

Proposition d'une adaptation française et premières validations de l'échelle d'utilisabilité Computer System Usability Questionnaire (F-CSUQ)

Proposal for a French adaptation and first validations of the Computer System Usability Questionnaire (F-CSUQ)

Guillaume Gronier
guillaume.gronier@list.lu

Luxembourg Institute of Science and Technology
Esch-sur-Alzette, Luxembourg

Laurence Johannsen
laurence.johannsen@list.lu

Luxembourg Institute of Science and Technology
Esch-sur-Alzette, Luxembourg

ABSTRACT

The measurement of the usability of computer systems is based on a set of methods, including the use of questionnaires. While there are many questionnaires in English, few exist or have been adapted into French. This article proposes a translation of the Computer System Usability Questionnaire (CSUQ), developed within IBM and validated in the 1990s. This 19-item scale is particularly well suited to the evaluation of expert systems, especially internal systems in companies. It measures three dimensions of usability: usefulness, quality of textual information, and quality of the interface. This proposal for a French translation of the CSUQ (named F-CSUQ) was carried out by a committee of three bilingual experts. The first psychometric validations show a very good internal consistency, with a Cronbach's alpha of 0.960. This shows that all the F-CSUQ items converge towards the same concept. However, the principal component analysis (PCA) does not allow the original structure, composed of the 3 dimensions, to be recovered. Only a smaller version of the F-CSUQ, with 13 items and proposed in this article, respects the three-dimensional structure. These results nevertheless converge with those obtained by other translations of the CSUQ.

RÉSUMÉ

La mesure de l'utilisabilité des systèmes informatiques repose sur un ensemble de méthodes dont fait partie la passation de questionnaires. S'il existe de nombreux questionnaires en anglais, peu existent ou ont été adaptés en français. Cet article propose une traduction de l'échelle Computer System Usability Questionnaire (CSUQ), développée au sein d'IBM et validée dans les années 1990. Cette échelle à 19 items est particulièrement bien adaptée pour l'évaluation des systèmes experts, notamment les systèmes internes aux entreprises. Elle mesure trois dimensions de l'utilisabilité : l'utilité, la qualité des informations textuelles, et la qualité de l'interface. Cette proposition de traduction en version française du

CSUQ (nommée F-CSUQ) a été réalisée par un comité de 3 experts bilingues. Les premières validations psychométriques présentent notamment une très bonne consistance interne, avec un alpha de Cronbach de 0,960. Cela montre que tous les items du F-CSUQ convergent vers le même concept. Toutefois, l'analyse factorielle par composantes principales (ACP) ne permet pas de retrouver la structure originale, composée des 3 dimensions. Seule une version plus réduite du F-CSUQ, à 13 items et proposée dans cet article, respecte la structure tridimensionnelle. Ces résultats convergent néanmoins vers ceux obtenus par d'autres traductions du CSUQ.

CCS CONCEPTS

• **Human-centered computing** → Usability testing.

KEYWORDS

Questionnaire, Usability, Translation, Psychometric validation

MOTS CLÉS

Questionnaire, Utilisabilité, Traduction, Validation psychométrique

ACM Reference Format:

Guillaume Gronier and Laurence Johannsen. 2022. Proposition d'une adaptation française et premières validations de l'échelle d'utilisabilité Computer System Usability Questionnaire (F-CSUQ): Proposal for a French adaptation and first validations of the Computer System Usability Questionnaire (F-CSUQ). In *IHM '22: Proceedings of the 33rd Conference on Interaction Human-Machine (IHM '22)*, April 5–8, 2022, Namur, Belgium. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3500866.3516379>

1 INTRODUCTION

L'évaluation de l'utilisabilité [2] est une étape importante et incontournable dans le processus de Design de l'expérience utilisateur (UX). Elle consiste à recueillir l'avis des utilisateurs sur un produit, un système ou un service, de manières qualitative et quantitative. Cela permet ainsi de s'assurer de la qualité du produit, ou de proposer des recommandations pour son amélioration.

Parmi les méthodes d'évaluation, les questionnaires standardisés, aussi appelés échelles, sont des outils de mesure qui permettent de quantifier l'UX ou utilisabilité, et d'en établir un score qui doit être ensuite interprété. Ces questionnaires sont prédéfinis, contiennent un nombre d'items prédéterminés, et nécessitent des analyses particulières. Leur validation par une démarche scientifique en fait des outils fidèles et valides. La fidélité réfère au degré de constance et

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

IHM '22, April 5–8, 2022, Namur, Belgium

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9189-4/22/04...\$15.00

<https://doi.org/10.1145/3500866.3516379>

d'exactitude avec lequel un questionnaire mesure le concept qu'il est supposé mesurer. La validité réfère au degré selon lequel le questionnaire mesure le concept ce qu'il est supposé mesurer.

Les avantages des questionnaires d'utilisabilité sont nombreux : rapidité de passation, possibilité de diffusion à distance, prise en charge de larges échantillons d'utilisateurs, possibilité de benchmarks entre différents produits ou entre différentes versions d'un même produit, facilité de compréhension des résultats.

A ce jour, une dizaine d'échelles d'utilisabilité existe et a fait l'objet d'une démarche de validation. On recensera par exemple le System Usability Scale (SUS) [6, 7], le Usability Metric for User Experience (UMUX) [12], l'UMUX-LITE [28], le Design-Oriented Evaluation of Perceived Usability (DEEP) [34], le Software Usability Measurement Inventory (SUMI) [19], le Web Analysis and Measurement Inventory (WAMMI) [18], le Questionnaire for User Satisfaction (QUIS) [8], le modular evaluation of key Components of User Experience (meCUE) [30], et d'autres questionnaires que nous abordons dans cet article.

Néanmoins, la plupart de ces échelles sont originaires de pays anglo-saxons, et très peu d'entre elles ont été traduites en langue française et validées scientifiquement. Par exemple, l'AttrakDiff [22] et le SUS [14] possèdent une version française qui a été validée, mais bien d'autres échelles mériteraient d'intégrer la boîte à outils des UX designers ou des chercheurs en Interactions Humain-Machine (IHM).

C'est pourquoi nous nous sommes intéressés à la traduction d'une nouvelle échelle réputée outre-Atlantique, le Computer System Usability Questionnaire (CSUQ), que nous présentons dans la partie suivante. Malgré sa publication assez ancienne, en 1993, le CSUQ nous semble toujours être une échelle intéressante pour évaluer l'utilisabilité d'un système. En effet, non seulement le CSUQ comporte plusieurs dimensions qui donnent des indications sur le génotype du système, c'est-à-dire ce qui cause les problèmes d'utilisabilité [16], ou plus précisément les éléments de l'interface qui doivent être corrigés [34], mais il s'agit aussi d'une échelle encore très utilisée aujourd'hui pour différents types d'études, que nous décrivons plus précisément dans la section 2.3.

2 L'ÉCHELLE COMPUTER SYSTEM USABILITY QUESTIONNAIRE (CSUQ)

2.1 Présentation du CSUQ

L'échelle « Computer Usability Satisfaction Questionnaires » (CSUQ) [24] est une version dérivée de l'échelle PSSUQ (Post Study System Usability Questionnaire) [25], dont elle reprend exactement les mêmes items mais dans une formulation au temps présent (par exemple : It is simple to use this system/Il est simple d'utiliser ce système), mieux adaptée à une passation générique par questionnaire auto-administré. Le PSSUQ utilise quant à lui une formulation au temps passé (par exemple : It was simple to use this system/Il était simple d'utiliser ce système) afin de recueillir l'avis des utilisateurs juste après la passation de scénarios d'usage, principalement lors du recueil de données lors d'un test utilisateur.

Le CSUQ, tout comme le PSSUQ, est issu des travaux internes d'IBM menés par Suzanne Henry dans les années 1980 pour la

mesure de l'utilisabilité des systèmes, leur performance et la satisfaction qu'ils procurent à leurs utilisateurs (projet SUMS : System Usability MetricS).

2.2 Passation et calcul des scores

Le CSUQ se présente sous deux formes : une version originale à 19 items (tableau 1), et une version courte à 16 items. Les réponses des utilisateurs sont recueillies à l'aide d'une échelle de Likert à 7 points, allant de 1 = « Tout à fait d'accord » à 7 = « Pas du tout d'accord », associée à une possibilité de réponse « NA » (Non Applicable). Il n'y a pas d'items inversés.

La passation reste classique : après qu'un utilisateur ait interagi avec un système, que ce soit ponctuellement ou de manière régulière, il lui est demandé de répondre le plus spontanément et honnêtement possible à chacune des affirmations, en se positionnant sur l'échelle de 1 à 7. Dans le cas où il estimerait qu'une affirmation ne correspondrait pas au système qu'il doit évaluer, l'utilisateur peut cocher la case N.A.

Les scores du CSUQ sont regroupés en quatre catégories :

- un score global (OVERALL), qui réunit les items de 1 à 19, c'est-à-dire tous les items ;
- un score sur l'utilité du système (SYSUSE), qui réunit les items de 1 à 8 ;
- un score sur la qualité de l'information (INFOQUAL), qui réunit les items de 9 à 15 ;
- un score sur la qualité de l'interface (INTERQUAL), qui réunit les items de 16 à 18.

Les catégories SYSUSE, INFOQUAL et INTERQUAL représentent les trois dimensions du CSUQ (figure 1).

Le score de chaque catégorie est obtenu en réalisant une moyenne des items concernés. Puisque l'échelle peut être soumise à plusieurs reprises (aux mêmes sujets ou à des sujets différents), selon par exemple les différentes versions d'un système ou pour confronter plusieurs systèmes entre eux, les moyennes des évaluations successives peuvent être comparées entre elles et représentées graphiquement.

2.3 Utilisations et traductions actuelles du CSUQ

La formulation générique des items permet au CSUQ d'être appliquée à différents types de système. Par exemple, la CSUQ a récemment été utilisée pour mesurer l'utilisabilité d'un système de CAPTCHA [29], un tableau électronique pour l'éducation [4], une application mobile pour l'évaluation de la quantité d'alcool prise par les jeunes adultes [11], un site web de présentation des scores personnels de santé [9], un système de sécurité sociale électronique [17], ou encore un outil d'aide à la décision en matière d'anticoagulation pour la fibrillation auriculaire [33]. Ces cas d'application illustrent bien le caractère générique du CSUQ, et l'intérêt de ce questionnaire pour l'évaluation de systèmes actuels.

A notre connaissance, le CSUQ a pour le moment été traduit et validé en turque [10] (T-CSUQ), en espagnol [15] et en arabe [3] (A-CSUQ). Une version française a été proposée par Lallemand et Gronier [21], mais elle n'a ni été faite sur la base d'un comité de traduction, comme il est souvent d'usage dans une démarche

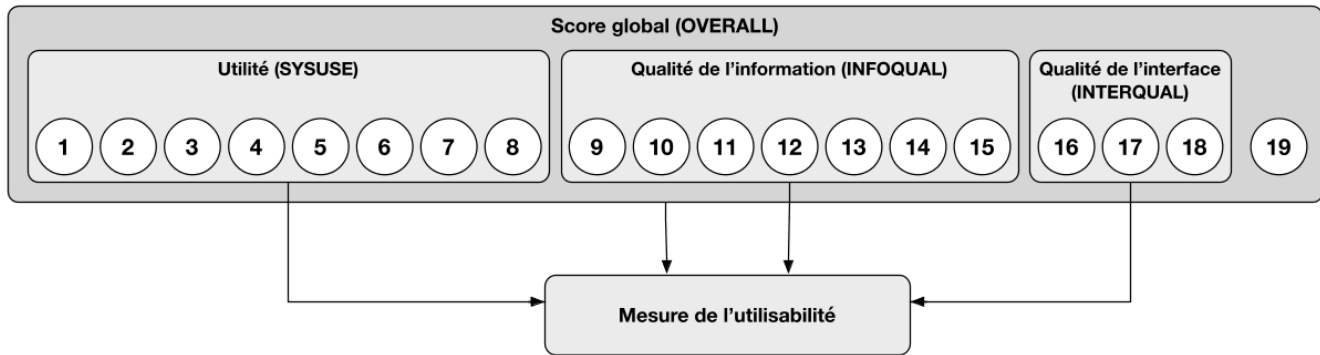


Figure 1: Répartition des 19 items du CSUQ par dimension.

scientifique de traduction, ni était validée du point de vue psychométrique. Il s'agit d'une traduction libre, comme il est précisé dans l'ouvrage des deux auteurs.

3 TRADUCTION FRANÇAISE DU CSUQ : LE F-CSUQ

Pour assurer la traduction du French-Computer System Usability Questionnaire (F-CSUQ), nous nous sommes inspirés de la démarche proposée par Vallerand [32]. Cet auteur propose une méthodologie de validation transculturelle de questionnaires psychologiques. Cette méthodologie comprend sept étapes qui permettent la traduction et la validation de questionnaires anglophones vers le français. Pour les besoins de notre traduction, nous avons retenu 3 étapes distinctes :

- (1) la préparation d'une version expérimentale ;
- (2) le prétest de la version expérimentale ;
- (3) la validation psychométrique du F-CSUQ.

Chacune de ces étapes est décrite dans les sections suivantes.

3.1 Préparation d'une version expérimentale

La première étape a consisté à préparer une version expérimentale du questionnaire original dans la langue cible (ici en langue française). Plusieurs méthodes sont possibles, parmi la traduction traditionnelle où le chercheur prépare seul une traduction, la méthode du comité et la méthode de la traduction inversée. Pour notre étude, et afin de minimiser les biais de la traduction traditionnelle que relate Vallerand [32], nous avons opté pour une traduction de type comité. Dans ce cadre, trois chercheurs bilingues, de nationalité française, ont été sollicités pour proposer une traduction des 19 items du CSUQ. Dans un premier temps, chaque traducteur a procédé à une traduction de manière individuelle. Puis dans un second temps, les chercheurs ont confronté leur traduction afin de se mettre d'accord sur une traduction commune.

À l'issue de la session par comité, un accord a donc été trouvé sur la traduction de chacun des 19 items (tableau 1). Notons que certains items n'ont pas été traduits mot à mot, mais qu'une formulation plus usuelle en langue française a souvent été privilégiée. Par exemple, l'item "Overall, I am satisfied with how easy it is to use this system", qui correspondrait pour une traduction mot à mot "Dans

l'ensemble, je suis satisfait de la facilité d'utilisation de ce système", a été traduit par le comité par "De manière générale, je trouve ce système facile à utiliser". En effet, les experts ont jugé que la formulation originale "je suis satisfait de la facilité d'utilisation" n'était pas naturelle en français, mais qu'il était plus courant de dire qu'un système était "facile à utiliser". De même, l'item original "I believe I became productive quickly using this system" se traduirait mot à mot par "Je crois que je suis devenu rapidement productif en utilisant ce système". Néanmoins, les experts ont préféré retenir une formulation qui ne garde pas le début de phrase "Je crois que", estimant que cela nuirait à la bonne compréhension de la phrase dans son intégralité. Aussi ont-ils préféré la traduction "Je suis rapidement devenu productif-ive en utilisant ce système".

En accord avec les traducteurs, nous pensons que la traduction d'une échelle dans une langue doit être adaptée aux formulations courantes de cette langue. Ce point a notamment été relevé pour la traduction française du SUS par Gronier et Baudet [14], qui ont comparé les propriétés psychométriques de deux traductions : une traduction mot-à-mot et traduction « littéraire ». Les formulations de la version littéraire ne reprenaient pas exactement celles des items originaux du SUS, mais s'affranchissaient des débuts de phrases comme « je pense que... », « je trouve que... », « j'ai trouvé que... » ou « j'ai pensé que... ». Après plusieurs analyses sur la compréhension des items et la robustesse des traitements psychométriques, les auteurs ont sélectionné la version littéraire, qui présente ainsi selon eux une structure rédactionnelle plus valide et mieux adaptée à la culture francophone.

Pour finir, notons que les traducteurs ont choisi une écriture inclusive, qui permet d'assurer une égalité des représentations entre les femmes et les hommes.

Suite à la traduction par comité, une traduction inversée a été réalisée auprès de cinq autres chercheurs bilingues français-anglais, de langue maternelle française. La traduction inversée consiste, à partir de la traduction vers une langue cible, de réaliser une traduction « aveugle » (c'est-à-dire sans connaître la version originale) vers la langue source. Pour notre étude, l'objectif était de vérifier que la traduction en français permettait de retrouver le sens des items originaux en anglais. Toutes les traductions inversées ont permis de valider que le sens des items avait été conservé.

Table 1: Traduction du CSUQ en français.

Items originaux (CSUQ)	Traduction française (F-CSUQ)
1. Overall, I am satisfied with how easy it is to use this system.	1. De manière générale, je trouve ce système facile à utiliser.
2. It is simple to use this system.	2. Ce système est simple à utiliser.
3. I can effectively complete my work using this system.	3. Je peux faire mon travail efficacement en utilisant ce système.
4. I am able to complete my work quickly using this system.	4. Je suis capable de faire mon travail rapidement en utilisant ce système.
5. I am able to efficiently complete my work using this system.	5. Je suis capable de faire mon travail de manière efficiente en utilisant ce système.
6. I feel comfortable using this system.	6. Je suis à l'aise avec l'utilisation de ce système.
7. It was easy to learn to use this system.	7. Cela a été facile d'apprendre à utiliser ce système.
8. I believe I became productive quickly using this system.	8. Je suis rapidement devenu productif-ive en utilisant ce système.
9. The system gives error messages that clearly tell me how to fix problems.	9. Les messages d'erreurs générés par le système expliquent clairement comment résoudre les problèmes.
10. Whenever I make a mistake using the system, I recover easily and quickly.	10. Lorsque je fais une erreur en utilisant le système, je peux les rattraper facilement et rapidement.
11. The information (such as on-line help, on-screen messages, and other documentation) provided with this system is clear.	11. Les informations (telles que l'aide en ligne, les messages à l'écran, et autre documentation) fournies avec le système sont claires.
12. It is easy to find the information I need	12. Il est facile de trouver l'information dont j'ai besoin.
13. The information provided with the system is easy to understand.	13. Les informations fournies avec le système sont faciles à comprendre.
14. The information is effective in helping me complete my work.	14. Les informations sont efficaces pour m'aider à faire mon travail.
15. The organization of information on the system screens is clear.	15. L'organisation des informations sur les écrans du système est claire.
16. The interface of this system is pleasant.	16. L'interface de ce système est agréable.
17. I like using the interface of this system.	17. J'ai plaisir à utiliser l'interface de ce système.
18. This system has all the functions and capabilities I expect it to have.	18. Ce système possède toutes les fonctionnalités et capacités que j'en attends.
19. Overall, I am satisfied with this system.	19. De manière générale, je suis satisfait-e de ce système.

3.2 Prétest de la version expérimentale

La deuxième étape a consisté à prétester la version expérimentale du F-CSUQ, afin de déterminer si les items étaient clairs, rédigés sans ambiguïté et ce, dans un langage qui s'apparente bien à celui de la population cible [32]. Pour ce faire, nous avons soumis la version expérimentale à un panel de praticiens de l'expérience utilisateur (UX). Nous leur demandions d'indiquer le niveau de compréhension de chacun des items sur une échelle de 1 (pas du tout compréhensible) à 5 (tout à fait compréhensible). 5 praticiens ont contribué à l'évaluation du F-CSUQ. Les moyennes et les écarts-types pour chaque item sont présentés dans le tableau 2.

Les résultats du prétest montrent qu'aucun item ne semble poser de problème de compréhension, l'item le moins bien noté ayant un score de 3,60 (item 5), ce que nous avons estimé acceptable puisque ce score est supérieur au score moyen de 3. Néanmoins, notons que cet item comprend le terme « efficient », qui selon nous peut être source d'ambiguïté car c'est un terme difficile à définir. Cela pourrait donc expliquer son score de compréhension plus faible que les tous les autres items. Rappelons que l'efficience est un composant du concept d'utilisabilité, défini selon la norme ISO 9241-11 [1] comme « le rapport entre les ressources dépensées et la précision et le degré d'achèvement selon lesquels l'utilisateur atteint des objectifs spécifiés ».

3.3 Étape 3 : évaluation psychométrique du F-CSUQ

La troisième étape a consisté à procéder à une évaluation psychométrique du F-CSUQ. Plusieurs analyses statistiques ont été réalisées, afin de mesurer la fidélité, la structure factorielle, la sensibilité et la validité de contenu.

La taille minimale recommandée de l'échantillon pour les tests psychométriques, et plus spécifiquement pour l'analyse factorielle, est d'avoir au moins cinq participants par item, ce qui, pour la CSUQ à 19 items, correspond un minimum de 90 participants [31].

Pour notre étude, nous avons administré le F-CSUQ à 215 utilisateurs. Les utilisateurs étaient des étudiants bénévoles en sciences humaines et sociales, à qui il était demandé de penser à un système qu'ils connaissaient bien, ou qu'ils pouvaient afficher à l'écran sur leur ordinateur ou leur smartphone. Les étudiants étaient alors invités à évaluer ce système en complétant le F-CSUQ, mais aussi le System Usability Scale (SUS) [6] qui servirait ensuite de comparateur. Nous ne nous sommes volontairement pas intéressés au système évalué. En effet, nous avons considéré que le système importait peu, mais qu'il était primordial que ce soit toujours le même système qui soit évalué pour tous les items du CSUQ. Les utilisateurs ont été sensibilisés sur ce point précis. Cette approche méthodologique, qui consiste à ne pas s'arrêter sur l'évaluation

Table 2: Moyennes et écart-types des scores de compréhension aux items du F-CSUQ sur une échelle de 1 (Pas du tout compréhensible) à 5 (Tout à fait compréhensible).

Items du F-CSUQ	Moyenne (Ecart-type)
1. De manière générale, je trouve ce système facile à utiliser.	4,40 (0,80)
2. Ce système est simple à utiliser.	4,40 (0,80)
3. Je peux faire mon travail efficacement en utilisant ce système.	4,00 (0,89)
4. Je suis capable de faire mon travail rapidement en utilisant ce système.	4,20 (0,75)
5. Je suis capable de faire mon travail de manière efficiente en utilisant ce système.	3,60 (0,80)
6. Je suis à l'aise avec l'utilisation de ce système.	4,40 (0,80)
7. Cela a été facile d'apprendre à utiliser ce système.	4,60 (0,80)
8. Je suis rapidement devenu productif-ive en utilisant ce système.	4,20 (0,98)
9. Les messages d'erreurs générés par le système expliquent clairement comment résoudre les problèmes.	4,40 (0,80)
10. Lorsque je fais une erreur en utilisant le système, je peux les rattraper facilement et rapidement.	4,20 (0,75)
11. Les informations (telles que l'aide en ligne, les messages à l'écran, et autre documentation) fournies avec le système sont claires.	4,20 (0,80)
12. Il est facile de trouver l'information dont j'ai besoin.	4,40 (0,89)
13. Les informations fournies avec le système sont faciles à comprendre.	4,00 (0,89)
14. Les informations sont efficaces pour m'aider à faire mon travail.	4,00 (0,98)
15. L'organisation des informations sur les écrans du système est claire.	4,00 (0,89)
16. L'interface de ce système est agréable.	4,20 (0,98)
17. J'ai plaisir à utiliser l'interface de ce système.	4,20 (0,98)
18. Ce système possède toutes les fonctionnalités et capacités que j'en attends.	4,20 (0,80)
19. De manière générale, je suis satisfait-e de ce système.	4,40 (0,89)

d'un système précis, a été appliquée pour la validation de la version française du SUS par Gronier et Baudet [14]. Les analyses psychométriques portent alors sur la cohérence interne de l'échelle et sur l'articulation des items les uns par rapport aux autres.

Au final, après avoir retenu les questionnaires entièrement et correctement complétés, nous avons retenu 167 réponses, correspondant à 100 femmes (âge moyen = 19,82 ans ; écart-type = 3,95 ; âge minimum = 17 ; âge maximum = 42) et 67 hommes (âge moyen = 21,70 ans ; écart-type = 2,42 ; âge minimum = 18 ; âge maximum = 30).

3.3.1 Mesure de la fidélité. La mesure de la fidélité permet d'estimer à quel point le score observé des items d'une échelle se rapproche du score vrai, c'est-à-dire d'un score qui mesure le même concept. Afin de mesurer la fidélité de notre échelle, nous avons calculé un alpha de Cronbach qui constitue l'un des tests de mesure de la fidélité les plus utilisés en sciences humaines et sociales. Le coefficient alpha de Cronbach du F-CSUQ atteste d'une fidélité satisfaite, avec un score de 0,960 nettement supérieur au seuil de 0,70 recommandé par Landauer [23] ou Kline [20]. Ce score est très proche de celui de l'échelle originale, qui est de 0,95 [24].

Aucune suppression d'item ne permet d'obtenir un alpha supérieur, ce qui démontre que tous les items convergent vers la mesure du même concept. Notre traduction dispose donc d'une bonne cohérence interne.

Afin de pouvoir comparer la fidélité du F-CSUQ à celle de la version originale, nous avons calculé les scores de fidélité pour les trois dimensions : SYSUSE, INFOQUAL et INTERQUAL. Les résultats sont présentés dans le tableau 3, en comparaison avec les alphas du CSUQ [24] du T-CSUQ [10] et du A-CSUQ [3].

Les résultats montrent que la consistance interne du F-CSUQ pour chaque dimension est la traduction qui s'approche le plus des scores de la version originale anglophone.

3.3.2 Analyse factorielle confirmatoire (AFC). Nous avons souhaité nous assurer que les items du F-CSUQ étaient statistiquement liés aux trois dimensions du CSUQ (SYSUSE, INFOQUAL et INTERQUAL).

Pour cela, nous avons calculé une analyse factorielle confirmatoire (AFC), qui a pour objectif de vérifier la structure d'un questionnaire en fonction de variables prédéfinies. Les variables sont ici les dimensions qui composent le questionnaire, le but étant alors de juger si l'appartenance des items à leur dimension est cohérente et valide.

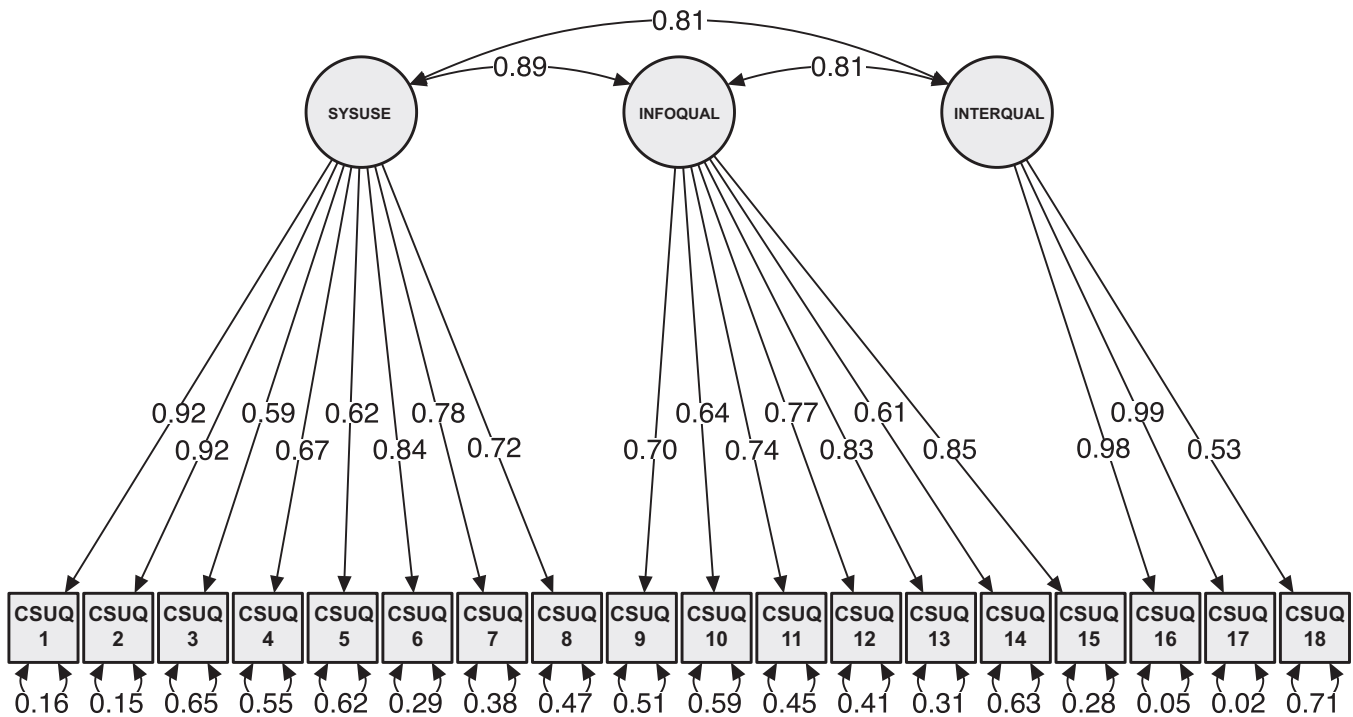
Les résultats présentés dans la figure 2 montrent que les items qui composent chacune des dimensions sont fortement corrélés à leur dimension, et ce de manière significative ($p < .001$). Autrement dit, la structure factorielle du F-CSUQ est cohérente avec celle de l'échelle originale.

L'AFC indique ainsi que le F-CSUQ peut être utilisé et interprété de la même manière que le CSUQ : les items expliquent chacune des dimensions à laquelle ils sont associés, ce qui signifie qu'il est possible de calculer un score par dimension et de l'interpréter comme avec le CSUQ.

3.3.3 Analyse factorielle par composantes principales (ACP) basée sur la valeur propre. Nous nous sommes ensuite intéressés à la manière dont était structuré le F-CSUQ, sans chercher à imposer une structure identique au CSUQ en termes de nombre de dimensions. Pour cela, nous avons appliqué une analyse factorielle par composantes principales (ACP). Cette méthode, à caractère exploratoire,

Table 3: Comparaison des alphas de Cronbach pour chaque dimension du F-CSUQ.

Dimensions	Alpha de Cronbach du CSUQ	Alpha de Cronbach du F-CSUQ	Alpha de Cronbach du T-CSUQ	Alpha de Cronbach du A-CSUQ
SYSUSE (1-8)	0,93	0,926	0,88	0,949
INFOQUAL (9-15)	0,91	0,896	0,71	0,925
INTERQUAL (16-18)	0,89	0,866	0,73	0,891

**Figure 2: Résultat de l'analyse factorielle confirmatoire du F-CSUQ, reprenant l'organisation par dimension des items du CSUQ.**

visé à mettre en évidence les dimensions qui organisent les relations entre différentes variables, les variables étant dans notre cas les items du questionnaire F-CSUQ.

Le test des éboulis de Cattell (1966) (Eigenvalues) suggère une structure à 2 composantes, avec chacune une valeur propre supérieure à 1 (figure 3). Autrement dit, les résultats révèlent que le F-CSUQ comporte naturellement 2 dimensions, que nous décrivons ci-dessous.

Le tableau 4 présente les scores de la rotation Varimax à deux composantes pour les 19 items du F-CSUQ, comprenant les items 1, 2, 6, 7, 11, 13, 15, 16, 17 et 19 pour la première composante, et les items 3, 4, 5, 8, 9, 10, 12, 14 et 18 pour la seconde.

Selon Nunnally [31], les items doivent avoir un poids supérieur à .05 pour être conservés dans l'analyse. Dans notre cas, l'item 18 n'atteint pas ce score ni pour la composante 1, ni pour la composante 2. Il s'agit donc d'un item qui peut être exclu du F-CSUQ.

On constate que les deux composantes regroupent chacune une catégorie d'items bien distincte. La composante 1 fait référence aux items qui traitent d'un système de façon générale, et qui en mesure

la satisfaction globale. Par exemple, on retrouve des items comme « *J'ai plaisir à utiliser l'interface de ce système* », « *Je suis à l'aise avec l'utilisation de ce système* », ou bien encore « *De manière générale, je suis satisfait-e de ce système* ». Ces affirmations s'appliquent à n'importe quel type de système, dans un contexte d'utilisation indéterminé, et pour des tâches indéfinies.

La composante 2 regroupe quant à elle des items relatifs à un système qui sert des tâches bien spécifiques, qui ont généralement attiré au travail. Par exemple, les items « *Je suis capable de faire mon travail de manière efficace en utilisant ce système* », « *Je suis capable de faire mon travail rapidement en utilisant ce système* » ou « *Les informations sont efficaces pour m'aider à faire mon travail* » sont clairement orientés vers la réalisation de tâches au travail. Dans ce cadre, les affirmations de la composante 2 s'appliquent davantage à l'évaluation d'un système dans un contexte professionnel, où les erreurs d'utilisation doivent pouvoir être facilement corrigées dans un soucis de performance et de productivité : « *Lorsque je fais une erreur en utilisant le système, je peux les rattraper facilement et rapidement* », « *Les messages d'erreurs générés par le système*

Table 4: Rotation Varimax à 2 composantes pour les 19 items du F-CSUQ. Les chiffres en gras indiquent une charge de l'item supérieure à .05 sur un seul facteur. Les chiffres en gras et italique indiquent une charge de l'item supérieure à .05 sur plusieurs facteurs.

Items du F-CSUQ	Composante 1	Composante 2
17. J'ai plaisir à utiliser l'interface de ce système.	0.871	
1. De manière générale, je trouve ce système facile à utiliser.	0.860	
2. Ce système est simple à utiliser.	0.860	
16. L'interface de ce système est agréable.	0.844	
15. L'organisation des informations sur les écrans du système est claire.	0.841	
19. De manière générale, je suis satisfait-e de ce système.	0.799	0.408
6. Je suis à l'aise avec l'utilisation de ce système.	0.787	
7. Cela a été facile d'apprendre à utiliser ce système.	0.746	
13. Les informations fournies avec le système sont faciles à comprendre.	0.743	
11. Les informations (telles que l'aide en ligne, les messages à l'écran, et autre documentation) fournies avec le système sont claires.	0.533	0.495
8. Je suis rapidement devenu productif-ive en utilisant ce système.	0.521	0.618
12. Il est facile de trouver l'information dont j'ai besoin.	0.494	0.591
9. Les messages d'erreurs générés par le système expliquent clairement comment résoudre les problèmes.	0.459	0.537
18. Ce système possède toutes les fonctionnalités et capacités que j'en attends.	0.410	0.468
5. Je suis capable de faire mon travail de manière efficiente en utilisant ce système.		0.919
3. Je peux faire mon travail efficacement en utilisant ce système.		0.890
4. Je suis capable de faire mon travail rapidement en utilisant ce système.		0.859
14. Les informations sont efficaces pour m'aider à faire mon travail.		0.760
10. Lorsque je fais une erreur en utilisant le système, je peux les rattraper facilement et rapidement.		0.602

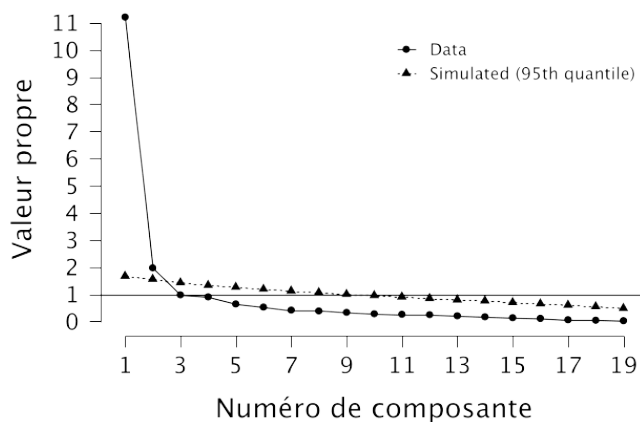


Figure 3: Tracé d'effondrement du F-CSUQ, qui suggère une structure à deux composantes.

expliquent clairement comment résoudre les problèmes », « Je suis rapidement devenu productif-ive en utilisant ce système ».

Notons que l'item 8 « Je suis rapidement devenu productif-ive en utilisant ce système » sature à 0.521 sur la composante 1 et 0.618 sur la seconde. Aussi, en raison de sa saturation plus élevée et du sens de l'item qui correspond bien à la réalisation de tâches au travail, avons-nous choisi d'associer cet item sur la composante 2.

3.3.4 Analyse factorielle par composantes principales (ACP) à 3 facteurs. Puisque le CSUQ est composé de 3 dimensions (SYSUSE, INFOQUAL et INTERQUAL), et que nous n'en avons trouvées que

2 lors d'une ACP basée sur la valeur propre, nous avons souhaité analyser la façon dont le F-CSUQ était structuré autour du même nombre de dimensions que l'échelle originale. Pour ce faire, nous avons calculé une ACP à 3 composantes, en ne gardant que les 18 items qui sont répartis dans une des trois dimensions (c'est-à-dire en excluant l'item 19 qui n'appartient à aucune dimension). En d'autres termes, nous avons cherché à savoir si la traduction du CSUQ permettait de retrouver la structure originale du questionnaire, c'est-à-dire le regroupement des items en fonction des dimensions SYSUSE, INFOQUAL et INTERQUAL.

Le tableau 5 présente les scores suite à la rotation Varimax à trois composantes pour les 18 items du F-CSUQ, comprenant les items 1, 2, 6, 13, 15, 16 et 17 pour la première composante, les items 3, 4, 5, 8, 14 et 18 pour la deuxième, et les items 9, 10, 11 et 12 pour la troisième.

On remarque ainsi que nous ne retrouvons pas exactement la même répartition des items par dimensions que le CSUQ. Cela était également le cas avec la traduction en turque. (T-CSUQ), qui observait une dispersion des items dans 3 composantes différentes, ne correspondant pas aux dimensions du questionnaire original [10].

Aussi, comme l'ont fait Erdiç et Lewis pour le T-CSUQ [10], nous avons cherché à réduire le nombre d'items du F-CSUQ pour qu'ils correspondent davantage aux dimensions SYSUSE, INFOQUAL et INTERQUAL. Ce procédé de réduction de l'échelle originale, qui s'appuie sur les résultats de l'analyse par composantes principales, ne nous semble pas dénaturer le sens de l'échelle, mais rend au contraire la version traduite plus efficiente et mieux adaptée à la langue cible. A ce titre, Erdiç et Lewis soulignent que le T-CSUQ n'est

Table 5: Rotation Varimax à 3 composantes pour les 18 items du F-CSUQ. Les chiffres en gras indiquent une charge de l’item supérieure à .05 sur un seul facteur. Les chiffres en gras et italique indiquent une charge de l’item supérieure à .05 sur plusieurs facteurs.

Dimensions	Items du F-CSUQ	Composante 1	Composante 2	Composante 3
SYSUSE	1	0.847		
	2	0.845		
	3		0.889	
	4		0.835	
	5		0.896	
	6	0.789		
	7	0.750		
	8	0.493	0.556	
INFOQUAL	9			0.774
	10			0.630
	11			0.745
	12			0.604
	13	0.639		0.522
	14		0.675	
	15	0.789		
INTERQUAL	16	0.809		
	17	0.832		
	18	0.457	0.516	

pas une simple traduction item par item du CSUQ, mais que seuls 13 des 19 items originaux ont survécu au processus de traduction et de validation. Par conséquent, le T-CSUQ est décrit comme une version optimisée du CSUQ en langue turque. Dans notre cas, en tenant compte du nombre d’items et de leur charge (supérieure à .05) par composante et par dimension, nous avons ainsi considéré que la composante 1 correspondait à la dimension INTERQUAL avec les items 16 et 17 ; la composante 2 à la dimension SYSUSE avec les items 3, 4, 5, et 8 ; et la composante 3 à la dimension INFOQUAL avec les items 9, 10, 11, 12 et 13. Notons ainsi que nous retrouvons pour chaque dimension du F-CSUQ (SYSUSE, INFOQUAL et INTERQUAL), une sélection des items qui composaient chacune des dimensions du CSUQ. Par conséquent, nous obtenons une échelle à 11 items, au lieu de 18, auxquels il faut rajouter l’item générique 19, qui n’appartient à aucune dimension de l’échelle originale.

Afin de vérifier la cohérence et la stabilité de cette échelle à 12 items, nous avons de nouveau calculé une ACP à 3 facteurs en ne conservant que les items mentionnés ci-dessus (tableau 6).

Les résultats montrent qu’à chaque composante correspond une dimension du F-CSUQ. Cela nous permet ainsi de conclure qu’il est possible de conserver la structure originale du questionnaire CSUQ, à condition de réduire le nombre d’items à 12.

3.3.5 Mesure de la sensibilité. La mesure de la sensibilité consiste à calculer, à l’aide de tests paramétriques de comparaison de moyennes, si une échelle est en mesure de différencier les scores selon différentes modalités d’une même variable indépendante. Bien souvent, on s’intéresse aux différences de scores obtenus selon le genre (femme-homme), ainsi qu’à l’impact de l’âge. Dans notre cas, tout comme pour le CUSQ [24] et le T-CSUQ [10], le calcul d’un *t-test* ne permet pas de révéler de différences significatives en

fonction du genre, et ce quel que soit les dimensions du F-CSUQ (tous les *p-value* sont supérieures à 0.2).

De même, il n’y a pas de corrélation entre l’âge et les résultats au F-CSUQ (tous les *p-value* sont supérieures à 0.1).

3.3.6 Validation concourante par comparaison du F-CSUQ avec le F-SUS. La validation concourante d’un questionnaire consiste à vérifier que les scores à une échelle donnée correspondent aux scores à une autre échelle censée mesurer le même concept. Dans notre cas, nous avons comparé les scores au F-CSUQ avec ceux obtenus avec le System Usability Scale (SUS) [6] dans sa version française (F-SUS) validée par Gronier et Baudet [14]. Les résultats sont présentés dans le tableau 7.

On observe ainsi que le F-SUS est corrélé de façons positive et significative (tous les *p-value* sont inférieurs à .001) avec toutes les dimensions du F-CSUQ, et avec son score global. Cela montre que le F-CSUQ mesure bien le même concept que le F-SUS, c’est-à-dire l’utilisabilité d’un système.

4 DISCUSSION

4.1 Le F-CSUQ : une échelle de mesure de l’utilisabilité des systèmes du quotidien et au travail ?

L’analyse factorielle par composantes principales basée sur la valeur propre a révélé une structure du F-CSUQ à deux facteurs, dont nous avons souligné qu’un facteur regroupait des items qui permettaient d’évaluer l’utilisabilité des systèmes du quotidien, et l’autre facteur des items davantage orientés vers l’évaluation des systèmes dédiés au travail. Le premier facteur s’intéressait ainsi à la simplicité d’utilisation, de compréhension et à la satisfaction générale d’un système ; le second s’intéressait à l’efficacité et la productivité de

Table 6: Rotation Varimax à 3 composantes pour 12 items du F-CSUQ. Les chiffres en gras indiquent une charge de l'item supérieure à .05 sur un seul facteur. Les chiffres en gras et italique indiquent une charge de l'item supérieure à .05 sur plusieurs facteurs.

Dimensions	Items du F-CSUQ	Composante 1	Composante 2	Composante 3
SYSUSE	3	0.894		
	4	0.870		
	5	0.906		
	8	0.573		0.516
INFOQUAL	9		0.741	
	10		0.659	
	11		0.772	
	12	0.401	0.691	
	13		0.655	0.503
INTERQUAL	16			0.896
	17			0.905
	19			0.765

Table 7: Coefficient de corrélation et degré de significativité entre les scores au F-SUS et au F-CSUQ.

		F-CSUQ (SYSUSE)	F-CSUQ (INFOQUAL)	F-CSUQ (INTERQUAL)	F-CSUQ (OVERALL)
F-SUS	Pearson's r	0.700	0.662	0.673	0.738
	p-value	< .001	< .001	< .001	< .001

tâches à réaliser dans un contexte de travail. En tenant compte de la suppression de l'item 18 qui n'était associé à aucun des deux facteurs, nous disposons alors d'une échelle à 18 items, répartis sur deux dimensions.

Cette nouvelle configuration du CSUQ n'est toutefois pas sans poser quelques questions. Le premier questionnement porte sur l'intérêt d'utiliser une échelle qui s'éloigne des différentes mesures du CSUQ, à travers les dimensions SYSUSE, INFOQUAL et INTERQUAL. En effet, la particularité du CSUQ est de pouvoir différencier plusieurs facteurs d'utilisabilité, à travers l'utilité, la qualité des informations, et la qualité de l'interface. Chacune de ces mesures permet ainsi de guider le diagnostic d'un système. En accord avec Yang, Linder et Bolchini [34], un problème d'utilisabilité, défini comme un obstacle à une utilisation satisfaisante, efficace et efficace du système, comporte invariablement deux composantes : la manifestation perçue de l'obstacle sur l'expérience réelle de l'utilisateur (appelée le phénotype d'utilisabilité) et le(s) défaut(s) réel(s) de conception du système qui cause(nt) le problème (appelée le génotype d'utilisabilité). Les trois dimensions du CSUQ permettent ainsi d'identifier le génotype d'utilisabilité, et constituent donc un moyen d'identifier les points faibles ou forts d'un système. Il convient donc d'en conserver la structure phénotypale à trois dimensions.

Le second questionnement porte sur la traduction même du F-CSUQ. Peut-être faut-il comprendre le terme « work », que l'on retrouve dans plusieurs items (« *I can effectively complete my work using this system* », « *I am able to complete my work quickly using this system* », « *I am able to efficiently complete my work using this system* »), plutôt dans le sens de « tâche à accomplir » que dans le sens de « travail » ? Dans ce cadre, le contexte professionnel serait

moins présent en utilisant le terme de tâche, ce qui changerait ainsi probablement l'interprétation des items pour les rendre plus adaptés à tout type de système. Cela changerait alors très certainement la structure factorielle du F-CSUQ.

4.2 Le F-CSUQ : une échelle plus courte que le CSUQ ?

L'analyse factorielle par composantes principales à 3 facteurs a quant à elle permis de retrouver les dimensions initiales du CSUQ. En supprimant les items dont la charge était inférieure à .05, nous avons dégagé une échelle à 12 items que nous reprenons dans son intégralité dans le tableau 8.

Cette version du F-CSUQ nous semble la plus cohérente et la plus respectueuse de l'échelle originale. En effet, les dimensions SYSUSE, INFOQUAL et INTERQUAL sont conservées, et la réduction du nombre d'items répond à une problématique actuelle qui vise à réduire le plus possible la taille des questionnaires d'utilisabilité, afin de faciliter leur passation. Cette réflexion sur la réduction des échelles a notamment été menée par Finstad [12] avec l'UMUX (comportant 4 items), qui se veut une échelle plus rapide à faire passer que le SUS (comportant 10 items). Lewis, Utesch et Maher [28] sont allés encore plus loin en proposant l'UMUX-LITE à 2 items.

Pour finir, comme nous l'avons souligné précédemment, la version turque du CSUQ propose également une échelle réduite à 13 items, répondant ainsi à un souci d'efficacité dans la passation du questionnaire. Le F-CSUQ à 12 items nous semble donc la meilleure version à retenir pour la traduction en français du CSUQ.

Table 8: Présentation du F-CSUQ dans sa version à 12 items.

Dimensions	Items du F-CSUQ
SYSUSE	1. Je peux faire mon travail efficacement en utilisant ce système.
	2. Je suis capable de faire mon travail rapidement en utilisant ce système.
	3. Je suis capable de faire mon travail de manière efficiente en utilisant ce système.
	4. Je suis rapidement devenu productif-ive en utilisant ce système.
INFOQUAL	5. Les messages d'erreurs générés par le système expliquent clairement comment résoudre les problèmes.
	6. Lorsque je fais une erreur en utilisant le système, je peux les rattraper facilement et rapidement.
	7. Les informations (telles que l'aide en ligne, les messages à l'écran, et autre documentation) fournies avec le système sont claires.
	8. Il est facile de trouver l'information dont j'ai besoin.
INTERQUAL	9. Les informations fournies avec le système sont faciles à comprendre.
	10. L'interface de ce système est agréable.
	11. J'ai plaisir à utiliser l'interface de ce système.
	12. De manière générale, je suis satisfait-e de ce système.

5 CONCLUSION ET PERSPECTIVES DE RECHERCHE

Cette recherche se donnait pour objectif de proposer une version française du questionnaire de mesure de l'utilisabilité CSUQ. Après une traduction par comité, une contre-traduction et différentes analyses psychométriques, nous avons adapté le questionnaire original dans une version française à 13 items, reprenant la structure initiale du CSUQ qui comprend 3 dimensions. Pour aller plus loin dans la validation de cette traduction, plusieurs études pourraient être menées et sont développées dans les sections suivantes.

5.1 Application du F-CSUQ sur différents types de systèmes

Le CSUQ a été appliqué à de nombreux systèmes de la vie quotidienne, dont notamment Google, Gmail, Amazon, Microsoft Word et Excel [27]. Ces évaluations ont permis de poursuivre les recherches sur la validation psychométrique du CSUQ, en comparaison avec d'autres questionnaires d'utilisabilité. Il nous semble pertinent de compléter ces études en confrontant le F-CSUQ aux mêmes systèmes, afin de comparer les résultats entre l'application de l'échelle originale et notre traduction.

De plus, notre étude n'a pas porté sur l'évaluation d'un système spécifique, puisque nous ne nous sommes pas intéressés à celui qu'avaient choisi nos participants. Or, l'évaluation d'un même système pourrait peut-être avoir des résultats différents de ceux que nous avons obtenus. Dès lors, nous avons d'ores et déjà mené une étude complémentaire sur un système interne à une grande administration, auprès d'une dizaine d'utilisateurs, dont les données sont en cours d'analyse. Le faible échantillon de cette étude ne permet toutefois pas de conduire des tests psychométriques de façon fiable, mais l'étude apporte un nouvel éclairage sur la manière dont le F-CSUQ peut être appliqué sur le terrain.

5.2 Traduction et validation du PSSUQ et du ASQ

Le CSUQ est intimement lié au PSSUQ et au ASQ, deux autres échelles conçues en même temps que le CSUQ [24]. Même si ces

échelles sont moins souvent utilisées que le CSUQ, elles représentent d'autres alternatives qui peuvent faire sens en fonction de l'objectif de l'étude en utilisabilité qui est menée. Par exemple, l'ASQ s'applique très bien après la réalisation de chaque scénario dans le cadre d'un test utilisateur. Le PSSUQ prend également tout son sens après la démonstration d'un produit. Ainsi, en complément du F-CSUQ, nous pensons que ces deux échelles pourraient être traduites en français, afin de couvrir toutes les situations d'évaluation possibles.

5.3 Faciliter l'interprétation du score au F-CSUQ

Pour finir, quelques recherches ont récemment engagé une réflexion sur l'interprétation du score du CSUQ [26, 27], tout comme cela a été le cas pour le SUS [5]. Ces recherches se donnent ainsi pour objectif de faire correspondre un adjectif qualificatif à un score obtenu à partir d'une échelle, afin que ce score puisse facilement attribuer une qualité à un système (bon, très bon, mauvais, horrible, etc.). Des recherches similaires pourraient être menées sur le F-CSUQ, tout comme cela a déjà été fait sur la version française du SUS [13].

REFERENCES

- [1] ISO 9241-11:1998. 1998. Exigences ergonomiques pour travail de bureaux avec terminaux à écrans de visualisation. Partie 11 : Lignes directrices concernant l'utilisabilité.
- [2] ISO 9241-210:2010. 2010. Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems.
- [3] Abeer A Al-Hassan, Bareeq AlGhannam, Mohammad Bin Naser, and Haneen Alabdulrazzaq. 2021. An Arabic Translation of the Computer System Usability Questionnaire (CSUQ) with Psychometric Evaluation Using Kuwait University Portal. *International Journal of Human-Computer Interaction* 37, 20 (2021), 1981–1988.
- [4] Obead Alhadreti. 2021. Assessing academics' perceptions of blackboard usability using SUS and CSUQ: A case study during the COVID-19 pandemic. *International Journal of Human-Computer Interaction* 37, 11 (2021), 1003–1015.
- [5] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.
- [6] John Brooke. 1996. Sus: a "quick and dirty" usability. *Usability evaluation in industry* 189, 3 (1996), 189–194.
- [7] John Brooke. 2013. SUS: a retrospective. *Journal of usability studies* 8, 2 (2013), 29–40.

- [8] John P Chin, Virginia A Diehl, and Kent L Norman. 1988. Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 213–218.
- [9] Alexandra Doggett, Dustin T Weiler, and Jason J Saleem. 2017. Variations in the usability of independent web-based personal health records. *International Journal of Human-Computer Interaction* 33, 4 (2017), 280–286.
- [10] Oğuzhan Erdinç and James R Lewis. 2013. Psychometric evaluation of the T-CSUQ: The Turkish version of the computer system usability questionnaire. *International Journal of Human-Computer Interaction* 29, 5 (2013), 319–326.
- [11] Tera L Fazzino, Corby K Martin, and Kelsie Forbush. 2018. The remote food photography method and smartintake app for the assessment of alcohol use in young adults: feasibility study and comparison to standard assessment methodology. *JMIR mHealth and uHealth* 6, 9 (2018), e10460.
- [12] Kraig Finstad. 2010. The usability metric for user experience. *Interacting with Computers* 22, 5 (2010), 323–327.
- [13] Guillaume Gronier. 2021. Mesure de la sensibilité et de la signification de la version française du System Usability Scale: Measuring the sensitivity and significance of the French version of the System Usability Scale. In *32e Conférence Francophone sur l'Interaction Homme-Machine*. 1–13.
- [14] Guillaume Gronier and Alexandre Baudet. 2021. Psychometric evaluation of the F-SUS: Creation and validation of the French version of the System Usability Scale. *International Journal of Human-Computer Interaction* 37, 16 (2021), 1571–1582.
- [15] AMI Hedlefs, GA de la Garza, MMP Sánchez, and VAA Garza. 2015. Adaptación al español del Cuestionario de Usabilidad de Sistemas Informáticos CSUQ. *RECI Revista Iberoamericana de las Ciencias Computacionales e Informática* 4, 8 (2015).
- [16] Erik Hollnagel. 1993. The phenotype of erroneous actions. *International Journal of Man-Machine Studies* 39, 1 (1993), 1–32.
- [17] Reza Khajouei and Fatemeh Farahani. 2019. The evaluation of users' satisfaction with the social security electronic system in Iran. *Health and Technology* 9, 5 (2019), 797–804.
- [18] Jurek Kirakowski and Bozena Cierlik. 1998. Measuring the usability of web sites. In *Proceedings of the Human Factors and Ergonomics Society annual meeting*, Vol. 42. SAGE Publications Sage CA: Los Angeles, CA, 424–428.
- [19] J Kirakowski, M Corbett, and M Sumi. 1993. The software usability measurement inventory. *Br J Educ Technol* 24, 3 (1993), 210–2.
- [20] Rex B Kline. 2005. *Principles and practice of structural equation modeling* (2nd ed. ed.). Guilford publications.
- [21] Carine Lallemand and Guillaume Gronier. 2015. *Méthodes de design UX: 30 méthodes fondamentales pour concevoir et évaluer les systèmes interactifs*. Editions Eyrolles.
- [22] Carine Lallemand, Vincent Koenig, Guillaume Gronier, and Romain Martin. 2015. Création et validation d'une version française du questionnaire AttrakDiff pour l'évaluation de l'expérience utilisateur des systèmes interactifs. *European Review of Applied Psychology* 65, 5 (2015), 239–252.
- [23] Thomas K Landauer. 1997. *Behavioral research methods in human-computer interaction*. Elsevier, Chapter Chapter 9, 203–227.
- [24] James R Lewis. 1995. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction* 7, 1 (1995), 57–78.
- [25] James R Lewis. 2002. Psychometric evaluation of the PSSUQ using data from five years of usability studies. *International Journal of Human-Computer Interaction* 14, 3-4 (2002), 463–488.
- [26] James R Lewis. 2018. Measuring perceived usability: The CSUQ, SUS, and UMUX. *International Journal of Human-Computer Interaction* 34, 12 (2018), 1148–1156.
- [27] James R Lewis. 2019. Measuring perceived usability: SUS, UMUX, and CSUQ ratings for four everyday products. *International Journal of Human-Computer Interaction* 35, 15 (2019), 1404–1419.
- [28] James R Lewis, Brian S Utesch, and Deborah E Maher. 2013. UMUX-LITE: when there's no time for the SUS. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2099–2102.
- [29] Kapil Chalil Madathil, Joel S Greenstein, and Kristin Horan. 2019. Empirical studies to investigate the usability of text-and image-based CAPTCHAs. *International Journal of Industrial Ergonomics* 69 (2019), 200–208.
- [30] Michael Minge, Manfred Thüring, and Ingmar Wagner. 2016. Developing and validating an English version of the meCUE questionnaire for measuring user experience. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 60. SAGE Publications Sage CA: Los Angeles, CA, 2063–2067.
- [31] Jum C Nunnally. 1978. *Psychometric theory*. McGraw-Hill New York.
- [32] Robert J Vallerand. 1989. Vers une méthodologie de validation trans-culturelle de questionnaires psychologiques: Implications pour la recherche en langue française. *Canadian Psychology/Psychologie Canadienne* 30, 4 (1989), 662.
- [33] Mark L Wess, Jason J Saleem, Joel Tsevat, Sara E Luckhaupt, Joseph A Johnston, Ruth E Wise, Jonathan E Kopke, and Mark H Eckman. 2011. Usability of an atrial fibrillation anticoagulation decision-support tool. *Journal of primary care & community health* 2, 2 (2011), 100–106.
- [34] Tao Yang, Jared Linder, and Davide Bolchini. 2012. DEEP: design-oriented evaluation of perceived usability. *International Journal of Human-Computer Interaction* 28, 5 (2012), 308–346.