Taylor & Francis
Taylor & Francis Group

Behaviour & Information Technology

ISSN: 0144-929X (Print) 1362-3001 (Online) Journal homepage: http://www.tandfonline.com/loi/tbit20

Shaking the usability tree: why usability is not a dead end, and a constructive way forward

Simone Borsci, Stefano Federici, Alessio Malizia & Maria Laura De Filippis

To cite this article: Simone Borsci, Stefano Federici, Alessio Malizia & Maria Laura De Filippis (2018): Shaking the usability tree: why usability is not a dead end, and a constructive way forward, Behaviour & Information Technology, DOI: 10.1080/0144929X.2018.1541255

To link to this article: https://doi.org/10.1080/0144929X.2018.1541255

Published online: 02 Nov 2018.

BEHAVIOUR & INFORMATION TECHNOLOGY
https://doi.org/10.1080/0144929X.2018.1541255

Taylor & Francis
Taylor & Francis Group

Shaking the usability tree: why usability is not a dead end, and a constructive way forward

Simone Borsci [a,b], Stefano Federici [c], Alessio Malizia [d,e] and Maria Laura De Filippis [f]

[a]Department of Cognitive Psychology and Ergonomics, Faculty of Behavioral Management and Social sciences, University of Twente, Enschede, Netherlands; [b]National Institute for Health Research, London IVD Co-operative, Faculty of Medicine, Department of Surgery & Cancer, Imperial College, London, UK; [c]Department of Philosophy, Social and Human Sciences and Education, University of Perugia, Perugia, Italy; [d]Schools of Creative Arts, University of Hertfordshire, Hertfordshire, UK; [e]Faculty of Logistics, Molde University College, Molde, Norway; [f]CognitiveLab Research Group, Perugia, Italy

ABSTRACT

A recent contribution to the ongoing debate concerning the concept of usability and its measures proposed that usability reached a dead end – i.e. a construct unable to provide stable results and to unify scientific knowledge. Extensive commentaries rejected the conclusion that researchers need to look for alternative constructs to measure the quality of interaction. Nevertheless, several practitioners involved in this international debate asked for a constructive way to move forward the usability practice. In fact, two key issues of the usability field were identified in this debate: (i) knowledge fragmentation in the scientific community, and (ii) the unstable relationship among the usability metrics. We recognise both the importance and impact of these key issues, although, in line with others, we may not agree with the conclusion that the usability is a dead end. Under the light of the international debate, this work discusses the strengths and weaknesses of usability construct and its application. Our discussion focuses on identifying alternative explanations to the issues and to suggest mitigation strategies, which may be considered the starting point to move forward the usability field. However, scientific community actions will be needed to implement these mitigation strategies and to harmonise the usability practice.

ARTICLE HISTORY
Received 6 June 2018
Accepted 22 October 2018

KEYWORDS
Human–machine interface; interaction design; ISO 9241-11; usability; usability factors; usability testing

1. Introduction

The formal umbrella construct of usability was defined in 1998 by the International Standard for Organisation, as follows: 'The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use' (ISO 9241-11, 1998). This definition aims to bridge and connect different definitions and standards of interaction design. Moreover, this umbrella provides a common framework of usability engineering (Bevan 1995; Bevan et al. 2016; ISO 9241-11, 1998). Before and after the formalisation of the ISO 9241-11 (1998), experts have attempted to critically discuss and extend the theoretical framework of usability (e.g. Quesenbery 2003).

The history of interaction design and evaluation is populated by alternative, and sometimes conflicting, constructs and definitions of usability. A detailed revision of the issues associated to the conceptualisation of usability was recently proposed by Tractinsky (2018). This author suggested that the framework of the ISO 9241-11 is collapsing and that the usability is a dead end. Tractinsky identified two main theoretical reasons to claim that the construct is falling apart, that we may summarise as follows:

(i) **The broad characterisation of the usability construct**. Usability is not well characterised by the ISO 9241-11 definition, namely the umbrella construct. In fact, umbrella constructs are usually too broad and weak to adequately serve the scope of a scientific community. The uncertainty around the construct of usability is a barrier to accumulate and to properly organise the disciplinary knowledge. For this reason, the scientific usability community is fragmented and experts are still debating the construct of usability and its assessment measures – i.e. issues of the fragmentation of knowledge.

CONTACT Simone Borsci s.borsci@utwente.nl Department of Cognitive Psychology and Ergonomics, University of Twente, 5 Drienerlolaan, 7522 NB Enschede, Netherlands

(ii) **The unclear representation and operationalisation of the usability construct**. Experts tend to operationalise their evaluation (protocols) in different ways by using a wide range of methods to assess different variables. This is a limitation to perform appropriate replication of findings and comparative analyses of usability evaluation studies (Hornbæk et al. 2014). Evidence gathered in comparative studies (Frøkjær, Hertzum, and Hornbæk 2000; Hornbæk 2006; Hornbæk and Law 2007; Sauro and Lewis 2009; Hornbæk et al. 2014) do not completely support the theoretical construct of usability. In fact, studies reported an unstable and weak relationship among the dimensions of efficiency, effectiveness, and satisfaction – i.e. the issue of the relationship among the usability metrics.

Several authors (Bardzell 2018; Bertelsen 2018; Hertzum 2018; Hornbæk 2018; Reeves 2018; Roto 2018; Stage 2018) commented on Tractinsky's proposal by mostly rejecting the conclusion that usability is a dead end, but also recognising the need of a way to progress in a constructive way the usability field.

We acknowledge that both the issues (the fragmentation of knowledge and the unstable relationship among the usability metrics) are key challenges for the field usability. Tractinsky's perspective (2018) is that these issues originate from the uncertainty associated to the weak umbrella construct of the ISO 9241-11. Tractinsky suggests that experts have to look for an alternative (and hopefully more) robust set of evaluation metrics yet to be identified. Therefore, this author proposes to abandon or to significantly revise the current construct.

The present work will take a different approach from the commentaries cited above; in fact, we will not reject or agree with the reasoning and the conclusion of Tractinsky, but we will attempt to define a constructive way forward.

To achieve this aim the present work critically discusses the two issues identified by Tractinsky and seeks for evidence to support alternative perspectives on the causes of knowledge fragmentation and the unstable relationship between the dimensions of the construct. Mitigation strategies against the issues identified above will be proposed to harmonise the fragmented field of

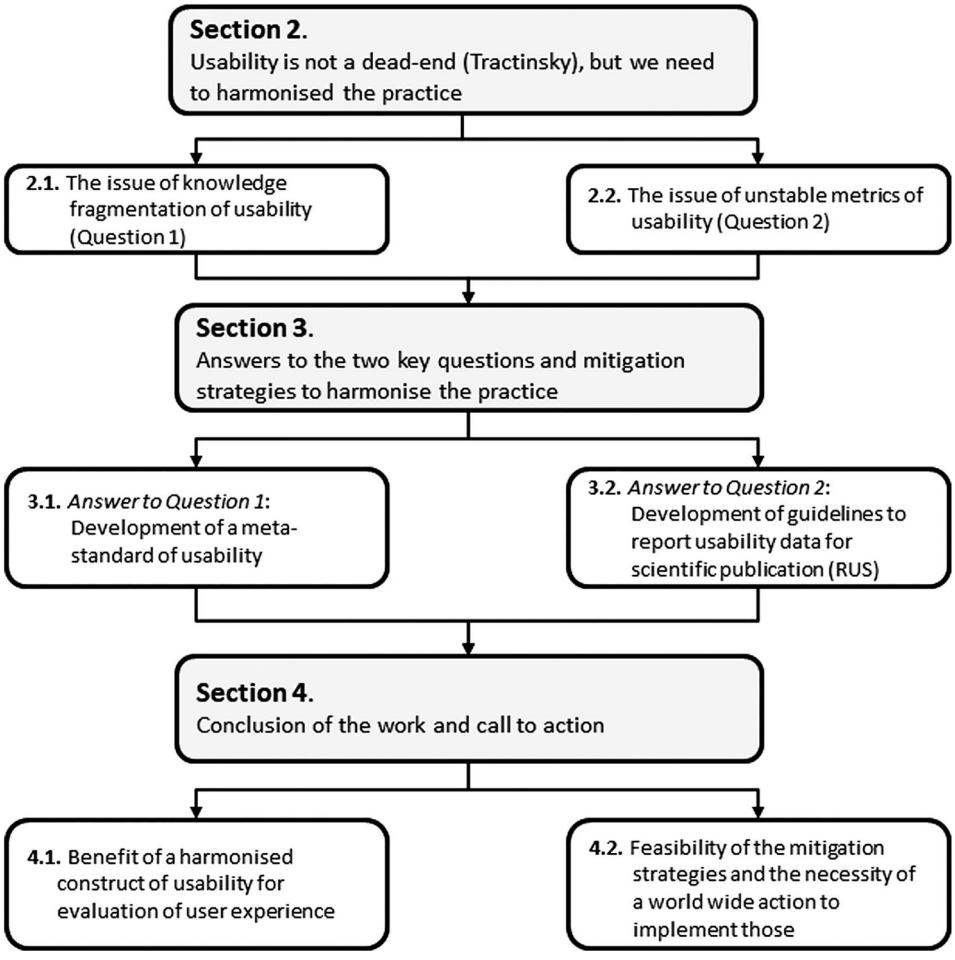

Figure 1. Structure of the work and key contents of the four main sections.

usability and to enable appropriate ways to replicate and compare usability studies.

Figure 1 provides a graphic insight of the contents of the four main sections of this work.

2. Theoretical issues

2.1. The broad characterisation of the usability construct: knowledge fragmentation

Umbrella definitions and terms are typical of multidisciplinary fields. These definitions usually aim to describe complex phenomena to bring together the perspective of experts from different backgrounds. Umbrella constructs are developed by communities to provide experts with a common set of objectives within a shared theoretical framework. Examples of umbrella definitions, terms, and frameworks are quite common in literature of different domains e.g. pervasive developmental disorders (Caronna, Milunsky, and Tager-Flusberg 2008); disability (WHO 2001); advanced nursing practice (Lowe et al. 2012); psychology of attitude (Eagly and Chaiken 2007); Public Service Motivation (Vandenabeele, Scheepers, and Hondeghem 2006), etc.

Hirsch and Levin (1999) suggested that umbrella definitions are usually abstract and not measurable. These types of definitions, after different phases of expansions, usually reach a natural phase in which the construct may collapse. Tractinsky (2018), by referring to the model of analysis proposed by Hirsch and Levin (1999), argued that the usability construct is collapsing because umbrella terms are usually too weak and broad. To summarise Tractinsky's assumption: the uncertainty created by the broad framework of usability has prevented the scientific community from accumulating and sharing common knowledge. In the present day, the usability umbrella has reached a critical point. In fact, experts compensate the fragility of the umbrella (offered by the ISO 9241-11) and are often forced to measure additional dimensions to the one specified in the construct – i.e. efficiency, effectiveness, and satisfaction.

Nevertheless, before we accept this assumption, we need to find an answer to the following question: Why in other multidisciplinary communities umbrella constructs and definitions are functional tools able to serve the aim of professionals and to unify their scientific communities? (Question 1).

2.2. The unclear representation and operationalisation of the usability construct: unstable metrics

Comparative studies and meta-analyses have identified a variable correlation among usability dimensions of efficiency, effectiveness, and satisfaction with a span on average from 0.25 to 0.6 e.g. in one study the level of correlation among the metrics to measure efficiency, effectiveness, and satisfaction may result lower (or higher) than the correlation reported in other studies (Frøkjær, Hertzum, and Hornbæk 2000; Hornbæk 2006; Hornbæk and Law 2007; Sauro and Lewis 2009; Hornbæk et al. 2014). The variability of the relationship among the usability dimensions is presented by Tractinsky (2018) as the ultimate evidence of failure of the usability construct. However, authors of these comparative studies have underlined that the variability in the relationship among the usability dimensions is due to multiple factors, such as: which methods are used and how, and also how data are reported by professionals to ensure that results can be replicated and compared (Frøkjær, Hertzum, and Hornbæk 2000; Hornbæk 2006; Hornbæk and Law 2007; Sauro and Lewis 2009). In tune with Hornbaek and colleagues (2014), 'Replicability' is defined as the possibility to confirm, expand, or generalise previous findings.

Tractinsky's (2018) assumption is that experts tend to operationalise the usability assessment in different ways because there is no clear relationship between the construct and its measures. We agree that the variability of the relationship among the usability dimensions may indicate an open and inherent methodological challenge in the field.

Nevertheless, before we accept the assumption that the usability construct is a dead end because its dimensions have an unstable relationship we need to find an answer to the following question: How is replicability fostered in usability and in other research fields? (Question 2).

In fact, an alternative explanation could be that this issue is not due to a deficiency in the ISO 9241-11 framework but to a lack in the way in which studies are reported and communicated. This resonates with initiatives such as RepliCHI (Wilson et al. 2013) which aimed to cumulate and share knowledge and previous experience within the human–computer interaction (HCI) community about methods and practices to enable replication of studies and results.

3. From issues to potential mitigation strategies

We identified in the sections above (2.1 and 2.2) two questions (Question 1 and 2) that need answers before we may declare that usability is a dead end:

- Question 1. Why are umbrella constructs and definitions are functional tools in other multidisciplinary communities to aid professionals and to unify their scientific communities? This question

resonates also with the points raised by previous commentaries (Bertelsen 2018; Hertzum 2018; Hornbæk 2018) about the fact that umbrella constructs are flexible, while rigid constructs may not serve the empirical instances of multiple communities. Indeed, as Reeves (2018) suggested, the nature of problems faced by usability experts are different and product-related, and practitioners usually are referring to usability in different ways. Therefore, probably, there is need for a better and unifying knowledge of the meaning of usability (Bertelsen 2018) but there is also the need to maintain a flexible and adaptive framework to serve multiple empirical purposes. Section 3.1 will attempt to answer Question 1.

- Question 2. How is replicability fostered in usability and in other research fields? This question resonates with the fact that empirical instances always require usability practitioners to associate design and evaluation (Hertzum 2018; Stage 2018) so that practitioners my apply different approaches and consider different factors when assessing usability. This variability of approaches aims to serve the scope and the purposes of products under assessment, despite being detrimental to the purposes of formal comparable analysis of evaluation results. Section 3.2. will attempt to answer Question 2.

3.1. Answer to Question 1. Umbrella definitions in other communities: from knowledge fragmentation to a meta-standard of usability

Among the different umbrella terms and definitions applied in other sectors, one of the most important and well-known is that of disability. Disability is defined as 'complex phenomenon, reflecting the interaction between features of a person's body and features of the society in which he or she lives' (WHO 2001). The definition of disability conveys the construct of the biopsychosocial model in which people functioning is evaluated in terms of body functions and structures, activities, and participation (Jette 2006; Federici et al. 2017; Federici and Scherer 2017).

Similarly to the usability construct, which provides a common framework to a variety of communities (e.g. designers, engineers, computer scientists, HCI and human factors experts, psychologists, etc.), the umbrella definition of disability has connected different perspectives (medical, psychological, and sociological) around the 'complex, dynamic, multidimensional, and contested' (WHO & World Bank 2011) characteristic of the individual functioning.

Evaluators who are dealing with individual functioning usually assess people disability throughout subjective and objective measures (Federici et al. 2017). The evaluation of people functioning is operationalised throughout a common evaluation framework known as the International Classification of Functioning, Disability and Health (ICF, WHO 2001).

The ICF provides (i) a classification of health and health-related domains for measuring health and disability at both individual and population levels, and (ii) a set of guidelines on how to share, communicate, and report the evaluation of functioning performed by experts from different fields.

The definition of disability, and its associated ICF framework, is about the same age as ISO 9241-11, but this umbrella is still standing and supporting professionals and researchers to evaluate people functioning. In fact, the biopsychosocial model of individual functioning and the definition of disability, thanks to the ICF, prevent fragmentation by providing a set of normative on what has to be evaluated and how to report the assessment.

Conversely, the ISO 9241-11 appears to be less able to convey a unified perspective. Usability researchers and practitioners from different communities (e.g. web design, software development, industrial design, manufacturing, etc.) mainly communicate their results and theoretical advancements primarily within their own fields. As suggested by Hornbæk (2018) experts of usability are well aware of the umbrella, the methods and standardised tools for evaluation, therefore the variability among the evaluation studies and their results, and among the knowledge communities, cannot be found only in the construct of usability.

We are not questioning that the usability field is fragmented; maybe, as suggested by Lewis (2014), still only few practitioners shared their research by attempting to reduce knowledge fragmentation across different communities. Definitively, however, looking at the contribution of umbrella constructs in other sectors, we cannot find Tractinsky's argument persuasive. In fact, just because umbrella terms are usually 'broad, blurred, and transient' (Tractinsky 2018, 35) this does not explain why in certain fields umbrella definitions bring to a cohesion of different perspectives and to solid and shared methods of evaluation, while in usability field the umbrella construct led to knowledge fragmentation.

Before we may accept or reject the Tractinsky's proposal (i.e. discarding the usability framework and looking for new unifying paradigms) we need to deeply understand the reasons behind this fragmentation. In fact, we could be exposed to the risk of replacing the

Table 1. Standards of product design and evaluation associated to ISO 9241-11.

Categories of products	Types of products	Standards which refer to ISO 9241-11
Medical products	(i) Devices/tests/Health Technology, Tele/e-health	IEC 62366; IEC 60601-1-6; ISO/TS 13131
Software applications	(ii) Interface design and Engineering, Data and biometrics, Information technology.	ISO/IEC 29156; ISO/IEC 25063; ISO/IEC 25064; ISO25060; ISO/IEC 25062; ISO/TR 16982; ISO/IEC TR 29196; ISO/IEC 29138 ISO/IEC 29136; ISO 27500; ISO/IEC 26514; ISO/IEC 25062; ISO/IEC 25051; ISO/IEC 25040; ISO/IEC 25030; ISO/IEC 25023 ISO/IEC 25022; ISO/IEC TS 25011; ISO/IEC 25010; ISO/IEC 25000 ISO/IEC 24779-4; ISO/IEC 24756; ISO/IEC 24751-1; ISO/IEC 24748; ISO/IEC TR 24714-1; ISO/IEC 23988; ISO 23185:2009 ISO/IEC TR 20748-1; ISO/IEC 19796-3; ISO/IEC 19773; ISO/IEC 17549-2; ISO 14915-3; ISO/IEC 14598-6; ISO/IEC 12207; ISO/IEC 10779; ISO 10075-3; ISO/IEC 25066; ISO 9241-11 family (part: 960; 420; 400; 307; 304; 210; 171; 151; 143; 129; 110; 100; 20)
other/miscellaneous applications	(iii) Everyday products	ISO 20282
	(iv) Building	ISO 21542; ISO 15686-10; ISO 11863
	(v) Design control centres	ISO 11064 family (parts: 5 and 7)
	(vi) Documents	ISO/TR 28118
	(vii) Machines industrial robots, office machines, earth-moving machinery, transportation	ISO 1503
	(viii) Manikins and body templates	ISO 15536
	(ix) Packaging	ISO 17480
	(x) Ships and marine technology	ISO 17894
	(xi) Vehicles	ISO 29061 ISO 17287
	(xii) Work systems	ISO 6385

usability construct with a new one without reducing the fragmentation of the different communities of practitioners.

3.1.1. Fragmentation of knowledge or adaptability to the innovation?

To explore the reasons behind the fragmentation, and to observe how the umbrella is applied in the real world, we researched the Online Browsing Platform of the ISO (https://www.iso.org/obp/ui). In particular, we searched for all those standards, which referenced the ISO 9241-11 to define the rules to assess the usability of different types of technologies. Out of 79 published standards (and partial standards) which referenced the ISO 9241-11, we manually:

(i) Grouped all the parts of the same standard in a family e.g. ISO 11064 part 5 and ISO 11064 part 7 were grouped in the same family: ISO 11064.
(ii) Excluded those standards that propose general guidelines of design and ergonomics principles without proposing a product-specific standard, e.g. ISO/TR 16982; ISO/TR 22411; ISO/TR 18529; ISO/TS 18152; ISO 26800.

The final dataset was composed of 54 standards. All these standards refer to the usability umbrella as the framework used to assess the interaction with 19 different types of products.

Table 1 summarises all the product-specific standards that usability experts have to comply with to evaluate the usability of different types of products. For our convenience, Table 1 presents the standards into three categories of products: medical products, software applications, other/miscellaneous applications.

Experts evaluating different types of products may be asked to look at the usability (characterisation and representation) in different ways – i.e. each product-specific standard may carry out variations of the ISO 9241-11. Three cases in which different communities have accommodated the umbrella construct to the inherent characteristics of different products they have to evaluate and can be exemplified by three standards for software application (ISO 25010, 2011), medical device (IEC 62336, 2015), and everyday products (ISO 20282, 2006), as follows:

- **Extension of the umbrella**: ISO 25010 considers the usability as a subset of the quality in use. Evaluators of software interfaces are requested to look not only at effectiveness, efficiency, and satisfaction, but also at the 'freedom from risk' and at the coverage of context – i.e. usability in the specified context of use and in contexts beyond those initially identified. In this

ISO, usability is not only measured in tune with the ISO 9241-11 definition, but it is also measured as a set of characteristics related to the quality: appropriateness recognisability, learnability, operability, user error protection, user interface aesthetics, and accessibility (ISO 25010, 2011). The ISO 25010 adds elements both to the usability construct characterisation and representation proposed in the ISO 9241-11 i.e. extension of the umbrella.

- **Integration of the umbrella**: ISO 62336 proposes a usability engineering process of medical device (IEC 62336, 2015). This product-specific standard conveys a process in which usability evaluation is strongly associated with the management and estimation of residual risk in use – i.e. the residual likelihood that harm may occur after the usability evaluation (ISO 14971, 2007). ISO 62336 recommends measuring usability only in tune with the key dimensions of the umbrella construct, although a particular relevance is posed on the definition of the contextual variables, which may produce unexpected risks in use. ISO 62366 does not add elements to the construct representation; however, the management of the risk in use is an add-on to the characterisation of the umbrella. In fact, for a medical device to be considered usable it must be efficient, effective, and satisfactory, and have an acceptable residual risk in use i.e. integration of the umbrella characteristics.
- **Reduction of the umbrella**: ISO 20282 suggests that to assess everyday products only considering the main dimensions of the umbrella construct. However, because the tasks performed with everyday products are usually of low complexity and risk, experts do not have to evaluate (residual or freedom from) risk. In particular, this standard suggests that experts have to prioritise the measure of effectiveness over the other dimensions of usability. Therefore, ISO 20282 does not add elements to the construct characterisation, but the dimension of effectiveness is prioritised in terms of operationalisation of the evaluation i.e. reduction of the umbrella.

These examples show that product-specific standards may propose extensions, integration and even reductions of the usability construct. Communities of experts, by working on specific products, have agreed on and formalised at international level amendments to the usability construct: (i) characterisation i.e. what usability is; and (ii) representation and operationalisation i.e. how to assess the interaction (Stolterman 2008) and how to report and communicate the results of a usability study (e.g. Theofanos and Quesenbery 2005). These amendments to the ISO 9241-11 accommodate inherent differences among the products (i.e. features, contexts, risk in use, etc.) and enable experts to evaluate (under the same umbrella) different aspects of the interaction with a product.

The primary cause of the knowledge fragmentation, hence, is not the construct of usability per se but the need of each community to have a flexible construct which may be modified and adapted (product-specific standards) to the need of a large technology landscape.

3.1.2. Umbrella organisation: advantages and disadvantages

The usability construct appears to be organised in a hierarchical way. The first level is composed by the ISO 9241-11, which proposes a one-size-fits-all construct. The second level is composed by product-specific standards, which amend the construct. This second level is continuously evolving to answer the specific evaluation needs of each type of technology.

The disadvantage of this hierarchical organisation is the fragmentation of knowledge e.g. experts who are working in the medical device field have a different perspective on (and know-how about) usability compared to experts who are working on software.

The main advantage of this organisation of the usability construct is the possibility to support developers and practitioners, independently from the type of product, to evaluate usability; even when a new type of technology emerges from the landscape and specific standards are not available. For instance, human–robot interaction (HRI) is becoming a reality in specialised sectors – manufacturing, surgery etc. The collaborative interaction with robots to perform tasks was recently regulated (ISO/TS 15066, 2016). Nevertheless, currently, a product-specific standard to operationalise the usability evaluation of collaborative HRI is not available. To cover this lack in regulation, experts are exploring ways to adapt the ISO 9241-11 for the specific characteristics of these type of technology (Kiesler and Hinds 2004; Yanco, Drury, and Scholtz 2004). The umbrella definition is acting as a safety net for experts, by enabling the generation of evaluation protocols under a unified framework, until product-specific standards will provide a regulation related to usability evaluation – see as an example of product-specific standard (under development) for robot-assisted surgery: IEC/DIS 80601-2-77 (under development).

The ISO 9241-11 proposes a common and broad ground, which can be applied when new technology and new forms of interaction emerge from the innovation landscape. Therefore, we cannot fully support the idea that the usefulness of the usability umbrella seems exhausted today (Tractinsky 2018, 37). Although,

we agree that fragmentation is an issue, our analysis also suggests that:

(i) The usability definition is solid enough to serve the scope of multiple communities of manufacturers and practitioners to develop and evaluate a variety of regulated and unregulated products;
(ii) The fragmentation of knowledge in the usability communities is the main disadvantage of the hierarchical organisation of the regulatory framework. However, this fragmentation results from the need to have a one-size-fits-all construct, which is flexible enough to react to the innovation changes.

This alternative perspective on the fragmentation issue resonates and extends the reasoning of other experts in their answers to Tractinsky (Bardzell 2018; Bertelsen 2018; Hertzum 2018; Hornbæk 2018; Reeves 2018; Stage 2018). By rejecting the assumption that the fragmented knowledge is due to the definition of usability, we are also proposing that fragmentation is an inherent aspect of the empirical need of different communities to adapt the usability evaluation to different products and technological innovation. This issue, therefore, could be handled without changing the construct but with a new international agreement and guidelines, which may reflect the different practices and knowledge gathered by different communities about the usability.

3.1.3. Reduce the fragmentation without losing the adaptability: first mitigation strategy

The fragmentation of knowledge and the diversity among communities of experts who work on specialised sectors is a main challenge in the usability field. As Lewis (2014) suggested, experts need to: (i) look at the lesson learned (and yet to be learned) within their own and other communities about usability; and (ii) experiment methods, and try to better exchange their know-how. However, encouraging experts to better communicate and acquire knowledge may not be enough to break the barriers among the communities. Nor could be enough to acknowledge and raise awareness about the fact that usability is a set of different of (sometimes conflicting) images; intended as the focus, mindset and perspective that practitioners are referring to when thinking and applying usability in their fields (Hertzum 2010).

A solution to the fragmentation issue needs to bring together the different images to enable a cross-disciplinary communication and share of knowledge. However, this requires time and a long process of sedimentation to enable a proficient exchange among the communities. In fact, to introduce methods of collaboration among such a heterogeneous collective of disciplines can bring with it some communication challenges between those parties involved, where difficulties making their concepts explicit can arise. A growing challenge to human-centred design practice involving usability is the need to bridge the communication gap between various professions, designers, other stakeholders, and end user groups involved in the design process. Even within sub-departments of organisations people have 'unique perspectives' on aims and tasks causing conflict. Alignment and recognition of the perspective and need of different communities require time for a new culture to be effective and productive. For example, Strober (2011) discovered that at Bio X, an interdisciplinary science centre at Stanford, it took two years of weekly meetings to learn the culture and habits of mind of each other's disciplines. Therefore, time effective solutions need to be explored to begin to address some aspects of developing productive communication and concepts about usability between various and diverse communities.

Recognising the fragmentation of knowledge as an inherent challenge of different usability communities (which often do not talk to each other) involves the opportunity to call for international actions among experts to build a cross-disciplinary regulation/guideline (meta-standard) of usability.

This meta-standard will not propose a change in the current characterisation of the usability definition, but it will reflect the hierarchical organisation of the construct and its product-specific adaptations. This may diffuse among communities the know-how about the different ways in which the construct is characterised, represented, and operationalised to test different and unregulated products. Moreover, this meta-standard may act in the long run as a point of reference for practitioners to enable the different communities to recognise each other, and to share their similarities and differences, and their cumulated knowledge in the application of the usability construct.

A key barrier to develop this meta-standard, however, is the issue related to the representation and operationalisation of the usability construct, i.e. how evaluation is performed and reported. Indeed, as also explained by Tractinsky (2018), if the usability construct is operationalised in different ways how can we ensure that people are measuring the same construct? And how can we compare and replicate different usability studies to enable adequate knowledge accumulation?

In fact, for such meta-standard to be useful and successful this has also to be able to provide a way for different communities to share, replicate, compare, and effectively communicate results.

3.2. Answer to Question 2. Replicability: from unstable metrics to guidelines to report usability studies to enable cross-discipline communication

The flexibility of the ISO 9241-11 is well reflected in the adaptability of the evaluation methods. Indeed, a variety of assessment methods (for a review on evaluation methods, see Wilson and Sharples 2015) are available to serve the evaluation objectives at multiple stages of product development. Practitioners, in each community, apply and adapt these common methods to satisfy their different scopes due to the variable contexts of use and the inherent difference of the products. For instance, in healthcare, usability is part of the outcome-based evaluations that include assessment of the perspective of patients, economic impact, accuracy, safety, reliability, etc. (Kushniruk and Patel 2004; Borsci, Buckle, and Hanna 2016; Polisena et al. 2018). Therefore, usability methods in healthcare are one of the multiple evidence needed to assess services and tools, and to support the adoption of new systems and processes (Green and Mercer 2001; Brown et al. 2013; Federici, Meloni, and Borsci 2016; Borsci et al. 2017; Roberts et al. 2017; Roberts et al. 2018). Clinicians, biomedical engineers and experts of health technology assessment are increasingly recognising the importance and demanding for adapted usability methods to evaluate the quality of interaction and its relationship with value and likelihood of technologies' adoption (Huddy et al. 2015; Polisena et al. 2018). Concurrently in other fields, such as the one of virtual and augmented reality field, usability methods are often used to also investigate the relationship between the quality in use and key factors such as: acceptance, trust, and presence (Borsci, Lawson, and Broome 2015; Borsci et al. 2016).

Video games have different design considerations and usability issues than other types of software. To ensure the satisfaction of game players, considerable care is required in the game design process and could be better guaranteed with the use of formal usability evaluation procedures by game developers (Federoff 2002). Not only video games require a careful selection and application of usability guidelines by practitioners (e.g. game designers and developers) but also they pose new challenges in terms of usability for Networked Multiplayer Games. Indeed, Networked Multiplayer Games need to allow a huge variety of interactions compared to single-player games. Think about the network implications of communication and coordination among players. Therefore, additional and relevant usability issues must be considered to address issues of group play whilst there are no current usability engineering heuristics specifically oriented to multiplayer games (Pinelle et al. 2009).

Even serious games – games designed for a primary purpose other than pure entertainment (e.g. games used by industries such as defence, education, etc.) – require a particular focus from practitioners to adopting the right usability guidelines for their purpose. The process of learning in an immersive multi-user environment often requires an increased time and continuity to achieve formal learning outcomes when compared with more traditional (face-to-face) learning environments. Nevertheless, the design process used in serious games is generally less controlled than in video games and involves mainly three stages: design, development, and research implementation with live players. Testing is somehow reduced in the development of such games, possibly due not addressing an audience as wide as the pure entertainment market. This can lead to negative (unintended) consequences such as learners building incorrect mental models of complex systems that are difficult to change (Warren, Jones, and Lin 2011).

In line with the heterogeneity of purposes, each usability evaluation protocol is designed by practitioners in tune with the product-specific standard or by considering the specific features of the device i.e. unregulated product. In addition to the type of product, experts select the methods for the assessment by taking into account several factors, such as: evaluator's expertise, stage of development, available budget and aim of evaluation – formative or summative (Hartson, Andre, and Williges 2001; Hertzum and Jacobsen 2003; Liljegren 2006; Hertzum 2010). Moreover, scenarios and tasks used during a usability evaluation usually represent inherent and specific contextual conditions of the interaction with a product. Each evaluation protocol contains, hence, a tailored way to measure usability – i.e. there are multiple correct ways to assess efficiency, effectiveness, and satisfaction.

Tailored evaluation protocols are based on practitioners experiential and utilitarian constructs of usability (Hertzum and Clemmensen 2012). This customised way of approaching usability evaluation is a common practice in the usability field. For instance, Hornbæk (2006) showed how practitioners who measured the construct of usability in 180 studies have applied 12 different measures to assess effectiveness, 13 to gather data related to efficacy, and 15 measures of satisfaction. The main advantage of the tailored evaluation protocol is that practitioners may select the most appropriate methods to deal with inherently different products and contexts of use. The main disadvantage is that each evaluation protocol may contain substantially different (and sometimes not comparable) measures to assess usability and other related aspects. As international experts highlighted (Frøkjær, Hertzum, and Hornbæk 2000; Hornbæk 2006; Hornbæk and Law 2007; Sauro and Lewis 2009;

Hornbæk et al. 2014) the scientific publications of usability studies often suffer the following issues:

(i) Researchers do not always provide enough information (data, context, methods) to compare or replicate the findings e.g. often only partial data are reported, or usability dimensions are only partially investigated (Hornbæk 2006);
(ii) Researchers rarely perform correlation analysis among efficiency, effectiveness, and satisfaction, e.g. a lack in terms of quality control (Frøkjær, Hertzum, and Hornbæk 2000; Hornbæk and Law 2007; Sauro and Lewis 2009);
(iii) Researchers rarely describe the criteria to select, and the validity or reliability of, evaluation methods. Therefore, measures applied could be inappropriate, or hard to be compared, e.g. the use of home-made questionnaires instead of validated scales to measure satisfaction, or the evaluation of user's performances in terms of user's perception of interaction (subjective performance) instead of using objective indexes such as errors and task achievement (Hornbæk 2006).

In line with these issues, comparative analyses of usability studies are always affected by (i) poorly reported data, and (ii) lack of information about applied methods and different conditions of assessment.

The weak correlation among the dimensions of usability appears to be due to inappropriate ways of reporting usability results in scientific studies more than to a lack of the framework. Therefore, also in this case, Tractinsky's solution (2018) to identify other (more reliable and rigid) metrics to unify the practice of interaction evaluation does not seem entirely convincing. This is mainly because changing the construct will not solve a problem due to practitioners' way of communicating and sharing results. In fact, researchers by measuring different products will always need a certain level of adaption of the metrics. Until research protocols and evaluated dimensions will not be reported appropriately, practitioners will always struggle to replicate and compare data and results independently from the type of usability construct.

3.2.1. Common guidelines to report usability data for cross-disciplinary communication: second mitigation strategy

Differently from the scientific community, industry has solved this problem by developing a Common Industry Format (CIF) for usability report of summative data. CIF was initially conceptualised in the early 2000s by a collaboration with several industries (Bevan 2009), and it originated the ISO/IEC 25062 (2006). This standard describes the components and how to use the CIF.

CIF aims to (i) promote best practice to report usability methods and data independently from the type of product evaluated by experts (Bevan 2009), (ii) ensure comparability, and (iii) replicate the evaluation protocols where possible (ISO/IEC 25062, 2006). This reporting approach proposes a systematic way to report data and the main factors which could differentiate one study from another, such as: contextual factors, analysis of performance, and methodological variables (Bevan 1995; Bevan et al. 2016). In addition, the CIF proposed in the ISO/IEC 25062 can be used as a checklist to ensure that all the required content is included in a report to enable results replication and facilitate comparative analysis (Bevan et al. 2016).

Unfortunately, ISO/IEC 25062 (2006) and its associated concept of a unified way to report data were not successfully adopted within the scientific community. We analysed the number of conference and peer-reviewed articles, which referenced the ISO 25062 in the last 10 years (databases: Scupus, IEEEXplore and Science Direct). Figure 2 shows that from 2007 to 2017 only 30 articles about product design and development have applied CIF or, at least, referenced the ISO/IEC 25062 to report their evaluation data.

The reason for such a low level of adoption in scientific literature of a unified way to report usability data for summative and formative studies (Theofanos and Quesenbery 2005; ISO IEC 25062, 2006) has to be further investigated. This may be related to the fact that the template for reporting is quite demanding, or it is perceived as very industry-oriented to be used or adapted for scientific reporting purposes. Nevertheless, the lack of success of CIF in the scientific literature resonates quite well with the indication that replication is rarely performed in usability research field, appearing only in 3.1% out of 891 reviewed studies (Hornbæk et al. 2014).

The CIF and its adaptations to report formative test (Theofanos and Quesenbery 2005) may offer an off-the-shelf tool which may be adapted to build a guideline on how to report, in scientific publications, a minimal set of information and data to enable replication and comparative analysis. This is not an uncommon practice in science. For instance in medical research practitioners are requested to report their studies in scientific publication by using common guidelines (see, for instance, CONSORT and COREQ guidelines for reporting: Moher, Schulz, and Altman 2001; Tong, Sainsbury, and Craig 2007). These standardised ways to report studies are applied to facilitate the communication among wide communities, and to enable replication and comparability of the studies in critical fields.

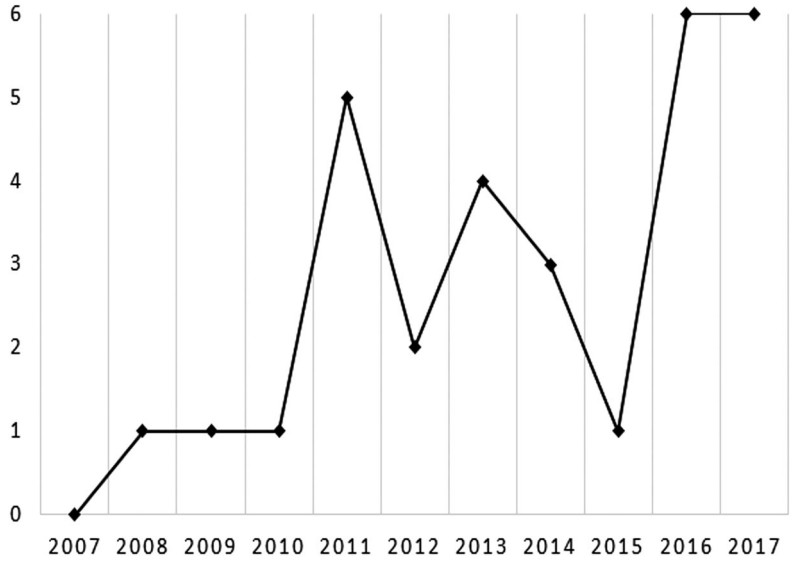

Figure 2. Number of publications (journal and conference articles) between 2007 and 2017 which referenced the ISO IEC 25062. Online databases: Scopus, IEEEXplore and Science Direct; keywords: Usability, ISO IEC 25062; ISO 25062; Boolean operator AND/OR.

We are proposing to adopt an approach similar to the one adopted in other field; for instance designing guidelines to report usability data for scientific publication (RUS) may substantially reduce the replicability issues. The success of such guidelines will require, however: (i) an international action of usability communities, and (ii) support from publishers to enforce the compliance to a unified way of summarising and presenting data in journals and conferences in tune with emerging initiatives of open data (e.g. https://www.epsrc.ac.uk/about/standards/researchdata/). In addition to promoting replicability and comparability, RUS will also enable better assessment of the quality of each usability study i.e. methods, variable and data and will provide enhanced information about differences among studies due to contextual factors and products' characteristics.

4. Conclusion

4.1. Why we should keep the usability construct in the era of user experience?

User experience (UX) proposes to focus the interaction evaluation on 'person's perceptions and responses resulting from the use and/or anticipated use of a product' (ISO 9241-210, 2010, 3). UX includes the dimensions of usability (Law and van Schaik 2010) and concurrently attempts to enlarge the assessment's factors with a focus on cognitive, aesthetics, and qualitative aspects of interaction measured throughout time (Tractinsky 1997; Hassenzahl 2005; Hassenzahl and Tractinsky 2006; Petrie and Bevan 2009; Borsci et al. 2013; Borsci et al. 2015; Borsci et al. 2016; Bussolon 2016). UX practices were built upon usability and interaction methodology, but also dealt with methodological challenges due to lack of consistent measures to reliably evaluate factors such as people expectation (anticipated use), emotional reactions, etc. (Bargas-Avila and Hornbæk 2011). From a historical point of view, UX represented an attempt of practitioners from different communities to go beyond the limits of traditional usability practice too focused on efficiency and effectiveness of work-related systems, contexts, and activities (Bargas-Avila and Hornbæk 2012). Nevertheless, it could be argued that building upon the strengths of usability practice, UX also inherited its limitations. In fact, fragmentation and heterogeneity of practices are recognised in literature as issues that affect the UX field (Lallemand, Gronier, and Koenig 2015).

Despite its limitations, the UX concept proposes an enlargement of the way of measuring the human–artefacts interaction by focusing the evaluation on the quality of the experience, which includes but it is not limited to the quality of interaction i.e. usability. This quite recent new concept had, in a short period of time, a pervasive impact and diffusion across different communities of practice. From that practitioners may derive two lessons. First, it is possible to bring together different communities to work together and to diffuse a new set of practices. Therefore, this effort is feasible, and it could be implemented to deal with the inherent challenges of the usability practice. Second, the legitimate attempt to react to the established practice of usability by shifting to a broader construct of UX is struggling to produce a harmonised and unified set of practice. In fact, UX was built to include and enlarge usability without a real effort to mitigate its limitations. This juxtaposition of

the UX concept over the one of usability resulted on a new construct of interaction built on unstable foundations.

We are arguing that usability, intended as the measure of quality of interaction, is the core of the UX, and it will be the key part of any emerging construct of interaction in the future. Therefore, to solve the fragmentation of knowledge and know-how and to find solution to harmonise the way of reporting evaluation data is an essential step to progress the entire interaction field and may strongly benefit also the UX field and any future paradigm of interaction evaluation. For this reason, we are proposing, and concluding the present paper with, a call to action to harmonise the usability practice across communities.

4.2. A call to action

This work presents an alternative perspective on: (i) knowledge fragmentation; and (ii) the unstable relationship among the usability metrics affecting the scientific community in the usability field.

The following two mitigation strategies were proposed to harmonise the complexity in the usability field:

- Our hypothesis is that the issue of knowledge fragmentation is not due to the weakness and lack of usefulness of the ISO 9241-11. Evidence suggests that the ISO 9241-11 provides a useful, solid, and flexible construct to assess different types of products. The umbrella definition serves as a basis to evaluate new (unregulated) products or it is adapted by product-specific standards to fit the specific product requirements. Knowledge fragmentation is originated by the fact that specialised communities refer to the same construct (ISO 9241-11) adapted by product-specific standards. Change or improve the construct will not enable a unified knowledge around usability. In fact, each specialised community will continue to seek adaptation (to enlarge or reduce the variables) of the construct to fit the product requirements and its context of use (Hertzum and Clemmensen 2012). To reconcile this fragmentation, we proposed the development of a meta-standard. This may provide a unified framework for all the experts independently from their community by also raising awareness across the communities about how usability is characterised, represented, and operationalised in other fields. The last 20 years of literature and adaptations of the ISO 9241-11 into product-specific standards may be considered a solid cumulated knowledge, which belongs to the industrial and scientific communities. Currently, a significant amount of practitioners and researchers are only aware of (and apply) a small amount of this knowledge i.e. the one related to their specific area of interest.
- Our second hypothesis is that the unstable relationship among the dimensions of ISO9241-11 is not due to weaknesses in the construct to convey adequate ways to represent and operationalise the usability evaluation. In fact, it is a direct consequence of the lack of adoption in the scientific field of a unified way of data reporting (Theofanos and Quesenbery 2005; ISO IEC 25062, 2006). We proposed the development of a RUS guideline to: (i) promote a unified way to report usability studies in scientific literature; and (ii) promote study replication and comparative analysis. This unified way of reporting has not to be intended as constraint, but as in other field it is only a way to enable replicability by appropriately present: methods, scenarios, (observed and unobserved) variables, limitations, etc.

We are not proposing that the meta-standard (first mitigation strategy) and the RUS (second mitigation strategy) are the only possible solutions to harmonise knowledge and know-how in the usability field without enforcing an unnecessary change of construct. However, we believe that these solutions (or the identification of alternative ones) are feasible only if different communities of practice pool together their efforts, as it happened for the definition and diffusion of the UX.

For instance, once that different community will agree on a common agenda to deal with the usability issues we discussed in the present paper and on how to operationalise (our or alternative) mitigation strategies it will be possible to perform in each community an Action Research – intended as an approach to involve multiple key player of a specific context/field in the resolution of issues and in the definition and test of the solutions (Argyris and Schön 1989; Marshall and Rossman 2014).

The real complexity of what we are proposing stands from the fact that we are calling for a worldwide action that has to involve as many as possible communities of practice and operators e.g. academics, industries, national and international associations, publishers etc. We do believe that the solutions we proposed could be considered a starting point for an international discussion among communities to define an agenda and to identify (or implement new) solutions to the usability issues that each community alone may not achieve or may fail to diffuse.

Only a large adhesion of multiple communities to an international agenda could enable different key players from different fields to: (i) work together to solve the inherent issues of usability practice, (ii) agree and realise

common solutions, and (iii) promote the use of those solution. The development of a common framework (Meta-standard and RUS) supported by a large number of communities may, indeed, guarantee an impact in terms of diffusion of appropriate solutions to harmonise the usability field, and this may also have significant benefit for emerging and unstable construct such as the one of UX.

Acknowledgements

We acknowledge that this work was inspired by the original contribution of Professor Noam Tractinsky. We thank this author for his ability to identify two pervasive issues in the usability field. We also extensively used and referred to the commentaries of other authors (Bardzell 2018; Bertelsen 2018; Hertzum 2018; Hornbæk 2018; Reeves 2018; Roto 2018; Stage 2018) to support our research of theoretical and formal evidence to identify alternative explanations and to define potential mitigation strategies for the issues. Finally, we would also like to acknowledge the precious work of Dr. Nigel Bevan, who recently passed away. The work of this author massively shaped our contribution in particular for aspects related to the CIF.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Simone Borsci http://orcid.org/0000-0002-3591-3577
Stefano Federici http://orcid.org/0000-0001-5681-0633
Alessio Malizia http://orcid.org/0000-0002-2601-7009

References

Argyris, Chris, and Donald A. Schön. 1989. "Participatory Action Research and Action Science Compared: A Commentary." *American Behavioral Scientist* 32 (5): 612–623. doi:10.1177/0002764289032005008.

Bardzell, Jeffrey. 2018. "Commentary: Appreciative Disagreements." *Human–Computer Interaction* 33 (2): 201–204. doi:10.1080/07370024.2017.1326826.

Bargas-Avila, Javier A., and Kasper Hornbæk. 2011. "Old Wine in New Bottles or Novel Challenges: A Critical Analysis of Empirical Studies of User Experience." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2689–2698. Vancouver, BC, Canada: ACM.

Bargas-Avila, Javier A., and Kasper Hornbæk. 2012. "Foci and Blind Spots in User Experience Research." *Interactions* 19 (6): 24–27.

Bertelsen, Olav W. 2018. "Commentary: Usability and the Primacy of Practice." *Human–Computer Interaction* 33 (2): 182–185. doi:10.1080/07370024.2017.1321991.

Bevan, Nigel. 1995. "Measuring Usability as Quality of Use." *Software Quality Journal* 4 (2): 115–130. doi:10.1007/bf00402715.

Bevan, Nigel. 2009. "International Standards for Usability Should be More Widely Used." *Journal of Usability Studies* 4 (3): 106–113.

Bevan, Nigel, Jim Carter, Jonathan Earthy, Thomas Geis, and Susan Harker. 2016. "New ISO Standards for Usability, Usability Reports and Usability Measures." In *Human-Computer Interaction. Theory, Design, Development and Practice: 18th International Conference, HCI International 2016, Toronto, ON, Canada, July 17–22, 2016. Proceedings, Part I*, edited by Masaaki Kurosu, 268–278. Cham: Springer International Publishing.

Borsci, Simone, Peter Buckle, and George B. Hanna. 2016. "Why you Need to Include Human Factors in Clinical and Empirical Studies of in Vitro Point of Care Devices? Review and Future Perspectives." *Expert Review of Medical Devices* 13 (4): 405–416. doi:10.1586/17434440.2016.1154277.

Borsci, Simone, Peter Buckle, Jeremy Huddy, Zenia Alaestante, Zhifang Ni, and George B. Hanna. 2017. "Usability Study of pH Strips for Nasogastric Tube Placement." *PLOS ONE* 12 (11): e0189013. doi:10.1371/journal.pone.0189013.

Borsci, Simone, Stefano Federici, Silvia Bacci, Michela Gnaldi, and Francesco Bartolucci. 2015. "Assessing User Satisfaction in the Era of User Experience: Comparison of the SUS, UMUX, and UMUX-LITE as a Function of Product Experience." *International Journal of Human–Computer Interaction* 31 (8): 484–495. doi:10.1080/10447318.2015.1064648.

Borsci, Simone, Jasna Kuljis, Julie Barnett, and Leandro Pecchia. 2016. "Beyond the User Preferences: Aligning the Prototype Design to the Users' Expectations." *Human Factors and Ergonomics in Manufacturing & Service Industries* 26: 16–39.

Borsci, Simone, Masaaki Kurosu, Stefano Federici, and Maria Laura Mele. 2013. *Computer Systems Experiences of Users with and Without Disabilities: An Evaluation Guide for Professionals*. Boca Raton, FL: CRC Press.

Borsci, Simone, Glyn Lawson, and Simon Broome. 2015. "Empirical Evidence, Evaluation Criteria and Challenges for the Effectiveness of Virtual and Mixed Reality Tools for Training Operators of car Service Maintenance." *Computers in Industry* 67: 17–26. https://doi.org/10.1016/j.compind.2014.12.002.

Borsci, Simone, Glyn Lawson, Bhavna Jha, Mark Burges, and Davide Salanitri. 2016. "Effectiveness of a Multidevice 3D Virtual Environment Application to Train car Service Maintenance Procedures." *Virtual Reality* 20 (1): 41–55. doi:10.1007/s10055-015-0281-5.

Brown, William, Po-Yin Yen, Marlene Rojas, and Rebecca Schnall. 2013. "Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for Evaluating Mobile Health (mHealth) Technology." *Journal of Biomedical Informatics* 46 (6): 1080–1087. doi:10.1016/j.jbi.2013.08.001.

Bussolon, Stefano. 2016. "The X Factor." Paper presented at the International Conference of Design, User Experience, and Usability.

Caronna, E. B., J. M. Milunsky, and H. Tager-Flusberg. 2008. "Autism Spectrum Disorders: Clinical and Research Frontiers." *Archives of Disease in Childhood* 93 (6): 518–523. doi:10.1136/adc.2006.115337.

Eagly, Alice H., and Shelly Chaiken. 2007. "The Advantages of an Inclusive Definition of Attitude." *Social Cognition* 25 (5): 582–602. doi:10.1521/soco.2007.25.5.582.

Federici, Stefano, Marco Bracalenti, Fabio Meloni, and Juan V. Luciano. 2017. "World Health Organization Disability Assessment Schedule 2.0: An International Systematic Review." *Disability and Rehabilitation* 39 (23): 2347–2380. doi:10.1080/09638288.2016.1223177.

Federici, Stefano, Fabio Meloni, and Simone Borsci. 2016. "The Abandonment of Assistive Technology in Italy: A Survey of Users of the National Health Service." *European Journal of Physical and Rehabilitation Medicine* 52 (4): 516–526.

Federici, Stefano, and Marcia Scherer. 2017. *Assistive Technology Assessment Handbook*. 2nd ed. Boca Raton, FL: CRC Press.

Federoff, Melissa A. 2002. "Heuristics and Usability Guidelines for the Creation and Evaluation of Fun in Video Games." PhD dissertation, Indiana University.

Frøkjær, Erik, Morten Hertzum, and Kasper Hornbæk. 2000. Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated?" Paper presented at the Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI '00). ACM, New York, NY, USA, 345–352. doi:10.1145/332040.332455.

Green, Lawrence W., and Shawna L. Mercer. 2001. "Can Public Health Researchers and Agencies Reconcile the Push From Funding Bodies and the Pull From Communities?" *American Journal of Public Health* 91 (12): 1926–1929. doi:10.2105/ajph.91.12.1926.

Hartson, H. Rex, Terence S. Andre, and Robert C. Williges. 2001. "Criteria for Evaluating Usability Evaluation Methods." *International Journal of Human-Computer Interaction* 13 (4): 373–410.

Hassenzahl, Marc. 2005. "The Thing and I: Understanding the Relationship Between User and Product." In *Funology: From Usability to Enjoyment*, edited by Mark Blythe, Kees Overbeeke, Andrew Monk, and Peter Wright, 31–42. Berlin, DE: Springer.

Hassenzahl, Marc, and Noam Tractinsky. 2006. "User Experience – A Research Agenda." *Behaviour & Information Technology* 25 (2): 91–97. doi:10.1080/01449290500330331.

Hertzum, Morten. 2010. "Images of Usability." *International Journal of Human–Computer Interaction* 26 (6): 567–600. doi:10.1080/10447311003781300.

Hertzum, Morten. 2018. "Commentary: Usability – A Sensitizing Concept." *Human–Computer Interaction* 33 (2): 178–181. doi:10.1080/07370024.2017.1302800.

Hertzum, Morten, and Torkil Clemmensen. 2012. "How do Usability Professionals Construe Usability?" *International Journal of Human-Computer Studies* 70 (1): 26–42. doi:10.1016/j.ijhcs.2011.08.001.

Hertzum, Morten, and Niels Ebbe Jacobsen. 2003. "The Evaluator Effect: A Chilling Fact about Usability Evaluation Methods." *International Journal of Human-Computer Interaction* 15 (1): 183–204.

Hirsch, Paul M, and Daniel Z Levin. 1999. "Umbrella Advocates Versus Validity Police: A Life-Cycle Model." *Organization Science* 10 (2): 199–212.

Hornbæk, Kasper. 2006. "Current Practice in Measuring Usability: Challenges to Usability Studies and Research." *International Journal of Human-Computer Studies* 64 (2): 79–102.

Hornbæk, Kasper. 2018. "Commentary: Usability and Theory Building." *Human–Computer Interaction* 33 (2): 186–189. doi:10.1080/07370024.2017.1321992.

Hornbæk, Kasper, and Effie Lai-Chong Law. 2007. "Meta-analysis of Correlations among Usability Measures." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 617–626. San Jose, California, USA: ACM.

Hornbæk, Kasper, Søren S. Sander, Javier Andrés Bargas-Avila, and Jakob Grue Simonsen. 2014. Is Once Enough?: On the Extent and Content of Replications in Human-computer Interaction. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14). ACM, New York, NY, USA, 3523–3532. doi:10.1145/2556288.2557004.

Huddy, Jeremy R., Melody Ni, Stella Mavroveli, James Barlow, Doris-Ann Williams, and George B. Hanna. 2015. "A Research Protocol for Developing a Point-Of-Care Key Evidence Tool 'POCKET': A Checklist for Multidimensional Evidence Reporting on Point-of-Care in Vitro Diagnostics." *BMJ Open* 5 (7): e007840. doi:10.1136/bmjopen-2015-007840.

IEC. 2015. "IEC 62366-1:2015 Medical Devices – Part 1: Application of Usability Engineering to Medical Devices." In. Brussels, BE: CEN.

IEC/DIS. under development. "80601-2-77 Medical Electrical Equipment – Part 2-77: Particular requirements for the basic safety and essential performance of medical robots for surgery."

ISO. 1998. "ISO 9241-11:1998 Ergonomic Requirements for Office Work with Visual Display Terminals – Part 11: Guidance on Usability." In. Brussels, BE: CEN.

ISO. 2006. "ISO 20282-1:2006 Ease of Operation of Everyday Products – Part 1: Design Requirements for Context of Use and User Characteristics." In. Brussels, BE: CEN.

ISO. 2007. "ISO 14971:2007 Medical Devices – Application of Risk Management to Medical Devices." In. Brussels, BE: CEN.

ISO. 2010. "ISO 9241-210:2010 Ergonomics of Human-System Interaction – Part 210: Human-centred Design for Interactive Systems." In. Brussels, BE: CEN.

ISO. 2016. "ISO/TS 15066 – Robots and Robotic Devices – Collaborative Robots." In. Brussels, BE: CEN.

ISO/IEC. 2006. "ISO/IEC 25062 Software Engineering – Software Product Quality Requirements and Evaluation (SQuaRE) – Common Industry Format (CIF) for Usability Test Reports." In. Brussels, BE: CEN.

ISO/IEC. 2011. "ISO/IEC 25010:2011 Systems and Software Engineering – Systems and Software Quality Requirements and Evaluation (SQuaRE) – System and Software Quality Models." In. Brussels, BE: CEN.

Jette, Alan M. 2006. "Toward a Common Language for Function, Disability, and Health." *Physical Therapy* 86 (5): 726–734.

Kiesler, Sara, and Pamela Hinds. 2004. "Introduction to This Special Issue on Human-Robot Interaction." *Human–Computer Interaction* 19 (1–2): 1–8. doi:10.1080/07370024.2004.9667337.

Kushniruk, Andre W., and Vimla L. Patel. 2004. "Cognitive and Usability Engineering Methods for the Evaluation of Clinical Information Systems." *Journal of Biomedical Informatics* 37 (1): 56–76. doi:10.1016/j.jbi.2004.01.003.

Lallemand, Carine, Guillaume Gronier, and Vincent Koenig. 2015. "User Experience: A Concept Without Consensus? Exploring Practitioners' Perspectives Through an International Survey." *Computers in Human Behavior* 43: 35–48. doi:10.1016/j.chb.2014.10.048.

Law, Effie L. C., and Paul van Schaik. 2010. "Modelling User Experience – An Agenda for Research and Practice." *Interacting with Computers* 22 (5): 313–322. doi:10.1016/j.intcom.2010.04.006.

Lewis, James R. 2014. "Usability: Lessons Learned … and Yet to Be Learned." *International Journal of Human–Computer Interaction* 30 (9): 663–684. doi:10.1080/10447318.2014.930311.

Liljegren, Erik. 2006. "Usability in a Medical Technology Context Assessment of Methods for Usability Evaluation of Medical Equipment." *International Journal of Industrial Ergonomics* 36 (4): 345–352.

Lowe, Grainne, Virginia Plummer, Anthony Paul O'Brien, and Leanne Boyd. 2012. "Time to Clarify – the Value of Advanced Practice Nursing Roles in Health Care." *Journal of Advanced Nursing* 68 (3): 677–685.

Marshall, Catherine, and Gretchen B. Rossman. 2014. *Designing Qualitative Research*. 6th edition. London: Sage.

Moher, David, Kenneth F. Schulz, and Douglas G. Altman. 2001. "The CONSORT Statement: Revised Recommendations for Improving the Quality of Reports of Parallel-Group Randomised Trials." *The Lancet* 357 (9263): 1191–1194. doi:10.1016/S0140-6736(00)04337-3.

Petrie, Helen, and Nigel Bevan. 2009. "The Evaluation of Accessibility, Usability, and User Experience." In *The Universal Access Handbook*, edited by Constantine Stephanidis, 299–314. Boca Raton, FL: CRC Press.

Pinelle, David, Nelson Wong, Tadeusz Stach, and Carl Gutwin. 2009. Usability Heuristics for Networked Multiplayer Games. In Proceedings of the ACM 2009 international conference on Supporting group work (GROUP '09). ACM, New York, NY, USA, 169–178. doi:10.1145/1531674.1531700.

Polisena, Julie, Rossana Castaldo, Oriana Ciani, Carlo Federici, Simone Borsci, Matteo Ritrovato, Daniel Clark, and Leandro Pecchia. 2018. "Health Technology Assessment Methods Guidelines for Medical Devices: How Can We Address the Gaps? The International Federation of Medical and Biological Engineering Perspective." *International Journal of Technology Assessment in Health Care* 34 (3): 276–289. doi:10.1017/S0266462318000314.

Quesenbery, Whitney. 2003. *The Five Dimensions of Usability. Vol. 20*. Mahwah, NJ: Lawrence Erlbaum Associates.

Reeves, Stuart. 2018. "Commentary: Usability in Vivo." *Human–Computer Interaction* 33 (2): 190–194. doi:10.1080/07370024.2017.1324306.

Roberts, Harry W, James Myerscough, Simone Borsci, Melody Ni, and David P S O'Brart. 2017. "Time and Motion Studies of National Health Service Cataract Theatre Lists to Determine Strategies to Improve Efficiency." *British Journal of Ophthalmology*, doi:10.1136/bjophthalmol-2017-310452.

Roberts, Harry W., Vijay K. Wagh, Isabella J. M. Mullens, Simone Borsci, Melody Z. Ni, and David P. S. O'Brart. 2018. "Evaluation of a Hub-and-Spoke Model for the Delivery of Femtosecond Laser-Assisted Cataract Surgery Within the Context of a Large Randomised Controlled Trial." *British Journal of Ophthalmology* 102 (11): 1556–1563.

Roto, Virpi. 2018. "Commentary: Actionable Constructs for Practitioners." *Human–Computer Interaction* 33 (2): 198–200. doi:10.1080/07370024.2017.1325326.

Sauro, Jeff, and James R. Lewis. 2009. Correlations Among Prototypical Usability Metrics: evidence for the Construct of Usability. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09). ACM, New York, NY, USA, 1609–1618. doi:10.1145/1518701.1518947.

Stage, Jan. 2018. "Commentary: Usability in Theory and Practice." *Human–Computer Interaction* 33 (2): 195–197. doi:10.1080/07370024.2017.1333908.

Stolterman, Erik. 2008. "The Nature of Design Practice and Implications for Interaction Design Research." *International Journal of Design* 2 (1): 55–65.

Strober, Myra H. 2011. *Interdisciplinary Conversations Challenging Habits of Thought, Stanford, California*. Stanford, CA: Stanford University Press.

Theofanos, Mary, and Whitney Quesenbery. 2005. "Towards the Design of Effective Formative Test Reports." *Journal of Usability Studies* 1 (1): 27–45.

Tong, Allison, Peter Sainsbury, and Jonathan Craig. 2007. "Consolidated Criteria for Reporting Qualitative Research (COREQ): A 32-Item Checklist for Interviews and Focus Groups." *International Journal for Quality in Health Care* 19 (6): 349–357. doi:10.1093/intqhc/mzm042.

Tractinsky, Noam. 1997. "Aesthetics and Apparent Usability: Empirically Assessing Cultural and Methodological Issues." Paper presented at the Conference on Human Factors in Computing Systems: CHI '97, Atlanta, GA, Mar 22–27.

Tractinsky, Noam. 2018. "The Usability Construct: A Dead End?" *Human–Computer Interaction* 33 (2): 131–177. doi:10.1080/07370024.2017.1298038.

Vandenabeele, Wouter, Sarah Scheepers, and Annie Hondeghem. 2006. "Public Service Motivation in an International Comparative Perspective: The UK and Germany." *Public Policy and Administration* 21 (1): 13–31. doi:10.1177/095207670602100102.

Warren, Scott, Greg Jones, and Lin Lin. 2011. "Usability and Play Testing." In *Serious Educational Game Assessment: Practical Methods and Models for Educational Games, Simulations and Virtual*, edited by Annetta Leonard, and Bronack Stephen, 131–146. Rotterdam: Sense Publishers. doi:10.1007/978-94-6091-329-7_8.

WHO (World Health Organization). 2001. *ICF: International Classification of Functioning, Disability and Health*. Geneva, CH: WHO.

WHO (World Health Organization), and World Bank. 2011. "World Report on Disability." In Geneva, CH: World Health Organization.

Wilson, Max L.L., Paul Resnick, David Coyle, and Ed H. Chi. 2013. "RepliCHI: the Workshop." In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, 3159–3162. Paris: ACM.

Wilson, John R, and Sarah Sharples. 2015. *Evaluation of Human Work*. 4th. London: CRC press.

Yanco, Holly A., Jill L. Drury, and Jean Scholtz. 2004. "Beyond Usability Evaluation: Analysis of Human-Robot Interaction at a Major Robotics Competition." *Human–Computer Interaction* 19 (1): 117–149. doi:10.1207/s15327051hci1901&2_6.