Usabilidade percebida e atributos de qualidade em um sistema acadêmico: reanálise do SAGRES/UESB e uma estrutura para organização de evidências do usuário

Camilo A. Mascarenhas de Almeida
Universidade Estadual do Sudoeste da Bahia
Vitória da Conquista – BA, Brasil
camiloalvesm@gmail.com

Francisco dos Santos Carvalho†
Universidade Estadual do Sudoeste da Bahia
Vitória da Conquista – BA, Brasil
franciscoscarvalho@uesb.edu.br

Gidevaldo Novais dos Santos
Universidade Estadual do Sudoeste da Bahia
Vitória da Conquista – BA, Brasil
gnovais@uesb.edu.br

Resumo—Este artigo retoma os dados de uma pesquisa realizada em 2022 sobre o SAGRES, sistema acadêmico utilizado na Universidade Estadual do Sudoeste da Bahia (UESB), para examinar como seus usuários perceberam a experiência de uso naquele período. A reanálise parte de uma distinção importante: dificuldades de interação, lentidão, estabilidade, confiança no serviço e cobertura funcional podem aparecer juntas na experiência do usuário, mas correspondem a atributos de natureza diferente. A partir das respostas de 43 participantes, o estudo revisita o instrumento original, separa os itens mais diretamente ligados à interação daqueles associados ao comportamento percebido do serviço e confronta esses resultados com os pontos fortes e fracos relatados livremente. Os achados apontam para uma experiência de interação predominantemente desfavorável, ao mesmo tempo em que os participantes reconhecem valor nas funções acadêmicas oferecidas pelo sistema. Para tornar essa leitura mais precisa, os resultados são organizados em três eixos de evidência: interação/usabilidade percebida, comportamento percebido do serviço e adequação funcional percebida. Os eixos não substituem as características da ISO/IEC 25010; funcionam como uma camada intermediária para organizar evidências produzidas pelos usuários antes de relacioná-las aos modelos de qualidade.

Palavras-chave— Usabilidade percebida. Sistemas acadêmicos. Qualidade de software. ISO/IEC 25010. SAGRES.

I. INTRODUÇÃO

Sistemas acadêmicos concentram atividades que fazem parte da rotina de estudantes e docentes, como consulta e lançamento de notas, registro de frequência, acesso a horários, emissão de documentos e acompanhamento de informações curriculares. Em muitas instituições, o uso desse tipo de sistema é necessário para executar tarefas acadêmicas e administrativas. Por isso, dificuldades de navegação, compreensão, desempenho ou disponibilidade podem afetar atividades obrigatórias mesmo quando as funções essenciais estão presentes.

A ISO 9241-11:2018 [1] define usabilidade em relação à capacidade de usuários específicos atingirem objetivos específicos com eficácia, eficiência e satisfação em determinado contexto de uso. A avaliação pode combinar observação de tarefas e medidas de autorrelato. Este trabalho se concentra na segunda perspectiva: examina o que os participantes declararam sobre o uso do SAGRES. Não foram realizados testes de tarefa, medições técnicas de tempo de resposta ou inspeção heurística.

O estudo deriva da monografia de Almeida [2], desenvolvida no curso de Ciência da Computação da UESB. A pesquisa original buscava conhecer a percepção dos usuários sobre a usabilidade do SAGRES. A reanálise preserva esse problema, mas distingue os itens mais diretamente relacionados à interação daqueles que tratam de características vizinhas de qualidade, como desempenho, confiabilidade e segurança. Essa separação evita tratar como uma única medida de usabilidade um conjunto de afirmações que cobre fenômenos diferentes.

A questão de pesquisa é: como os participantes avaliaram a usabilidade percebida do SAGRES e quais outros atributos de qualidade aparecem associados aos pontos fortes e fracos relatados? O objetivo geral é descrever a experiência de interação percebida no contexto investigado. Como objetivos específicos, busca-se identificar os itens e agrupamentos com avaliações mais favoráveis e desfavoráveis, analisar os pontos fortes e fracos relatados livremente, examinar a convergência entre as respostas fechadas e abertas e organizar essas evidências em uma estrutura analítica que diferencie interação, comportamento percebido do serviço e adequação funcional.

A monografia formulou a hipótese de que os usuários pesquisados não estavam satisfeitos com a usabilidade do SAGRES. Nesta reanálise, essa hipótese é examinada por um subconjunto de 23 itens semanticamente mais próximos de interação, facilidade de uso, aprendizagem, navegação, apresentação e satisfação subjetiva. A composição desse subconjunto foi definida post hoc a partir do conteúdo dos itens e das fronteiras conceituais da ISO/IEC 25010:2011 [3]; portanto, o resultado oferece apoio empírico à hipótese no caso estudado, mas não constitui teste confirmatório pré-registrado nem permite generalização para todos os usuários da UESB.

Como contribuição analítica, o artigo não propõe novas características de qualidade nem uma alternativa à ISO/IEC 25010. A contribuição está na forma de organizar evidências de percepção do usuário produzidas por um instrumento heterogêneo. Os três eixos funcionam como uma camada intermediária entre os relatos dos participantes e as características normativas: separam evidências sobre a interação, sobre o comportamento percebido do serviço e sobre a adequação funcional antes de sua interpretação à luz

[1] †In memoriam. O professor Francisco Carvalho faleceu em 20 de maio de 2025. A pesquisa da qual resultou este trabalho foi realizada com a sua participação em todas as etapas. Fica, nesta publicação, uma homenagem e a lembrança do docente e pesquisador que muito contribuiu para a formação professional de muitos bacharéis em Ciência da Computação.

dos modelos de qualidade. Essa organização permite mostrar o que a análise original de 2022 não distinguia explicitamente.

Os dados se referem especificamente a março de 2022. O artigo deve, portanto, ser lido como registro da experiência dos participantes naquele período e como linha de base histórica para avaliações posteriores do sistema.

II. TRABALHOS RELACIONADOS

Demirkol et al. [4] avaliaram a usabilidade percebida de um sistema de informação acadêmica em uma universidade turca com 324 estudantes, utilizando o System Usability Scale (SUS) e questões sobre aspectos apreciados e criticados. No Brasil, Costa, Santos e Santiago [5] aplicaram o SUS a 101 usuários do Q-Acadêmico e obtiveram escore de 67,22. Khalida e Pamungkas [6] também recorreram ao SUS em um estudo de sistema acadêmico orientado ao redesenho da solução.

Estudos mais recentes mostram que a avaliação de sistemas acadêmicos continua recorrendo a instrumentos padronizados, mas com resultados bastante distintos entre contextos. Al-Tahat [7] utilizou A-UMUX para avaliar o Student Information System da Arab Open University e reportou escore global de 75,37. Nambo et al. [8] combinaram diferentes técnicas alinhadas à ISO/IEC 25010 na avaliação de um sistema acadêmico, empregando SUS especificamente para usabilidade e ferramentas distintas para desempenho, confiabilidade, segurança e outras características. Caroline et al. [9], por sua vez, obtiveram SUS de 60,69 na avaliação do SIAM da Universitas Mulia e identificaram necessidade de melhorias de interface apesar de o sistema atender às funções acadêmicas.

Esses resultados não constituem benchmarks diretamente comparáveis ao SAGRES. SUS e A-UMUX têm regras próprias de composição e interpretação, enquanto o questionário analisado neste artigo foi construído especificamente para a pesquisa de 2022. A literatura recente é relevante, sobretudo, por reforçar duas decisões metodológicas desta reanálise: não converter o escore próprio em SUS e separar problemas de interação de características técnicas ou de qualidade que exigem formas de medição distintas.

As revisões de Vlachogianni e Tselios [10], [11] reforçam a ampla utilização de instrumentos padronizados na avaliação percebida de tecnologias educacionais. Em conjunto, os trabalhos relacionados mostram que medidas de autorrelato são úteis para registrar a experiência do usuário, mas devem ser interpretadas de acordo com o instrumento aplicado e, quando possível, complementadas por medidas observacionais ou técnicas.

III. REFERENCIAL TEÓRICO

A. *Usabilidade e avaliação percebida*

A ISO 9241-11:2018 [1] define usabilidade a partir de três resultados do uso: eficácia, eficiência e satisfação, sempre relacionados a usuários, objetivos e contexto específicos. Essa definição é importante porque impede tratar usabilidade como uma propriedade isolada da interface. Um mesmo sistema pode produzir experiências diferentes conforme a tarefa, o perfil do usuário e as condições em que o uso ocorre.

A percepção do usuário é uma fonte de evidência relevante, mas não substitui medições observacionais. Quando um participante afirma que o sistema é lento, por exemplo, o estudo registra uma percepção de lentidão. Sem telemetria ou medição direta, não é possível concluir qual era o tempo efetivo de resposta do serviço. De modo semelhante, confiança ou insegurança declarada em relação aos dados não constitui auditoria técnica de segurança. Essa distinção orienta a terminologia adotada neste artigo.

B. *Qualidade de produto e ISO/IEC 25010*

A ISO/IEC 25010:2011 [3] organizava a qualidade de produto de software em oito características: Adequação Funcional, Eficiência de Desempenho, Compatibilidade, Usabilidade, Confiabilidade, Segurança, Manutenibilidade e Portabilidade. O questionário foi aplicado em 2022, quando essa edição ainda era a referência vigente, e por isso ela é usada como base principal para interpretar os atributos citados nas respostas abertas.

A edição de 2023 [12] atualizou o modelo de qualidade de produto e passou a trabalhar com nove características. A antiga característica Usability (Usabilidade) do modelo de produto foi renomeada para Interaction Capability (Capacidade de Interação). Entre suas subcaracterísticas estão reconhecimento de adequação, capacidade de aprendizagem, operabilidade, proteção contra erros do usuário, engajamento do usuário, inclusividade, assistência ao usuário e autodescrição. A revisão é útil para interpretar os dados coletados anteriormente, mas não altera retrospectivamente o conteúdo do questionário aplicado em 2022.

A associação dos itens e das categorias qualitativas à ISO/IEC 25010 é usada neste estudo como referência interpretativa, não como prova de que o instrumento mede formalmente cada característica da norma. Alguns itens têm correspondência mais clara do que outros. As afirmações sobre tempo de resposta, por exemplo, aproximam-se de Eficiência de Desempenho; já itens como “mentalmente estimulante” representam uma experiência subjetiva mais ampla e não têm equivalência direta com uma única subcaracterística da norma.

C. *Questionários de usabilidade e satisfação*

QUIS, SUMI, WAMMI e SUS são instrumentos conhecidos de avaliação percebida de sistemas. Eles diferem em número de itens, fatores avaliados, escalas de resposta e procedimentos de interpretação. O SUS, proposto por Brooke [13], tornou-se especialmente difundido por sua brevidade e por possuir ampla documentação psicométrica [14].

O questionário do SAGRES foi construído a partir da discussão de Filardi e Traina [15] sobre fatores contemplados por instrumentos desse tipo. Entretanto, não se trata de tradução ou aplicação de SUS, QUIS, SUMI ou WAMMI. Suas 28 afirmações foram redigidas para o contexto do SAGRES. Consequentemente, os escores apresentados neste artigo não podem ser convertidos em escores SUS nem herdam automaticamente as propriedades de validade e confiabilidade desses instrumentos.

D. *Sistemas de uso mandatório*

O contexto de uso também influencia a interpretação dos resultados. O Technology Acceptance Model [16] e suas extensões discutem relações entre utilidade percebida, facilidade de uso, intenção comportamental e uso. Estudos como Brown et al. [17], Venkatesh e Davis [18] e Dečman [19] mostram que ambientes de uso mandatório exigem

cautela, porque o uso efetivo pode ocorrer mesmo quando a atitude ou a intenção do usuário é desfavorável.

O SAGRES se aproxima desse contexto: estudantes e docentes dependem do sistema para atividades acadêmicas específicas. Essa literatura é utilizada para interpretar a diferença entre usar o sistema e avaliá-lo favoravelmente. O presente estudo não aplicou uma escala TAM ou UTAUT e, portanto, não mede diretamente utilidade percebida, facilidade de uso percebida ou intenção comportamental.

E. Três eixos para organização de evidências percebidas

A ISO/IEC 25010 já distingue características como Usabilidade, Eficiência de Desempenho, Confiabilidade, Segurança e Adequação Funcional. O problema tratado aqui, portanto, não é a ausência dessas distinções no modelo normativo. Ele surge na passagem entre o relato do usuário e a interpretação do pesquisador: uma mesma avaliação pode reunir dificuldade de navegação, lentidão, indisponibilidade e ausência de funcionalidades, e um questionário construído sem delimitação explícita pode agregar esses fenômenos sob um único rótulo de "usabilidade". A organização em três eixos foi adotada para preservar a natureza das evidências antes de associá-las às características da norma.

O Eixo 1 – Interação/usabilidade percebida reúne evidências sobre facilidade de uso, aprendizagem, navegação, apresentação, ajuda, mensagens e experiência subjetiva durante a interação. O Eixo 2 – Comportamento percebido do serviço reúne percepções sobre tempo de resposta, estabilidade, disponibilidade, persistência de dados, comportamento anômalo e confiança declarada em segurança; essas evidências não substituem medições técnicas. O Eixo 3 – Adequação funcional percebida reúne menções sobre presença, ausência, completude e utilidade prática das funções acadêmicas. Menções relacionadas a características que não se encaixam nesses três focos, como Portabilidade e Compatibilidade, permanecem classificadas separadamente e não são forçadas em um dos eixos.

Os três eixos não constituem solução fatorial, escala psicométrica ou simplificação exaustiva da ISO/IEC 25010. A função é metodológica: responder, separadamente, a três perguntas que exigem evidências diferentes — como o usuário interage com o sistema; como percebe o comportamento do serviço durante o uso; e em que medida reconhece que o sistema oferece as funções necessárias. No questionário fechado de 2022, os dois primeiros focos estavam representados pelas 28 afirmações, enquanto a adequação funcional apareceu principalmente nas questões abertas. Essa assimetria, invisível quando todo o material é tratado genericamente como usabilidade, torna-se explícita na reanálise.

IV. PROCEDIMENTOS METODOLÓGICOS

A. Delineamento

A pesquisa é exploratória e descritiva, de corte transversal, e utiliza um levantamento por questionário aplicado em um único contexto institucional. O caso empírico é o SAGRES na UESB. A análise combina dados quantitativos das questões fechadas com análise qualitativa das respostas abertas sobre pontos fortes e fracos.

Esta versão reanalisa duas fontes preservadas da pesquisa de 2022: a exportação original do Google Formulários, que contém as 43 respostas individuais aos 28 itens e às questões complementares, e a planilha de codificação auditável das respostas abertas. A monografia de origem [2] foi consultada para recuperar o contexto, a formulação inicial do problema e a descrição do instrumento.

B. Participantes, coleta e aspectos éticos

A exportação original registra 43 respostas, recebidas entre 10 e 27 de março de 2022. A amostragem foi não probabilística e baseada em disponibilidade. O convite circulou principalmente em grupos vinculados ao curso de Ciência da Computação. Por essa razão, a amostra não representa estatisticamente a população de usuários do SAGRES na UESB. A participação voluntária também permite viés de autoseleção, já que pessoas com experiências particularmente positivas ou negativas podem ter maior motivação para responder.

O formulário não solicitava nome, matrícula, endereço eletrônico ou outro identificador direto, nem exigia autenticação. Como o instrumento solicitava avaliação de um sistema/serviço sem identificação individual dos participantes, a pesquisa foi tratada, nos termos da Resolução CNS nº 510/2016 [20], como pesquisa de opinião pública com participantes não identificados. A ausência de autenticação impede excluir completamente respostas repetidas; como verificação limitada, não foram encontrados dois registros com o mesmo padrão completo de respostas aos 28 itens.

C. Instrumento

O instrumento reunia 28 afirmações em escala de concordância de cinco pontos, questões objetivas complementares e duas questões abertas. Nesta reanálise, as questões objetivas utilizadas descrevem a importância atribuída a esse tipo de sistema, a autoavaliação de familiaridade com o SAGRES e uma nota geral de 0 a 10. As questões abertas pediam até três pontos fortes e até três pontos fracos.

As 28 afirmações eram balanceadas por polaridade: 14 positivas e 14 negativas. Os itens negativos foram invertidos na análise para que, em todos os casos, valores maiores representassem avaliação mais favorável. O instrumento foi revisado pelo orientador antes da aplicação, mas não passou por painel de especialistas, estudo piloto formal ou validação psicométrica prévia.

Para organizar os resultados, os itens foram agrupados posteriormente em seis dimensões:

- D1 – Desempenho e comportamento temporal: itens 1, 3 e 10;
- D2 – Consistência, previsibilidade e navegação: itens 7, 8, 12, 13, 16, 17, 21 e 22;
- D3 – Aprendizagem, ajuda e documentação: itens 2, 4, 5, 14, 18, 19, 25 e 28;
- D4 – Estética e conforto de uso: itens 6, 9, 11, 15, 20, 23 e 24;
- D5 – Segurança percebida: item 26;
- D6 – Facilidade de uso global: item 27.

Esse agrupamento é analítico e post hoc. Ele não corresponde a uma estrutura fatorial previamente testada. Para a organização em três eixos, adotou-se como regra o núcleo semântico predominante de cada afirmação. Itens cujo foco principal é a ação do usuário e a interação com a

interface foram classificados no Eixo 1; itens que descrevem o comportamento percebido do sistema ou do serviço, no Eixo 2; e menções sobre presença, ausência ou completude de funções, no Eixo 3. Quando uma evidência admite mais de uma leitura plausível, a ambiguidade é registrada em vez de tratada como correspondência normativa inequívoca.

As 28 afirmações foram redigidas para o contexto do SAGRES a partir da cobertura temática discutida por Filardi e Traina [15], sem constituir tradução ou reprodução de itens padronizados. Como essa origem é comum ao conjunto, o Quadro 1 se concentra na dimensão de análise, na polaridade, no eixo analítico e na correspondência teórica principal de cada afirmação.

Quadro 1 – As 28 afirmações do questionário, dimensão, polaridade, eixo analítico e correspondência teórica

Nº	Afirmação	Dim.	Pol.	Eixo	Correspondência principal no referencial
1	O SAGRES responde muito demoradamente às entradas de dados.	D1	Neg.	E2	ISO 25010: Eficiência de Desempenho / comportamento temporal
2	Aprender a operar o SAGRES é muito complicado.	D3	Neg.	E1	ISO 25010: Usabilidade / apreensibilidade
3	Demora muito para executar as funções existentes no SAGRES.	D1	Neg.	E2	ISO 25010: Eficiência de Desempenho / comportamento temporal
4	As informações do SAGRES são apresentadas de modo claro e compreensível.	D3	Pos.	E1	ISO 25010: Usabilidade / reconhecimento de adequabilidade
5	A documentação de ajuda (help) do SAGRES é muito clara e objetiva.	D3	Pos.	E1	ISO 25010: Usabilidade / apreensibilidade; Nielsen H10
6	Operar as funções do SAGRES é mentalmente estimulante.	D4	Pos.	E1	ISO 9241-11: satisfação e experiência subjetiva
7	O SAGRES é inconsistente.	D2	Neg.	E1	ISO 25010: Usabilidade / operacionalidade; Nielsen H4
8	O SAGRES é complicado quando quero fazer algo que não é padrão.	D2	Neg.	E1	ISO 25010: Usabilidade / operacionalidade; Nielsen H7
9	Usar o SAGRES é frustrante.	D4	Neg.	E1	ISO 9241-11: satisfação
10	A velocidade de processamento das inclusões de dados e/ou consultas no SAGRES é satisfatória.	D1	Pos.	E2	ISO 25010: Eficiência de Desempenho / comportamento temporal
11	Há momentos que me sinto tenso(a) ao usar o SAGRES.	D4	Neg.	E1	ISO 9241-11: satisfação
12	A organização das informações no SAGRES é bastante lógica.	D2	Pos.	E1	ISO 25010: Usabilidade / reconhecimento de adequabilidade; Nielsen H6
13	Há muitas etapas necessárias para que algo funcione no SAGRES.	D2	Neg.	E1	ISO 25010: Usabilidade / operacionalidade; Nielsen H7
14	As mensagens de erro não são adequadas.	D3	Neg.	E1	ISO 25010: Usabilidade / proteção contra erros; Nielsen H9
15	O SAGRES apresenta-se de uma forma muito atrativa.	D4	Pos.	E1	ISO 25010: Usabilidade / estética da interface; Nielsen H8
16	O SAGRES ocasionalmente se comporta de uma maneira que não pode ser entendida.	D2	Neg.	E2	ISO 25010: Confiabilidade / comportamento percebido; Nielsen H4
17	É fácil visualizar rapidamente quais são as opções em cada função do SAGRES.	D2	Pos.	E1	ISO 25010: Usabilidade / reconhecimento de adequabilidade; Nielsen H6
18	Eu tenho que procurar assistência (suporte técnico) na maioria das vezes quando uso o SAGRES.	D3	Neg.	E1	ISO 25010: Usabilidade / operacionalidade; Nielsen H10
19	O SAGRES precisa de mais explicações introdutórias.	D3	Neg.	E1	ISO 25010: Usabilidade / apreensibilidade; Nielsen H10
20	As telas do SAGRES são muito atraentes.	D4	Pos.	E1	ISO 25010: Usabilidade / estética da interface; Nielsen H8
21	É muito fácil encontrar o que procuro no SAGRES.	D2	Pos.	E1	ISO 25010: Usabilidade / operacionalidade; Nielsen H6
22	O SAGRES tem um ótimo sequenciamento das suas telas.	D2	Pos.	E1	ISO 25010: Usabilidade / operacionalidade; Nielsen H4
23	O tamanho das letras no SAGRES é adequado.	D4	Pos.	E1	ISO 25010: Usabilidade / acessibilidade
24	As cores das telas do SAGRES são adequadas.	D4	Pos.	E1	ISO 25010: Usabilidade / estética da interface
25	As mensagens de ajuda (help) no SAGRES são insuficientes.	D3	Neg.	E1	ISO 25010: Usabilidade / apreensibilidade; Nielsen H10
26	Confio na segurança do SAGRES.	D5	Pos.	E2	ISO 25010: Segurança / confiança declarada
27	Considero fácil usar o SAGRES.	D6	Pos.	E1	ISO 25010: Usabilidade / operacionalidade
28	Precisei aprender com terceiros como usar as funções do SAGRES.	D3	Neg.	E1	ISO 25010: Usabilidade / apreensibilidade; Nielsen H10

Fonte: elaboração própria a partir do questionário original e do referencial. E1 = Interação/usabilidade percebida; E2 = Comportamento percebido do serviço. O E3 = Adequação funcional percebida não possui item fechado específico e é recuperado principalmente pelas questões abertas. As correspondências teóricas são interpretativas e não implicam validação formal de cada item como medida da característica indicada. As heurísticas de Nielsen utilizadas no quadro seguem a formulação apresentada em [22].

Quadro 2 – Critérios operacionais para organização das evidências nos três eixos

Eixo / situação	Critério operacional	Base teórica utilizada	Limite ou ambiguidade principal
E1 – Interação/usabilidade percebida	Núcleo semântico centrado na interação usuário-interface: compreensão, aprendizagem, navegação, etapas, ajuda, mensagens, apresentação e resposta subjetiva durante o uso.	ISO 9241-11 (satisfação e contexto de uso) e Usabilidade/Capacidade de Interação da ISO/IEC 25010.	Itens afetivos podem ultrapassar a operabilidade estrita; são mantidos quando a experiência é evocada diretamente durante a interação.
E2 – Comportamento percebido do serviço	Núcleo semântico centrado no modo como o sistema/serviço se comporta: velocidade, estabilidade, disponibilidade, persistência, anomalias e confiança declarada em segurança.	Eficiência de Desempenho, Confiabilidade e Segurança da ISO/IEC 25010.	Pode haver sobreposição com previsibilidade da interação; a classificação segue o objeto principal da frase e não implica diagnóstico técnico.
E3 – Adequação funcional percebida	Menção à existência, ausência, completude ou utilidade prática de funções necessárias às atividades acadêmicas.	Adequação Funcional da ISO/IEC 25010.	Não equivale à utilidade percebida do TAM. Dificuldade de encontrar uma função existente é E1; ausência ou incompletude da função é E3.
Fora dos três eixos	Evidência relevante para outra característica que não corresponde diretamente aos três focos analíticos.	Demais características da ISO/IEC 25010.	Portabilidade e Compatibilidade são preservadas separadamente; os três eixos não pretendem cobrir todo o modelo de qualidade.

Fonte: elaboração própria. Os eixos organizam evidências percebidas e não substituem a classificação normativa da ISO/IEC 25010.

D. Tratamento quantitativo

As cinco alternativas da escala foram codificadas de 1 a 5. Nos itens de polaridade negativa, aplicou-se a transformação 6 - resposta, de modo que valores maiores sempre representassem avaliação mais favorável. A média dos 28 itens compôs o escore global de favorabilidade ao sistema. Esse índice é mantido como descrição geral, mas não é denominado escala de usabilidade porque inclui itens de Eficiência de Desempenho, Confiabilidade e Segurança percebida.

Para examinar a hipótese original sobre usabilidade, o Eixo 1 – Interação/usabilidade percebida foi operacionalizado com 23 itens. Os itens 1, 3 e 10, relacionados predominantemente a comportamento temporal, o item 16, relacionado ao comportamento percebido de confiabilidade, e o item 26, referente à confiança declarada em segurança, foram mantidos separadamente no Eixo 2 – Comportamento percebido do serviço. A classificação é post hoc e não representa uma estrutura fatorial validada. O Eixo 3 – Adequação funcional percebida não recebe escore fechado porque o questionário não continha afirmações específicas para esse domínio; sua análise decorre das respostas abertas.

Alguns itens exigem registro explícito de ambiguidade. O item 6 (“mentalmente estimulante”) não corresponde diretamente a uma subcaracterística específica da ISO/IEC 25010:2011 [3]; foi mantido no Eixo 1 porque expressa uma resposta subjetiva produzida durante a interação, aproximando-se da dimensão de satisfação da ISO 9241-11. O item 7 (“o SAGRES é inconsistente”) foi mantido no Eixo 1 por ter sido interpretado como consistência da interação, embora a redação seja ampla. O item 16 (“se comporta de uma maneira que não pode ser entendida”) foi classificado no Eixo 2 porque o objeto gramatical da avaliação é o comportamento do sistema, mas também pode refletir previsibilidade e consistência da interação. O item 26 foi mantido no Eixo 2 porque registra confiança percebida em segurança, e não usabilidade ou segurança técnica medida.

Como checagem de robustez dessas decisões, o núcleo de interação foi recalculado em três cenários alternativos: incluindo o item 16 no Eixo 1; retirando o item 6 do Eixo 1; e realizando as duas alterações simultaneamente. Em todos os cenários, 33 dos 43 participantes permaneceram abaixo do ponto neutro, o alfa de Cronbach permaneceu entre 0,92 e 0,93 e a comparação com o ponto neutro continuou significativa ($p < 0,001$). Assim, a interpretação substantiva da experiência de interação não depende da classificação adotada para esses dois itens limítrofes.

Foram calculadas frequências, médias, medianas e desvios padrão. A consistência interna foi avaliada pelo alfa de Cronbach para os agrupamentos com múltiplos itens, para o núcleo de 23 itens e para o conjunto de 28 itens. O alfa é interpretado apenas como indicador de consistência interna, não como evidência de unidimensionalidade ou validade de construto.

O núcleo de 23 itens e o escore global foram comparados ao ponto neutro 3 pelo teste de postos sinalizados de Wilcoxon; teste “t” de uma amostra e “d” de Cohen foram apresentados como análises complementares. A relação entre os escores e a nota geral de 0 a 10 foi examinada por correlações de Pearson e Spearman. Como todas as medidas foram obtidas no mesmo instrumento e no mesmo momento, as correlações são interpretadas como coerência entre formas de avaliação, e não como validação externa.

E. Tratamento qualitativo

As respostas abertas foram submetidas à análise de conteúdo com base em Bardin [21]. A unidade de registro foi a menção temática. Uma resposta contendo três atributos distintos podia gerar três categorias. Para cada respondente, uma mesma categoria foi contada no máximo uma vez, ainda que aparecesse repetidamente no texto.

As categorias de primeiro nível foram produzidas a partir do conteúdo das respostas e, em seguida, associadas de forma interpretativa às características da ISO/IEC 25010:2011 [3]. Para tornar o procedimento mais transparente e reprodutível, as categorias utilizadas na reanálise foram formalizadas em um codebook apresentado no Apêndice A. Para cada categoria, o codebook registra definição operacional, regra de inclusão e distinção em relação a categorias próximas, exemplo de resposta e associação teórica.

A codificação foi realizada por um único codificador. A formalização do codebook ocorreu retrospectivamente, a partir das categorias e decisões empregadas na reanálise, com a finalidade de explicitar o procedimento e permitir auditoria; ela não constitui validação independente da codificação. As ambiguidades foram resolvidas por quatro regras: segmentar respostas com mais de um atributo em menções temáticas; preferir a categoria mais específica; permitir codificação múltipla quando a mesma menção expressasse atributos diferentes de forma inseparável; e manter categoria residual apenas quando não houvesse atributo interpretável. Também se distinguiu ausência de funcionalidade (E3) de dificuldade para localizá-la ou utilizá-la (E1), e lentidão (Eficiência de Desempenho) de indisponibilidade/interrupção (Confiabilidade).

Nas citações diretas apresentadas nos resultados, a grafia e a construção sintática dos participantes foram preservadas. Eventuais erros de escrita fazem parte do registro original e não foram corrigidos.

F. Uso de ferramentas de inteligência artificial na reanálise

Ferramentas de inteligência artificial generativa foram utilizadas como apoio na organização dos procedimentos de reanálise, conferência de consistência dos cálculos e revisão da redação. As decisões sobre classificação, interpretação dos resultados, limitações e versão final do texto permanecem sob responsabilidade do autor.

V. RESULTADOS

A. Caracterização geral do caso

O SAGRES é usado para atividades acadêmicas de estudantes e docentes. Entre as funções descritas na monografia estão consulta de notas, frequência, horários e disciplinas, emissão ou acesso a documentos e solicitações acadêmicas; docentes também realizam lançamentos e registros relacionados às disciplinas.

Na questão sobre importância, 21 participantes (48,8%) classificaram esse tipo de sistema como extremamente importante e 19 (44,2%) como importante. Três respostas (7,0%) ficaram nas opções de pouca ou nenhuma importância. Quanto à familiaridade com o SAGRES, 4 participantes (9,3%) se declararam muito experientes e técnicos, 11 (25,6%) experientes mas não técnicos, 25

(58,1%) afirmaram conseguir lidar com a maioria das funções e 3 (7,0%) avaliaram a maioria das funções como difícil de usar. Esses resultados descrevem a amostra, mas não são usados como medida de utilidade percebida no sentido do TAM.

B. Escores e consistência interna

A Tabela 1 resume os agrupamentos analíticos, o núcleo de interação/usabilidade e o conjunto completo de 28 itens. Entre as seis dimensões post hoc, a menor média ocorreu em D2, Consistência, previsibilidade e navegação (2,19), seguida de D1, Desempenho e comportamento temporal (2,33). D3, Aprendizagem, ajuda e documentação, apresentou a maior média entre os agrupamentos com múltiplos itens (2,79), ainda abaixo do ponto neutro.

Tabela 1 – Escores e consistência interna (n = 43)

Agrupamento	Itens (k)	Alfa de Cronbach	Média	DP	Mediana
D1 – Desempenho e comportamento temporal	3	0,79	2,33	1,03	2,00
D2 – Consistência, previsibilidade e navegação	8	0,88	2,19	0,84	2,00
D3 – Aprendizagem, ajuda e documentação	8	0,77	2,79	0,74	2,88
D4 – Estética e conforto de uso	7	0,76	2,45	0,81	2,29
D5 – Segurança percebida	1	n.a.	2,44	1,40	2,00
D6 – Facilidade de uso global	1	n.a.	2,60	1,24	2,00
Núcleo de interação / usabilidade	23	0,92	2,49	0,71	2,48
Escore global de favorabilidade	28	0,93	2,46	0,72	2,39

Fonte: elaboração própria a partir da exportação original do questionário. Para D5 e D6, compostas por um único item, alfa não se aplica. Valores arredondados a duas casas decimais.

A consistência interna foi elevada tanto no núcleo de 23 itens (alfa = 0,92) quanto no conjunto completo (alfa = 0,93). Esses resultados não demonstram que os itens formem um único fator latente. O questionário foi construído para o estudo e as dimensões não foram definidas por análise fatorial.

C. Resultados por item

A Tabela 2 apresenta os 28 itens recalculados diretamente da exportação original do questionário, incluindo a distribuição completa das cinco alternativas de resposta. As colunas Favorável, Neutro e Desfavorável são apresentadas adicionalmente para facilitar a leitura e já consideram a polaridade do item: em uma afirmação negativa, discordar é contabilizado como avaliação favorável.

Tabela 2 – Distribuição completa das respostas e escores de favorabilidade por item (n = 43)

Item	Conteúdo abreviado	DT	DP	NN	CP	CT	Favorável	Neutro	Desfavorável	Média	DP
1	Responde demoradamente às entradas de dados	1	7	3	19	13	8 (18,6%)	3 (7,0%)	32 (74,4%)	2,16	1,11
2	Aprender a operar é complicado	5	11	11	12	4	16 (37,2%)	11 (25,6%)	16 (37,2%)	3,02	1,18
3	Demora para executar as funções	4	2	5	17	15	6 (14,0%)	5 (11,6%)	32 (74,4%)	2,14	1,23
4	Informações claras e compreensíveis	7	10	8	11	7	18 (41,9%)	8 (18,6%)	17 (39,5%)	3,02	1,35
5	Documentação de ajuda clara e objetiva	8	6	26	2	1	3 (7,0%)	26 (60,5%)	14 (32,6%)	2,58	0,93
6	Operar o sistema é mentalmente estimulante	21	9	8	4	1	5 (11,6%)	8 (18,6%)	30 (69,8%)	1,98	1,14
7	Sistema inconsistente	3	3	9	8	20	6 (14,0%)	9 (20,9%)	28 (65,1%)	2,09	1,27
8	Complicado para realizar algo fora do padrão	0	0	6	16	21	0 (0,0%)	6 (14,0%)	37 (86,0%)	1,65	0,72
9	Uso frustrante	3	4	7	14	15	7 (16,3%)	7 (16,3%)	29 (67,4%)	2,21	1,23
10	Velocidade de processamento satisfatória	9	14	6	9	5	14 (32,6%)	6 (14,0%)	23 (53,5%)	2,70	1,34
11	Tensão durante o uso	7	5	3	10	18	12 (27,9%)	3 (7,0%)	28 (65,1%)	2,37	1,53
12	Organização lógica das informações	6	16	10	8	3	11 (25,6%)	10 (23,3%)	22 (51,2%)	2,67	1,15
13	Muitas etapas para concluir uma ação	1	6	9	12	15	7 (16,3%)	9 (20,9%)	27 (62,8%)	2,21	1,15
14	Mensagens de erro inadequadas	0	3	19	11	10	3 (7,0%)	19 (44,2%)	21 (48,8%)	2,35	0,92
15	Apresentação visual atrativa	23	11	2	4	3	7 (16,3%)	2 (4,7%)	34 (79,1%)	1,86	1,17
16	Comportamento por vezes incompreensível	3	6	5	15	14	9 (20,9%)	5 (11,6%)	29 (67,4%)	2,28	1,26
17	Opções facilmente visualizáveis	13	19	3	3	5	8 (18,6%)	3 (7,0%)	32 (74,4%)	2,26	1,29
18	Necessidade frequente de suporte técnico	17	11	6	6	3	28 (65,1%)	6 (14,0%)	9 (20,9%)	3,77	1,31
19	Necessidade de mais explicações introdutórias	1	4	7	13	18	5 (11,6%)	7 (16,3%)	31 (72,1%)	2,00	1,09
20	Telas atrativas	22	9	8	4	0	4 (9,3%)	8 (18,6%)	31 (72,1%)	1,86	1,04
21	Facilidade para encontrar o que procura	10	19	4	9	1	10 (23,3%)	4 (9,3%)	29 (67,4%)	2,35	1,13
22	Bom sequenciamento de telas	18	14	6	4	1	5 (11,6%)	6 (14,0%)	32 (74,4%)	1,98	1,08
23	Tamanho das letras adequado	6	6	8	11	12	23 (53,5%)	8 (18,6%)	12 (27,9%)	3,40	1,40
24	Cores adequadas	3	8	12	7	13	20 (46,5%)	12 (27,9%)	11 (25,6%)	3,44	1,30
25	Mensagens de ajuda insuficientes	2	4	20	7	10	6 (14,0%)	20 (46,5%)	17 (39,5%)	2,56	1,10
26	Confiança na segurança	15	10	7	6	5	11 (25,6%)	7 (16,3%)	25 (58,1%)	2,44	1,40
27	Facilidade de uso global	9	15	5	12	2	14 (32,6%)	5 (11,6%)	24 (55,8%)	2,60	1,24
28	Necessidade de aprender com terceiros	13	5	2	15	8	18 (41,9%)	2 (4,7%)	23 (53,5%)	3,00	1,57

Fonte: elaboração própria a partir da exportação original do Google Formulários. DT = discordo totalmente; DP = discordo parcialmente; NN = nem discordo nem concordo; CP = concordo parcialmente; CT = concordo totalmente. As cinco primeiras colunas preservam a orientação original da resposta. Favorável, Neutro e Desfavorável consideram a polaridade do item; nos itens negativos, discordar representa avaliação favorável. Média e DP referem-se aos escores de favorabilidade recodificados de 1 a 5.

O item mais desfavorável foi o 8, relativo à dificuldade para realizar algo fora do padrão do sistema (média 1,65). Nenhum participante discordou dessa afirmação; 37 concordaram total ou parcialmente e 6 permaneceram neutros. Também se destacaram negativamente a atratividade da apresentação e das telas (itens 15 e 20; média 1,86), o sequenciamento de telas (item 22; 1,98) e a

necessidade de mais explicações introdutórias (item 19; 2,00).

Os resultados mais favoráveis ocorreram no item 18, sobre não precisar procurar suporte técnico na maioria das vezes (3,77), na adequação das cores (3,44), no tamanho das letras (3,40) e nos itens 2 e 4, ambos com média 3,02. O item 28 ficou exatamente no ponto neutro (3,00).

A neutralidade não se distribuiu uniformemente. Ela foi particularmente alta na documentação de ajuda (item 5: 60,5%), nas mensagens de ajuda (item 25: 46,5%) e nas mensagens de erro (item 14: 44,2%). Esses percentuais podem representar tanto ausência de opinião quanto baixo contato dos participantes com esses recursos; o desenho não permite distinguir as duas explicações.

Entre os 23 itens do núcleo de interação/usabilidade, 17 apresentaram média abaixo de 3, um ficou exatamente em 3 e cinco ficaram acima de 3. Como a exportação original preserva as respostas individuais, o subconjunto também pôde ser analisado por participante; seus resultados agregados são apresentados na Seção V.D.

D. Núcleo de usabilidade, escore global e nota geral

O núcleo de interação/usabilidade percebida, composto por 23 itens, apresentou média de 2,49 (DP = 0,71), mediana de 2,48, mínimo de 1,22 e máximo de 4,09. Trinta e três dos 43 participantes (76,7%) ficaram abaixo do ponto neutro e dez ficaram acima. O teste de Wilcoxon indicou diferença em relação a 3 ($p < 0,001$). O teste "t" de uma amostra apresentou resultado compatível, $t(42) = -4,69$; $p < 0,001$, com "d" de Cohen = -0,72. O item 27, que pergunta diretamente se o participante considera o SAGRES fácil de usar, também apresentou avaliação predominantemente desfavorável: média 2,60 e 24 participantes (55,8%) no lado desfavorável da escala.

Esses resultados oferecem apoio empírico à hipótese formulada na monografia de que os usuários pesquisados estavam insatisfeitos com a usabilidade do sistema. A interpretação, contudo, é limitada ao caso: a seleção dos 23 itens foi realizada nesta reanálise, a amostra é não probabilística e o instrumento não possui validação prévia como escala de usabilidade.

O escore global de favorabilidade, formado pelos 28 itens, apresentou média de 2,46 (DP = 0,72), mediana de 2,39, mínimo de 1,21 e máximo de 4,14. Trinta e quatro participantes (79,1%) ficaram abaixo de 3, dois exatamente em 3 e sete acima. A comparação com o ponto neutro também resultou em $p < 0,001$; no teste t, $t(42) = -4,92$, com $d = -0,75$. Esse escore é interpretado como avaliação ampla do sistema, pois incorpora desempenho, confiabilidade e segurança percebidos além da interação.

A nota geral atribuída ao SAGRES teve média de 4,93 em 10 (DP = 2,22), mediana igual a 5 e amplitude de 0 a 9. Portanto, o item de nota única se concentrou em torno do ponto médio da escala, sendo menos claramente desfavorável do que os escores compostos. O núcleo de 23 itens correlacionou-se positivamente com a nota geral ($r = 0,647$; $rho = 0,577$; $p < 0,001$ em ambos os casos), e o escore global apresentou associação ainda maior ($r = 0,697$; $rho = 0,621$; $p < 0,001$). As associações mostram coerência entre os formatos de avaliação, sem constituir validação externa independente.

E. Pontos fortes e fracos nas respostas abertas

As 43 respostas sobre pontos fortes produziram 67 ocorrências codificadas. Cinco participantes afirmaram explicitamente não identificar ponto forte. As respostas sobre pontos fracos produziram 114 ocorrências codificadas, incluindo duas respostas sem atributo específico. A diferença de volume é descritiva e depende do protocolo de segmentação e codificação; não é tratada como uma escala quantitativa de satisfação.

Tabela 3 – Categorias codificadas nos pontos fortes

Categoria codificada	Menções	% dos respondentes	Associação interpretativa à ISO/IEC 25010:2011 [3]
Operacionalidade e facilidade de uso básica	9	20,9	Usabilidade
Registro e consulta de notas e frequência	8	18,6	Adequação Funcional
Concentração das informações acadêmicas necessárias	8	18,6	Adequação Funcional
Nenhum ponto forte identificado	5	11,6	não aplicável
Adequação das cores	5	11,6	Usabilidade
Horários e calendário acadêmico	4	9,3	Adequação Funcional
Emissão e acesso a documentos	4	9,3	Adequação Funcional
Cumpre a maioria das funções a que se propõe	3	7,0	Adequação Funcional
Possibilidade de personalizar a tela	3	7,0	Usabilidade
Ser um sistema web	2	4,7	Portabilidade*
Planos de curso e conteúdo das aulas	2	4,7	Adequação Funcional
Facilidade de login e acesso inicial	2	4,7	Usabilidade
Segurança percebida do sistema	2	4,7	Segurança
Privacidade e sigilo dos dados acadêmicos	2	4,7	Segurança
Outros aspectos estéticos da interface	2	4,7	Usabilidade
Armazenamento e importação de dados de períodos anteriores	1	2,3	Adequação Funcional
Atendimento a solicitações acadêmicas	1	2,3	Adequação Funcional
Fluxograma do curso	1	2,3	Adequação Funcional
Disponibilidade	1	2,3	Confiabilidade
Bom desempenho condicionado a estar operante	1	2,3	Confiabilidade
Rapidez e eficiência	1	2,3	Eficiência de Desempenho

*A associação de "ser um sistema web" à Portabilidade é apenas aproximativa e não implica medição formal da característica normativa.

Tabela 4 – Categorias codificadas nos pontos fracos

Categoria codificada	Menções	% dos respondentes	Associação interpretativa à ISO/IEC 25010:2011 [3]
Instabilidade e interrupções durante o uso	14	32,6	Confiabilidade
Lentidão e tempo de resposta	14	32,6	Eficiência de Desempenho
Estética da interface e legibilidade	13	30,2	Usabilidade
Sistema pouco intuitivo e confuso	13	30,2	Usabilidade
Navegação, excesso de etapas e dificuldade de localizar funções	12	27,9	Usabilidade
Indisponibilidade e congestionamento do serviço	7	16,3	Confiabilidade
Funcionalidades ausentes ou incompletas	6	14,0	Adequação Funcional
Poluição e repetição de informações	6	14,0	Usabilidade
Ajuda, documentação e suporte ao usuário	5	11,6	Usabilidade
Insegurança percebida	4	9,3	Segurança
Perda de dados e erros de persistência	4	9,3	Confiabilidade
Pré-requisitos, prazos e limites de tempo de acesso	4	9,3	Adequação Funcional
Versão móvel deficiente	3	7,0	Portabilidade
Inconsistência do sistema	3	7,0	Usabilidade
Demora na atualização das informações	2	4,7	Eficiência de Desempenho
Resposta sem atributo especificado	2	4,7	não aplicável
Design não simplificado para inclusão de dados	1	2,3	Usabilidade
Falta de integração de dados acadêmicos	1	2,3	Compatibilidade

Os comentários ajudam a compreender o conteúdo das categorias. Entre os pontos fortes aparecem afirmações como “Facilidade na consulta das notas, facilidade em emitir comprovante de matrícula e só tem o horário das aulas logo no começo. É fácil acessar o fluxograma do curso.”, “Dá pra ver notas frequência de aulas e planos de curso” e “Tem todas as informações acadêmicas que eu preciso”. Também foram citados privacidade e sigilo dos dados.

Nos pontos fracos, são recorrentes descrições de indisponibilidade, lentidão e navegação: “O sagres recorrentemente tem os serviços indisponíveis.”, “Lentidão extrema; volta e meia não funciona, cai e não se consegue acessar ou operar.” e “As janelas que as opções abrem são um pouco confusas e geralmente eu preciso dá muita volta para achar o que eu quero”. Há ainda críticas à estética e à documentação, além de relatos de perda de dados e problemas de persistência.

Para observar o padrão em nível mais amplo, as categorias foram agregadas pelas características de qualidade às quais foram associadas. Excluindo as respostas sem atributo classificável, foram contabilizadas 62 ocorrências positivas e 112 negativas.

Tabela 5 – Distribuição das ocorrências codificadas por característica de qualidade

Característica	Pontos fortes	% dos fortes identificáveis	Pontos fracos	% dos fracos identificáveis
Adequação Funcional	32	51,6	10	8,9
Usabilidade	21	33,9	53	47,3
Segurança	4	6,5	4	3,6
Portabilidade	2	3,2	3	2,7
Confiabilidade	2	3,2	25	22,3
Eficiência de Desempenho	1	1,6	16	14,3
Compatibilidade	0	0,0	1	0,9
Total	62	100,0	112	100,0

Fonte: elaboração própria. A associação à ISO/IEC 25010:2011 [3] é interpretativa e não corresponde a uma escala normativa aplicada aos participantes.

A Adequação Funcional concentrou 51,6% das ocorrências positivas com atributo identificável. Já Usabilidade, Confiabilidade e Eficiência de Desempenho somaram 83,9% das ocorrências negativas. O padrão mostra que os participantes reconheceram funções acadêmicas úteis ao mesmo tempo em que relataram problemas importantes na interação e no comportamento do serviço.

VI. DISCUSSÃO

A. A hipótese original e a experiência de interação

A disponibilidade da matriz original de respostas permite separar a análise da interação de outros atributos de qualidade. No núcleo de 23 itens, a média foi 2,49 e 33 dos 43 participantes ficaram abaixo do ponto neutro. Além disso, o item 27, que pergunta diretamente sobre facilidade de uso, teve média de 2,60 e maioria de respostas no lado desfavorável. Em conjunto, esses resultados sustentam a interpretação de que a experiência de interação foi predominantemente negativa entre os participantes pesquisados em 2022.

Esse resultado deve ser lido junto à natureza da medida. O núcleo de 23 itens foi definido nesta reanálise, depois da coleta, para excluir afirmações cujo conteúdo se aproxima mais de desempenho, confiabilidade ou segurança. Seu alfa de Cronbach elevado indica consistência interna, mas não transforma o subconjunto em uma escala validada. A conclusão mais segura é que diferentes indicadores do próprio questionário convergem para uma percepção desfavorável de facilidade de uso e interação nesta amostra.

A nota global de 0 a 10 oferece uma nuance importante. Sua média foi 4,93 e a mediana 5, isto é, em torno do centro da escala. Isso evita interpretar toda forma de avaliação como igualmente negativa. O contraste sugere que um item único de avaliação geral e um conjunto detalhado de questões podem captar níveis diferentes de julgamento. Ainda assim, as correlações positivas mostram que participantes com avaliações mais favoráveis nos itens tenderam a atribuir notas gerais maiores.

B. Desempenho e confiabilidade na experiência relatada

Os participantes também relataram problemas que não pertencem exclusivamente à usabilidade em sentido estrito. Os itens de comportamento temporal apresentaram média de 2,33, e as respostas abertas destacaram lentidão, instabilidade, indisponibilidade e problemas de persistência. Essas ocorrências ajudam a explicar por que um usuário pode avaliar mal a experiência global mesmo quando consegue compreender determinadas telas ou executar parte das tarefas.

É importante manter a distinção entre percepção e medição técnica. O estudo não registrou latência, disponibilidade do serviço, erros de banco de dados ou indicadores de segurança. Portanto, os dados não demonstram que o SAGRES tinha determinado nível técnico de desempenho ou confiabilidade; demonstram que esses problemas foram percebidos e relatados com frequência pelos participantes. Essa separação é consistente com abordagens recentes que avaliam usabilidade e outras características da ISO/IEC 25010 por instrumentos distintos, como no trabalho de Nambo et al. [8].

C. Os três eixos como camada de organização das evidências

A combinação entre respostas fechadas e abertas permite aplicar ao caso a estrutura proposta na Seção III.E. No Eixo 1 – Interação/usabilidade percebida, o núcleo de 23 itens apresentou média de 2,49 e 33 participantes ficaram abaixo do ponto neutro. As respostas abertas reforçam esse padrão ao concentrar críticas em navegação, excesso de etapas, baixa intuitividade, estética e ajuda ao usuário.

No Eixo 2 – Comportamento percebido do serviço, os itens de desempenho temporal apresentaram média de 2,33; o item sobre comportamento incompreensível do sistema teve média de 2,28; e o item sobre confiança em segurança, 2,44. Nas respostas abertas, lentidão, instabilidade, indisponibilidade e perda de dados também aparecem de forma recorrente. Esses resultados registram percepções dos participantes e não substituem telemetria, teste de carga, monitoramento de disponibilidade ou auditoria de segurança.

O Eixo 3 – Adequação funcional percebida aparece de forma diferente porque não havia afirmações fechadas específicas sobre cobertura funcional. Ele emerge sobretudo nas questões abertas: Adequação Funcional concentrou 51,6% das ocorrências positivas com atributo de qualidade identificável e 8,9% das negativas. Foram valorizadas funções como consulta de notas e frequência, acesso a horários e documentos e concentração de informações acadêmicas. Os percentuais qualitativos não são diretamente comparáveis às médias dos escores, mas tornam visível um domínio que o bloco fechado não media.

A utilidade dos três eixos não está em criar novas características de qualidade, pois essas distinções já são contempladas pela ISO/IEC 25010. O ganho está em evitar que evidências heterogêneas sejam agregadas antes da interpretação normativa. No caso do SAGRES, essa organização mostra que a avaliação desfavorável do bloco fechado se concentra principalmente em interação e comportamento percebido do serviço, enquanto a principal fonte de avaliações positivas nas respostas abertas está na cobertura funcional. Portabilidade e Compatibilidade, quando mencionadas, permanecem fora dos três eixos e são preservadas como características próprias. Assim, os eixos funcionam como uma camada de organização do diagnóstico, e não como novo modelo de qualidade.

D. O que a reanálise acrescenta à análise de 2022

A comparação com a monografia de 2022 ajuda a delimitar o avanço produzido pela reanálise. O conjunto empírico é o mesmo, mas a forma de tratar e relacionar as evidências foi alterada. A monografia apresentou as frequências dos itens e categorizou pontos fortes e fracos dentro de uma discussão ampla de usabilidade; a reanálise recupera a matriz original, recodifica os itens por polaridade, separa atributos que não pertencem ao mesmo construto e explicita as regras da análise qualitativa.

Quadro 3 – Comparação entre a análise da monografia de 2022 e a reanálise atual

Aspecto	Análise de 2022	Reanálise do artigo / conhecimento adicional
Leitura dos 28 itens	Os resultados foram apresentados principalmente por frequências e discutidos de forma ampla como usabilidade.	Os itens são recodificados por polaridade e separados entre interação e comportamento percebido do serviço, evitando atribuir desempenho, confiabilidade e segurança à usabilidade.
Hipótese sobre usabilidade	A hipótese foi considerada confirmada pelo predomínio de avaliações desfavoráveis no conjunto do questionário.	A hipótese é examinada por um núcleo de 23 itens semanticamente mais próximo de interação/usabilidade, com análise de consistência, testes estatísticos e checagem de sensibilidade para itens limítrofes.
Respostas abertas	Pontos fortes e fracos foram categorizados e descritos como parte do diagnóstico geral.	O codebook explicita regras de classificação e a agregação mostra que a Adequação Funcional concentra as avaliações positivas, enquanto interação, desempenho e confiabilidade concentram as negativas.
Relação entre evidências	Os diferentes tipos de achado eram discutidos no mesmo enquadramento geral.	A reanálise mostra a coexistência entre reconhecimento das funções acadêmicas e experiência desfavorável de interação/comportamento do serviço.
Contribuição	Diagnóstico descritivo da percepção dos participantes em 2022.	Organização metodológica mais precisa do mesmo corpus, mostrando como evidências percebidas distintas podem ser separadas antes de sua associação aos modelos de qualidade.

Fonte: elaboração própria a partir de Almeida [2] e da reanálise dos dados originais.

O novo conhecimento produzido não está na identificação de atributos inexistentes na ISO/IEC 25010, mas na demonstração empírica de que o diagnóstico muda quando as evidências são preservadas por natureza antes da agregação. No conjunto original, uma medida única poderia sugerir apenas "baixa usabilidade"; na reanálise, torna-se possível distinguir uma experiência de interação desfavorável, percepções negativas sobre o comportamento do serviço e, simultaneamente, reconhecimento da adequação de várias funções acadêmicas.

E. Relação com avaliações recentes de sistemas acadêmicos

Os resultados de estudos recentes mostram ampla variação na avaliação percebida de sistemas acadêmicos. Costa, Santos e Santiago [5] reportaram SUS de 67,22 no Q-Acadêmico. Al-Tahat [7] encontrou A-UMUX de 75,37 no SIS da Arab Open University. Nambo et al. [8] obtiveram SUS de 68 na avaliação do SIAT e separaram a usabilidade de testes específicos de outras características da ISO/IEC 25010. Caroline et al. [9] reportaram SUS de 60,69 para o

SIAM da Universitas Mulia, com indicação de melhorias necessárias na interface.

Esses valores não permitem posicionar numericamente o SAGRES acima ou abaixo dos demais sistemas, porque o questionário desta pesquisa não é SUS nem A-UMUX. A comparação relevante é metodológica: estudos recentes favorecem instrumentos padronizados quando o objetivo é produzir escores comparáveis e, em avaliações mais amplas de qualidade, combinam medidas de usabilidade com testes específicos de desempenho, confiabilidade e segurança. Isso reforça a interpretação adotada nesta reanálise e indica um caminho para futuras avaliações do SAGRES.

F. Implicações para novas avaliações e para melhoria do sistema

A organização das evidências em três eixos oferece um roteiro para novas avaliações do SAGRES sem substituir os modelos de qualidade existentes. O Eixo 1 pode ser investigado com instrumento padronizado de usabilidade e testes observacionais de tarefas; o Eixo 2 requer telemetria e testes técnicos de tempo de resposta, disponibilidade, persistência e segurança; e o Eixo 3 demanda levantamento funcional com diferentes perfis de usuário para verificar cobertura, completude e adequação das funções. Características não cobertas pelos três eixos, como Portabilidade e Compatibilidade, devem continuar sendo avaliadas por métodos próprios quando fizerem parte do objetivo do estudo.

Os dados de 2022 não permitem definir qual intervenção teria maior retorno, pois não foram medidos custos, esforço de implementação ou impacto causal de mudanças. Eles permitem, contudo, identificar problemas percebidos que merecem investigação. A alta neutralidade nos itens de ajuda e mensagens também sugere uma questão específica para estudos futuros: verificar se os usuários conhecem esses recursos, quando os utilizam e se eles de fato auxiliam na conclusão das tarefas.

VII. LIMITAÇÕES

A principal limitação é a amostragem. Os 43 participantes foram recrutados por disponibilidade, com concentração no contexto de Ciência da Computação. Os percentuais observados não podem ser generalizados para toda a população de usuários do SAGRES. A participação voluntária também introduz possibilidade de autoseleção.

O instrumento é outro limite importante. As 28 afirmações foram construídas especificamente para a pesquisa, sem validação psicométrica prévia, painel formal de especialistas ou estudo piloto com amostra independente. As seis dimensões e o núcleo de 23 itens são agrupamentos post hoc. Embora apresentem consistência interna elevada em alguns casos, os coeficientes alfa não demonstram validade de construto nem unidimensionalidade. Os três eixos também resultam de uma reorganização analítica posterior à coleta e não pretendem constituir um modelo de qualidade, escala ou partição exaustiva da ISO/IEC 25010. O estudo tampouco dispõe de amostra adequada para estabelecer estrutura fatorial dos itens.

A exportação original do questionário permite recalcular as análises quantitativas e confirma a completude dos 28 itens, mas o formulário não exigia autenticação. Não foram identificados padrões completos de resposta duplicados, porém essa verificação não exclui a possibilidade de uma mesma pessoa ter respondido mais de uma vez com respostas diferentes.

Não houve teste de tarefa, inspeção heurística nem medição técnica de tempo de resposta, disponibilidade ou segurança. Assim, lentidão, instabilidade, perda de dados e insegurança são percepções declaradas, não diagnósticos técnicos do sistema. A análise qualitativa foi realizada por um único codificador, sem medida de concordância entre avaliadores. O codebook apresentado no Apêndice A aumenta a transparência das definições e decisões de classificação, mas não substitui uma avaliação independente da confiabilidade da codificação.

Por fim, os dados foram coletados entre 10 e 27 de março de 2022. Alterações posteriores no SAGRES não estão refletidas nos resultados. O período de início de semestre também pode ter influenciado a frequência de relatos sobre congestionamento ou indisponibilidade. A contribuição do estudo é, assim, documentar a experiência percebida naquele contexto e fornecer uma base para comparações futuras, não caracterizar o estado atual do sistema.

VIII. CONSIDERAÇÕES FINAIS

A reanálise dos dados originais indica que a experiência de interação com o SAGRES foi predominantemente desfavorável entre os 43 participantes pesquisados em março de 2022. O núcleo de 23 itens mais diretamente relacionados à interação/usabilidade apresentou média de 2,49 em 5, e 33 participantes ficaram abaixo do ponto neutro. O item que perguntava diretamente sobre facilidade de uso também apresentou maioria de respostas desfavoráveis. Esses resultados oferecem apoio empírico, no caso estudado, à hipótese formulada na monografia original.

A análise mais ampla mostra que a avaliação do sistema não pode ser reduzida à interface. Desempenho temporal, instabilidade e disponibilidade aparecem com frequência nas respostas, enquanto os pontos fortes se concentram sobretudo nas funções acadêmicas oferecidas. Adequação Funcional reuniu 51,6% das ocorrências positivas classificáveis; Usabilidade, Confiabilidade e Eficiência de Desempenho somaram 83,9% das negativas.

A nota geral de 0 a 10 ficou em torno do ponto médio da escala, o que recomenda cautela ao descrever a avaliação global como uniformemente negativa. O resultado mais consistente é a presença simultânea de funções acadêmicas reconhecidas como úteis e de problemas percebidos na interação e no comportamento do serviço.

Como contribuição analítica, a reanálise utiliza três eixos para organizar evidências percebidas antes de relacioná-las às características da ISO/IEC 25010: interação/usabilidade percebida, comportamento percebido do serviço e adequação funcional percebida. A contribuição não está em criar novas características de qualidade, mas em explicitar a natureza de cada evidência e evitar que resultados diferentes sejam agregados prematuramente sob o rótulo de usabilidade. A comparação com a monografia de 2022 mostra que essa reorganização produz um diagnóstico mais específico do mesmo corpus.

O estudo não permite generalização para todos os usuários da UESB nem descreve o SAGRES atual. Seu valor está em registrar de forma auditável a experiência de uma amostra em 2022, mostrar como decisões de

classificação alteram a precisão do diagnóstico e oferecer um procedimento documentado para organizar evidências percebidas de natureza diferente. Avaliações futuras podem avançar com instrumentos padronizados, amostras mais amplas, múltiplos codificadores, testes observacionais de tarefas e métricas objetivas de desempenho, disponibilidade e segurança.

REFERÊNCIAS

[1] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 9241-11:2018 – Ergonomics of human-system interaction – Part 11: Usability: definitions and concepts. Geneva: ISO, 2018.

[2] ALMEIDA, C. A. M. de. Conhecimentos de interação em busca de melhorias no processo de desenvolvimento de sistemas de software: um estudo de caso à luz da análise de usabilidade. 2022. Monografia (Bacharelado em Ciência da Computação) – Universidade Estadual do Sudoeste da Bahia, Vitória da Conquista, 2022.

[3] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION; INTERNATIONAL ELECTROTECHNICAL COMMISSION. ISO/IEC 25010:2011 – Systems and software engineering – SQuaRE – System and software quality models. Geneva: ISO/IEC, 2011.

[4] DEMIRKOL, D.; SENELER, C.; DAIM, T.; SHAYGAN, A. Measuring perceived usability of university students towards a student information system (SIS): a Turkish university case. Technology in Society, v. 62, 101281, 2020. DOI: 10.1016/j.techsoc.2020.101281.

[5] COSTA, R. P. da; SANTOS, A. F. S. dos; SANTIAGO, C. P. Análise de usabilidade do sistema Q-Acadêmico utilizando o método System Usability Scale (SUS): um estudo de caso. In: ENCONTRO UNIFICADO DE COMPUTAÇÃO DO PIAUÍ (ENUCOMPI), 14., 2021. Anais [...]. Porto Alegre: SBC, 2021. p. 231-238. DOI: 10.5753/enucompi.2021.17776.

[6] KHALIDA, R.; PAMUNGKAS, R. W. P. Enhancing usability of the academic information system at Bhayangkara University: a design thinking and system usability approach. PIKSEL: Penelitian Ilmu Komputer Sistem Embedded and Logic, v. 11, n. 2, p. 373-382, 2023. DOI: 10.33558/piksel.v11i2.7239.

[7] AL-TAHAT, K. S. Using A-UMUX to Measure Perceived Usability of the Student Information System (SIS) at Arab Open University. In: 2025 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA). IEEE, 2025. DOI: 10.1109/ACDSA65407.2025.11166652.

[8] NAMBO, I.; PAKAYA, N.; AMALI, L. N.; OLII, S. Evaluasi kualitas sistem informasi akademik terpadu (SIAT) menggunakan ISO/IEC 25010. Jambura Journal of Informatics, v. 7, n. 2, p. 143-153, 2025. DOI: 10.37905/jji.v1i2.33720.

[9] CAROLINE, F. B.; ATIKAH, A. N. N.; SERVANDA, Y.; GANDARRITYAZ, R. Evaluasi Usability SIAM Universitas Mulia dengan Metode SUS. Jurnal Teknologi Informasi dan Ilmu Komputer, v. 2, n. 1, p. 18-25, 2026. DOI: 10.65258/jutekom.v2.i1.40.

[10] VLACHOGIANNI, P.; TSELIOS, N. Perceived usability evaluation of educational technology using the System Usability Scale (SUS): a systematic review. Journal of Research on Technology in Education, v. 54, n. 3, p. 392-409, 2022. DOI: 10.1080/15391523.2020.1867938.

[11] VLACHOGIANNI, P.; TSELIOS, N. Perceived usability evaluation of educational technology using the Post-Study System Usability Questionnaire (PSSUQ): a systematic review. Sustainability, v. 15, n. 17, 12954, 2023. DOI: 10.3390/su151712954.

[12] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION; INTERNATIONAL ELECTROTECHNICAL COMMISSION. ISO/IEC 25010:2023 – Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Product quality model. Geneva: ISO/IEC, 2023.

[13] BROOKE, J. SUS: a "quick and dirty" usability scale. In: JORDAN, P. W. et al. (Eds.). Usability Evaluation in Industry. London: Taylor & Francis, 1996. p. 189-194.

[14] LEWIS, J. R. The System Usability Scale: past, present, and future. International Journal of Human-Computer Interaction, v. 34, n. 7, p. 577-590, 2018. DOI: 10.1080/10447318.2018.1455307.

[15] FILARDI, A. L.; TRAINA, A. J. M. Montando questionários para medir a satisfação do usuário: avaliação de interface de um sistema que utiliza técnicas de recuperação de imagens por conteúdo. In: SIMPÓSIO BRASILEIRO DE FATORES HUMANOS EM SISTEMAS COMPUTACIONAIS, 8., 2008, Porto Alegre. Anais [...]. Porto Alegre: SBC, 2008. p. 176-185.

[16] DAVIS, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, v. 13, n. 3, p. 319-340, 1989. DOI: 10.2307/249008.

[17] BROWN, S. A.; MASSEY, A. P.; MONTOYA-WEISS, M. M.; BURKMAN, J. R. Do I really have to? User acceptance of mandated technology. European Journal of Information Systems, v. 11, n. 4, p. 283-295, 2002. DOI: 10.1057/palgrave.ejis.3000438.

[18] VENKATESH, V.; DAVIS, F. D. A theoretical extension of the technology acceptance model: four longitudinal field studies. Management Science, v. 46, n. 2, p. 186-204, 2000. DOI: 10.1287/mnsc.46.2.186.11926.

[19] DEČMAN, M. Modeling the acceptance of e-learning in mandatory environments of higher education: the influence of previous education and gender. Computers in Human Behavior, v. 49, p. 272-281, 2015. DOI: 10.1016/j.chb.2015.03.022.

[20] BRASIL. Ministério da Saúde. Conselho Nacional de Saúde. Resolução nº 510, de 7 de abril de 2016. Dispõe sobre as normas aplicáveis a pesquisas em Ciências Humanas e Sociais. Diário Oficial da União, Brasília, DF, 24 maio 2016, seção 1, p. 44-46.

[21] BARDIN, L. Análise de conteúdo. Lisboa: Edições Setenta, 1994.

[22] NIELSEN, J. Heuristic evaluation. In: NIELSEN, J.; MACK, R. L. (Eds.). Usability Inspection Methods. New York: John Wiley & Sons, 1994. p. 25-62.

APÊNDICE A - CODEBOOK DA ANÁLISE QUALITATIVA

O codebook formaliza retrospectivamente as categorias utilizadas na reanálise das respostas abertas. As definições abaixo documentam as regras de inclusão e distinção entre categorias e tornam explícitas decisões que, na análise original, estavam apenas implícitas. O codebook não substitui concordância entre codificadores, pois a classificação permaneceu sob responsabilidade de um único avaliador.

Quadro A1 – Codebook das categorias de pontos fortes

Categoria	Definição operacional	Regra de inclusão / distinção	Exemplo	Associação / eixo
Operacionalidade e facilidade de uso básica	Avaliação positiva da facilidade geral para executar operações ou interagir com o sistema.	Incluir elogios genéricos à facilidade/praticidade; excluir elogios a uma função específica, que vão para E3.	“Facilidade, visualização, interação.”	Usabilidade / E1
Registro e consulta de notas e frequência	Valorização de funções de consultar ou registrar notas, faltas e frequência.	Incluir somente quando a função é mencionada como recurso útil; dificuldade para localizá-la é E1.	“Dá pra ver notas frequência de aulas...”	Adequação Funcional / E3
Concentração das informações acadêmicas necessárias	Reconhecimento de que o sistema reúne informações acadêmicas relevantes em um único ambiente.	Incluir quando o valor está na disponibilidade/concentração da informação; excluir organização visual da informação.	“Tem todas as informações acadêmicas que eu preciso.”	Adequação Funcional / E3
Nenhum ponto forte identificado	Negação explícita da existência de ponto forte ou resposta equivalente.	Incluir “nenhum”, “nada”, “não me ocorre”; não usar para célula vazia ou texto ilegível.	“Nenhum.”	não aplicável
Adequação das cores	Avaliação positiva da paleta ou das cores da interface.	Incluir cor/paleta; outros elogios visuais ficam em estética geral.	“Cores agradáveis.”	Usabilidade / E1

Categoria	Definição operacional	Regra de inclusão / distinção	Exemplo	Associação / eixo
Horários e calendário acadêmico	Valorização do acesso a horários, calendário ou organização temporal acadêmica.	Incluir quando a função em si é elogiada.	“Tem o horário das aulas logo no começo.”	Adequação Funcional / E3
Emissão e acesso a documentos	Valorização de comprovantes, documentos ou acesso documental.	Incluir emissão/consulta de documentos; excluir facilidade genérica de navegação.	“Acesso a documentação.”	Adequação Funcional / E3
Cumpre a maioria das funções a que se propõe	Avaliação de cobertura funcional geral suficiente.	Incluir juízos sobre atender às funções esperadas; não usar para facilidade de uso.	“O sagres consegue realizar a maioria das funcionalidades necessárias.”	Adequação Funcional / E3
Possibilidade de personalizar a tela	Valorização de recursos de personalização visual.	Incluir apenas quando a personalização é citada como recurso.	“Dá pra personalizar a tela.”	Usabilidade / E1
Ser um sistema web	Valorização do acesso via web como característica de implantação/acesso.	Manter separado dos três eixos; associação à Portabilidade é apenas aproximativa.	“Sistema web.”	Portabilidade / fora dos eixos
Planos de curso e conteúdo das aulas	Valorização do acesso ou registro de planos, conteúdos e informações de aula.	Incluir quando o recurso funcional é o foco.	“...planos de curso.”	Adequação Funcional / E3
Facilidade de login e acesso inicial	Avaliação positiva do acesso inicial/autenticação.	Incluir facilidade de entrar no sistema; excluir navegação após o login.	“Fácil acesso para logar.”	Usabilidade / E1
Segurança percebida do sistema	Declaração positiva de segurança ou confiança geral no sistema.	Incluir percepção/confiança; não interpretar como medida técnica de segurança.	“Útil, descritivo e seguro.”	Segurança / E2
Privacidade e sigilo dos dados acadêmicos	Valorização do resguardo de notas ou informações pessoais/acadêmicas.	Incluir menções explícitas a privacidade/sigilo.	“A privacidade das minhas notas e informações.”	Segurança / E2
Outros aspectos estéticos da interface	Elogio visual que não se limita às cores.	Incluir design, aparência, tamanho/tela quando positivos; excluir personalização e cor quando específicos.	“Design...”	Usabilidade / E1
Armazenamento e importação de dados de períodos anteriores	Valorização de persistência/reuso funcional de dados acadêmicos anteriores.	Incluir recurso de importar/recuperar dados; perda de dados é categoria negativa de Confiabilidade.	“Importação de conteúdo de outros períodos.”	Adequação Funcional / E3
Atendimento a solicitações acadêmicas	Valorização de recurso para solicitações ou da resposta a essas solicitações.	Incluir solicitações acadêmicas como funcionalidade.	“Boa resposta às solicitações.”	Adequação Funcional / E3
Fluxograma do curso	Valorização do acesso ao fluxograma ou percurso curricular.	Incluir menção explícita ao fluxograma.	“É fácil acessar o fluxograma do curso.”	Adequação Funcional / E3
Disponibilidade	Avaliação positiva de o sistema estar acessível/operante.	Incluir disponibilidade como atributo positivo; facilidade de login é E1.	“Praticidade, disponibilidade, utilidade.”	Confiabilidade / E2
Bom desempenho condicionado a estar operante	Avaliação positiva condicionada ao fato de o sistema estar funcionando.	Incluir expressões como “quando funciona”; manter separada de rapidez.	“Quando funciona é bom!”	Confiabilidade / E2
Rapidez e eficiência	Avaliação positiva da velocidade de resposta ou execução.	Incluir rapidez/desempenho; excluir facilidade de navegação.	“Rapidez, eficiência...”	Eficiência de Desempenho / E2

Fonte: elaboração própria a partir das respostas abertas e das categorias utilizadas na reanálise.

Quadro A2 – Codebook das categorias de pontos fracos

Categoria	Definição operacional	Regra de inclusão / distinção	Exemplo	Associação / eixo
Instabilidade e interrupções durante o uso	Funcionamento irregular, quedas ou interrupções durante uma sessão.	Incluir “cai”, “instável”, interrupções; se o foco for impossibilidade total de acesso, usar indisponibilidade.	“Instável (sai do ar várias vezes).”	Confiabilidade / E2
Lentidão e tempo de resposta	Demora percebida para carregar, salvar, processar ou responder.	Incluir demora/latência; excluir indisponibilidade total e demora de atualização de informação já registrada.	“Lentidão extrema.”	Eficiência de Desempenho / E2
Estética da interface e legibilidade	Crítica à aparência, design, fontes, cores ou legibilidade.	Incluir aspectos visuais; excluir confusão de navegação quando o problema é localizar funções.	“É muito feio, parece um site dos anos 2000.”	Usabilidade / E1
Sistema pouco intuitivo e confuso	Dificuldade de compreender como o sistema funciona ou o significado das opções.	Incluir “não intuitivo”, “confuso”; excluir apenas excesso de etapas/localização, que é navegação.	“Interface pouco intuitiva.”	Usabilidade / E1
Navegação, excesso de etapas e dificuldade de localizar funções	Dificuldade para encontrar caminhos, funções ou concluir ações por excesso de passos.	Incluir “muita volta”, “muitos passos”, dificuldade de localizar; ausência da função é E3.	“Preciso dá muita volta para achar o que eu quero.”	Usabilidade / E1
Indisponibilidade e congestionamento do serviço	Impossibilidade de acessar ou utilizar o sistema por indisponibilidade/congestionamento.	Incluir “fora do ar”, “serviço indisponível”, congestionamento; quedas intermitentes durante uso ficam em instabilidade.	“O sagres recorrentemente tem os serviços indisponíveis.”	Confiabilidade / E2
Funcionalidades ausentes ou incompletas	Recurso esperado que não existe, não está disponível ou não atende completamente à tarefa.	Incluir ausência/incompletude; se a função existe mas é difícil de encontrar, usar E1.	“Falta de recursos como solicitação de trancamento...”	Adequação Funcional / E3
Poluição e repetição de informações	Excesso, redundância ou repetição que prejudica a leitura ou operação.	Incluir avisos/informações repetidas; excluir mera feiura da interface.	“Poluição de informação.”	Usabilidade / E1
Ajuda, documentação e suporte ao usuário	Ausência, insuficiência ou baixa qualidade de ajuda, instruções ou suporte.	Incluir help/documentação/instruções/suporte; mensagens de erro específicas podem ser E1, mas ficam aqui quando tratadas como ajuda.	“A ajuda do Sagres que é inexistente.”	Usabilidade / E1
Insegurança percebida	Desconfiança ou percepção negativa de segurança.	Incluir percepção subjetiva; não inferir vulnerabilidade técnica.	“Insegurança.”	Segurança / E2
Perda de dados e erros de persistência	Dados não salvos, desaparecidos ou perdidos após registro.	Incluir desaparecimento/falha de persistência; lentidão ao salvar, sem perda, é desempenho.	“Os dados desaparecem sem motivo.”	Confiabilidade / E2
Pré-requisitos, prazos e limites de tempo de acesso	Restrições de fluxo, prazo, pré-requisito ou sessão que dificultam a execução da tarefa.	Incluir travas/regras temporais ou burocráticas; ausência da função é E3 e excesso de etapas de interface é E1.	“Trava de ajuste de datas.”	Adequação Funcional / E3
Versão móvel deficiente	Crítica específica à experiência ou disponibilidade em dispositivo móvel.	Incluir comparação explícita com versão móvel/celular; não forçar nos três eixos.	“A versão para celular é inferior...”	Portabilidade / fora dos eixos

Categoria	Definição operacional	Regra de inclusão / distinção	Exemplo	Associação / eixo
Inconsistência do sistema	Comportamentos, padrões ou elementos percebidos como inconsistentes.	Incluir menção explícita a inconsistência; quando a fala descreve queda/erro operacional, priorizar Confiabilidade.	“Interface feia, inconsistente e não intuitivo.”	Usabilidade / E1
Demora na atualização das informações	Defasagem entre uma alteração e sua aparição no sistema.	Incluir demora de atualização de notas, disciplinas ou informações; demora de carregamento é tempo de resposta.	“Demora para atualizar as informações...”	Eficiência de Desempenho / E2
Resposta sem atributo especificado	Resposta que expressa negatividade, mas não identifica atributo analisável.	Usar apenas quando não houver conteúdo temático classificável.	“...” / “Todos”.	não aplicável
Design não simplificado para inclusão de dados	Dificuldade de preenchimento atribuída ao desenho da interface de entrada.	Incluir problema específico de preencher/inserir dados; se o foco for excesso de passos, usar navegação.	“Ruim de preencher.”	Usabilidade / E1
Falta de integração de dados acadêmicos	Ausência de integração ou reaproveitamento entre dados/módulos acadêmicos.	Incluir necessidade de reinserção ou falta de integração; não confundir com perda de dados já gravados.	“...sempre necessário o retrabalho da inserção dos dados.”	Compatibilidade / fora dos eixos

Fonte: elaboração própria a partir das respostas abertas e das categorias utilizadas na reanálise. Exemplos preservam a redação original dos participantes quando reproduzidos entre aspas.