



# **Principes de conception et de validation d'un outil d'évaluation de l'utilisabilité perçue adapté à des enfants**

Charlotte Baraudon

## **► To cite this version:**

Charlotte Baraudon. Principes de conception et de validation d'un outil d'évaluation de l'utilisabilité perçue adapté à des enfants. Septièmes Rencontres Jeunes Chercheurs en EIAH (RJC EIAH 2018) , Apr 2018, Besançon, France. <hal-01769535>

**HAL Id: hal-01769535**

**<https://hal.science/hal-01769535v1>**

Submitted on 18 Apr 2018

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

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



HAL Authorization

# Principes de conception et de validation d'un outil d'évaluation de l'utilisabilité perçue adapté à des enfants

Charlotte Baraudon

Université de Lorraine, Metz, France  
charlotte.baraudon@reseau-canope.fr

**Résumé.** Un living lab correspond à un laboratoire ouvert où plusieurs acteurs peuvent dans un esprit de co-création venir tester ou faire tester des outils dans le but de développer de nouveaux produits conformes aux besoins et désirs des usagers. En lien avec le projet e-FRAN - e-TAC, nous avons conçu le Li'l@b comme un living lab adapté aux particularités du milieu scolaire. Le Li'l@b a pour objectif la co-conception d'interfaces numériques ciblées apprentissage s'inscrivant dans une approche dite centrée-utilisateur. Or, les outils de co-conception et d'évaluation d'interfaces numériques spécifiquement adaptés à des enfants restent rares comparativement à ceux destinés aux adultes. Face à ce constat, il apparaît nécessaire de concevoir et valider des outils favorables à l'implication d'élèves dans la co-création des environnements d'apprentissage qui leurs sont directement destinés. Dans ce contexte, cet article expose les principes de conception et d'évaluation d'un questionnaire d'évaluation de l'utilisabilité perçue, pensé pour être renseigné par des enfants dès 7 ans. Dans une première partie seront présentés le choix et la reformulation des items correspondant aux dimensions jugées pertinentes. Dans une deuxième partie, la mise en place d'un pré-test permettant de s'assurer de la compréhensibilité des items et de l'échelle de mesure et celle d'un test pour vérifier sa fiabilité seront abordés. Ces travaux préliminaires s'intéressent à l'évaluation globale de la perception des élèves de l'utilisabilité d'une interface numérique. Ainsi, ce questionnaire d'évaluation dite ergonomique a été conçu selon un mode simple et rapide d'administration.

**Mots clefs :** Utilisabilité perçue, Evaluation centrée enfant, Validation d'un outil psychométrique.

## 1 Introduction

Un living Lab correspond à une méthodologie d'innovation ouverte portée par l'utilisateur. L'approche consiste à promouvoir la co-création de solutions répondant aux enjeux sociétaux actuels [1]. C'est une approche nouvelle à l'innovation, car elle repose sur des collaborations non naturelles liant des usagers, des chercheurs, des ingénieurs, entreprises pour développer de nouveaux produits [1]. Phénomène initié à la fin des années 90, la mise en place de living labs prend actuellement de l'ampleur notamment depuis la création du réseau européen des living labs (ENoLL) en

2005 [2]. Ainsi, on compte actuellement près d'une centaine de living labs en France. Cependant, parmi ces derniers, il n'existe encore à notre connaissance, aucun living lab spécifiquement dédié à la co-conception d'environnements numériques d'apprentissage par des enseignants et des élèves de premier et second degré.

Dans le contexte en lien avec le projet e-FRAN - e-TAC [3], nous avons donc conçu le Li'l@b (espace du réseau Canopé et Université de Lorraine) comme un living lab adapté aux particularités du milieu scolaire. Le Li'l@b a pour objectif la co-conception d'interfaces numériques ciblées apprentissages formel et informel selon une approche dite centrée-utilisateur. Cette dernière vise à impliquer l'utilisateur final (ici élèves et/ou enseignants) à toutes les étapes de la conception. Or, si les outils de co-conception d'interfaces numériques permettant la participation des enfants se développent (e.g. [4-6]), les outils leur permettant de les évaluer restent rares [7, 8] comparativement à ceux destinés aux adultes (e.g. [9-11]). Face à ce constat, il apparaît nécessaire de développer des outils spécifiquement pensés et validés pour être renseignés par des enfants, et ainsi permettre leur implication directe dans l'évaluation des environnements d'apprentissage qui leurs sont destinés.

Cet article expose donc les principes de conception, actuellement en cours, et de validation d'un questionnaire d'évaluation empirique de l'utilisabilité perçue, spécifiquement pensé pour être renseigné par des enfants dès 7 ans. Nous faisons l'hypothèse que, à l'issue du processus de conception décrit ici, le questionnaire K-Uses sera un outil facilement compréhensible des enfants. Ceci dans le but permettre aux concepteurs d'interfaces numériques d'apprentissage d'avoir une évaluation fiable de l'utilisabilité perçue par des élèves et pouvoir y apporter des améliorations éventuelles directement en lien avec les besoins des apprenants.

## 2 Evaluation de l'utilisabilité et questionnaires d'évaluation

Dans le domaine des Interactions Homme-Machine (IHM) et des environnements informatiques pour l'apprentissage humain (EIAH), l'utilisabilité perçue d'une interface, d'un système numérique (ordinateur, tablette, logiciel, etc.) est un concept qui est apparu dans les années 80 [12, 13]. Si les questions de perception de l'utilisabilité par des utilisateurs adultes est très largement référencée [14], il n'en est pas de même pour les enfants. Bien qu'il pose parfois question quant aux critères qu'il recouvre [15, 16], la définition de l'utilisabilité telle que référencée dans la norme ISO 9241-11 peut servir de consensus. Elle définit l'utilisabilité comme : « Le degré selon lequel un produit peut être utilisé, par des utilisateurs identifiés, pour atteindre des buts définis avec efficacité, efficience et satisfaction dans un contexte d'utilisation spécifié » (ISO 9241-11,1998). Ici, l'efficacité se définit comme « la précision ou le degré d'achèvement selon lequel l'utilisateur atteint des objectifs spécifiés. ». L'efficience concerne « le rapport entre les ressources dépensées, la précision et le degré d'achèvement selon lequel l'utilisateur atteint des objectifs spécifiés. », et la satisfaction représente « l'absence d'inconfort, et les attitudes positives dans l'utilisation du produit ». On distingue généralement deux types de méthodes complémentaires d'évaluation de l'utilisabilité : les méthodes analytiques et les méthodes

empiriques [17]. Les approches analytiques évaluent la conception du système. Les interfaces sont ici étudiées selon un ensemble de référents afin de contrôler qu'elles possèdent bien certaines qualités et de détecter les problèmes qu'elles peuvent poser (e.g. *via* inspection heuristique, examen d'experts). Les évaluations empiriques consistent, quant à elles, à recueillir des données relatives aux comportements de l'utilisateur final lors de l'utilisation du système. Ce type d'évaluation (e.g. *via* tests d'utilisabilité, questionnaires) nécessite l'existence du système réel (ou de son prototype), et la présence d'utilisateurs. Dans le cadre de cette étude, nous nous focalisons sur cette dernière approche car nous souhaitons recueillir les perceptions des jeunes utilisateurs en vue de mieux connaître et comprendre les atouts ou les freins générés par les interfaces elles-mêmes au cours des apprentissages. Ces tests empiriques d'utilisabilité reposent souvent sur des questionnaires, car ils sont faciles à gérer, fiables, statistiquement objectifs, et économiques. Actuellement, de nombreux questionnaires d'utilisabilité sont disponibles [18]. Certains possèdent de nombreux items [19], d'autres ont été conçus spécifiquement pour des sites WEB [20, 21], notamment des sites WEB complexes [22]. L'un des plus utilisés reste le *System Usability Scale* ou SUS [9], car il est reconnu dans le monde des IHM comme fiable, rapide et facile d'administration. Cependant, tous ces questionnaires ont été conçus et validés pour des adultes et non pour de jeunes utilisateurs. Nous proposons donc de construire et valider une nouvelle échelle d'évaluation de l'utilisabilité par des enfants, nommée K-Uses, pour « Kids et Utilisabilité des systèmes et environnements scolaires » pour tout type de système numérique.

### **3 Moyens et méthodes de conception d'un questionnaire d'évaluation de l'utilisabilité perçue par des enfants - K-Uses**

Pour élaborer le questionnaire K-Uses, nous avons dans un premier temps recherché les avantages exposés par Bangor et al. (2008) [23] vis-à-vis du SUS. Aussi, nous visons la construction d'un outil d'évaluation de l'utilisabilité d'un système destiné aux enfants dès 7 ans, qui se veut être i) facile à mettre en œuvre, ii) facile d'exécution pour les répondants, iii) s'adaptant à tous les types d'interfaces, iv) gratuit, v) avec un score global facilement interprétable et qui v) soit validé et fiable.

#### **3.1 Définition des contenus du questionnaire**

Inspirée des méthodologies déjà employées pour des questionnaires d'utilisabilité existants (e.g. DEEP [22] et ICF-US [24]), la procédure établie pour développer le K-Uses est exposée ci-dessous (voir Fig. 1). La première étape basée sur le recensement de questionnaires d'utilisabilité existants (Annexe 1), permet d'identifier des dimensions pertinentes de l'utilisabilité concernant tous les types d'interfaces (e.g. ordinateur classique, tablette, interface tangible). Le choix des dimensions pertinentes s'est réalisé sur la base de la norme ISO 9241-11.

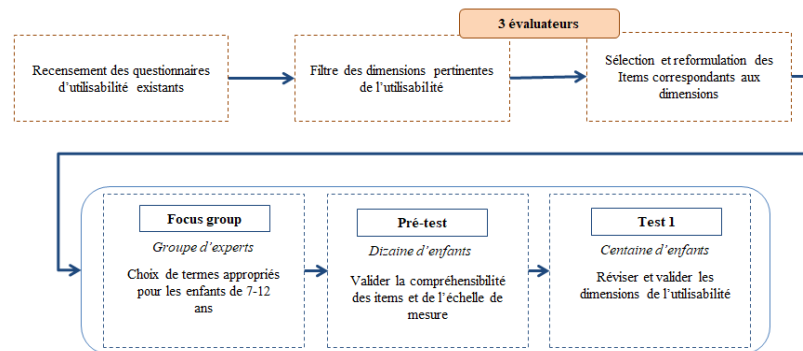


Fig. 1. Procédure utilisée pour valider et développer le K-Uses.

Puis, la deuxième étape effectuée par trois experts en IHM de façon croisée permet de sélectionner les items choisis en fonction de leur pertinence perçue par chacun et correspondant aux trois facteurs de l'utilisabilité associés à la norme ISO 9241-11. Les items décrivant des éléments trop précis de l'interface, ou ceux étant élaborés pour un seul type d'interface (e.g. site web) n'ont pas été retenus. Les experts prennent une décision à l'unanimité.

Puis, ces items sélectionnés sont soit maintenus tels quels, soit reformulés, soit supprimés. Cette étape permet d'obtenir un nombre réduit d'items susceptibles d'être retenus pour notre questionnaire pour arriver à atteindre un nombre de 4, 5 questions par dimension. Dans une troisième étape, un focus group impliquant des linguistes, des enseignants et des experts en IHM s'assurera que les termes soient appropriés à des enfants de 7-12 ans. Les modifications et ajustements nécessaires du questionnaire seront ensuite effectués.

### 3.2 Conception adaptée aux spécificités de réponses de l'enfant

Il est généralement admis que la recherche par sondage est réalisable avec des enfants dès l'âge de sept ans, mais avec des questionnaires soigneusement adaptés [25, 26]. Bell [27] recommandant d'éviter les questions négatives dans les questionnaires pour enfants, le questionnaire comprendra uniquement des affirmations. De plus, un texte d'introduction clair sera fourni et l'échelle sera entièrement numérotée pour aider les enfants dans leur compréhension [25, 27, 28]. Plusieurs études se sont consacrées à trouver un format d'échelle adapté aux enfants pour mesurer au mieux leurs jugements (e.g. [7, 8]). L'utilisation des échelles de type Likert, nécessitant une réponse graduelle à une série d'énoncés, est un moyen commun d'évaluer les attitudes, les valeurs, les états internes et les jugements [29]. Elles peuvent être proposées à des enfants à partir de 7 ans car préférées par ces derniers comparativement à d'autres typologies (e.g. VAS [30]). De plus, d'autres travaux mettent en évidence que les échelles en 5 points sont à privilégier pour ce type de public [7, 27, 29]. De ce fait, K-Uses demandera l'avis des jeunes utilisateurs selon une échelle de Likert en 5 points.

Prenant en compte le stade de développement des enfants et de leur maîtrise du vocabulaire, le questionnaire évaluera l'utilisabilité selon des aspects pragmatiques concrets et avec un vocabulaire explicite validé au cours du focus group, puis par les pré-tests.

#### **4 Processus de Validation**

A l'issue de ce processus de prototypage, un pré-test sur une dizaine d'élèves de cycle 3 servira à valider en situation réelle la compréhensibilité des items et celle de l'échelle de mesure. Les différentes affirmations du questionnaire seront lues par les enfants accompagnées d'une question portant sur l'évaluation de la compréhensibilité de ces affirmations [10, 31, 32]. Afin de simuler avec réalisme l'utilisation réelle d'outil, une consigne sera donnée aux participants d'évaluer l'utilisabilité d'un système bien connu des élèves (e.g. téléphone portable, tablette). Lors de la phase test qui aura lieu au Li-l@b, nous utiliserons deux systèmes numériques avec deux niveaux de difficultés d'utilisation que les élèves ont l'habitude d'utiliser pour éviter un effet de nouveauté, afin d'apporter un effet expérimental significatif et contrasté en termes de pouvoir de différenciation et de classification [33]. Une centaine d'enfants de 7-12 ans, à partir de scénarios, devront utiliser chacun des deux systèmes, répondre ensuite à notre questionnaire et à un item qui permet de mesurer l'utilisabilité globale du système afin de vérifier que leurs réponses soient cohérentes [10]. Pour savoir si le questionnaire est fiable, nous utiliserons le coefficient Alpha [34] qui permet d'évaluer le degré avec lequel les items corréleront entre eux. Le questionnaire sera considéré comme fiable si l'alpha est compris entre .70 et .90, valeurs qui montrent un niveau de consistance interne élevé. Enfin, pour analyser les données collectées, une analyse factorielle exploratoire et une analyse de fiabilité seront faites pour prouver et modifier les dimensions d'utilisabilité que nous avons supposées ou pour obtenir de nouvelles dimensions, dont le résultat sera justifié et révisé en utilisant l'analyse factorielle confirmatoire.

#### **5 Bilan et ouverture**

Nous avons exposé les principes et méthodes associés à la construction d'un nouvel outil d'évaluation de l'utilisabilité globale conçu pour tous les types de systèmes et spécifiquement pensés pour des enfants dès l'âge de 7 ans, aspect qui en fait son originalité. Cet outil s'inspire des avantages du SUS [9] qui se veut "*Quick and Dirty*". Au-delà des étapes de conception se basant sur de nombreux travaux antérieurs, la participation de linguistes, d'enseignants et experts IHM et le pré-test sur les utilisateurs finaux permettra de s'assurer qu'il soit simple, clair et approprié aux enfants. Enfin, l'analyse des données permettra de valider scientifiquement l'outil. Cependant, il nécessitera d'effectuer plusieurs études pour juger de sa valeur réelle à long terme. Cette étude est un point de départ pour de prochaines études permettant de tester l'efficacité du K-Uses pour des tâches de nature différente (e.g. dispositif pédagogique

non-numérique) et sur des populations d'âges différents (e.g. adolescents, adultes novices).

## Références

1. Dubé, P., Sarrailh, J., Billebaud, C., Grillet, C., Zingraff, V., Kostecki, I.: Le livre Blanc des Living Labs. Umwelt Service Design, Montréal, Canada (2014)
2. <http://www.openlivinglabs.eu/>
3. (<http://www.caissedesdepots.fr/espaces-de-formation-de-recherche-et-danimation-numeriques-e-fran>)
4. Iversen, O.S., Smith, R.C., Blikstein, P.: Special Issue on Digital Fabrication in Education. *International Journal of Child-Computer Interaction* (2016)
5. Read, J.: Children as participants in design and evaluation. *interactions* 22, 64-66 (2015)
6. Blikstein, P.: Digital fabrication and 'making' in education: The democratization of invention. *FabLabs: Of machines, makers and inventors* 4, (2013)
7. Hall, L., Hume, C., Tazzyman, S.: Five Degrees of Happiness: Effective Smiley Face Likert Scales for Evaluating with Children. In: *Proceedings of the The 15th International Conference on Interaction Design and Children*, pp. 311-321. ACM, (Year)
8. Yusoff, Y.M., Ruthven, I., Landoni, M.: The fun semantic differential scales. In: *Proceedings of the 10th International Conference on Interaction Design and Children*, pp. 221-224. ACM, (Year)
9. Brooke, J.: SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 4-7 (1996)
10. Lallemand, C., Koenig, V., Gronier, G., Martin, R.: Création et validation d'une version française du questionnaire AttrakDiff pour l'évaluation de l'expérience utilisateur des systèmes interactifs. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology* 65, 239-252 (2015)
11. Lewis, J.R., Mayes, D.K.: Development and psychometric evaluation of the emotional metric outcomes (EMO) questionnaire. *International Journal of Human-Computer Interaction* 30, 685-702 (2014)
12. Bennett, J., Case, D., Sandelin, J., Smith, M.: *Visual display terminals: usability issues and health concerns*. Prentice-Hall Englewood Cliffs New Jersey (1984)
13. Eason, K.D.: Towards the experimental study of usability. *Behaviour & Information Technology* 3, 133-143 (1984)
14. Tricot, A., Plégat-Soutjis, F., Camps, J.-F., Amiel, A., Lutz, G., Morcillo, A.: Utilité, utilisabilité, acceptabilité: interpréter les relations entre trois dimensions de l'évaluation des EIAH. In: *Environnements Informatiques pour l'Apprentissage Humain 2003*, pp. 391-402. ATIEF; INRP, (Year)
15. Bastien, J., Scapin, D.: Évaluation des systèmes d'information et Critères Ergonomiques. *Environnements évolués et évaluation de l'IHM, Interaction homme-machine pour les SI* 2, 53-80 (2001)
16. Hassenzahl, M., Platz, A., Burmester, M., Lehner, K.: Hedonic and ergonomic quality aspects determine a software's appeal. In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 201-208. ACM, (Year)
17. Senach, B.: *Évaluation ergonomique des interfaces homme-machine: une revue de la littérature*. INRIA (1990)
18. <http://vu.fr/annexe1>

19. Lin, H.X., Choong, Y.-Y., Salvendy, G.: A proposed index of usability: a method for comparing the relative usability of different software systems. *Behaviour & information technology* 16, 267-277 (1997)
20. Aladwani, A.M., Palvia, P.C.: Developing and validating an instrument for measuring user-perceived web quality. *Information & management* 39, 467-476 (2002)
21. Wang, J., Senecal, S.: Measuring perceived website usability. *Journal of Internet Commerce* 6, 97-112 (2007)
22. Yang, T., Linder, J., Bolchini, D.: DEEP: design-oriented evaluation of perceived usability. *International Journal of Human-Computer Interaction* 28, 308-346 (2012)
23. Bangor, A., Kortum, P.T., Miller, J.T.: An empirical evaluation of the system usability scale. *Intl. Journal of Human-Computer Interaction* 24, 574-594 (2008)
24. Martins, A.I., Rosa, A.F., Queirós, A., Silva, A., Rocha, N.P.: Definition and Validation of the ICF-Usability Scale. *Procedia Computer Science* 67, 132-139 (2015)
25. Borgers, N., De Leeuw, E., Hox, J.: Children as respondents in survey research: Cognitive development and response quality 1. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 66, 60-75 (2000)
26. Scott, J.: Children as respondents: Methods for improving data quality. *Survey measurement and process quality* 331-350 (1997)
27. Bell, A.: Designing and testing questionnaires for children. *Journal of Research in Nursing* 12, 461-469 (2007)
28. Hox, J., Borgers, N., Sikkel, D.: Response quality in survey research with children and adolescents: the effect of labeled response options and vague quantifiers. *International Journal of Public Opinion Research* 15, 83-94 (2003)
29. Mellor, D., Moore, K.A.: The use of Likert scales with children. *Journal of pediatric psychology* 39, 369-379 (2013)
30. Laerhoven, H.V., Zaag-Loonen, H.V.D., Derkx, B.H.: A comparison of Likert scale and visual analogue scales as response options in children's questionnaires. *Acta paediatrica* 93, 830-835 (2004)
31. Sharfina, Z., Santoso, H.B.: An Indonesian adaptation of the System Usability Scale (SUS). In: *Advanced Computer Science and Information Systems (ICACSIS), 2016 International Conference on*, pp. 145-148. IEEE, (Year)
32. Vallerand, R.J.: Vers une méthodologie de validation trans-culturelle de questionnaires psychologiques: Implications pour la recherche en langue française. *Canadian Psychology/Psychologie Canadienne* 30, 662 (1989)
33. Cai, M., Lin, Y., Zhang, W.: Study of the optimal number of rating bars in the Likert scale. In: *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services*, pp. 193-198. ACM, (Year)
34. Cronbach, L.J.: Coefficient alpha and the internal structure of tests. *psychometrika* 16, 297-334 (1951)