurs lacunes et leurs opportunités (par exemple, la nécessité de plus d'énoncés contextuels). Une possibilité connexe consiste à améliorer les procédures de validation en s'assurant que les pratiques de validation appropriées sont systématiquement saisies pour leurs types de validité respectifs. Enfin, les chercheurs peuvent étudier les applications du cadre dans les projets de science de la conception et rendre compte des résultats liés à l'utilité et à l'impact à long terme de l'utilisation du cadre sur la maturité de la science de la conception et son intégration avec d'autres traditions de recherche.

CONCLUSION

Cette recherche définit la validité de la science de la conception et propose le cadre de validité de la science de la conception ainsi qu'un processus pour son utilisation. Le cadre, issu d'une analyse approfondie de la littérature sur la validité,

identifie et organise les types de validité implicites et explicites en trois catégories principales : *critère*, *causal* et *contexte*. Les catégories de validité sont destinées à aider les chercheurs à obtenir et à présenter rigoureusement des preuves soutenant leurs arguments. La mise en correspondance des énoncés de connaissances avec le cadre de validité devrait aider les chercheurs et favoriser la mise en relation des connaissances scientifiques avec les artefacts des systèmes d'information. Le cadre lui-même a été validé en évaluant ses propres énoncés de connaissances et en fournissant des preuves des types de validité pertinents qui soutiennent ces propositions.

Remerciements. Les auteurs souhaitent remercier le rédacteur en chef, Balaji Padmanabhan, le rédacteur associé, Stefan Seidel, et les relecteurs anonymes pour leur travail sur les versions précédentes de notre article. Nous apprécions les commentaires reçus des participants aux ateliers du *HICSS* et de l'*ECIS* ainsi que de l'Université du Colorado, de l'Université d'Arizona, de l'Université d'État de l'Arizona et de l'Université de Sydney.

Financement. Cette recherche a été financée en partie par un financement du *National Institutes of Health* des États-Unis à Kai R. Larsen [Grant 3U24AG052175-08S1] et par un financement du Conseil de sciences naturelles et de génie du Canada à Jeffrey Parsons [Subvention RGPIN-2020-04916].

Références

- Abbasi, A., Albrecht, C., Vance, A., and Hansen, J. 2012. "Metafraud: A Meta-Learning Framework for Detecting Financial Fraud," *MIS Quarterly*, pp. 1293–1327.
- Abbasi, A., and Chen, H. 2008. "CyberGate: A Design Framework and System for Text Analysis of Computer-Mediated Communication," *MIS Quarterly* (32:4), pp. 1–30.
- Abbasi, A., Zhou, Y., Deng, S., and Zhang, P. 2018. "Text Analytics to Support Sense-Making in Social Media: A Language-Action Perspective," *MIS Quarterly* (42:2), pp. 1–38.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, and Psychological Testing (US). 2014. *Standards*

- for Educational and Psychological Testing, Amer Educational Research Assn.
- APA. 2020. "Validity," American Psychological Association. (<https://dictionary.apa.org/validity>).
- Arazy, O., Kumar, N., and Shapira, B. 2010. "A Theory-Driven Design Framework for Social Recommender Systems," *Journal of the Association for Information Systems* (11:9), pp. 455–490.
- Avdiji, H., and Winter, R. 2019. "Knowledge Gaps in Design Science Research," in *DESRIST*, pp. 1–9.
- Avoine, G., Bingol, M. A., Kardas, S., Lauradoux, C., and Martin, B. 2009. "A Formal Framework for Cryptanalyzing Rfid Distance Bounding Protocols," *This Work Is Partially Funded by FP7-Project ICE under the Grant Agreement* (206546), Citeseer.
- Baskerville, R., Kaul, M., and Storey, V. C. 2015. "Genres of Inquiry in Design-Science Research: Justification and Evaluation of Knowledge Production.," *MIS Quarterly* (39:3), pp. 541–564.
- Boudreau, M.-C., Gefen, D., and Straub, D. W. 2001. "Validation in Information Systems Research: A State-of-the-Art Assessment," *MIS Quarterly* (25:1), pp. 1–16.
- vom Brocke, J., Gau, M., and Mädche, A. 2021. "Journaling the Design Science Research Process. Transparency About the Making of Design Knowledge," in *DESRIST 2021*, Springer, pp. 131–136.
- vom Brocke, J., Hevner, A. R., and Maedche, A. 2020. *Design Science Research: Cases*, Berlin / Heidelberg: Springer.
- vom Brocke, J., Winter, R., Hevner, A., and Maedche, A. 2020. "Accumulation and Evolution of Design Knowledge in Design Science Research: A Journey through Time and Space," *Journal of the Association for Information Systems*.
- Burton-Jones, A., Boh, W. F., Oborn, E., and Padmanabhan, B. 2021. "Editor's Comments: Advancing Research Transparency at MIS Quarterly: A Pluralistic Approach," *MIS Quarterly* (45:2), pp. iii–xviii.
- Carter, J. W. 2019. *Aristotle on Earlier Greek Psychology: The Science of Soul*, Cambridge, UK: Cambridge University Press.
- Chan, E. K. 2014. "Standards and Guidelines for Validation Practices: Development and Evaluation of Measurement Instruments," in *Validity and Validation in Social, Behavioral, and Health Sciences*, Springer, pp. 9–24.
- Chatterjee, S., Sarker, S., Lee, M. J., Xiao, X., and Elbanna, A. 2021. "A Possible Conceptualization of the Information Systems (IS) Artifact: A General Systems Theory Perspective 1," *Information Systems Journal* (31:4), Wiley Online Library, pp. 550–578.
- Chowdary, G. J., and Kanhangad, V. 2022. "A Dual-Branch Network for Diagnosis of Thorax Diseases From Chest X-Rays," *IEEE Journal of Biomedical and Health Informatics* (26:12), IEEE, pp. 6081–6092.
- Cohen, L., Manion, L., and Morrison, K. 2013. *Research Methods in Education*, London England: Routledge.
- Collier-Reed, B., and Ingerman, Å. 2013. "Phenomenography: From Critical Aspects to Knowledge Claim," in *Theory and Method in Higher Education Research* (Vol. 9), Emerald Group Publishing Limited, pp. 243–260.
- Collins, J., Hall, N., and Paul, L. A. 2004. *Causation and Counterfactuals*, Cambridge, MA: MIT Press.
- Cook, T. D., and Campbell, D. T. 1979. *Quasi-Experimentation : Design & Analysis Issues for Field Settings*, Chicago: Rand McNally College Pub. Co.
- Creswell, J. W., and Miller, D. L. 2000. "Determining Validity in Qualitative Inquiry," *Theory into Practice* (39:3), Taylor & Francis, pp. 124–130.
- Ethayarajh, K., and Jurafsky, D. 2020. "Utility Is in the Eye of the User: A Critique of NLP Leaderboards," *ArXiv Preprint ArXiv:2009.13888*.
- Etudo, U., Yoon, V., and Liu, D. 2017. "Financial Concept Element Mapper (FinCEM) for XBRL Interoperability: Utilizing the M3 Plus Method," *Decision Support Systems* (98), Elsevier, pp. 36–48.
- Fan, W., Liu, X., Lu, P., and Tian, C. 2018. *Catching Numeric Inconsistencies in Graphs*, presented at the Proceedings of the 2018 International Conference on Management of Data, pp. 381–393.
- Galison, P. L., and D'Agostino, S. 1987. *How Experiments End*, Chicago, IL: University of Chicago Press Chicago.
- Germonprez, M., Hovorka, D. S., and Collopy, F. 2007. "A Theory of Tailorable Technology Design," *Journal of the Association for Information Systems* (8:6), pp. 351–367.
- Gonzalez-Huerta, J., Boubaker, A., and Mili, H. 2017. "A Business Process Re-Engineering Approach to Transform BPMN Models to Software Artifacts," in *E-Technologies*, pp. 170–184.

- Gregor, S., and Hevner, A. R. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37:2), pp. 337–355.
- Gregor, S., and Jones, D. 2007. "The Anatomy of Design Theory," *Journal of the Association for Information Systems* (8:5), pp. 312–335.
- Gregor, S., Kruse, L. C., and Seidel, S. 2020. "The Anatomy of a Design Principle," *Journal of the Association for Information Systems* (21:6), pp. 1622–1652.
- Gregory, R. W., and Muntermann, J. 2014. "Research Note—Heuristic Theorizing: Proactively Generating Design Theories," *Information Systems Research* (25:3), INFORMS, pp. 639–653.
- Guba, E. G., and Lincoln, Y. S. 1994. "Competing Paradigms in Qualitative Research," *Handbook of Qualitative Research* (2:163–194), California, Sage Publications, p. 105.
- Gumport, P. J. 2007. *Sociology of Higher Education: Contributions and Their Contexts*, Washington, DC: Johns Hopkins University Press+ ORM.
- Hevner, A., March, S., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75–105.
- Hevner, A., Parsons, J., Brendel, A. B., Lukyanenko, R., Tiefenbeck, V., Tremblay, M. C., and vom Brocke, J. 2024. "Transparency in Design Science Research," *Decision Support Systems* (182:1), pp. 1–19.
- Hoyningen-Huene, P. 2013. *Systematicity: The Nature of Science*, Oxford, UK: Oxford University Press.
- Iivari, J. 2015. "Distinguishing and Contrasting Two Strategies for Design Science Research," *European Journal of Information Systems* (24:1), Springer, pp. 107–115.
- Iivari, J., Rotvit Perlt Hansen, M., and Haj-Bolouri, A. 2021. "A Proposal for Minimum Reusability Evaluation of Design Principles," *European Journal of Information Systems* (30:3), Taylor & Francis, pp. 286–303.
- Johannesson, P., and Perjons, E. 2014. *An Introduction to Design Science*, Berlin / Heidelberg: Springer.
- Koornneef, H., Verhagen, W. J., and Curran, R. 2020. "A Decision Support Framework and Prototype for Aircraft Dispatch Assessment," *Decision Support Systems* (135), Elsevier, p. 113338.
- Kuechler, W., and Vaishnavi, V. 2012. "A Framework for Theory Development in Design Science Research: Multiple Perspectives.," *Journal of the Association for Information Systems* (13:6), pp. 395–423.
- Larsen, K., and Bong, C. H. 2016. "A Tool for Addressing Construct Identity in Literature Reviews and Meta-Analyses," *MIS Quarterly* (40:3), pp. 1–23.
- Larsen, K. R., and Becker, D. S. 2020. *Automated Machine Learning for Business: An Introduction to Accurate, Easy, and Fast Analytics*, New York NY: Oxford, UK.
- Larsen, K. R., Hovorka, D., Dennis, A., and West, J. D. 2019. "Understanding the Elephant: The Discourse Approach to Boundary Identification and Corpus Construction for Theory Review Articles," *Journal of the Association for Information Systems* (20:7), p. 15.
- Li, J., Larsen, K. R., and Abbasi, A. 2020. "TheoryOn: A Design Framework and System For Unlocking Behavioral Knowledge Through Ontology Learning," *MIS Quarterly*, pp. 1–55.
- Lincoln, Y. S., and Guba, E. G. 1985. *Naturalistic Inquiry* (Vol. 75), Thousand Oaks, CA: SAGE Publications.
- Lowry, P. B., Gaskin, J., Humpherys, S. L., Moody, G. D., Galletta, D. F., Barlow, J. B., and Wilson, D. W. 2013. "Evaluating Journal Quality and the Association for Information Systems Senior Scholars' Journal Basket via Bibliometric Measures: Do Expert Journal Assessments Add Value?," *MIS Quarterly*, pp. 993–1012.
- Lukyanenko, R., and Parsons, J. 2020. "Design Theory Indeterminacy: What Is It, How Can It Be Reduced, and Why Did the Polar Bear Drown?," *Journal of the Association for Information Systems* (21:5), pp. 1–30.
- Lukyanenko, R., Parsons, J., Wiersma, Y., and Maddah, M. 2019. "Expecting the Unexpected: Effects of Data Collection Design Choices on the Quality of Crowdsourced User-Generated Content," *MIS Quarterly* (43:2), pp. 634–647.
- Mandviwalla, M. 2015. "Generating and Justifying Design Theory," *Journal of the Association for Information Systems* (16:5), p. 314.
- March, S. T., and Smith, G. F. 1995. "Design and Natural Science Research on Information Technology," *Decision Support Systems* (15:4), pp. 251–266.
- Matadamas-Hernández, J., Román-Alonso, G., Rojas-González, F., Castro-García, M. A., Boukerche, A., Aguilar-Cornejo, M., and Cordero-Sánchez, S. 2012. "Parallel Simulation of Pore Networks Using Multicore Cpus," *IEEE Transactions on Computers* (63:6), IEEE, pp. 1513–1525.
- Maxwell, J. 1992. "Understanding and Validity in Qualitative Research," *Harvard Educational*

- Review (62:3), Harvard Education Publishing Group, pp. 279–301.
- Meth, H., Mueller, B., and Maedche, A. 2015. “Designing a Requirement Mining System,” *Journal of the Association for Information Systems* (16:9), p. 2.
- Moules, N. J., McCaffrey, J., and Field, J. 2015. *Conducting Hermeneutic Research: From Philosophy to Practice*, New York NY: Peter Lang Publishing.
- National Academies of Sciences, E., and Medicine. 2022. *Ontologies in the Behavioral Sciences: Accelerating Research and the Spread of Knowledge: Digest Version*, Washington, DC: The National Academies Press.
- Newell, A. 1975. “A Tutorial on Speech Understanding Systems,” *Speech Recognition*, pp. 4–54.
- Newton, P., and Shaw, S. 2014. *Validity in Educational and Psychological Assessment*, Hoboken, NJ: Sage.
- Nunamaker, J. F., Chen, M., and Purdin, T. D. 1991. “Systems Development in Information Systems Research,” *Journal of Management Information Systems* (7:3), pp. 89–106.
- Onwuegbuzie, A. J., and Leech, N. L. 2007. “Validity and Qualitative Research: An Oxymoron?,” *Quality & Quantity* (41:2), Springer, pp. 233–249.
- Pääkkönen, T., Kekäläinen, J., Keskustalo, H., Azzopardi, L., Maxwell, D., and Järvelin, K. 2017. “Validating Simulated Interaction for Retrieval Evaluation,” *Information Retrieval Journal* (20), Springer, pp. 338–362.
- Padmanabhan, B., Fang, X., Sahoo, N., and Burton-Jones, A. 2022. “Machine Learning in Information Systems Research,” *MIS Quarterly* (46:1), pp. iii–xix.
- Peffer, K., Tuunanen, T., and Niehaves, B. 2018. “Design Science Research Genres: Introduction to the Special Issue on Exemplars and Criteria for Applicable Design Science Research,” *European Journal of Information Systems* (27:2), pp. 129–139.
- Peffer, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. “A Design Science Research Methodology for Information Systems Research,” *Journal of Management Information Systems* (24:3), pp. 45–77.
- Piramuthu, S., and Doss, R. 2017. “On Sensor-Based Solutions for Simultaneous Presence of Multiple RFID Tags,” *Decision Support Systems* (95), Elsevier, pp. 102–109.
- Prat, N., Comyn-Wattiau, I., and Akoka, J. 2015. “A Taxonomy of Evaluation Methods for Information Systems Artifacts,” *Journal of Management Information Systems* (32:3), pp. 229–267.
- Ramakrishnan, M., Gregor, S., Shrestha, A., and Soar, J. 2023. “Design Principles for Platform-Enabled Knowledge Commons with an Expository Instantiation,” *Journal of the Association for Information Systems* (24:5), pp. 1413–1438.
- Refsgaard, J. C., Van der Sluijs, J. P., Brown, J., and Van der Keur, P. 2006. “A Framework for Dealing with Uncertainty Due to Model Structure Error,” *Advances in Water Resources* (29:11), Elsevier, pp. 1586–1597.
- Roberts, P., and Priest, H. 2006. “Reliability and Validity in Research,” *Nursing Standard* (20:44), Royal College of Nursing Publishing Company (RCN), pp. 41–46.
- Rosemann, M., and Vessey, I. 2008. “Toward Improving the Relevance of Information Systems Research to Practice: The Role of Applicability Checks,” *MIS Quarterly* (32:1), pp. 1–22.
- Salmon, W. C. 1998. *Causality and Explanation*, Oxford University Press.
- Sedrakyan, G., Poelmans, S., and Snoeck, M. 2017. “Assessing the Influence of Feedback-Inclusive Rapid Prototyping on Understanding the Semantics of Parallel UML Statecharts by Novice Modellers,” *Information and Software Technology* (82), Elsevier, pp. 159–172.
- Sein, M., Henfridsson, O., Purao, S., Rossi, M., and Lindgren, R. 2011. “Action Design Research,” *MIS Quarterly* (35:1), p. 37.
- Shultz, K. S., Riggs, M. L., and Kottke, J. L. 1998. “The Need for an Evolving Concept of Validity in Industrial and Personnel Psychology: Psychometric, Legal, and Emerging Issues,” *Current Psychology* (17), Springer, pp. 265–286.
- Taylor, C. S. 2013. *Validity and Validation*, Series in Understanding Statistics, New York NY: Oxford University Press USA.
- Thomas, M. A., Li, Y., and Lee, A. S. 2022. “Generalizing the Information Systems Artifact,” *Information Systems Research* (33:4), INFORMS, pp. 1452–1466.
- Tiefenbeck, V., Goette, L., Degen, K., Tasic, V., Fleisch, E., Lalive, R., and Staake, T. 2016. “Overcoming Salience Bias: How Real-Time Feedback Fosters Resource Conservation,” *Management Science* (64:3), pp. 1458–1476.
- Tuunanen, T., Winter, R., and Vom Brocke, J. 2024. “Dealing with Complexity in Design Science Research,” *MIS Quarterly* (48:2), pp. 427–458.
- Umapathy, K., Purao, S., and Barton, R. R. 2008. “Designing Enterprise Integration Solutions:

- Effectively,” *European Journal of Information Systems* (17:5), Springer, pp. 518–527.
- Valecha, R., Sharman, R., Rao, H. R., and Upadhyaya, S. 2013. “A Dispatch-Mediated Communication Model for Emergency Response Systems,” *ACM Transactions on Management Information Systems (TMIS)* (4:1), ACM New York, NY, USA, pp. 1–25.
- Venable, J., Pries-Heje, J., and Baskerville, R. 2016. “FEDS: A Framework for Evaluation in Design Science Research,” *European Journal of Information Systems* (25:1), pp. 77–89.
- Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. 2003. “User Acceptance of Information Technology: Toward a Unified View,” *MIS Quarterly*, pp. 425–478.
- Weinhardt, C., Kloker, S., Hinz, O., and van der Aalst, W. M. 2020. “Citizen Science in Information Systems Research,” *Business & Information Systems Engineering* (62), Springer, pp. 273–277.
- Winter, S., Berente, N., Howison, J., and Butler, B. 2014. “Beyond the Organizational ‘Container’: Conceptualizing 21st Century Sociotechnical Work,” *Information and Organization* (24:4), pp. 250–269.
- Zaitsev, A., and Mankinen, S. 2022. “Designing Financial Education Applications for Development: Applying Action Design Research in Cambodian Countryside,” *European Journal of Information Systems* (31:1), Taylor & Francis, pp. 91–111.

BIOGRAPHIES

Kai R. Larsen [0000-0002-8812-9866] est professeur de systèmes d'information à la *Leeds School of Business* de l'Université du Colorado à Boulder. Il est membre du corps professoral du département des sciences de l'information, professeur à l'université d'Agder et conseiller de recherche chez Gallup. Kai est surtout connu pour avoir fourni une solution pratique au Jingle Fallacy d'Edward Thorndike (1904) et pour ses contributions à la théorie sémantique des réponses aux enquêtes (STSR). Il a reçu le prix *INFORMS Design Science Award* de 2019, le *prix Herbert A. Simon Design Science Award* de 2020 et le *prix AIS Education Innovation Award* de 2023.

Roman Lukyanenko [0000-0001-8125-5918] est maître de conférence à la *McIntire School of Commerce* de l'Université de Virginie. Ses recherches portent sur la gestion des données et les méthodes de

recherche (validité et intelligence artificielle dans les revues de littérature). Roman développe activement des idées, des outils et des méthodes pour améliorer la gestion des données et les pratiques de recherche. Ces solutions ont reçu des prix importants, notamment le prix *INFORMS Design Science Award*, la médaille d'or du Gouverneur général du Canada et le prix Hebert A. Simon *Design Science Award*. Les recherches de Roman ont été publiées dans *Nature*, *MIS Quarterly*, *Information Systems Research* et *ACM Computing Surveys*. Son article de 2019 sur la qualité des données collaboratives a reçu le prix du meilleur article du *MIS Quarterly*.

Roland M. Mueller [0000-0002-8706-7763] est professeur de systèmes d'information à la *Berlin School of Economics and Law* en Allemagne. Il a obtenu son doctorat en systèmes d'information à l'Université libre de Berlin. Ses recherches portent sur les ontologies et les grands modèles de langage pour les revues de littérature, l'analyse métathéorique, l'ingénierie des ontologies, les méthodes de science de la conception et l'innovation pilotée par l'utilisateur dans les projets d'IA. Il est l'auteur de trois livres et de plus de 100 articles universitaires, et ses travaux ont reçu de nombreuses distinctions, dont le Herbert A. Simon *Design Science Award*. Au-delà du monde universitaire, Roland a contribué à l'innovation en tant que fondateur et membre du conseil consultatif de plusieurs startups. Il détient un brevet en apprentissage automatique.

Véda C. Storey [0000-0002-8753-1553] est professeure d'université distinguée et professeure Tull de systèmes informatiques et d'informatique au *J. Mack Robinson College of Business* de la *Georgia State University*. Ses recherches portent sur la gestion des données, la modélisation conceptuelle et la science de la conception. Elle s'intéresse particulièrement à l'évaluation de l'impact des nouvelles technologies sur les entreprises et la société du point de vue de la gestion des données. Le Dr Storey est membre du Collège des chercheurs seniors de l' AIS et du comité directeur de la Conférence internationale sur la modélisation conceptuelle. Elle est récipiendaire du prix Peter P. Chen, une *ER Fellow*, une *AIS Fellow* et une *INFORMS Fellow*.

Jeffrey Parsons [0000-0002-4819-2801] est professeur de systèmes d'information et professeur de recherche à la Faculté d'administration des affaires de l'Université *Memorial* de Terre-Neuve au Canada. Ses intérêts de recherche portent sur la manière de mieux représenter les conceptualisations humaines du monde dans les données. Ses travaux sur ce sujet et sur des

sujets connexes ont été publiés dans plusieurs disciplines. Les recherches de Jeff ont été reconnues de plusieurs manières, notamment en tant qu'article *MISQ* de l'année (2019), *AIS Senior Scholars Paper Award* (2020) et *INFORMS Design Science Award* (2014). Il est membre de l'*Association for Information Systems*, *Distinguished Research Fellow* de la TU Dresden, *Schoeller Senior Fellow* et *ER Fellow*. Il a été rédacteur en chef du *MIS Quarterly* et est actuellement rédacteur en chef d'*Information Systems Research*.

Debra Vander Meer [0000-0002-5930-6667] est professeure de systèmes d'information à la *Florida International University (FIU)*. Ses recherches actuelles portent sur l'application de concepts issus de l'informatique et des systèmes d'information à des problèmes concrets. Ses travaux sont publiés dans des revues et des conférences de premier plan dans ces domaines. Avant de rejoindre la *FIU*, elle a travaillé pendant dix ans dans l'industrie, notamment dans une start-up financée par du capital-risque. Elle est titulaire d'une licence de l'université de Georgetown, d'une maîtrise en systèmes d'information de gestion de l'université d'Arizona et d'un doctorat en informatique du *Georgia Institute of Technology*.

Dirk S. Hovorka [0000-0001-7049-5617] est professeur en systèmes d'information d'entreprise à l'Université de Sydney, en Australie. Il a obtenu son doctorat en systèmes d'information de l'Université du Colorado et est titulaire d'une maîtrise en télécommunications interdisciplinaires ainsi que d'une maîtrise en géologie. Ses recherches actuelles explorent la manière dont les pratiques scientifiques et sociétales font émerger des « mondes » possibles à travers la théorie, la conception et la conceptualisation des futurs. Ses recherches se concentrent sur *les approches spéculatives* de la « connaissance » technologique, la société et les environnements biophysiques, tout en reconnaissant que l'avenir est intrinsèquement résistant à la connaissance empirique et que des changements sociotechniques sans précédent rendent notre connaissance du passé moins indicative des états futurs. Dirk est rédacteur en chef du *JAIS (Research Perspectives)*, membre du comité de rédaction de l'*ISR* et coprésident du mini-parcours *HICSS « Informing Research: Engaging with Futures »* et du parcours *ECIS « Future as a Site of Inquiry »*. Il est lauréat du prix Beta Gamma Sigma *Professor of the Year 2018*, du prix Wayne Loneragan pour l'enseignement exceptionnel (2018), du prix du doyen pour l'enseignement (2022) et a reçu de nombreuses distinctions honorifiques d'enseignement. Dirk est co-auteur du meilleur article *AIS* 2011

« *Secondary Design: A Case of Behavioral Design Research* ».

Wanda Brunel est une étudiante en Master de Linguistique, littérature et civilisation en langue étrangère Anglaise à l'université Lumière Lyon II. Elle a obtenu sa licence de la même université en 2023 après avoir fait une année au Trinity College Dublin en Erasmus. Elle a été en 2024-2025 lectrice de français à l'université de Virginie.

Lila Belahlou, normalienne, est titulaire d'un master de sciences sociales obtenu auprès de l'École Normale Supérieure de Paris (ENS-PSL) et de l'École de Hautes Études en Sciences Sociales (EHESS). *Procter Fellow* de *Princeton University* en 2023, lauréate du Certificat d'Aptitude au Professorat de l'Enseignement du Second degré (CAPES) de Lettres modernes en 2024, elle effectue un lectorat au sein du département de Français de l'Université de Virginie en 2024-2025.

Annexe A. Utilisation générale des types de validité de la science de la conception

Nous avons examiné dans quelle mesure le cadre de validité de la science de la conception, traduit dans les 70 définitions de validité existantes classées dans notre cadre dans la colonne de droite du tableau 7, a été utilisé dans 199 articles de science de la conception publiés dans le panier de huit revues *AIS Senior Scholars* entre avril 2004 et décembre 2017. Chaque article a été examiné au niveau de la phrase et comparé à notre liste de types de validité, en utilisant des expressions récurrentes au niveau du mot. Le premier auteur a lu toutes les phrases résultantes et a exclu les résultats de recherche où leurs expressions récurrentes ont généré des faux positifs excessifs. Par exemple, des termes tels que *précision*, *l'exhaustivité*, *la précision* et *le rappel* ont des significations spécifiques, mais polysémiques, dans la recherche sur la qualité des données et l'apprentissage automatique. Les termes restants ont donné lieu à des dizaines de milliers de phrases, les résultats rapportés sont donc des estimations prudentes.

Nous avons identifié le nombre de fois qu'une validité correspondant à une valeur dans le cadre (tableau 7) a été utilisée au moins une fois dans un article, selon l'année et le type de validité. Nous avons conclu que les articles de science de la conception n'utilisent pas les mêmes termes de validité que ceux utilisés dans d'autres disciplines, à l'exception de certains utilisés pour décrire les mesures de la matrice de confusion. Les dénominations de validité associées aux types de validité caractéristiques étaient rarement utilisées. Même dans la catégorie très utilisée des types de validité d'efficacité, la plupart des termes liés à la validité discutés provenaient de mesures de la matrice de confusion dans l'apprentissage automatique. Les exceptions concernaient l'utilisation des *types de validité caractéristiques de la méthode*, les types de validité les plus utilisés dans la littérature sur la science de la conception.

Nous avons constaté un manque de cohérence et un manque d'utilisation réelle des termes de validité dans les articles sur la science de la conception, ce qui implique une opportunité significative d'améliorer la communication de l'évaluation grâce à une plus grande cohérence dans le langage relatif aux énoncés de validité explicites. Une fois que les chercheurs s'engagent à respecter des normes de validité partagées, la cohérence de la communication et des rapports devrait s'améliorer.