

Assessing Quality Through Use: A Usability-Based Evaluation of the Conta gov.br Authentication Platform

Demétrius de Almeida Jubé
Secretaria de Tecnologia da Informação (Prodasen),
Senado Federal
Brasília, DF, Brasil
jube@senado.leg.br

Letícia Lopes Leite
Departamento de Ciência da Computação (CIC),
Universidade de Brasília (UnB)
Brasília, DF, Brasil
lleite@unb.br

ABSTRACT

This article presents an empirical study that evaluated the Conta gov.br authentication platform, a nationally adopted service in Brazil, focusing on usability as a dimension of software quality. Thirty participants took part in usability testing using standardized metrics: System Usability Scale (SUS), Single Ease Question (SEQ), and Task Completion Time (TCT). Results showed variations in perceived usability influenced by demographic factors such as age and education level. Statistical correlations between scores and user profiles were also analyzed. Based on the findings, a Practical Guide was developed to support public institutions in their adoption strategies. The results contribute evidence-based insights to inform quality-oriented decisions in digital public services.

KEYWORDS

usability, authentication, public digital services, software quality, Conta gov.br

1 Introdução

A transformação digital do setor público brasileiro tem avançado significativamente nos últimos anos, impulsionada por diretrizes como a Estratégia Nacional de Governo Digital (ENGD) [16]. A referida iniciativa estabelece que os serviços públicos digitais devem ser mais acessíveis, seguros e centrados no cidadão — incluindo a consolidação de mecanismos de autenticação que simplifiquem o acesso e reforcem a confiança dos usuários. Nesse contexto, o Governo Federal lançou, em 2016, a plataforma de login unificado Conta gov.br [13], que permite o acesso centralizado a diversos serviços públicos por meio de uma única credencial.

Essa arquitetura de autenticação contribui para reduzir a carga cognitiva dos usuários, ao minimizar a quantidade de senhas que precisam ser memorizadas, conforme apontado por Sun et al. [34]. Os autores também destacam outras vantagens dos mecanismos de Login Único, ou *Single Sign-On* (SSO): maior agilidade no preenchimento de cadastros, interoperabilidade entre sistemas por meio de padrões abertos e personalização de serviços com base em dados do usuário.

O modelo SSO também contribui para mitigar um problema recorrente nos serviços digitais: a “Grande Fadiga da Autenticação”, descrita por Sasse et al. [30], que demonstra como autenticações frequentes e fragmentadas podem interromper tarefas, reduzir a produtividade e comprometer a percepção de segurança. A Conta gov.br, ao adotar esse modelo, incorpora suas vantagens típicas e ainda apresenta características técnicas e institucionais que a diferenciam de soluções comerciais genéricas de autenticação em dois fatores (2FA): níveis progressivos de segurança (Bronze, Prata

e Ouro), integração com bases cadastrais oficiais e conformidade com a Lei Geral de Proteção de Dados (LGPD) [15]. Essas propriedades a tornam particularmente relevante para avaliação sob a perspectiva da usabilidade, uma vez que a qualidade da experiência de autenticação pode impactar diretamente a adesão, a confiança e a efetividade na prestação de serviços públicos digitais.

Apesar de sua ampla adoção institucional [18], ainda são escassos os estudos empíricos que avaliem a usabilidade da Conta gov.br em comparação com métodos tradicionais, como o login por e-mail e senha. A usabilidade exerce papel decisivo na aceitação de tecnologias digitais, por refletir o quão fáceis, eficazes e agradáveis essas soluções são na perspectiva do usuário [28]. Essa característica é ainda mais relevante em serviços públicos digitais, uma vez que os usuários possuem perfis diversos. Além de impactar diretamente a experiência de uso, a usabilidade é reconhecida como um atributo técnico essencial da qualidade de software, integrando modelos consolidados como o ISO/IEC 25010 [14]. Avaliar se a Conta gov.br melhora ou compromete a experiência de autenticação é, portanto, essencial para embasar políticas de adoção tecnológica mais eficazes. A qualidade dessa experiência também influencia decisões técnicas sobre arquitetura, interoperabilidade e *design* — aspectos centrais na avaliação da qualidade de software no setor público.

A análise considerou três aspectos principais: a Usabilidade Geral, mensurada pela *System Usability Scale* (SUS) [7]; a Facilidade Percebida de Uso, conceito proposto por Davis [10], avaliada por meio da *Single Ease Question* (SEQ) [31]; e o Tempo de Conclusão da Tarefa (TCT), citado por Rubin [29] como métrica de eficiência. Também foram examinadas correlações entre os resultados e variáveis demográficas. Os achados fortalecem discussões sobre *design* centrado no usuário e subsidiam decisões estratégicas sobre a adoção de tecnologias de autenticação em serviços públicos digitais. As informações obtidas foram consolidadas em um Guia Prático com recomendações para a integração consciente da Conta gov.br, ampliando o impacto aplicado da pesquisa.

O texto está organizado da seguinte forma: a Seção 2 apresenta os fundamentos teóricos e os principais conceitos necessários à compreensão do estudo, enquanto a Seção 3 discute trabalhos prévios relacionados à aferição de usabilidade em serviços digitais. A Seção 4 descreve a metodologia adotada, incluindo o protocolo experimental, o perfil dos participantes e as métricas utilizadas. A Seção 5 trata dos cuidados éticos aplicados durante a coleta dos dados. A Seção 6 apresenta os resultados quantitativos e qualitativos obtidos, enquanto a Seção 7 analisa suas implicações, limitações e contribuições para o *design* de sistemas públicos de autenticação. Por fim, a Seção 8 traz as conclusões do estudo e propostas

para pesquisas futuras, seguida pelos agradecimentos constantes na Seção 9.

2 Fundamentação Teórica

A literatura especializada tem enfatizado a importância de avaliar soluções digitais sob as perspectivas da usabilidade, segurança e interação humano-computador (IHC). Particularmente em sistemas que envolvem autenticação e proteção de dados sensíveis, autores como Korir et al. [23], Polakis et al. [27] e Wang et al. [39] demonstram que métodos estruturados de avaliação — como testes com usuários, análises heurísticas e questionários padronizados — são fundamentais para identificar falhas de interface e garantir que mecanismos de segurança não comprometam a acessibilidade e a eficácia das aplicações.

Entre as métricas mais consolidadas para avaliação da experiência do usuário, destaca-se a *System Usability Scale*, proposta por Brooke [7], composta por dez questões em escala Likert. Seus escores, que variam de 0 a 100, podem ser interpretados com base na escala adjetiva descrita por Bangor et al. [2], categorizando os sistemas conforme a percepção dos usuários — de “*Worst Imaginable*” a “*Best Imaginable*”. Essa simplicidade interpretativa amplia sua aplicabilidade, inclusive em contextos não especializados.

Estudos posteriores propuseram interpretações complementares para a SUS. Sauro e Lewis [31] desenvolveram uma análise baseada em percentis, enquanto Tullis e Stetson [36] evidenciaram a estabilidade de resultados mesmo em amostras pequenas. Essas contribuições foram incorporadas na ferramenta online *SUS Analysis Toolkit*, proposta por Blattgerste et al. [5], que permite calcular intervalos de confiança e comparar os resultados com *benchmarks* normativos.

Outra métrica encontrada na literatura é a *Single Ease Question*, composta por uma única pergunta em escala de sete pontos, aplicada logo após a execução de uma tarefa. Sua simplicidade e foco na percepção imediata tornam-na eficaz para identificar gargalos na interação, conforme demonstrado por Tedesco e Tullis [35] e Sauro e Lewis [31].

No que se refere à dimensão de eficiência — que, segundo Bevan [4], corresponde à quantidade de recursos exigidos do usuário, como tempo, esforço cognitivo ou repetição de ações — destaca-se o Tempo de Conclusão da Tarefa, métrica que expressa o tempo necessário para executar uma ação específica. Kwang Bok Lee e Grice [24] associam a eficiência também à facilidade de memorização e à produtividade no uso recorrente. Rubin [29], por sua vez, recomenda o uso de estatísticas como média, mediana e desvio-padrão na análise do TCT.

Esses instrumentos vêm sendo utilizados com frequência em estudos centrados no usuário e constituem uma base conceitual relevante para avaliações de qualidade de software orientadas à experiência de uso. A seguir, apresentamos trabalhos que empregaram essas e outras abordagens na análise de serviços digitais.

3 Trabalhos Relacionados

A usabilidade tem sido amplamente utilizada como critério para avaliação da qualidade de serviços públicos digitais. Monteiro et al. [26], por exemplo, analisaram dez *chatbots* governamentais, evidenciando dificuldades de navegação, falta de padronização e ausência de

feedback ao usuário. Vieira e Andrade [38] avaliaram o aplicativo Caixa Tem, observando melhorias significativas após ajustes em protótipos de alta fidelidade, com base em métricas como tempo de execução e número de interações. Aragão et al. [1], por sua vez, investigaram soluções voltadas a cidades inteligentes, ressaltando a importância de considerar o ambiente de uso e os dispositivos móveis como fatores que influenciam a experiência do usuário.

Além de evidenciarem problemas de interface, esses estudos aplicaram métricas de usabilidade para mensurar atributos de qualidade, contribuindo para a adoção de abordagens mais centradas no usuário no setor público.

Outro exemplo relevante é o estudo de Resende Júnior e Lopes Leite [20], que avaliou o Sistema Eletrônico de Informações (SEI) com base na *System Usability Scale* e em uma avaliação heurística. A análise identificou 99 problemas de usabilidade, como ausência de *feedback*, múltiplas telas redundantes e uma interface pouco intuitiva, reforçando a importância de combinar métodos qualitativos e quantitativos na avaliação de sistemas públicos.

Tais investigações reforçam o papel da usabilidade como dimensão-chave da qualidade de software, especialmente em contextos governamentais. Neste estudo, busca-se aprofundar essa perspectiva no escopo de sistemas que envolvem autenticação digital. A seguir, apresenta-se a metodologia adotada.

4 Metodologia

A metodologia desta pesquisa foi estruturada em quatro etapas complementares. A primeira consistiu na realização de uma Revisão Sistemática da Literatura (RSL), com o objetivo de mapear estudos sobre avaliação de usabilidade em sistemas de autenticação digital, com ênfase em soluções baseadas em *Single Sign-On*. Os resultados dessa etapa fundamentaram a elaboração de um protocolo experimental, que definiu as técnicas e métricas utilizadas na fase seguinte. Na terceira etapa, o referido protocolo foi aplicado em testes práticos com usuários, visando à coleta de dados empíricos sobre a experiência de uso dos métodos de autenticação analisados. Por fim, esses dados foram utilizados na confecção de um Guia Prático que reuniu as informações mais relevantes para subsidiar a decisão de utilizar a Conta gov.br por gestores públicos.

4.1 Revisão Sistemática da Literatura

A RSL teve como objetivo identificar os métodos mais recorrentes na avaliação de usabilidade em soluções de login único, bem como as métricas associadas, a fim de orientar o desenho do protocolo experimental adotado nesta pesquisa. As questões de pesquisa, critérios de inclusão e exclusão, *string* de busca e *checklist* de qualidade estão descritos em detalhe no estudo anterior dos autores [22], que analisou 141 artigos e, após as etapas de triagem e avaliação, resultou na seleção final de 18 estudos examinados em profundidade.

Além de embasar a escolha das métricas, a RSL contribuiu para alinhar o experimento às boas práticas observadas na literatura, especialmente no que diz respeito à estrutura dos testes e ao uso de diretrizes éticas. Essa base empírica ajudou a evitar decisões arbitrárias e assegurou a compatibilidade dos instrumentos com o contexto da autenticação digital.

A análise revelou a adoção frequente de práticas consolidadas, como diretrizes éticas explícitas com ênfase em privacidade e transparência (61%), roteiros estruturados para orientar os participantes (55%) e o oferecimento de compensações financeiras (55%). Além disso, 77% dos estudos aplicaram questionários baseados em escalas do tipo Likert, evidenciando uma preferência consolidada na área.

Entre as métricas mais utilizadas, destacaram-se a Facilidade Percebida de Uso e a *System Usability Scale*, ambas presentes em cerca de 39% dos estudos, além do Tempo de Conclusão da Tarefa, citado em aproximadamente 33%. Em razão de sua recorrência e aderência ao escopo do trabalho, essas métricas foram incorporadas ao protocolo experimental. Sua adoção também favorece a consistência do estudo e a comparabilidade dos resultados. Outras métricas, como taxa de sucesso ou número de erros, apareceram em alguns estudos, mas com baixa frequência, motivo pelo qual não foram selecionadas.

4.2 Criação do Protocolo Experimental

A criação do protocolo considerou os padrões metodológicos extraídos da literatura e as especificidades do contexto estudado. Sua elaboração seguiu os princípios estruturantes propostos por Bastien [3], cujas etapas são: (1) Definição dos objetivos, (2) Recrutamento dos participantes, (3) Seleção das tarefas, (4) Criação dos cenários, (5) Escolha das métricas, (6) Preparação do ambiente, (7) Definição do testador e do protocolo, (8) Elaboração dos questionários de satisfação e (9) Apresentação dos resultados.

O recrutamento dos participantes baseou-se nos resultados de uma enquête nacional enviada a 639.493 usuários do sistema e-Cidadania do Senado Federal [19], que resultou em 19.309 respostas. Desses, 5.900 manifestaram interesse em participar de testes futuros, por meio de uma pergunta específica no formulário. Apenas os e-mails daqueles que indicaram esse interesse foram utilizados, garantindo a preservação dos demais dados pessoais. Os convites foram enviados diretamente pelo pesquisador, com mensagens padronizadas e o Termo de Consentimento Livre e Esclarecido (TCLE) anexado. Os agendamentos das entrevistas foram organizados manualmente. Essa abordagem favoreceu a diversidade geográfica, etária e educacional entre os participantes.

As entrevistas foram realizadas por videoconferência, o que superou as limitações geográficas e viabilizou o registro do Tempo de Conclusão das Tarefas por meio da gravação de tela e áudio. Esse formato também permitiu o acompanhamento em tempo real, garantindo a padronização do protocolo e a fidelidade dos dados coletados. Nesse contato, eram apresentadas as tarefas que seriam realizadas pelos participantes, assim como os questionários que deveriam ser respondidos.

As tarefas foram organizadas em um formulário eletrônico contendo todas as instruções. O formulário era iniciado por um termo de consentimento que esclarecia o uso dos dados e permitia que o participante se retirasse caso se sentisse desconfortável. Em seguida, eram coletadas informações sociodemográficas e tecnológicas — como faixa etária, escolaridade, familiaridade com tecnologia e uso prévio da Conta gov.br. As versões anonimizadas do TCLE e do formulário utilizados na pesquisa estão disponíveis para consulta pública [21].

Os participantes, então, realizavam duas simulações de autenticação, uma com login por e-mail e senha e outra com a Conta gov.br, usando credenciais fictícias em ambiente de homologação. Após cada tarefa, os participantes responderam a um questionário composto pela SUS, para medir a usabilidade geral percebida. Utilizou-se a tradução para o português brasileiro validada por Lourenço et al. [25], cuja aplicação obteve um Alfa de Cronbach de 0,76, indicando confiabilidade satisfatória. Além disso, foi aplicada a *Single Ease Question*, destinada a aferir a Facilidade Percebida na realização da tarefa.

Uma vez que a pesquisa tinha a característica de um estudo intra-sujeitos, ou seja, o mesmo participante avaliava dois cenários distintos, era possível que ela sofresse influência do efeito de carregamento, conhecido como *carryover effect*. Este efeito é explicado por Bordens e Abbott [6] como um fenômeno em que a experiência adquirida na primeira tarefa pode influenciar o desempenho na segunda, introduzindo vieses relacionados à familiarização com o cenário ou à adaptação à dinâmica do teste nos dados obtidos. Para evitá-lo, foi adotada a estratégia sugerida por Sauro e Lewis [31], que consiste em alternar a ordem de aplicação das tarefas. Assim, metade dos participantes iniciou o teste utilizando o método tradicional de autenticação (e-mail e senha), enquanto a outra metade começou pela autenticação via Conta gov.br. Essa distribuição teve como objetivo trazer maior confiabilidade na comparação entre os métodos avaliados.

4.3 Aplicação dos Testes

Os testes de usabilidade foram realizados com participantes previamente recrutados, conforme descrito na Seção 4.2. A partir da base de contatos válida, foram enviados 1.000 convites individuais, resultando em 111 respostas. Dessas, 98 pessoas confirmaram interesse, e 30 participaram efetivamente da etapa de testes, conduzida entre os dias 30 de agosto e 29 de novembro de 2024.

A definição desse número de participantes foi definida com base na literatura. Sauro e Lewis [31] afirmam que, a partir de 30 observações, a distribuição das médias de usabilidade tende a se estabilizar e se aproximar de uma distribuição normal, mesmo quando os dados individuais não seguem essa distribuição. Tal comportamento está relacionado ao Teorema Central do Limite, segundo o qual a média de uma amostra tende à normalidade conforme o tamanho da amostra aumenta, independentemente da distribuição original dos dados. Isso permite o uso de análises estatísticas mais robustas, com maior precisão na estimativa de parâmetros populacionais. Essa amostra também está de acordo com as recomendações de Tullis e Stetson [36] vistas na Seção 2, que indicam que 12 participantes já são suficientes para obter estimativas consistentes com estudos maiores utilizando a SUS.

Não houve processo de seleção entre os interessados: participaram aqueles que responderam ao convite e estavam disponíveis para o agendamento. Ainda assim, a amostra final apresentou diversidade em termos demográficos. Em relação ao gênero, 60% dos participantes se identificaram como homens e 40% como mulheres. Quanto à faixa etária, 43,3% estavam entre 30 e 44 anos, 33,3% entre 45 e 59 anos, 10% entre 15 e 29 anos e 13,3% tinham 60 anos ou mais.

A amostra também apresentou ampla diversidade geográfica, considerando tanto a região de nascimento quanto a de residência

atual. A maioria dos participantes (53,3%) nasceu no Sudeste, seguida pelas regiões Nordeste (16,7%), Sul (13,3%), Norte (13,3%) e Centro-Oeste (3,3%). Quanto à residência atual, observou-se uma leve redistribuição: o Sudeste permaneceu predominante (46,7%), enquanto o Nordeste e o Sul concentraram 16,7% cada, o Norte manteve 13,3%, e o Centro-Oeste passou a representar 6,7% da amostra.

Quanto à escolaridade, 60% dos participantes eram pós-graduados, 33,3% possuíam graduação completa e 6,7% tinham até o ensino médio. Todos relataram já ter utilizado a Conta gov.br anteriormente, e a maioria se autoavaliou com alta familiaridade com tecnologia da informação, especialmente em relação ao uso de redes sociais, aplicativos, editores de texto e planilhas.

4.4 Confecção do Guia Prático

Com base nas métricas de usabilidade analisadas, nas correlações entre perfis de participantes e experiência de uso, e na documentação técnica da Conta gov.br [12, 17], foi elaborado um Guia Prático com orientações voltadas a gestores públicos sobre a adoção da solução. O material, em formato *markdown*, adota linguagem acessível e estrutura modular, facilitando sua disseminação e reutilização por diferentes instituições públicas.

O Guia traduz evidências empíricas em recomendações operacionais, articulando os achados do estudo com os requisitos técnicos da plataforma. No repositório disponível online [11], são apresentadas seções que abordam critérios de elegibilidade para integração, etapas de homologação, exigências técnicas de autenticação, estrutura de níveis e confiabilidades, além da análise de riscos associados à infraestrutura externa. Assim, o material busca extrapolar o escopo acadêmico, servindo como instrumento prático para apoiar decisões informadas sobre a adoção da Conta gov.br em serviços digitais públicos.

As etapas descritas — da revisão sistemática à elaboração do guia — seguiram um planejamento metodológico rigoroso, com foco na qualidade da pesquisa e na confiabilidade dos dados. Todas as atividades que envolveram participantes humanos foram conduzidas com base em princípios éticos, conforme detalhado na Seção 5 a seguir.

5 Cuidados Éticos

Esta pesquisa foi aprovada por um Comitê de Ética em Pesquisa, sob o CAAE 6.903.626, em conformidade com a Resolução nº 466/2012 do Conselho Nacional de Saúde. Todos os participantes foram informados sobre os objetivos do estudo, os procedimentos envolvidos e seus direitos, incluindo a possibilidade de se retirar a qualquer momento. O Termo de Consentimento Livre e Esclarecido foi apresentado no momento do agendamento da entrevista, garantindo uma participação voluntária e consciente.

Durante o planejamento, foram identificados seis possíveis riscos para os participantes: (1) exposição de dados pessoais, (2) exposição de credenciais, (3) vazamento de informações durante o compartilhamento de tela, (4) constrangimento ao ser observado e sensação de estar sendo julgado, (5) ansiedade com o tempo das tarefas e (6) insegurança quanto ao uso da imagem e da voz. Cada um desses riscos foi tratado com cuidados específicos, descritos a seguir.

A coleta de dados pessoais foi limitada ao mínimo necessário e incluiu: e-mail, faixa etária, gênero, grau de instrução, estado de

nascimento e de residência, familiaridade com tecnologia e uso prévio da Conta gov.br. Essas informações ficaram disponíveis apenas para os pesquisadores, em um repositório com autenticação de dois fatores. As credenciais utilizadas durante os testes eram fictícias e criadas para uso em ambiente de homologação, evitando que o participante usasse seus dados reais.

Para evitar o vazamento de informações, os participantes foram orientados a fechar outros aplicativos antes do início da gravação. O constrangimento foi reduzido com uma abordagem clara e acolhedora, deixando evidente que o sistema — e não o participante — estava sendo avaliado. Para lidar com a ansiedade, foi informado que não havia tempo limite e que o teste poderia ser pausado ou encerrado a qualquer momento. As gravações de tela e áudio, geradas por meio de uma plataforma de videoconferência com armazenamento em nuvem, foram transferidas para um armazenamento local com acesso restrito, e usadas apenas para os fins da pesquisa.

Todas as medidas vistas nessa Seção buscaram mitigar riscos aos participantes e seguir as orientações éticas estabelecidas, contribuindo para a condução responsável do estudo e fornecendo maior credibilidade à pesquisa. A seguir, são apresentados os principais resultados obtidos a partir dos testes de usabilidade realizados com os voluntários.

6 Resultados

Os testes de usabilidade aplicados permitiram comparar o desempenho e a percepção dos participantes em relação aos dois métodos de autenticação avaliados: e-mail e senha, e Conta gov.br. Cada uma das métricas coletadas será apresentada e discutida em subseções específicas, a fim de detalhar os resultados obtidos e facilitar a compreensão dos achados. Os dados e materiais suplementares associados aos testes estão disponíveis no Figshare [21].

6.1 Usabilidade Geral - SUS

A avaliação da usabilidade geral foi conduzida por meio da *System Usability Scale*, aplicada após a realização de cada tarefa de autenticação. As respostas dos participantes foram analisadas com o apoio da *SUS Analysis Toolkit*, mencionada na Seção 2, dadas as funcionalidades para interpretação dos escores.

Os dados foram organizados para submissão na ferramenta e deram origem ao gráfico ilustrado na Figura 1. Ele nos permite identificar características como a dispersão, mediana e valores fora do padrão (*outliers*), além de apresentar a classificação da Escala Adjetiva mencionada na Seção 2.

Os escores SUS obtidos indicaram avaliações bastante positivas para ambos os métodos de autenticação, com medianas próximas de 90 para o login tradicional por e-mail e senha, e de 88 para a Conta gov.br. Ambas as medianas situam-se na faixa “*Best Imaginable*”, segundo a Escala Adjetiva. No entanto, a dispersão dos resultados foi maior no caso da Conta gov.br, sugerindo maior variabilidade na experiência dos usuários — possivelmente relacionada ao nível de familiaridade com o serviço.

Para avaliar se essa diferença era estatisticamente significativa, foi estabelecido um protocolo que considerava quatro critérios principais: dependência entre as amostras, tamanho da amostra, normalidade das diferenças pareadas (resíduos) e presença de *outliers*. Como o estudo utilizou um desenho intra-sujeitos, as amostras

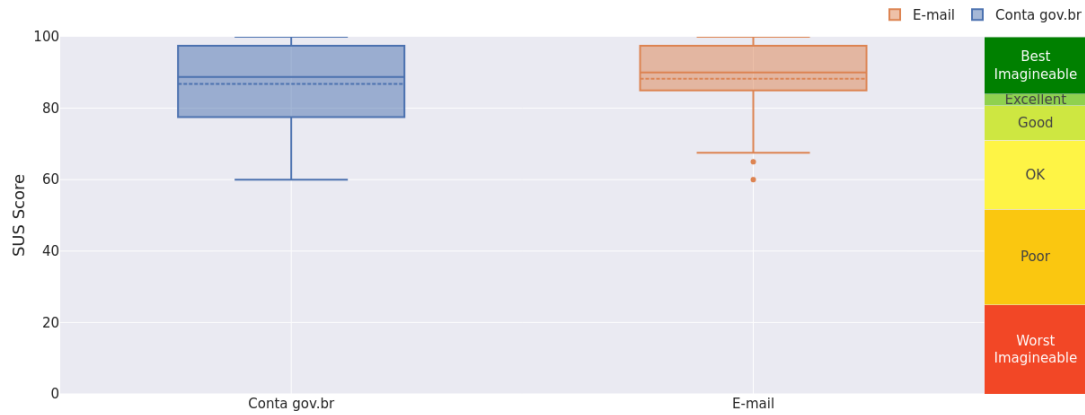


Figura 1: Dados estatísticos da medição da SUS.

foram consideradas dependentes. O tamanho da amostra (30 participantes) permitia a aplicação de um teste paramétrico, hipótese confirmada pelo teste de Shapiro-Wilk [33], que não indicou violação da normalidade ($W(30) = 0,98$, $p = 0,86$). Também não foram encontrados *outliers* relevantes nos resíduos.

Em um cenário como o descrito anteriormente, Sauro e Lewis [31] recomendam o uso do Teste T Pareado para comparação entre os métodos. O resultado do cálculo indicou que não houve diferença estatisticamente significativa entre os dois métodos ($t(29) = -0,64$, $p = 0,5258$). O Intervalo de Confiança para a diferença média vai de $-5,93$ a $3,09$, indicando que o intervalo observado pode variar dentro dessa faixa. O tamanho de efeito (d de Cohen = $0,12$) foi considerado Muito Pequeno [8], reforçando a interpretação de que os dois mecanismos foram percebidos de forma bastante similar em termos de usabilidade.

Em síntese, os resultados da escala SUS indicam que ambos os mecanismos de autenticação foram bem avaliados, sem diferenças estatisticamente relevantes. A próxima subseção apresenta os dados referentes à Facilidade Percebida, medida por meio da SEQ, com foco na avaliação imediata após cada tarefa.

6.2 Facilidade Percebida - SEQ

A *Single Ease Question* foi aplicada ao final de cada tarefa, com o objetivo de capturar a percepção imediata dos participantes sobre a facilidade de execução. O resultado da comparação entre os dois mecanismos de autenticação está apresentado na Figura 2.

Os resultados indicaram médias elevadas para ambos os métodos: aproximadamente 6,67 para a Conta gov.br e 6,77 para o login com e-mail e senha, em uma escala de 1 a 7. Essa diferença entre as médias foi mínima, e os intervalos de confiança indicaram baixa variabilidade nas respostas, sugerindo percepção de facilidade consistente entre os participantes.

Para verificar se essa diferença era estatisticamente significativa, foi aplicado o mesmo protocolo descrito na Seção 6.1. Como as condições de dependência entre as amostras e o tamanho da amostra permaneceram constantes, a verificação concentrou-se na normalidade dos resíduos, a fim de determinar o teste estatístico

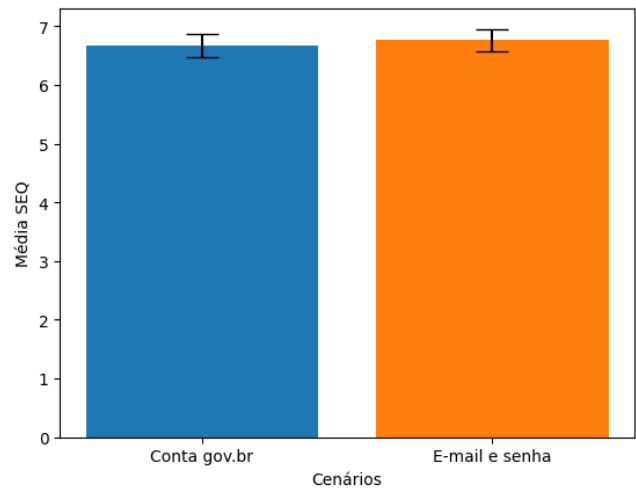


Figura 2: Média do SEQ para os cenários testados.

mais adequado para avaliar a hipótese de ausência de diferença significativa.

Foi aplicado o teste de Shapiro-Wilk, que indicou ausência de normalidade nos resíduos ($W = 0,7210$, $p = 0,0000032$). Diante dessa violação da normalidade, Bordens e Abbott [6] indicam o uso de testes não-paramétricos, em particular o Teste de Wilcoxon de Postos Sinalizados. Utilizando a biblioteca Pingouin [37], o teste retornou uma estatística $W = 15,0$ e *valor-p* de $0,3506$, o que indica que a diferença observada entre os métodos não é estatisticamente significativa. A ferramenta também analisa o efeito usando a Correlação Bisserial de Postos [9], reportando um tamanho de efeito pequeno ($RBC = -0,33$). Em outras palavras, os participantes não perceberam diferença prática na facilidade de uso entre os dois mecanismos de autenticação.

A próxima seção apresenta uma análise baseada em dados objetivos de desempenho, por meio da métrica de Tempo de Conclusão da Tarefa, permitindo uma avaliação mais abrangente da interação dos usuários com os dois métodos.

6.3 Tempo de Conclusão da Tarefa

Para medir o Tempo de Conclusão da Tarefa, foram analisadas as gravações das entrevistas realizadas por videoconferência. Como não era possível obter esse dado diretamente do sistema, os tempos foram extraídos manualmente a partir do instante em que a tela de login era apresentada até o carregamento da tela de sucesso da autenticação. Esse procedimento permitiu comparar a duração das tarefas em ambos os métodos de login com precisão de milissegundos. Os resultados estão apresentados na Figura 3.

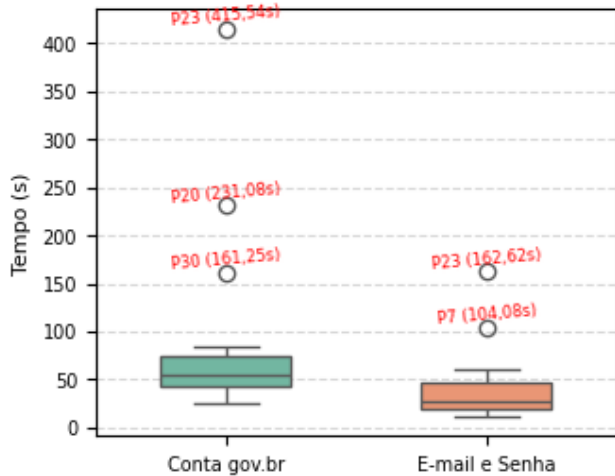


Figura 3: Distribuição dos Tempos de Conclusão da Tarefa.

O *boxplot* evidencia que o tempo médio foi significativamente maior para a Conta gov.br (76,74s) em relação ao método tradicional (38,32s). A mediana, considerada por Sauro e Lewis [31] como a medida mais adequada para comparação de tempos por ser menos sensível a valores extremos, também indica diferença expressiva: 55,06s para a Conta gov.br contra 27,46s para o login por e-mail. A variabilidade dos tempos foi maior na autenticação com a Conta gov.br, sugerindo maior inconsistência na experiência dos participantes. Ainda é possível destacar alguns *outliers*, com os participantes P23 (415,54s), P20 (231,08s) e P30 (161,25s), que levaram mais tempo com a Conta gov.br, e P7 (104,08s) e P23 (162,62s), com o método tradicional. Esses dados reforçam que, embora ambos os métodos tenham permitido a execução da tarefa, a Conta gov.br exigiu mais tempo e apresentou maior dispersão nos resultados.

A principal razão para a variação observada nos tempos de autenticação com a Conta gov.br está relacionada à sua natureza como solução de login único, que depende de uma infraestrutura externa para operar corretamente. Nos casos destacados como *outliers*, ocorreu indisponibilidade momentânea do serviço durante a realização dos testes, o que impactou diretamente os tempos de conclusão. Embora o ambiente de produção da Conta gov.br conte com mecanismos de alta disponibilidade, falhas pontuais ainda podem ocorrer e devem ser consideradas na análise de desempenho e nos critérios de adoção.

Mesmo que a diferença entre os métodos seja clara, foi feita a análise de significância estatística para fins de completude. Essa

diferença foi verificada por meio do Teste de Wilcoxon de Postos Sinalizados, já que o teste de Shapiro-Wilk indicou violação da normalidade nos resíduos ($W(30) = 0,69, p = 0,000010$). O resultado do Wilcoxon ($W = 34,0, p = 6,9179 \times 10^{-6}$) confirmou que a diferença observada é estatisticamente significativa. A magnitude do efeito ($RBC = 0,8538$) foi classificada como Muito Forte.

Em síntese, os resultados apontam similaridades no SUS e SEQ entre os métodos avaliados, mas uma discrepância relevante no tempo necessário para completar a autenticação. Esses achados reforçam a necessidade de considerar aspectos de desempenho, além da usabilidade percebida, ao decidir pela adoção de mecanismos de login unificado como a Conta gov.br. A próxima seção aprofunda a análise, explorando as correlações entre os resultados obtidos e os dados demográficos dos participantes, com o objetivo de identificar possíveis padrões que influenciem a experiência de uso.

6.4 Análise de Correlações

A análise de correlações buscou identificar possíveis associações entre as métricas de usabilidade (SUS, SEQ e TCT) e as variáveis demográficas dos participantes. Para isso, foi necessário escolher o coeficiente estatístico mais adequado, de acordo com a distribuição dos dados. Conforme apontado por Schober et al. [32], o coeficiente de Pearson é indicado para dados com distribuição normal e sem *outliers*, enquanto o de Spearman é mais apropriado quando essas condições não são atendidas. A conferência desse quesito foi feita com o teste de Shapiro-Wilk, e os resultados apontaram ausência de normalidade em todas as amostras, conforme mostra a Tabela 1.

Tabela 1: Dados do Teste de Shapiro-Wilk no conjunto de resultados.

Métrica	valor- <i>p</i>
SUS - Conta gov.br	$6,24 \times 10^{-3}$
SUS - E-mail e Senha	$3,03 \times 10^{-3}$
SEQ - Conta gov.br	$1,42 \times 10^{-7}$
SEQ - E-mail e Senha	$8,37 \times 10^{-9}$
TCT - Conta gov.br	$1,28 \times 10^{-8}$
TCT - E-mail e Senha	$3,17 \times 10^{-6}$

Como os dados não apresentaram distribuição normal, o coeficiente de Spearman foi utilizado para mensurar as correlações. A Figura 4 ilustra o mapa de calor obtido na análise de correlação das métricas de usabilidade coletadas: SUS, SEQ e TCT, considerando os dois métodos de autenticação analisados (Conta gov.br e E-mail e senha). As células da matriz são coloridas conforme a intensidade da correlação: tons mais quentes (vermelho) indicam correlações positivas mais fortes, enquanto tons mais frios (azul) indicam correlações negativas.

Dentre os achados, observou-se uma correlação positiva moderada entre SUS e SEQ (0,60 para a Conta gov.br e 0,52 para E-mail e Senha), indicando que participantes com melhor percepção geral de usabilidade também relataram maior facilidade no uso da interface. Em contrapartida, o TCT apresentou correlação negativa com o SUS no cenário da Conta gov.br (-0,26), sugerindo que tempos mais longos de autenticação podem reduzir ligeiramente a avaliação positiva de usabilidade.

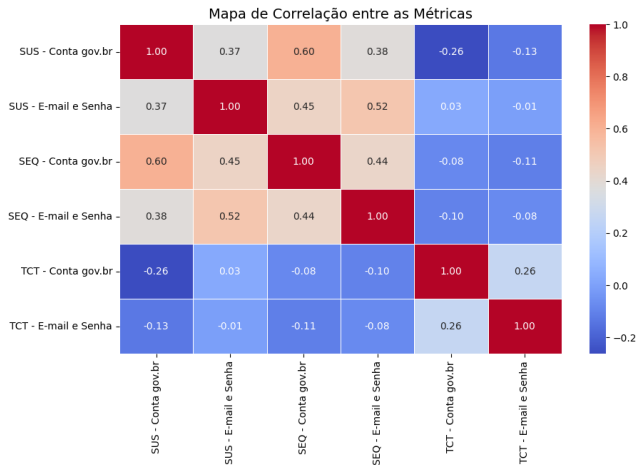


Figura 4: Correlação entre os resultados das métricas.

Também foi feita uma análise de correlação com as variáveis demográficas ordinais dos participantes (Faixa etária, Grau de instrução e Facilidade com tecnologia), com o objetivo de identificar possíveis padrões ou associações entre esses dados e as métricas de usabilidade. Os resultados dessa análise estão representados no mapa de calor da Figura 5, que apresenta o mesmo padrão de cores mencionado anteriormente e será discutido a seguir.

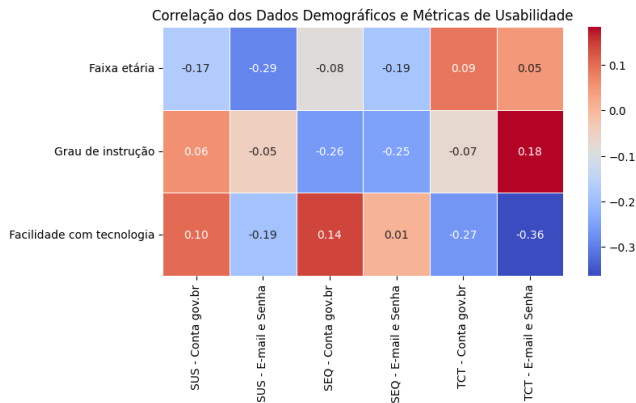


Figura 5: Correlação entre os dados demográficos ordinais e as métricas de usabilidade.

A variável que mais se destacou foi a “faixa etária”, que apresentou correlação negativa com o SUS no cenário de e-mail e senha (−0,29) e com o TCT da Conta gov.br (−0,19). Esses dados sugerem que participantes mais velhos tenderam a avaliar a usabilidade de forma menos favorável e a levar mais tempo para concluir as tarefas.

A familiaridade tecnológica também demonstrou influência sobre os resultados: apresentou correlação positiva com a SEQ da Conta gov.br (0,14) e correlação negativa com os tempos de autenticação nos dois métodos (até −0,36), indicando que usuários mais habituados ao uso de tecnologia perceberam maior facilidade

e agilidade. O grau de instrução, por sua vez, teve correlação negativa com o SUS e o SEQ da Conta gov.br (até −0,26), o que sugere avaliações mais exigentes por parte dos participantes com maior escolaridade.

Os achados apresentados reforçam a importância de considerar o perfil do público ao planejar estratégias de adoção do login unificado em serviços digitais — sobretudo no que diz respeito à comunicação e ao suporte oferecido a usuários com menor familiaridade tecnológica ou em faixas etárias mais elevadas. Na próxima seção, os resultados serão discutidos de forma integrada, considerando suas implicações práticas, limitações do estudo e sugestões para o aprimoramento do *design* de serviços públicos digitais.

7 Discussão

A análise dos dados demonstrou que a autenticação via Conta gov.br obteve avaliações de usabilidade e facilidade percebida semelhantes às do método tradicional de e-mail e senha. Esse resultado contrasta com o fluxo mais complexo da Conta gov.br, que envolve múltiplas etapas e, como evidenciado na análise do Tempo de Conclusão da Tarefa (Seção 6.3), requer mais tempo para ser concluído. Apesar disso, os participantes atribuíram escores elevados a ambos os métodos, indicando que o tempo adicional pode não ter comprometido de forma significativa a experiência geral. Essa percepção positiva, contudo, não elimina a necessidade de atenção ao desempenho da plataforma em diferentes contextos.

Ainda que a autenticação via Conta gov.br envolva um número maior de interações e transições de tela, os participantes atribuíram avaliações tão positivas quanto ao método por e-mail e senha. Essa aparente contradição sugere que a familiaridade prévia com a Conta gov.br pode ter neutralizado o impacto da complexidade do fluxo, indicando uma possível “naturalização” do esforço requerido. Alternativamente, pode-se considerar que as métricas utilizadas (SUS e SEQ), por mais consolidadas que sejam, não capturam nuances ligadas à carga cognitiva em fluxos já conhecidos pelos usuários. Essa hipótese reforça a importância de empregar métodos complementares, como observação comportamental ou carga de trabalho percebida, em futuras avaliações.

A triangulação entre dados subjetivos (SUS, SEQ), objetivos (TCT) e demográficos possibilitou a identificação de padrões relevantes. A idade dos participantes teve influência perceptível: usuários mais velhos atribuíram notas mais baixas de usabilidade e apresentaram maior tempo de conclusão, especialmente no uso da Conta gov.br. Já os participantes com maior familiaridade tecnológica avaliaram positivamente a solução de login único e foram mais rápidos na realização das tarefas. O grau de instrução também se mostrou relevante, sugerindo que indivíduos com maior escolaridade tendem a ser mais exigentes em suas avaliações. Esses achados indicam que o perfil do público-alvo pode interferir na experiência de uso e, portanto, deve ser considerado em estratégias de adoção.

Esses dados de correlação oferecem também subsídios práticos para decisões estratégicas. Por exemplo, a tendência de usuários mais velhos e com maior grau de instrução atribuírem escores ligeiramente mais baixos de usabilidade sugere que ações de sensibilização e comunicação adaptadas a diferentes perfis podem mitigar resistências. Do mesmo modo, o fato de usuários com maior fluência

digital avaliarem positivamente a solução SSO indica uma predisposição favorável entre públicos mais conectados — argumento que pode ser explorado por gestores na justificativa da integração.

Além disso, a coleta qualitativa de sugestões espontâneas por parte dos participantes evidenciou pontos importantes para aprimoramento. Foram mencionadas, por exemplo, a necessidade de explicações prévias ao login, a baixa visibilidade do botão “Sair”, o excesso de texto em algumas telas e a ausência de dicas visuais durante o processo de autenticação. Esses elementos reforçam que, mesmo diante de avaliações positivas, ajustes pontuais no fluxo e na comunicação da interface podem aprimorar significativamente a experiência do usuário.

Apesar da robustez estatística das análises, algumas ameaças à validade interna devem ser reconhecidas. A coleta em ambiente de homologação, com credenciais fictícias, pode ter reduzido a possibilidade de falhas reais de uso. A amostra, ainda que diversificada, não incluiu pessoas com escolaridade inferior ao ensino médio nem usuários totalmente inexperientes com a Conta gov.br, o que limita a validade externa e impede generalizações a públicos com menor letramento digital. Além disso, a percepção positiva registrada pode estar associada a viés de expectativa ou satisfação induzida por fatores como o contato direto com o pesquisador.

Finalmente, a investigação concentrou-se exclusivamente na etapa de autenticação, sem avaliar eventuais dificuldades nas etapas anteriores, como o cadastro na plataforma ou a recuperação de senha — aspectos que também influenciam a percepção de usabilidade. Outro ponto a ser considerado é a concentração dos participantes na região Sudeste do país, o que pode limitar a representatividade dos diferentes contextos de uso em outras regiões do Brasil. Essas limitações não invalidam os achados, mas reforçam a necessidade de replicação do estudo em cenários reais e com públicos mais amplos e diversos.

Do ponto de vista prático, os resultados sugerem que, quando direcionada a usuários já familiarizados com sua interface, a Conta gov.br pode ser integrada sem prejuízos relevantes à percepção de usabilidade. No entanto, o tempo adicional exigido para autenticação, aliado à complexidade do fluxo e à dependência de infraestrutura externa, impõe desafios que devem ser ponderados conforme o perfil do público-alvo e os requisitos do serviço digital. Ao reunir evidências sobre o desempenho comparado dos mecanismos e sobre os fatores que modulam a experiência de uso, este estudo oferece contribuições específicas e delimitadas, úteis para gestores públicos no processo de decisão sobre a adoção da Conta gov.br como solução de autenticação.

A análise dos dados também evidenciou uma lacuna importante: a documentação disponível acerca da Conta gov.br é predominantemente técnica, centrada na integração do mecanismo aos sistemas, sem incluir discussões baseadas em dados sobre os benefícios e desafios de sua adoção. Diante dessa ausência, foi elaborado o Guia Prático de Adoção da Conta gov.br, que reúne as informações obtidas de forma acessível e orientada a evidências, com o propósito de apoiar decisões mais informadas sobre o uso da plataforma em serviços digitais públicos. Conforme descrito na Seção 4.4, o material oferece aos gestores públicos um roteiro claro, fundamentado em resultados empíricos, para orientar o processo de integração da solução de autenticação em seus serviços.

8 Conclusão

Este estudo teve como objetivo principal avaliar a usabilidade da Conta gov.br em um serviço digital público, com vistas a subsidiar decisões estratégicas sobre sua adoção institucional. Diferentemente de análises exclusivamente técnicas ou normativas, a abordagem adotada buscou compreender a experiência real dos usuários em tarefas de autenticação, considerando não apenas métricas de desempenho, mas também fatores humanos e contextuais. Essa perspectiva contribui para um debate mais equilibrado sobre o uso de soluções de login único em ambientes governamentais.

Os resultados indicaram que a Conta gov.br apresentou usabilidade percebida semelhante à do método tradicional de e-mail e senha, mesmo exigindo um tempo significativamente maior para conclusão da tarefa — um aumento de 100,53%. Ainda assim, os participantes atribuíram escores elevados a ambos os métodos, sinalizando que o tempo adicional não comprometeu, por si só, a experiência geral. No entanto, esse achado ressalta a necessidade de atenção ao desempenho técnico da plataforma, especialmente em contextos de maior exigência operacional.

A condução da pesquisa envolveu o desenvolvimento de um protocolo experimental robusto, ancorado em revisão sistemática da literatura e nas boas práticas metodológicas da área. Esse protocolo pode ser reaproveitado em avaliações similares, inclusive adaptado a outros serviços digitais que venham a incorporar a Conta gov.br ou soluções tecnológicas equivalentes. Assim, o trabalho oferece não apenas um diagnóstico pontual, mas também uma base metodológica útil e replicável para investigações futuras.

Como principal desdobramento prático, foi desenvolvido um Guia para Adoção da Conta gov.br, estruturado com foco na clareza, aplicabilidade e concisão. O material reúne orientações técnicas, dados empíricos e representações visuais dos principais achados da pesquisa, visando apoiar gestores públicos na tomada de decisão. Disponibilizado em formato *markdown* para facilitar sua disseminação, o guia busca preencher a lacuna entre a produção científica e sua aplicação concreta, promovendo escolhas mais fundamentadas, seguras e centradas na experiência do usuário em serviços digitais públicos.

9 Agradecimentos

Agradecemos aos colegas do Senado Federal, em especial ao Instituto DataSenado, à Coordenação do Programa e-Cidadania e ao Serviço de Soluções para Portais, pelo apoio técnico e pela parceria que viabilizou parte essencial da pesquisa. Aos participantes dos testes de usabilidade, que gentilmente doaram seu tempo, deixo registrada minha sincera gratidão.

Este trabalho contou com o apoio de ferramentas de Inteligência Artificial Generativa para auxílio na redação, estruturação e revisão do texto. Em particular, foi utilizada a ferramenta ChatGPT da OpenAI, para apoiar a formulação de seções do artigo, adequação da linguagem acadêmica, refinamento da fluidez textual, elaboração de scripts estatísticos em Python e revisão de trechos com base em dados previamente analisados pelo autor.

REFERENCES

- [1] Alexandre Aragão, Lucas Machado, Nathyane Moreno, Davi Viana, Francisco Silva, Tércio Sousa, Luis Rivero, Ariel Teles, Arlindo F. Da Conceição, and Inaldo Costa. 2020. Evoluindo uma aplicação para Cidades Inteligentes através de Avaliação de Métricas de Qualidade e Usabilidade. *iSys - Brazilian Journal of Information Systems* 13, 3 (June 2020), 55–81. doi:10.5753/isys.2020.795
- [2] Aaron Bangor, Phil Kortum, and James Miller. 2009. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *J. Usability Stud.* 4 (April 2009), 114–123. <https://uxpajournal.org/determining-what-individual-sus-scores-mean-adding-an-adjective-rating-scale/>
- [3] J.M. Christian Bastien. 2010. Usability testing: a review of some methodological and technical aspects of the method. *International Journal of Medical Informatics* 79, 4 (April 2010), e18–e23. doi:10.1016/j.ijmedinf.2008.12.004
- [4] Nigel Bevan and Miles Macleod. 1994. Usability measurement in context. *Behaviour & Information Technology* 13, 1–2 (Jan. 1994), 132–145. doi:10.1080/01449299408914592
- [5] Jonas Blattgerste, Jan Behrends, and Thies Pfeiffer. 2022. A Web-Based Analysis Toolkit for the System Usability Scale. In *Proceedings of the 15th International Conference on Pervasive Technologies Related to Assistive Environments* (Corfu, Greece) (PETRA '22). Association for Computing Machinery, New York, NY, USA, 237–246. doi:10.1145/3529190.3529216
- [6] Kenneth S. Borden and Bruce B. Abbott. 2011. *Research design and methods: a process approach* (8th ed ed.). McGraw-Hill, New York. OCLC: ocn436028182.
- [7] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval. Ind.* 189 (11 1995).
- [8] Jacob Cohen. 2013. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed ed.). Taylor and Francis, Hoboken.
- [9] Edward E. Cureton. 1956. Rank-Biserial Correlation. *Psychometrika* 21, 3 (Sept. 1956), 287–290. doi:10.1007/BF02289138
- [10] Fred D. Davis. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* 13, 3 (Sept. 1989), 319. doi:10.2307/249008
- [11] Demétrius de Almeida Jubé. 2025. Guia Prático para Adoção da Conta gov.br em Serviços Públicos Digitais. <https://github.com/demetriusjube/guia-adocao-conta-govbr>. Disponível em: GitHub. Acesso em: out. 2025..
- [12] Secretaria de Governo Digital. 2024. Roteiro de Integração do Login Único – Roteiro de Integração do Login Único documentation. <https://acesso.gov.br/roteiro-tecnico/>
- [13] Departamento de Serviços Públicos Digitais. 2019. Por quem foi criado? – Dúvidas Frequentes da Conta gov.br. Disponível em: https://acesso.gov.br/faq/_perguntasdafaq/porquemfoicriado.html. Acesso em: 13 mai. 2025.
- [14] Nebojsa Djordjevic. 2017. Usability: Key characteristic of software quality. *Vojnotehnicki glasnik* 65, 2 (2017), 513–529. doi:10.5937/vojtehg65-11028 Publisher: Centre for Evaluation in Education and Science (CEON/CEES).
- [15] Governo Federal do Brasil. 2021. Termo de Uso e Política de Privacidade - gov.br. Disponível em: https://acesso.gov.br/faq/_downloads/d6442e668037599687c09290b6c5d6ef/TERMO_DE_USO_E_POLITICA_DE_PRIVACIDADE.pdf. Acesso em: 13 mai. 2025.
- [16] Governo Federal do Brasil. 2023. Estratégia de Governo Digital 2020–2022. Disponível em: <https://www.in.gov.br/web/dou/-/decreto-n-10.332-de-28-de-abril-de-2020-254430358>. Acesso em: 13 mai. 2025.
- [17] Governo Federal do Brasil. 2024. Serviço de Integração aos Produtos de Identidade Digital gov.br. <https://www.gov.br/governodigital/pt-br/transformacao-digital/servico-de-integracao-aos-produtos-de-identidade-digital-gov.br/servico-de-integracao-aos-produtos-de-identidade-digital-gov.br>
- [18] Governo Federal do Brasil. 2025. Onde usar a conta gov.br. <https://www.gov.br/governodigital/pt-br/identidade/conta-gov-br/onde-usar-a-conta-govbr/onde-usar-a-conta-gov-br>
- [19] Senado Federal. 2023. *Conheça os resultados e o funcionamento do Portal e-Cidadania*. Technical Report. <https://www.senado.gov.br/bi-arqs/Arquimedes/ecidadania/relatorio-simplificado-pdf.pdf>
- [20] Sinalva Ferreira Resende Júnior and Leticia Lopes Leite. 2024. Um estudo sobre problemas de usabilidade no software de processos administrativos eletrônicos do Governo Federal do Brasil. *iSys - Brazilian Journal of Information Systems* 17, 1 (June 2024). doi:10.5753/isys.2024.4141
- [21] Demétrius Jubé. 2025. Assessing Quality Through Use: A Usability-Based Evaluation of the Conta gov.br Authentication Platform - Material da pesquisa. doi:10.6084/M9.FIGSHARE.29631566 Artwork Size: 443910 Bytes Pages: 443910 Bytes.
- [22] Demétrius de Almeida Jubé and Leticia Lopes Leite. 2024. Usability in Single Sign-On Solutions: What to Evaluate? <https://doi.org/10.6084/m9.figshare.30278305>. Peer-reviewed paper accepted at CISTI 2024 (19th Iberian Conference on Information Systems and Technologies), Salamanca, Spain. Available in open access due to pending publication of conference proceedings.
- [23] Maina Korir, Simon Parkin, and Paul Dunphy. 2022. An Empirical Study of a Decentralized IdentityWallet: Usability, Security, and Perspectives on User Control. In *Proceedings of the 18th Symposium on Usable Privacy and Security*, SOUPS 2022. 195 – 211. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85140874129&partnerID=40&md5=d1ca37a97e5971895064184f9e41ce4> Type: Conference paper.
- [24] Kwang Bok Lee and R.A. Grice. 2004. Developing a new usability testing method for mobile devices. In *International Professional Communication Conference, 2004. IPCC 2004. Proceedings*. IEEE, Minneapolis, MN, USA, 115–127. doi:10.1109/IPCC.2004.1375285
- [25] Douglas Fabiano Lourenço, Elenice Carmona Valentim, and Maria Helena Bana de Moraes Lopes. 2022. Tradução e adaptação transcultural da System Usability Scale para o português do Brasil. *Aquichan* 22, 2 (maio 2022), e2228. doi:10.5294/aqui.2022.22.2.8
- [26] Mateus De Souza Monteiro, Gabrieli Oliveira Da Silva Batista, and Luciana Cardoso De Castro Salgado. 2023. Investigating usability pitfalls in Brazilian and Foreign governmental chatbots. *Journal on Interactive Systems* 14, 1 (July 2023), 331–340. doi:10.5753/jis.2023.3104
- [27] Iasonas Polakis, Panagiotis Ilia, Federico Maggi, Marco Lancini, Georgios Kontaxis, Stefano Zanero, Sotiris Ioannidis, and Angelos D. Keromytis. 2014. Faces in the Distorting Mirror: Revisiting Photo-Based Social Authentication. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS '14)*. Association for Computing Machinery, New York, NY, USA, 501–512. doi:10.1145/2660267.2660317 event-place: Scottsdale, Arizona, USA.
- [28] Jenny Preece, Yvonne Rogers, and Helen Sharp. 2002. *Interaction design: beyond human-computer interaction*. J. Wiley & Sons, New York, NY.
- [29] Jeffrey Rubin and Dana Chisnell. 2008. *Handbook of usability testing: how to plan, design, and conduct effective tests* (2nd ed ed.). Wiley Pub, Indianapolis, IN. OCLC: ocn212204392.
- [30] M. Angela Sasse, Michelle Steves, Kat Krol, and Dana Chisnell. 2014. The Great Authentication Fatigue – And How to Overcome It. In *Cross-Cultural Design*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Alfred Kobsa, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Demetri Terzopoulos, Doug Tygar, Gerhard Weikum, and P. L. Patrick Rau (Eds.). Vol. 8528. Springer International Publishing, Cham, 228–239. doi:10.1007/978-3-319-07308-8_23 Series Title: Lecture Notes in Computer Science.
- [31] Jeff Sauro and James R. Lewis. 2016. *Quantifying the user experience: practical statistics for user research* (2nd edition ed.). Elsevier, Morgan Kaufmann, Amsterdam Boston Heidelberg.
- [32] Patrick Schober, Christa Boer, and Lothar A. Schwarte. 2018. Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia* 126, 5 (May 2018), 1763–1768. doi:10.1213/ANE.0000000000002864
- [33] S. S. Shapiro and M. B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52, 3–4 (Dec. 1965), 591–611. doi:10.1093/biomet/52.3.4591
- [34] San-Tsai Sun, Eric Pospisil, Ildar Muslukhov, Nuray Dindar, Kirstie Hawkey, and Konstantin Beznosov. 2013. Investigating Users’ Perspectives of Web Single Sign-On: Conceptual Gaps and Acceptance Model. *ACM Transactions on Internet Technology* 13, 1 (Nov. 2013), 1–35. doi:10.1145/2532639
- [35] Donna P. Tedesco and Thomas S. Tullis. 2006. A Comparison of Methods for Eliciting Post-Task Subjective Ratings in Usability Testing. (2006).
- [36] Thomas Tullis and Jacqueline Stetson. 2006. A Comparison of Questionnaires for Assessing Website Usability. (June 2006).
- [37] Raphael Vallat. 2018. Pingouin: statistics in Python. *Journal of Open Source Software* 3, 31 (Nov. 2018), 1026. doi:10.21105/joss.01026
- [38] José Vieira and Rodrigo Andrade. 2024. On the investigation of Usability in the Caixa Tem Application. *Journal on Interactive Systems* 15, 1 (May 2024), 417–433. doi:10.5753/jis.2024.3889
- [39] Tianhao Wang, Huangyi Ge, Omar Chowdhury, Hemanta K. Maji, and Ninghui Li. 2016. On the Security and Usability of Segment-based Visual Cryptographic Authentication Protocols. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, Vienna Austria, 603–615. doi:10.1145/2976749.2978417