

Medição de Usabilidade e Eficiência de Dashboards: Framework Desenvolvida usando os Princípios de Desenho de Business Intelligence e Interação Homem-Máquina

Measuring Usability and Dashboards Efficiency: Framework Developed using the Principles of Design of Business Intelligence and Human Machine Interaction

Miguel Oliveira, ISCTE – Instituto Universitário de Lisboa, Portugal, mhdoa@iscte.pt

Elsa Cardoso, ISCTE – Instituto Universitário de Lisboa e INESC-ID, Portugal,
elsa.cardoso@iscte.pt

Marina Santana, Novabase Business Solutions, Portugal, marina.santana@novabase.pt

Resumo

A usabilidade e eficiência no processo de decisão são aspetos essenciais em interfaces de *Business Intelligence* (BI). Algumas organizações afirmam construir *dashboards* com estas características, mas sem forma efetiva de comprová-lo. Sendo uma temática pouco investigada em BI, é contudo largamente estudada na área científica de Interação Homem- Máquina (IHM). Este artigo apresenta o projeto de desenho e desenvolvimento de uma *framework* para medição da usabilidade e eficiência de *dashboards* utilizando os princípios de desenho de BI e IHM, com adaptação da ISO/IEC 9241-11. Este projeto académico foi realizado na empresa Portuguesa Novabase, na competência de *Business Analytics* (BA). Para a validação da *framework* recorreu-se a profissionais da Novabase nas duas áreas, a professores de IHM e a estudantes com conhecimentos nas duas áreas, observando-se diferenças significativas de usabilidade consoante o número de erros introduzidos e tipo de avaliadores.

Palavras chave: *business intelligence*, usabilidade, eficiência, *dashboards*

Abstract

The usability and efficiency in decision-making are essential aspects in business intelligence interfaces (BI). Some organizations claim to build dashboards with these characteristics, but have no effective way of proving it. Being a little researched topic in the BI field, it is nevertheless widely studied in the scientific area of Human- Machine Interaction (HMI). This paper presents the project of design and development of a framework for the measurement of dashboards' usability and efficiency using the principles of BI and HMI design, with adaptation of ISO / IEC 9241-11. This academic project was carried out in the Portuguese company Novabase, the competence of Business Analytics (BA). The framework was validated by Novabase professionals in both areas, and HMI teachers and students with knowledge in both areas, observing significant differences in usability depending on the number of errors introduced and type of evaluators.

Keywords: *business intelligence*, usability, efficiency, *dashboards*

1. INTRODUÇÃO

Os *dashboards* são atualmente uma ferramenta indispensável, sendo utilizadas pelos gestores para monitorizar a performance e decidir o rumo das organizações. Um *dashboard* é uma aplicação de *Business Intelligence* (BI), consolidada num único écran, que permite a visualização rápida da informação mais relevante para atingir um ou mais objectivos [Few 2013]. Um *dashboard* é uma ferramenta de comunicação que permite retirar conhecimento e conclusões de uma forma eficiente, apenas com um simples olhar.

Inicialmente os *dashboards* destinavam-se exclusivamente a pessoas com funções de chefia. Mas hoje em dia, os *dashboards* – devido às suas vantagens – são utilizados em vários níveis, permitindo coordenar o esforço dos colaboradores no cumprimento de objetivos de gestão na ótica da visão da organização [Eckerson 2010]. Contudo, nem sempre estes resultados são alcançados por falta de usabilidade e eficiência de algumas interfaces. Às vezes, os utilizadores rejeitam-nas por não cumprirem os requisitos ou por não serem fáceis de utilizar [Fonseca et al. 2012] e [White 2005], com conseqüente perda de valor dos projetos [Rudin 2007], e perda de eficiência na comunicação e desempenho dos negócios [Few 2006]. Todavia, seguindo as heurísticas de desenho, é possível desenvolver interfaces com maior usabilidade [Few 2006], [Fonseca et al. 2012] e [Nielsen 1993]. A área de Interação Homem-Máquina (IHM) estuda a usabilidade e eficiência de interfaces, podendo-se adaptar metodologias e métodos de avaliação para BI, identificando problemas existentes e possíveis soluções.

Durante este estudo questionou-se “Como medir a usabilidade e a eficiência de uma interface em *Business Intelligence*?” Desenvolveu-se uma *framework* adaptada à ISO/IEC 9241-11 para medir a usabilidade de um *dashboard*, utilizando heurísticas de desenho, aplicável em contexto organizacional, pretendendo-se que seja fácil e prática. Pretende-se testar quatro hipóteses relacionadas com a análise da usabilidade percebida em função da existência e do tipo de erros, do tipo de avaliador e do tipo de boas práticas.

Este artigo está estruturado em seis secções. Na secção 2 discutem-se as boas práticas de avaliação de usabilidade em BI e IHM. A *framework* desenvolvida para medir a usabilidade e eficiência de *dashboards* é apresentada na secção 3. O processo de validação da *framework* e a análise dos resultados são apresentados nas secções 4 e 5, respectivamente. Por fim, na secção 6 apresentam-se as conclusões e reflete-se sobre o trabalho futuro.

2. AVALIAÇÃO DE USABILIDADE

A definição de usabilidade não é consensual na literatura, como é demonstrado na Tabela 1. A usabilidade depende do propósito e dos requisitos dos *stakeholders*. Contudo, produtos com

usabilidade permitem poupar dinheiro, aumentar a qualidade e a satisfação do cliente [Fonseca et al. 2012]. A usabilidade deve ser considerada durante o processo de desenvolvimento, requerendo o envolvimento e participação dos *stakeholders* em todas as fases.

Boas práticas de desenho de *dashboards* em BI

Few [2010] afirma que o design do *dashboard* não é o único factor a considerar para este ser aceite. Mesmo que o *dashboard* seja concebido usando as boas práticas de desenho, se não satisfizer os requisitos pré-estabelecidos, o cliente rejeita-o. Few [2006] refere que um bom design destaca a informação essencial para a monitorização (*data-pixels*), retira a ênfase aos elementos auxiliares para mostrar a informação (*non-data pixels*) e possui simplicidade, mostrando-se claro, organizado e simples.

	BI	IHM		
	Few 2006	ISO/IEC 9241-11 1998	Nielsen 1993	Norman 1988
O que é usabilidade?	Agradabilidade estética; deteção de informação fora da norma; facilidade de uso; simplicidade no desenho; suportar as tarefas do utilizador	Eficácia; eficiência; satisfação	Eficiência; erros; facilidade de aprendizagem, de lembrança; satisfação	Facilidade de aprendizagem; eficácia no cumprimento do requisito (intenção vs. ação; percepção vs. interpretação)
Como atingi-la / medi-la?	Dar sentido e uso efetivo à informação; garantir agradabilidade estética; manter a consistência para uma melhor interpretação; dar uma visão global; testar o desenho com os utilizadores	Definição dos requisitos de usabilidade, indicadores para medir cada critério, metas para cumprir os indicadores e avaliar o desempenho face à meta	Definição dos requisitos de usabilidade, indicadores para medir cada critério, metas para cumprir os indicadores e avaliar o desempenho face à meta	Definição dum bom modelo concetual e uma filosofia baseada no utilizador para reduzir as discrepâncias entre os ciclos

Tabela 1 – Comparação usabilidade BI vs. IHM

Few [2006] identifica seis fatores que conjugados conferem usabilidade a um *dashboard*: (1) deve ser agradável ao utilizador, (2) deve promover a simplicidade na comunicação e exposição da informação; (3) deve ser fácil de usar, recorrendo a meios conhecidos pelo utilizador, (4) garantindo a consistência dos mesmos, exceto quando se pretende destacar certa informação; (5) deve fornecer informação necessária para detetar fenómenos fora da norma, auxiliando a tarefa do utilizador; e (6) o desenho deve ser mostrado aos utilizadores para ser aceite e validado.

Boas práticas de desenho de interfaces em IHM

Nielsen [1994] refere existirem diversos conjuntos de heurísticas utilizáveis no desenho de interfaces, por exemplo, os princípios de Norman [1988] e as heurísticas de Molich e Nielsen [1990]. Todavia, segundo Nielsen [1993] alguns conjuntos são apenas formalismos ou demasiado específicos e complexos ou extensivos para ser usados.

Fonseca et al. [2012] e Nielsen [1993] afirmam que uma boa interface possui duas características: eficácia no cumprimento das tarefas pré-definidas e atratividade para o utilizador, sublinhando que os elementos exibidos devem ser claros e explícitos (“*less is more*”), e a interface, eficiente e flexível para os utilizadores recorrentes.

Para agilizar e simplificar o desenvolvimento, Nielsen [1993] recomenda o uso de protótipos facilitadores da criação do modelo conceptual¹ e da obtenção de *feedback* em fases iniciais.

A norma ISO/IEC 9241-11 [1998] define usabilidade através de três dimensões: eficácia (exatidão e plenitude), eficiência (recursos necessários) e satisfação (atitudes e conforto). Utilizando uma *framework* composta por três elementos, avalia-a considerando o cumprimento das metas estabelecidas pelos requisitos. Consequentemente, identifica-se o contexto de uso e o produto tendo em conta o utilizador, o equipamento e a envolvência do sistema no meio. Deve ainda detalhar-se o resultado esperado, global ou específico (funcionalidade/tarefa), bem como escolher por cada requisito, pelo menos, três indicadores operacionais, um por dimensão, e o seu valor mínimo aceitável [Jokela et al. 2003].

Nielsen [1993] considera que usabilidade é uma componente da aceitação do sistema, medindo o grau de cumprimento dos requisitos estipulados pelos *stakeholders*, e a usabilidade, o desempenho dos mesmos. O autor propõe medir a usabilidade através de cinco fatores: (1) eficiência (produtividade); (2) capacidade de prevenção de erros (entre os requisitos e as ações tomadas); (3) aprendizagem (desempenho numa tarefa, utilizador inexperiente vs. experiente); (4) memorização (facilidade de reutilização sem reaprendizagem integral); e (5) satisfação (agradabilidade).

Norman [2002] decompõe usabilidade em facilidade de uso e eficácia no cumprimento da meta. O autor através de um modelo, as setes fases da ação, explica o funcionamento das ações dos seres humanos, recorrendo a dois ciclos interligados (execução e avaliação). Identificado o requisito, surge a necessidade de o satisfazer, iniciando-se, então, o ciclo de execução onde são planeadas as ações necessárias e a sua execução. Por oposição, o ciclo de avaliação inicia-se com a perceção de algo no meio que é interpretado e avaliado, daí resultando novos requisitos a serem satisfeitos.

¹ Modelo concetual ou de designer corresponde à compreensão que este possui do sistema, da sua organização e função [Fonseca et al. 2012] e [Johnson e Henderson 2002]

Assim, a falta de usabilidade corresponde a discrepâncias nos ciclos, i.e., a intenção de agir, diverge da ação tomada e a percepção do meio difere da sua avaliação

Outros estudos

Embora existam estudos que medem a usabilidade de uma interface ou avaliam a aplicação das boas práticas num *dashboard*, não foi encontrado nenhum estudo aplicado a BI que junte as duas componentes (i.e., BI e IHM). Green e Pearson [2006] e Joo, Lin, e Lu [2011] adaptaram a ISO/IEC 9241-11 para medir a usabilidade, respectivamente, de um *website* de comércio electrónico e de uma biblioteca, recorrendo a utilizadores que avaliaram subjetivamente o cumprimento dos requisitos. Lourenço [2013] desenvolveu uma *framework* que analisa o uso dos princípios de percepção visual no desenho de *dashboards* através de um conjunto de heurísticas.

3. FRAMEWORK PARA MEDIÇÃO DA USABILIDADE E EFICIÊNCIA DE DASHBOARDS

A *framework* para medição da usabilidade e eficiência de *dashboards* (com o acrónimo FMUED) foi desenvolvida no âmbito de uma dissertação de mestrado em Informática e Gestão do ISCTE – Instituto Universitário de Lisboa (ISCTE-IUL), em contexto empresarial na Novabase. O trabalho foi supervisionado por um *steering committee* composto por um orientador académico e dois consultores seniores de BI da Novabase. Com este trabalho, a empresa pretende disseminar internamente as boas práticas de desenho de *dashboards* a todos os colaboradores envolvidos no desenho de interfaces de BI, esperando obter no futuro uma solução que possa ser usada em projetos com clientes. Investigadas as soluções para medir a usabilidade de interfaces, decidiu-se adaptar a ISO/IEC 9241-11 para *dashboards*, dada a sua aceitação pela comunidade. Conceptualmente, a *framework* desenvolvida possui duas componentes: as heurísticas de desenho (organizadas em sete categorias) e o método de avaliação (ver Figura 1).

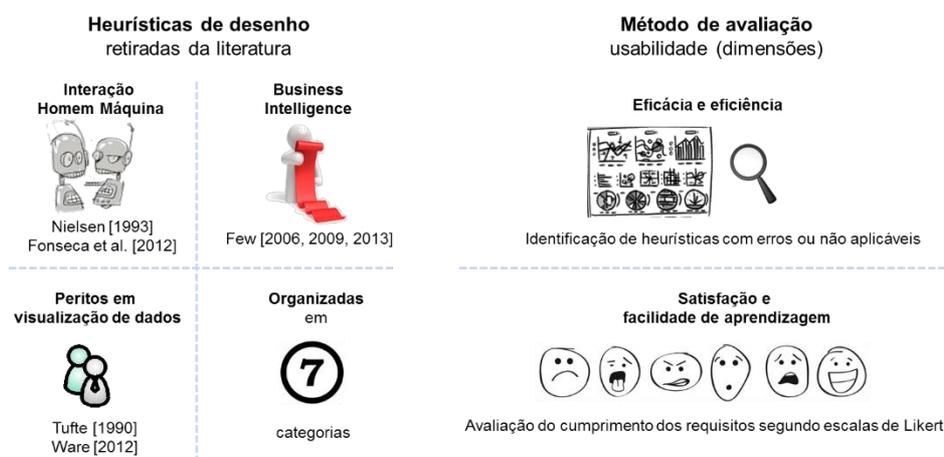


Figura 1 – Modelo conceptual da *framework* FMUED

Few [2006] identifica seis fatores que conjugados conferem usabilidade a um *dashboard*: (1) deve ser agradável ao utilizador, (2) deve promover a simplicidade na comunicação e exposição da informação; (3) deve ser fácil de usar, recorrendo a meios conhecidos pelo utilizador, (4) garantindo a consistência dos mesmos, exceto quando se pretende destacar certa informação; (5) deve fornecer informação necessária para detetar fenómenos fora da norma, auxiliando a tarefa do utilizador; e (6) o desenho deve ser mostrado aos utilizadores para ser aceite e validado.

Embora a norma delimite o contexto de uso do produto/interface, este não está ajustado para *dashboards*. O designer deve assim considerar os seguintes aspetos: utilizadores (proficiência em visualização de dados, experiência em *dashboards*, etc.); tarefa (propósito do *dashboard*); equipamento (*software* e *hardware*) e meio (frequência de atualização, departamento do utilizador, ligação a outras ferramentas).

A ISO/IEC 9241-11 operacionaliza os requisitos através da definição de indicadores. Em *dashboards*, o requisito é considerado completo quando é observável ou aceite pelo cliente.

Assim, decidiu-se remover a obrigatoriedade de determinar o tipo de requisito, os indicadores subjacentes e as metas para validá-lo, bastando definir os requisitos e categorizá-los de acordo com as dimensões de usabilidade da *framework*. Uma simplificação semelhante foi feita por Green e Pearson [2006] e Joo et al. [2011] para facilitar a avaliação.

Segundo a norma, usabilidade é composta por eficácia, eficiência e satisfação. Porém, a facilidade de aprendizagem do *dashboard* é também um fator relevante [Few 2006]. Consequentemente, adotou-se a facilidade de aprendizagem definida por Nielsen [1993], como uma dimensão para definir usabilidade. Sendo assim, as dimensões de usabilidade consideradas na *framework* FMUED são: **eficiência** (produtividade e possíveis análises); **eficácia** (funcionalidades, extração e qualidade da informação); **satisfação** (agradabilidade ao utilizador) e **facilidade de aprendizagem** (capacidade de compreensão).

Os requisitos pertencentes às dimensões de eficácia e eficiência devem ter representação direta e delimitada no *dashboard* (e.g., zonas), contrastando com os outros requisitos que são avaliados globalmente. Para facilitar a extração dos requisitos e aplicar uma abordagem que envolva os utilizadores no design, devem ser usados mock-ups no desenvolvimento de protótipos do *dashboard*.

Embora haja outros métodos para avaliar o desenho da interface, segundo Nielsen [1993] a avaliação por heurísticas é uma solução que requer baixos recursos e permite obter resultados rapidamente. Contudo, exige que os avaliadores tenham alguns conhecimentos em boas práticas

de desenho. A *framework* FMUED deve pois ser utilizada por especialistas no desenvolvimento de interfaces de BI.

Categorias	Tipo	Heurísticas
C1 - Considere quais as cores que escolhe	G	Definição de 2 paletes (standard e ênfase)
	G	Número de cores escolhidas (≤ 10)
	G	Uso de cores na natureza ou na roda das cores
	S	Consideração pelo daltonismo, cultura e contexto da cor
C2 - Considere a tipografia que usa	S	Escolha de cores no fundo (baças, não brilhantes e sem gradientes)
	G	Tipo de letra legível
	S	Uso preferencial de texto na horizontal
	S	Não usar todo o texto em letra maiúscula
C3 - Destaque a informação importante e promova a simplicidade	S	Uso apenas de dois tipos de letra distintos
	S	Uso de formas de destaque (cor, forma, movimento, etc.)
	G	Definição de zonas de ênfase no <i>dashboard</i>
	G	Organização da informação segundo um processo, desejo do utilizador ou lógica
C4 - Procure a consistência e considere perícia do utilizador	S	Colocar casas decimais ou unidades apenas se necessário
	S	Uso de alertas para destacar informação
	G	Repetição de cores, gráficos, etc.
	S	Utilização do tipo de gráfico mais correto para mostrar a informação
C5 - Forneça feedback do sistema e use uma linguagem compreensível pelo utilizador	S	Quebra da consistência apenas com o propósito de destacar
	S	Tipo de análise de acordo com a perícia do utilizador
	G	Permitir o <i>drill-down</i> / filtrar data / selecionar específicos <i>datasets</i>
	G	Uso de uma linguagem simples, compreensível e simpática
C6 - Reduza o ruído visual ao não enfatizar sem necessidade	G	Fornecer resposta face às ações do utilizador
	G	Mostrar o estado do <i>dashboard</i> (última atualização, online ou <i>frozen</i>)
	S	Desativar opções após uma ação (sombrear botões, usar opções <i>drop-down</i>)
	G	Existência de um botão de ajuda visível
C7 - Promova o reconhecimento em vez da lembrança	S	Limitar o número de <i>datasets</i> por gráfico
	S	Utilizar gráficos em 2-D
	S	Uso de linhas com grossura mínima e com cores clara e não saturadas
	G	Existência apenas de decorações ou embelezamentos necessários
C7 - Promova o reconhecimento em vez da lembrança	G	Correta definição e delimitação de blocos de informação
	G	Consegue-se ver o <i>dashboard</i> sem utilizar o <i>scroll</i>
	G	Uso de menus ou ícones para representar ações

Tabela 2 - Heurísticas utilizadas no estudo

Heurísticas de desenho

A *framework* FMUED inclui 32 boas práticas de desenho de interfaces (ver Tabela 2), classificadas em sete categorias (C1 a C7), de forma a facilitar a sua compreensão e evidenciar as áreas de análise. As boas práticas foram compiladas a partir da literatura existente nas áreas de desenho de *dashboards* em BI [Few 2006, 2009 e 2013], desenho de interfaces em IHM [Fonseca et al. 2012] e [Nielsen 1993], e visualização de dados [Tufte 1990] e [Ware 2012]. Para cada heurística foi definido o seu tipo, explicitando o nível da avaliação: global (validação do cumprimento em todo o *dashboard*) ou específico (em uma ou mais zonas do *dashboard*).

Método de avaliação

A ISO/IEC 9241-11 recomenda a avaliação da performance dos indicadores face ao *target*, sem contudo especificar um método de avaliação. A *framework* FMUED requer a avaliação dos requisitos explícitos através da identificação de erros nas heurísticas e o uso de escalas de *likert* para avaliar o cumprimento dos requisitos de facilidade de aprendizagem e satisfação.

Quando é identificado um erro numa boa prática do tipo específico, o avaliador deve indicar as zonas do *dashboard* onde foi mal aplicada. Se o erro for numa boa prática do tipo global, apenas deve assinalá-la. Nas situações em que não há informação suficiente para a avaliação ou quando a boa prática não se aplica ao *dashboard*, deve ser apenas assinalada como não aplicável. A usabilidade na FMUED é medida utilizando as fórmulas da Tabela 3, retirando-se previamente as boas práticas não aplicáveis.

Fórmulas de cálculo Usabilidade

$$\text{Usabilidade Total} = \frac{\sum \text{Usabilidade por dimensão}}{\text{Número de dimensões}}$$

$$\text{Usabilidade Dimensões Eficácia e Eficiência} = \frac{\text{Usabilidade por requisito explícito}}{\text{Número de requisitos}}$$

$$\text{Usabilidade por Requisito explícito} = \frac{\sum \text{Usabilidade por heurística}}{\text{Nº de heurísticas específicas} + \text{Nº de heurísticas globais} * \text{NIZ}}$$

$$\text{Usabilidade por Heurística} = \begin{cases} \text{Global} = \text{NIZ} \\ \text{Específica} = \frac{\text{NIZ com erro}}{\text{NIZ}} \end{cases}$$

$$\text{Usabilidade Dimensões Facilidade de aprendizagem e Satisfação} = \frac{\sum \text{valor da escala de likert}}{\text{Valor máximo da escala} * \text{Número de requisitos}}$$

Tabela 3 – Fórmulas de cálculo da usabilidade Legenda: NIZ = Número de zonas identificadas

4. VALIDAÇÃO DA FRAMEWORK

Para observar o desempenho do método de avaliação e a aceitação das heurísticas, desenhou-se uma experiência aplicada a um caso de estudo, onde para além do *dashboard* original foram desenhadas três versões suplementares com erros introduzidos. Os erros foram selecionados a partir das heurísticas (Tabela 2), variando apenas nas categorias e no número de erros introduzidos, resultando dois *dashboards* com cinco e um com sete erros, culminando em três situações de avaliação: zero, cinco e sete erros.

Caso de estudo

Optou-se por utilizar no processo de validação da FMUED num caso real, desenvolvido num projeto de Energia da Novabase, no controlo de perdas energéticas da organização que fornece energia em Cabo Verde. O objetivo do *dashboard* é a monitorização das perdas energéticas nas diversas ilhas.

A Figura 2 ilustra a distribuição energética, existindo uma central de produção (P), que produz e distribui pelos postos de transformação (PT) nas diversas ilhas, em média tensão (M.T.), sendo convertida em baixa tensão (B.T.), para abastecer a iluminação pública e os clientes. Contudo, durante a distribuição podem ocorrer dois tipos de perdas: (1) as de origem técnica, dispersão

por transferência na rede (11%), e.g., P para o PT; (2) as perdas não técnicas/fraude, devido a ligações à rede, não permitidas, para obtenção de energia sem custo. Assim, fluxos não previstos originam interrupções/sobrecargas na rede resultando na quebra de abastecimento a alguns sectores, afetando a economia, a qualidade de vida dos utentes e a rentabilidade da organização. Por existir exclusividade na distribuição e a resolução ser morosa, os clientes ficam sem energia até à reposição. Através deste *dashboard* pretende-se uma melhor monitorização destas perdas para definição de medidas de contingência de modo a aumentar a



Figura 2 – Ilustração do caso de estudo

Com o intuito de identificar os requisitos, utilizar uma abordagem centrada no utilizador e averiguar o aspeto final do *dashboard* aplicou-se a técnica de *mock-ups*, (ver Figura 3). O *dashboard* final (ver Figura 4) foi implementado usando a ferramenta Microstrategy©.

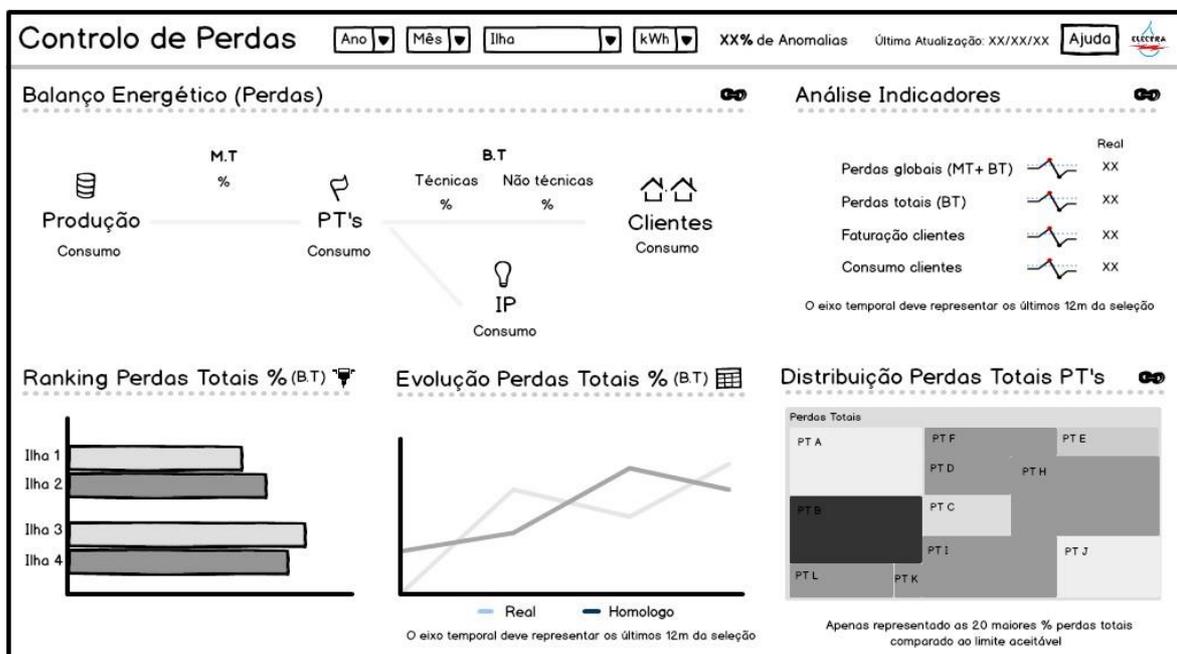


Figura 3 – Mock-up do caso de estudo, utilizando Balsamiq Mockups©

Protótipo e questionário

Para facilitar a avaliação da *framework* FMUED foi construído um protótipo em Microsoft Excel®, que inclui: o *dashboard* em avaliação; um temporizador para registrar o tempo da avaliação; as boas práticas de desenho; um gráfico com os resultados; e três botões, para observar os requisitos, avaliar os requisitos de facilidade de aprendizagem e satisfação e visualizar um resumo dos erros identificados. A Figura 5 apresenta um exemplo de uma avaliação real. As boas práticas sublinhadas a vermelho simbolizam erros que foram identificados pelo avaliador, afetando a usabilidade do *dashboard*.

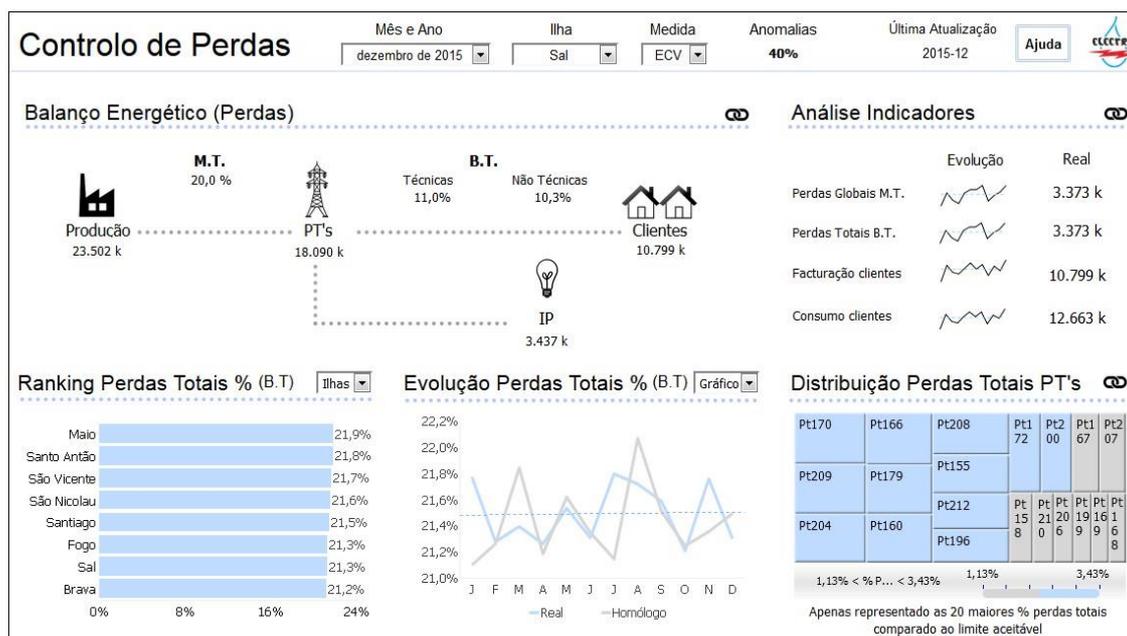


Figura 4 – Implementação do *mock-up*, utilizando Microstrategy®

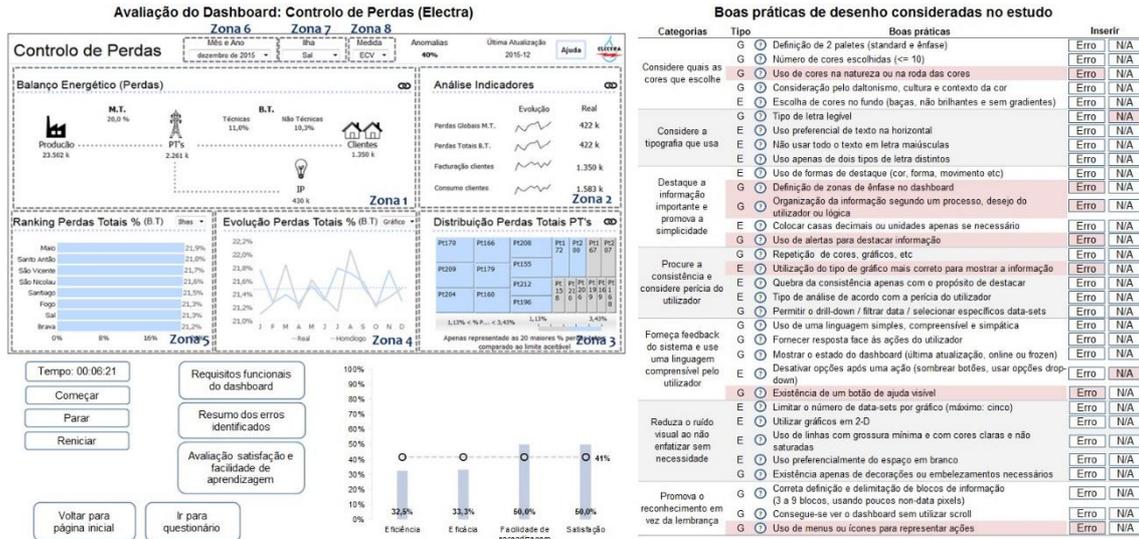


Figura 5 – Exemplo de uma avaliação utilizando o protótipo desenvolvido

De forma a contextualizar a amostra de inquiridos e obter feedback sobre o processo de avaliação, foi incorporado um questionário no protótipo, cujos detalhes são apresentados na Figura 6.

Questionário
Questionário Final

Como utilizar:

1 - Carregue na célula para visualizar as opções

Carregue nos botões para continuar

Volta para página inicial

Ir para Aplicação Framework

Caraterização da Amostra

- Qual é a instituição onde exerce atividade profissional?
- Qual é a área/curso dentro dessa instituição?
- Qual é a sua função na instituição?
Se escolheu Outra, por favor especifique
- Qual é o perfil da sua função na área de visualização de dados?
- Quantos anos tem de experiência em visualização de dados?
- Num ano colabora em quantos novos dashboards?
Desde do desenho até à implementação do dashboard

Caraterização do Processo de Avaliação

- Considerou excessivo o tempo de elaboração da avaliação?
Por favor, tenha em atenção o cronometro da folha excel (Aplicação Framework)
- Achou complexo o processo de avaliação?
Complexidade engloba o tempo de aprendizagem e a facilidade de utilização
- Considera existir mais benefícios do processo de avaliação face aos custos?
- Acha o processo de avaliação passível de reutilização em diferentes contextos?
Por exemplo, diferentes tipos de dashboards, boas práticas ou categorias de usabilidade
- Usaria esta framework num projeto para a sua instituição?
- Concorda que ao reutilizar este processo de avaliação irá demorar menos tempo?
- Por favor, deixe observações ou aspetos a melhorar do processo de avaliação

Figura 6 – Questionário embutido sobre a amostra e o processo de avaliação

Metodologia de distribuição do protótipo

A amostra de avaliadores da *framework* FMUED é composta por três perfis: (1) consultores de BI e IHM (competência de *Experience Design*) da Novabase; (2) professores de IHM; e (3) estudantes do mestrado em Engenharia Informática, com unidades curriculares em BI e IHM, e estudantes de licenciatura em Informática e Gestão, apenas com conhecimentos de BI, do ISCTE-IUL.

A amostra foi selecionada por conveniência, através de conhecimentos estabelecidos a nível académico com professores de várias instituições de Ensino Superior. A nível empresarial o protótipo foi enviado para todos os colaboradores da competência de *Business Analytics* envolvidos no desenho de interfaces. O método de distribuição pode ser observado na Figura 7. Cada inquirido selecionado recebeu por email duas versões do *dashboard* (escolhidas através de um gerador aleatório). As respostas foram todas devolvidas por email. A avaliação com os estudantes foi realizada em aulas práticas dedicadas, tendo sido distribuído aleatoriamente duas versões dos *dashboards* a cada estudante. De forma a garantir o anonimato das respostas, foi utilizado um extrator que eliminou a identificação do inquirido.

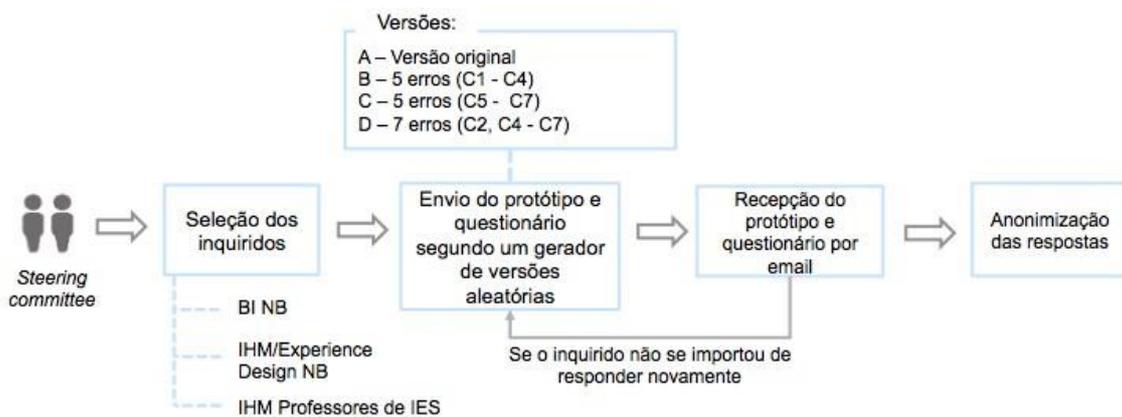


Figura 7 – Metodologia de distribuição do protótipo Legenda: Categorias (C#) da Tabela 2; NB = Novabase; IES = Instituições do Ensino Superior

5. ANÁLISE DE RESULTADOS

Foram identificados seguintes fatores que podem ter impacto nos resultados obtidos com a experiência:

- Nem todos os inquiridos possuíam igual experiência no desenho de interfaces;

- Alguns inquiridos tiveram uma explicação prévia (e.g., os estudantes), outros apenas viram um vídeo explicativo;
- Por não haver controlo durante a avaliação, os avaliadores podem ter tomado decisões com dúvidas pendentes;
- Por não existir uma ordem de preenchimento obrigatória, os inquiridos podem ter respondido a uma versão com erros e sem erros posteriormente e vice-versa;
- O cronómetro tem que ser iniciado pelo avaliador, pelo que podem existir desvios face ao verdadeiro tempo decorrido.

Obteve-se uma amostra de 166 inquiridos, dos quais 50 eram profissionais (40 BI e 10 IHM) e 116 estudantes (68 BI e 48 BI + IHM). O tempo médio de avaliação foi aproximadamente de 11 min 30 s. Os estudantes demoram 12 min e os profissionais 11 min.

Constataram-se alterações da usabilidade consoante o número de erros propositados (Figura 8), variando entre os 76,8% e os 67,2% e uma média de 70,5%. Contudo, os resultados variaram consoante o tipo de inquirido (Figura 9). Para os profissionais, verificou-se um decréscimo por incremento de erros. Os estudantes afirmaram que com cinco erros possuía menor usabilidade do que com sete erros. Os estudantes deram menor usabilidade em todas as situações comparativamente aos profissionais.

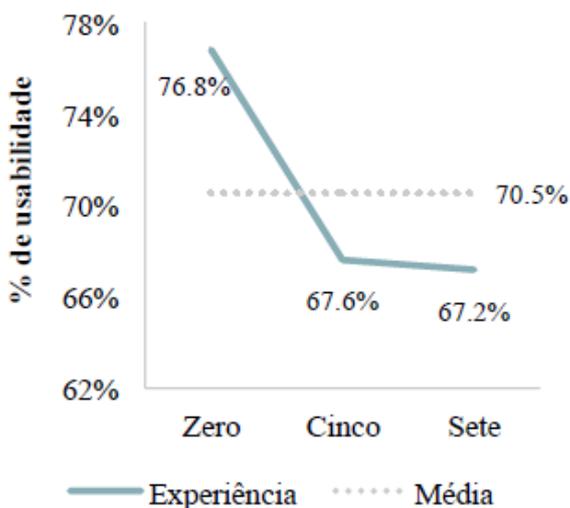


Figura 8 – Usabilidade por número de erros conforme o tipo de inquirido

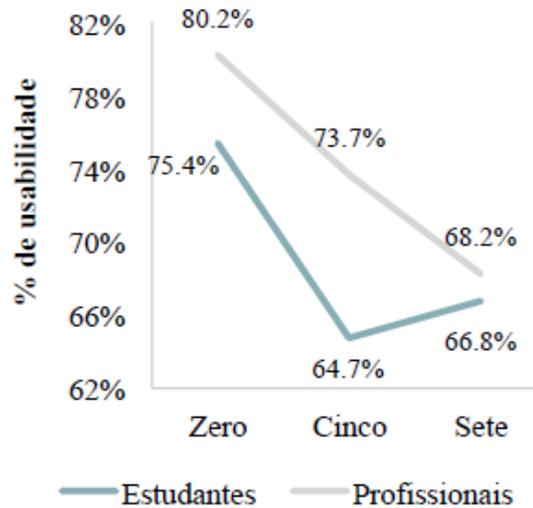


Figura 9 – Usabilidade por número de erros

A discrepância nas versões com cinco e sete erros decorre da diferença do número de erros específicos detetados entre as versões, conjugado com o acréscimo do número de boas práticas que foram consideradas não aplicáveis (Figura 10), e, ainda, por não existirem diferenças substanciais nos valores obtidos nas dimensões de satisfação e facilidade de aprendizagem com o incremento do número de erros (Figura 11).

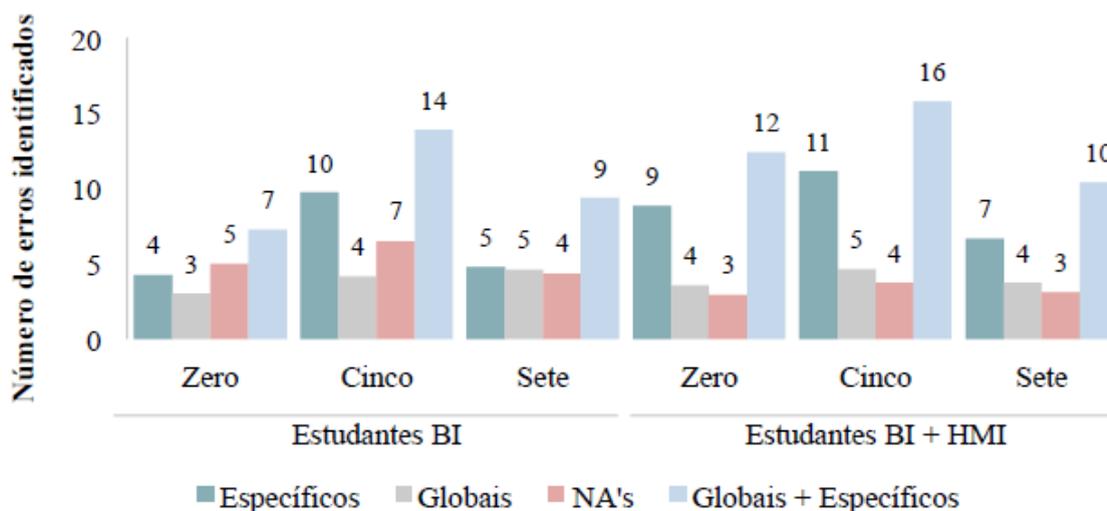


Figura 10 – Número de erros identificados pelos estudantes

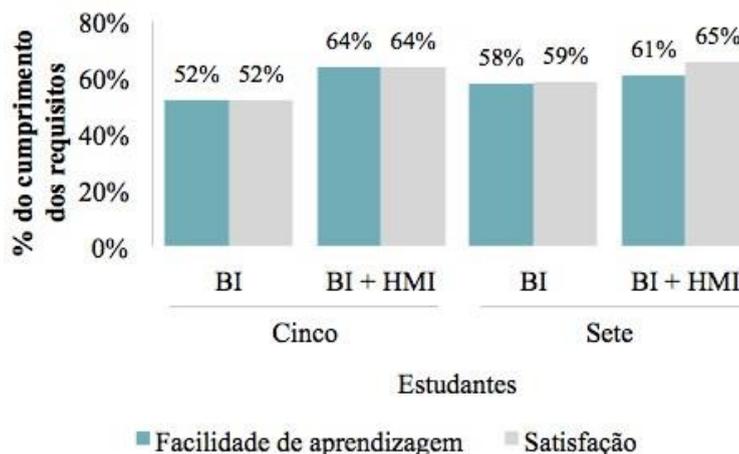


Figura 11 – Opinião em % dos estudantes no cumprimento dos requisitos de satisfação e facilidade de aprendizagem consoante o número de erros introduzidos

Analisando a percentagem de erros detetados (Figura 12), os profissionais, nas versões com cinco erros, apontaram uma percentagem superior média de identificação dos erros globais (69%), face aos estudantes (56%). Porém, nos erros específicos, verificou-se o oposto (35% vs. 55%). Na versão

com sete erros, os estudantes (39%) reconheceram uma maior percentagem nos erros globais comparativamente aos profissionais (28%) e igualmente nos específicos (respetivamente, 49% e 44%).

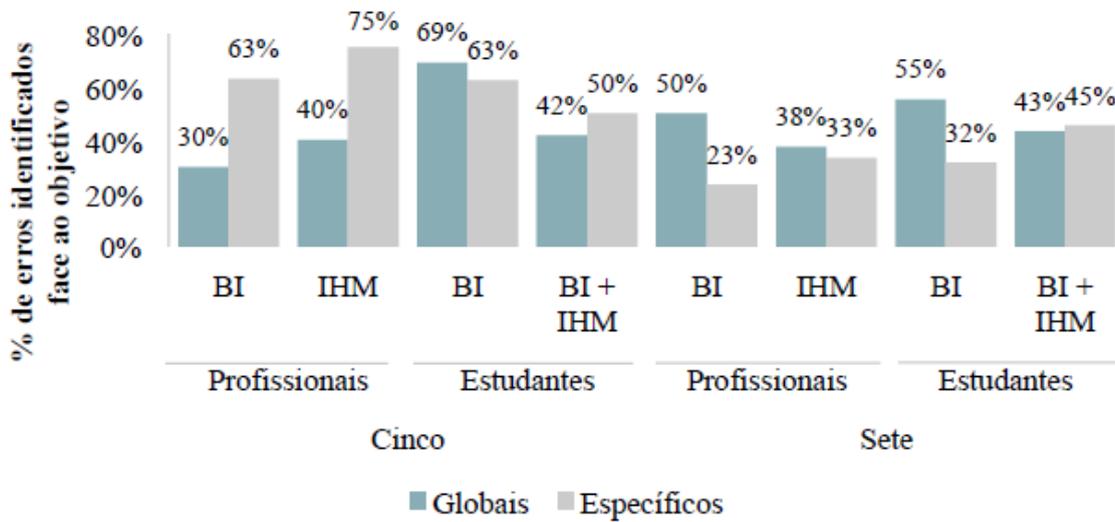


Figura 12 - % de erros identificados face ao objetivo por número de erros, avaliador e área

Relativamente à percentagem individual dos erros inseridos propositadamente, existem discrepâncias por tipo de avaliador.

Na versão B, apurou-se que, exceto no erro do texto em letras maiúsculas, os estudantes identificaram sempre em maior percentagem (Figura 13). Obteve-se uma média de 67%, concluindo-se que cada avaliador em média detetou 3,35 dos 5 erros introduzidos.

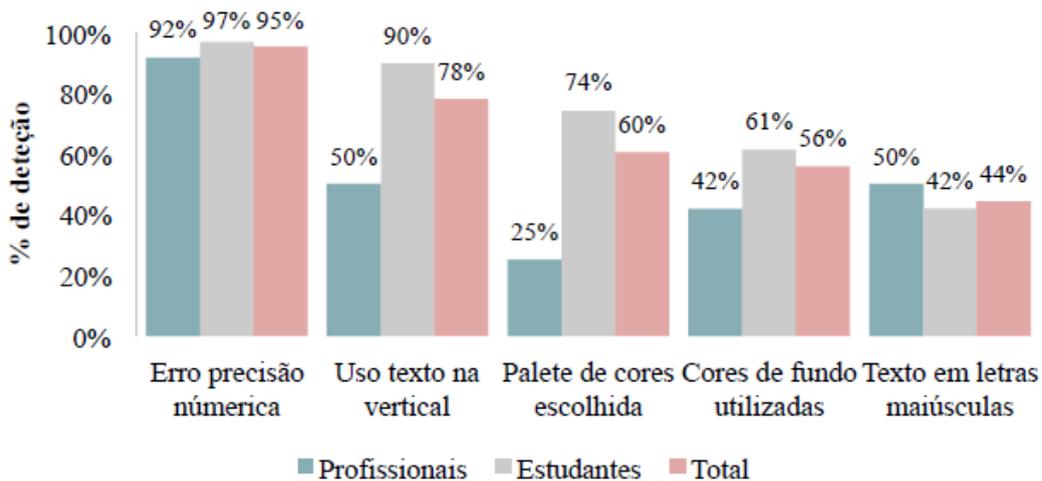


Figura 13 - % de erros detetados na versão B (5 erros)

Na versão C, existiam três casos (sem ajuda visível, aplicação de tabelas, decorações supérfluas) onde os profissionais conseguiram detetar uma maior percentagem (Figura 14). Porém, a utilização da barra de *scroll* e a separação de blocos, foi detetado em maior percentagem pelos estudantes. Concluiu-se que a percentagem total de identificação de erros foi de 41%, tendo cada avaliador detectado 2,05 dos 5 erros colocados.

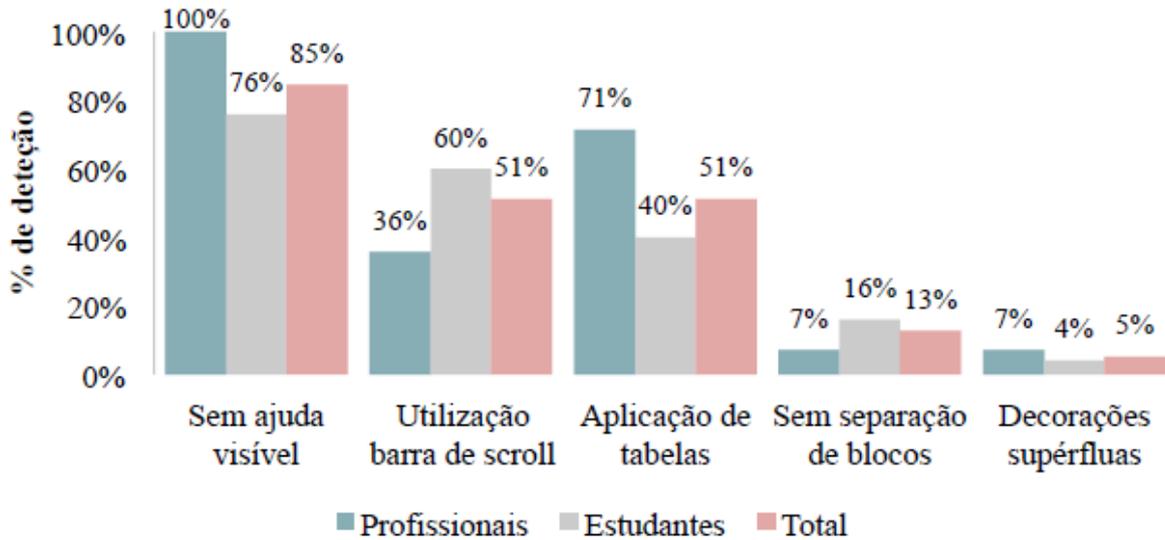


Figura 14 - % de erros detetados na versão C (5 erros) Por fim, na versão D, os estudantes detetaram uma maior percentagem de erros, exceto no uso do texto na vertical (Figura 15). Verificou-se uma percentagem de 44% no total, significando que cada avaliador detetou 3,09 erros face aos 7 inseridos.

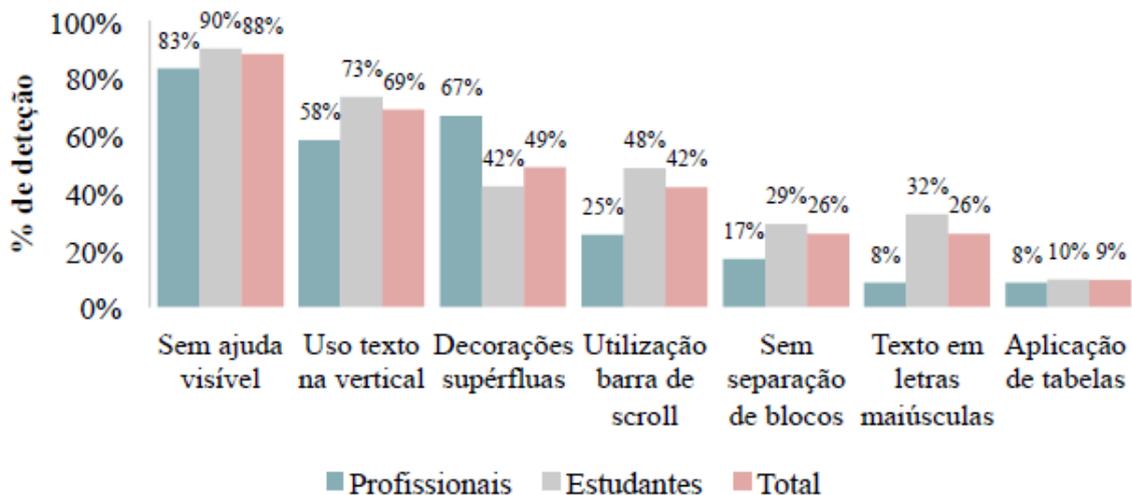


Figura 15 - % de erros detetados na versão D (7 erros)

Relativamente à complexidade do processo de avaliação (Figura 16), 47% dos inquiridos manifestaram opinião negativa e 18% uma opinião neutra sobre a complexidade do processo. Embora o protótipo desenvolvido tenha sido simplificado, ainda é necessário tornar o processo de avaliação mais intuitivo.

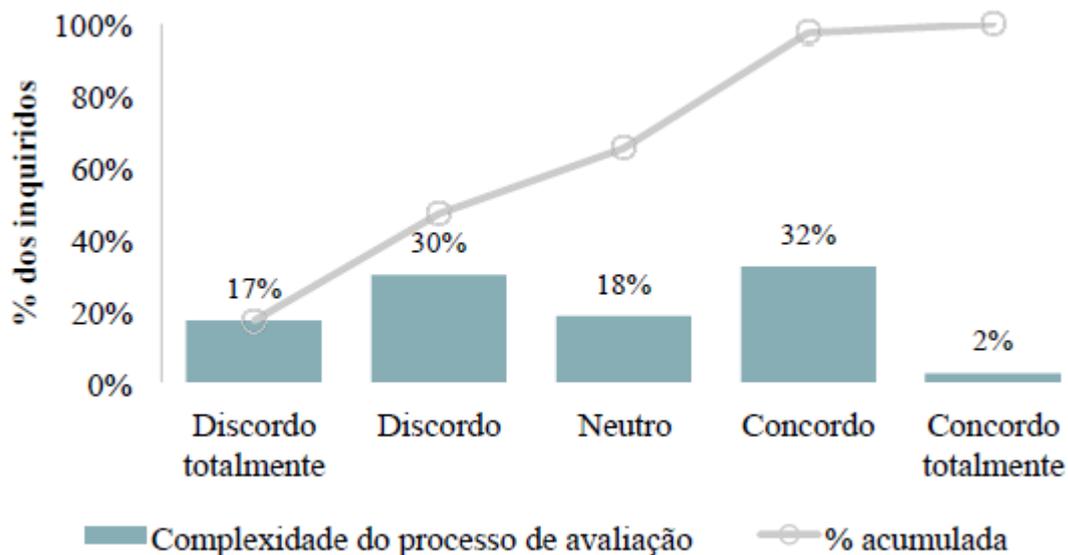


Figura 16 – Opinião sobre complexidade do processo de avaliação

Os profissionais consideram que a avaliação era reutilizável noutros contextos (diferentes tipos de dashboards, boas práticas e dimensões de usabilidade). Em relação à utilização do processo de avaliação em projetos organizacionais, 70% responderam positivamente, o que é um aspecto bastante relevante (Figura 17).

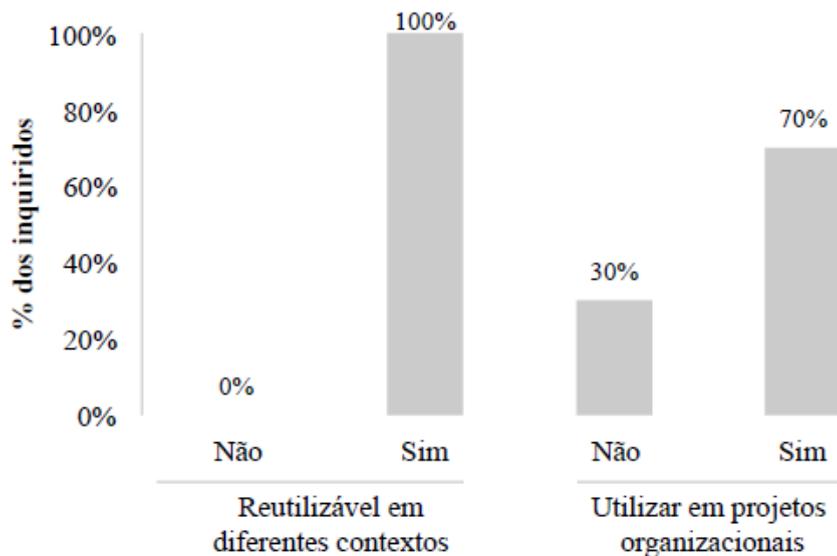


Figura 17 – Opinião dos profissionais sobre reutilização da *framework*

Análise das hipóteses de estudo

Nesta secção, será feita um análise detalhada das quatro hipóteses de estudo identificadas.

H1 - A usabilidade difere significativamente com a existência de erros?

De acordo com Few [2006], Fonseca et al. [2012] e Nielsen [1993] uma interface que não siga as boas práticas de desenho possui menor usabilidade. Assim, pretendeu-se averiguar se efetiva e significativamente se constatavam as diferenças referidas pelos autores.

Com esse fim, realizou-se o teste de Mann-Whitney U^2 , obtendo-se um valor de 1595,0, $z = -3571$ e $p = 0,000$, a 95% de confiança, verificando-se a existência de diferenças significativas consoante a existência de erros. Também se apurou que os avaliadores deram maiores rankings à versão sem erros, comparativamente com as que tinham erros, 106,10 e 75,36, respectivamente.

A Figura 18 demonstra que o *dashboard* sem erros possui, em termos medianos, maior usabilidade face aos que apresentam erros. Mais de 50% dos avaliadores consideraram que a versão sem erros tinha mais de 82% de usabilidade, diferenciando dos 67,8% obtidos nas outras versões. Nos marginais, verifica-se apenas a existência de um, diferenciando da versão com erros, onde foram identificados dois.

² Devido ao pressuposto da normalidade não ter sido garantido nos subgrupos, aplicou-se um teste não paramétrico.

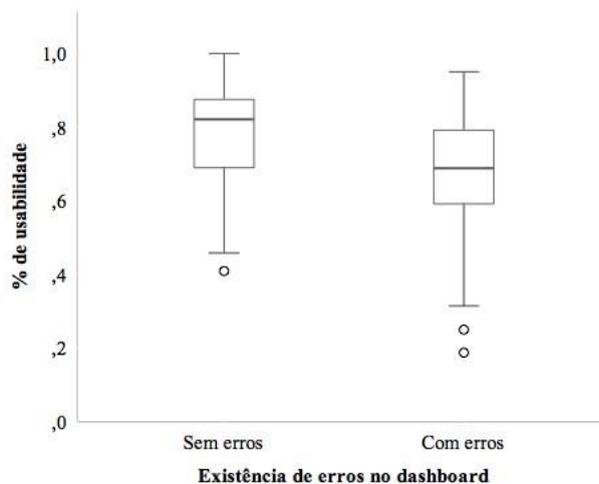


Figura 18 – Usabilidade consoante a existência de erros

H2 - A usabilidade diverge significativamente consoante os erros inseridos?

Para se apurar se as diferenças encontradas na H1 são visíveis consoante o número de erros introduzidos, aplicou-se o teste Kruskal-Wallis, $H_{(2)} = 12,766$, $p = 0,002$, a 95% de confiança, confirmando a existência das diferenças.

Na Figura 19 observa-se a distribuição para as diferentes situações, verificando-se que a versão com zero erros possui maiores valores nos quartis comparativamente aos restantes. Porém, nas versões com erros, as medianas são similares, possuindo a versão com cinco erros uma menor mediana, embora com uma distribuição menor e com valores superiores face aquela que tem sete erros. Todavia, possui 2 inquiridos marginais, diferenciando-se na versão com sete erros. Relativamente aos rankings, a versão com zero erros teve maior ranking (106,10), seguidamente, as versões com cinco erros (75,73) e, por fim, a com sete erros (74,66).

Para se apurar se as diferenças nas médias eram significativas, aplicou-se *a posteriori* o teste Scheffe³, confirmando-se apenas diferenças significativas com zero erros face à de cinco (0,92) e à de sete erros (0,96) e vice-versa. O teste evidenciou dois subgrupos homogêneos, com e sem erros.

³ Embora a ANOVA não seja robusta a violações da normalidade, o teste Scheffe é robusto, sendo aplicável nestas situações [Laureano 2013].

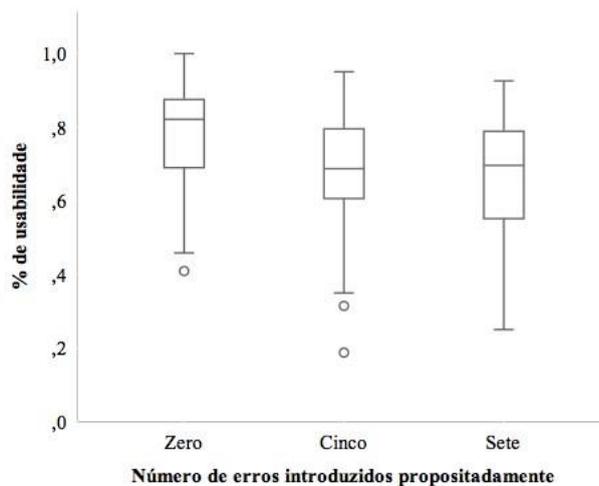


Figura 19 – Usabilidade consoante o número de erros introduzidos

H3 - A usabilidade diverge significativamente dependendo do tipo de avaliador?

Confirmadas as hipóteses H1 e H2 e sendo visíveis diferenças consoante o tipo de avaliador (Figura 9), questionou-se se estas são significativas. Assim, aplicou-se o t-test⁴ e confirmou-se existirem significativas diferenças entre os tipos de avaliadores⁵, $t_{(115,706)}$, $p = 0,014$, a 95% de confiança. Os profissionais (73,9%), em termos médios, avaliaram com maior usabilidade face aos estudantes (68%).

Na Figura 20 constatam-se as diferenças de usabilidade, tendo os profissionais uma mediana (75,9%), superior aos estudantes (68,5%). Contudo, a distribuição dos valores de usabilidade nos profissionais é inferior à dos estudantes, tendo sido identificados três profissionais considerados marginais e apenas um estudante.

⁴ Constatou-se normalidade nos subgrupos e inexistência de fortes assimetrias nas distribuições.

⁵ Observou-se inexistência de homogeneidade de variâncias, $F(4,204)$, $p = 0,04$

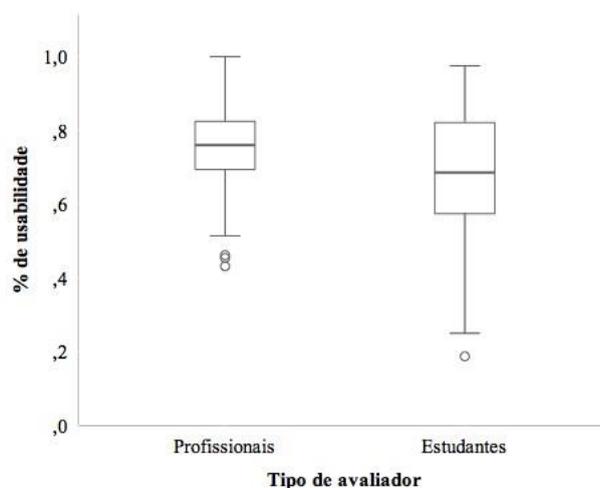


Figura 20 – Usabilidade por tipo de avaliador

H4 - Existem diferenças significativas na percentagem de boas práticas globais e específicas?

Durante a elaboração do estudo questionou-se se o tipo de boa prática (específica ou global) tinha influência na sua identificação por parte do avaliador. Com o intuito de confirmar se as diferenças eram significativas, aplicou-se o t-test de amostras emparelhadas⁶, $t_{(121)} = 0,336$, $p = 0,737$, a 95% de confiança.

Confirmou-se que a percentagem de erros identificados específicos (50,6%) e globais (49,2%), face ao objetivo, não são significativamente diferentes, apresentando os erros específicos uma ligeira superioridade face aos globais (1,4%).

Por fim, verificou-se correlação muito baixa, Pearson $(122) = 0,177$, sig = 0,051, concluindo-se que maior percentagem de erros específicos identificados não resulta numa maior percentagem de erros globais identificados e vice-versa.

Em síntese, das quatro hipóteses de estudo três foram suportadas: H1 (versão sem erros possui maior usabilidade), H2 (a usabilidade da versão sem erros difere significativamente da das versões com cinco e sete erros), H3 (estudantes atribuíram uma menor usabilidade face aos profissionais). Apenas a hipótese H4 não foi suportada (o tipo de boa prática não influencia a percentagem de identificação de erros).

Feedback recebido

No questionário enviado (ver Figura 6) foi pedido aos avaliadores *feedback* sobre o processo de avaliação. Apenas os profissionais enviaram observações e aspetos a melhorar, relativos a questões de interação,

⁶ Validaram-se os pressupostos das amostras emparelhadas.

heurísticas, requisitos de facilidade de aprendizagem e satisfação, e observações sobre os erros. Alguns dos comentários recebidos foram: “A utilização de uma imagem estática do *dashboard* não permite verificar se a interação segue as boas práticas”; “As heurísticas estão demasiado detalhadas, devendo ser removidas algumas e introduzidas outras (e.g., apresentação de informação hierarquicamente); “É difícil avaliar a facilidade de aprendizagem e satisfação por o avaliador não ser o utilizador final”; e “Apenas assinalar o erro não facilita a justificação do motivo, devendo ser possível comentar a decisão e afirmar se deriva de excesso ou defeito”.

6. CONCLUSÕES E TRABALHO FUTURO

O desenvolvimento de interfaces em BI está muitas vezes dissociado das boas práticas aceites em IHM, resultando em produtos finais com uma avaliação negativa por parte dos clientes de projetos de BI. O trabalho apresentado pretende quebrar essa tendência propondo uma *framework* que permite medir a usabilidade e eficiência de *dashboards*, utilizando boas práticas de desenho de BI e IHM, com adaptação da ISO/IEC 9241-11.

A *framework* desenvolvida possui duas componentes: um conjunto de heurísticas de desenho aplicadas a *dashboards* (organizadas em sete categorias, identificadas com base na revisão de literatura) e um método de avaliação. O trabalho foi desenvolvido no âmbito de uma dissertação de mestrado, mas inserido num contexto empresarial.

Foi desenhada uma experiência para validar a *framework*, incluindo a participação de avaliadores profissionais (consultores de BI e de IHM, e professores de IHM) e estudantes. Para tal, a Novabase forneceu um caso de estudo real aplicado ao sector energético em Cabo Verde, com vista ao desenvolvimento de um *dashboard* que permitisse controlar as perdas energéticas não técnicas, associadas a fraudes. Os objetivos principais da experiência consistiram na verificação das diferenças de usabilidade por número de erros introduzidos e na observação da aceitação da *framework* pela equipa de BI da Novabase. A amostra de avaliadores incluiu 166 inquiridos, dos quais 50 profissionais (40 BI e 10 IHM) e 116 estudantes (68 BI e 48 BI + IHM). O tempo médio de avaliação foi aproximadamente de 11 min 30 s, sem diferenças significativas entre estudantes e profissionais.

Os resultados demonstram as diferenças de usabilidade consoante o número de erros introduzidos e o tipo de inquiridos. Três das quatro hipóteses de estudo foram suportadas. Concluiu-se que a usabilidade do *dashboard* sem erros é maior (H1) e que a usabilidade da versão sem erros difere significativamente da das versões com cinco e sete erros (H2). Verificou-se também que os estudantes atribuíram uma menor usabilidade face aos profissionais (H3). Contudo, para os estudantes, um *dashboard* com sete erros não significa menor usabilidade face aos *dashboards* com cinco erros. Relativamente à percentagem de identificação de erros específicos e globais, não foram encontradas

diferenças significativas, tendo-se concluído que o tipo de boa prática não influencia a identificação do erro (H4, não suportada). Finalmente, um aspecto a realçar é a elevada percentagem de respostas positivas (70%) por parte dos profissionais quanto à reutilização da *framework* noutros contextos, nomeadamente em projetos com clientes.

Este estudo apresenta algumas limitações, que serão colmatadas com trabalho futuro. Apesar de se ter medido a usabilidade em termos médios, a avaliação das dimensões de satisfação e facilidade de aprendizagem não produziu resultados satisfatórios, de acordo com o *feedback* recebido pelos profissionais. Será necessário investigar e testar um novo método para validar estes requisitos, recorrendo em particular aos utilizadores finais [Nielsen, 1993]. O protótipo possui fragilidades devido ao software utilizado (Microsoft Excel©) que não permite testar a interatividade dos *dashboards* em diferentes sistemas operativos) e consequentemente a escolha de desenho feita (utilização de uma imagem estática).. Este problema será resolvido na solução final da *framework*, que deverá ser colocada em utilização na empresa patrocinadora do estudo. Atendendo ao *feedback* recebido acerca da complexidade do processo de avaliação, a solução final deverá contemplar um método mais intuitivo.

De forma a consolidar o trabalho desenvolvido, pretende-se realizar uma nova experiência de validação com uma maior amostra de avaliadores e estudando mais duas hipóteses que não puderam ser verificadas neste estudo: *Existem categorias de heurísticas onde a identificação foi significativamente menor?* e *Existe uma correlação negativa e significativa entre o número de erros detetados e a usabilidade?* Devido ao número limitado de erros selecionados, não foi possível incluir erros de todas as categorias. A ordem de avaliação dos dashboards pode também ter impacto nos resultados finais, pelo que a metodologia deve dividir a amostra em duas partes e prescrever para cada uma a ordem de avaliação, i.e., avaliar primeiro a versão sem erros (ou com menos erros) e vice-versa.

REFERÊNCIAS

- Eckerson, W. W., Performance Dashboards: Measuring, Monitoring, and Managing Your Business (2nd ed.), Wiley, 2010.
- Few, S., Information Dashboard Design: The Effective Visual Communication of Data, O'Reilly Media, 2006.
- Few, S., Now You See It: Simple Visualization Techniques for Quantitative Analysis, Analytics Press, 2009.
- Few, S., Visual Business Intelligence - Assessing the Effectiveness of a New Dashboard's Design, <http://www.perceptualedge.com/blog/?p=672>, (20 Dezembro 2014), 2010.
- Few, S., Information Dashboard Design: Displaying Data for At-a-Glance Monitoring (2nd ed.), Analytics Press, 2013.
- Fonseca, M. Campos, P. Gonçalves, D., Introdução ao Design de Interfaces (2nd ed.), FCA, 2012.

- Green, D. Pearson, J. M., “Development of a web site usability instrument based on ISO 9241-11”, *The Journal of Computer Information Systems*, 47, 1 (2006), 66–72.
- ISO/IEC 9241-11, Ergonomic requirements for office work with visual display terminals (VDT)- Part 11: Guidance on usability, 1998.
- Johnson, J. Henderson, A., “Conceptual models: begin by designing what to design”, *Interactions*, 9, 1 (2002), 25–32.
- Jokela, T. Iivari, N. Matero, J. Karukka, M., “The standard of user-centered design and the standard definition of usability: analyzing ISO 13407 against ISO 9241-11”, in *Proceedings of Latin American conference on Human-computer interaction*, ACM, NY, 2003, 53-60.
- Joo, S. Lin, S. Lu, K., “A Usability Evaluation Model for Academic Library Websites: Efficiency, Effectiveness and Learnability”, *Journal of Library and Information Studies*, 9, 2 (2011), 11–26.
- Lakatos, E. M. Marconi, M. de A., *Fundamentos de metodologia científica* (5th ed.), Atlas, 2003.
- Laureano, R. M. S., *Testes de Hipóteses com o SPSS* (2nd ed.), Edições Silabo, 2013.
- Lourenço, P., *Evaluation of Business Intelligence application design using the principles of Visual Perception*, ISCTE-IUL, Lisboa, 2013.
- Molich, R. Nielsen, J., “Improving a human-computer dialogue”, *Communications of the ACM*, 33, 3 (1990), 338–348.
- Nielsen, J., *Usability Engineering*, Morgan Kaufmann, 1993.
- Nielsen, J., “Enhancing the explanatory power of usability heuristics”, in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, 1994, 152-158.
- Norman, D. A., *The Design of Everyday Things* (2nd ed.), Basic Books, 2002.
- Rudin, K., “The Shift to On-Demand Business Intelligence”, *Business Intelligence Journal*, 12,1 (2007), 44–54.
- Tufte, E., *Envisioning Information* (2nd ed.), Graphics Press, 1990.
- Ware, C., *Information Visualization: Perception for Design* (3rd ed.), Morgan Kaufmann, 2012. White, C., *Improving Business Intelligence Usability*, <http://www.b-eye-network.co.uk/view/1557>, (22 Dezembro 2014), 2005.