

UNIVERSIDADE FEDERAL DO PAMPA

Adriel Herter Seffrin

Uma comparação entre técnicas de redesign
de interface baseadas em especialistas e
baseadas em usuários

Alegrete
2019

Adriel Herter Seffrin

Uma comparação entre técnicas de redesign de interface baseadas em especialistas e baseadas em usuários

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Software da Universidade Federal do Pampa como requisito parcial para a obtenção do título de Bacharel em Engenharia de Software.

Orientador: Prof. Me. Jean Felipe Patikowski Cheiran

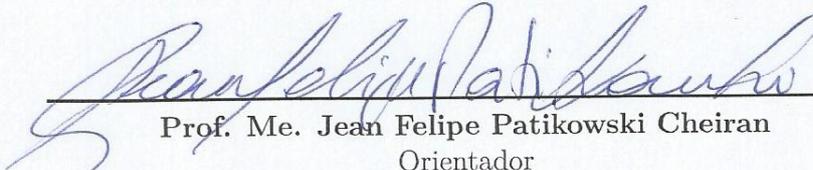
Alegrete
2019

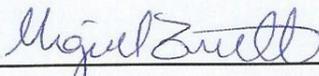
Adriel Herter Seffrin

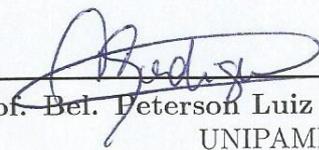
Uma comparação entre técnicas de redesign de interface baseadas em especialistas e baseadas em usuários

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Software da Universidade Federal do Pampa como requisito parcial para a obtenção do título de Bacharel em Engenharia de Software.

Trabalho de Conclusão de Curso defendido e aprovado em 29. de novembro de 2019.
Banca examinadora:


Prof. Me. Jean Felipe Patikowski Cheiran
Orientador
UNIPAMPA


Prof. Me. Miguel Zinelli da Costa Junior
UNIPAMPA


Prof. Bel. Peterson Luiz da Rosa Rodrigues
UNIPAMPA

AGRADECIMENTOS

Em primeiro lugar agradeço ao Prof. Me. Jean Felipe Patikowski Cheiran pelas suas sábias instruções e seu empenho na orientação desse trabalho e aos demais professores que fizeram parte de minha evolução nesta jornada acadêmica. Aproveito para agradecer aos membros da banca Miguel Zinelli e Peterson Rodrigues, pelas contribuições geradas para este trabalho.

Agradeço também a toda minha família, em especial minha mãe Elenir, que sempre me ajudou e nunca mediu esforços para que esse momento chegasse, também ao meu pai Mário, que me deu suporte quando precisei e ao meu primo Caio Alexandre, que me auxiliou muito no início do curso me passando alguns de seus conhecimentos.

Aproveito para agradecer as amigadas de Adriel Rodrigues, Peterson Rodrigues, Emanuel Krombauer, Christopher Krombauer, Leonardo Leonel, Guilherme Legramante, Magno Maia, Cauê Nonnenmacher, Dionatã Mafaldo, Steffani Minetto, Leonardo Lunkes, Jeisson Oliveira, Gabriel Sodré e Gabriel Pires, em especial esses 9 primeiros, que foram os que conheci nessa jornada, e os levarei como irmãos para toda a vida. Mas sem dúvida todos estes foram muito importantes durante esse longo caminho.

Por fim, agradeço a todas as pessoas que direta ou indiretamente contribuíram para a realização desse trabalho.

Meu muito obrigado a todos!

RESUMO

A usabilidade é uma das qualidades mais importantes em uma aplicação. Um produto que apresente problemas de usabilidade pode gerar experiências frustrantes para usuários e possivelmente cair em desuso. Uma alternativa para produtos com problemas de usabilidade é o *redesign*, processo que envolve redesenhar a aplicação para corrigir problemas. Existem diferentes técnicas para avaliar a usabilidade e redesenhar uma aplicação. Essas técnicas envolvem diferentes pessoas e possuem diferentes custos associados, sendo divididas em técnicas baseadas em especialistas em Interação Humano-Computador (IHC) e técnicas baseadas em não especialistas em IHC. Nesse contexto, este trabalho tem como objetivo realizar uma comparação entre as diferentes abordagens citadas anteriormente, isto é, comparar e verificar se há diferenças significativas entre utilizar especialistas e não especialistas no processo de avaliação e *redesign* de uma aplicação. Para isso, foram escolhidas as técnicas de percurso cognitivo e teste de usabilidade com usuário, representando as técnicas de avaliação utilizando especialistas e utilizando não especialistas, respectivamente. Para a realização do *redesign*, foi escolhida a prototipação de baixa fidelidade realizada separadamente por especialistas e não especialistas com base nos resultados das avaliações. A aplicação utilizada em todas as etapas deste trabalho foi um aplicativo móvel disponibilizado pela empresa de software DevPampa. Como principal contribuição desse trabalho, foi identificado que, com base em medições feitas durante as etapas de avaliação e *redesign*, não há diferenças estatísticas entre as duas abordagens.

Palavras-chave: *Redesign*. Interação Humano-Computador. Avaliação de Usabilidade. Comparação. Aplicativo Móvel.

ABSTRACT

Usability is one of the most important qualities in an application. A software product with usability issues is likely to foster frustrating experiences and possibly to fall into disuse. An alternative for products with usability problems is the redesign, a process that involves redraw the application to fix the issues. There are different techniques to evaluate the usability and redraw an application. These techniques involve different people and they have different associated costs, being divided into techniques based on Human-Computer Interaction (HCI) experts and techniques based on non-experts. In this context, this work aims to perform a comparison between the different aforementioned approaches, i.e., it aims to compare and verify whether there is any significant difference between using experts and non-experts in the evaluation and redesign process of an application. In order to do this, cognitive walkthrough and user testing techniques were chosen to represent evaluation techniques using experts and non-experts, respectively. For the redesign process, low fidelity prototyping was separately applied by both experts and non-experts based on reports of the previous evaluation. The application used in all stages of this work was a mobile app made available by DevPampa software company. As the main contribution of this study, it was identified that based on measurements made during evaluation and redesign phases there are no statistical differences between both approaches.

Key-words: Redesign. Human-Computer Interaction. Usability Evaluation. Comparison. Mobile App.

LISTA DE FIGURAS

Figura 1 – Desenho e organização Trabalho de Conclusão de Curso (TCC) 1	33
Figura 2 – Desenho e organização TCC 2	33
Figura 3 – Aplicativo Efipan versão original.	35
Figura 4 – Teste de normalidade de tempos para completar tarefas.	42
Figura 5 – Resultado do teste Mann-Whitney para tempos para completar tarefas.	43
Figura 6 – Teste de medianas sobre severidade dos problemas.	45
Figura 7 – Material utilizado para execução do <i>redesign</i>	46
Figura 8 – Análise de concordância Fleiss Kappa entre especialistas.	48
Figura 9 – Análise de concordância Fleiss Kappa entre não especialistas.	48
Figura 10 – Resultado do teste Mann-Whitney comparando diferentes propostas de solução de <i>redesign</i>	49
Figura 11 – Exemplo de telas do aplicativo desenvolvido baseado no <i>redesign</i> com especialistas	50
Figura 12 – Exemplo de telas do aplicativo desenvolvido baseado no <i>redesign</i> com não especialistas	51
Figura 13 – Teste de normalidade de dados dos escores do <i>System Usability Scale</i> (SUS).	52
Figura 14 – Resultado do teste Mann-Whitney para escores do SUS.	52

LISTA DE TABELAS

Tabela 1 – Atividades para aplicação do percurso cognitivo.	27
Tabela 2 – Compilação de dados da execução das avaliações com especialistas . . .	41
Tabela 3 – Compilação de dados da execução das avaliações com não especialistas	42
Tabela 4 – Compilação dos problemas encontrados por usuários especialistas . . .	44
Tabela 5 – Compilação dos problemas encontrados por usuários não especialistas .	45
Tabela 6 – Relação da execução do <i>redesign</i> para cada problema por cada especi- alista.	47
Tabela 7 – Relação da execução do <i>redesign</i> para cada problema por cada usuário não especialista.	47
Tabela 8 – Resultados da avaliação final SUS - Versão Especialistas	50
Tabela 9 – Resultados da avaliação final SUS - Versão Não especialistas	51

LISTA DE SIGLAS

IHC Interação Humano-Computador

SUS *System Usability Scale*

TCC Trabalho de Conclusão de Curso

SUMÁRIO

1	INTRODUÇÃO	19
1.1	Motivação	19
1.2	Objetivo	20
1.3	Organização deste trabalho	20
2	FUNDAMENTAÇÃO TEÓRICA	21
2.1	Interação Humano-Computador	21
2.2	Usabilidade	21
2.2.1	<i>Design</i> em IHC	21
2.2.1.1	Prototipação	22
2.2.2	Avaliação de IHC	22
2.2.2.1	<i>System Usability Scale</i> (SUS)	23
2.2.2.2	Avaliação Heurística	24
2.2.2.3	Percurso cognitivo	26
2.2.2.4	Teste de usabilidade com usuário	27
3	TRABALHOS RELACIONADOS	29
3.1	An experimental comparison of usage-based and checklist-based reading	29
3.2	Project management system review and <i>redesign</i> using user-centred <i>design</i> methodology	30
3.3	A comparison of usability techniques for evaluating <i>design</i>	31
3.4	A Survey of User-Centered Design Practice	31
3.5	Lições aprendidas	32
4	METODOLOGIA	33
4.1	Pesquisa bibliográfica	33
4.2	Definição de métricas para comparação das evoluções	34
4.3	Definição de técnicas de avaliação e de <i>redesign</i>	34
4.4	Definição do software para evolução	34
4.5	Execução de avaliação preliminar	34
4.6	Execução das avaliações	35
4.6.1	Percurso cognitivo	37
4.6.2	Teste de usabilidade com usuário	38
4.7	Documentação dos problemas encontrados	38
4.8	Aplicação das técnicas de <i>redesign</i>	38
4.9	Análise e escolha de soluções de <i>redesign</i>	38
4.10	Evolução do software	38
4.11	Aplicação do questionário de avaliação	39

4.12	Análise e conclusão dos resultados	39
5	RESULTADOS	41
5.1	Aplicação das avaliações	41
5.2	Compilação dos resultados das avaliações	43
5.3	Aplicação do <i>redesign</i>	44
5.4	Desenvolvimento das versões	48
5.5	Avaliação de satisfação	49
5.6	Análise e conclusão dos resultados	52
6	CONSIDERAÇÕES FINAIS	53
	REFERÊNCIAS	55
	APÊNDICES	57
	APÊNDICE A – TERMO DE CONSENTIMENTO COM DEV- PAMPA	59
	APÊNDICE B – REDESIGN COM ESPECIALISTAS	61
	APÊNDICE C – REDESIGN COM NÃO ESPECIALISTAS	111
	Índice	141

1 INTRODUÇÃO

A maioria das empresas de software podem ter noção da importância da usabilidade em uma aplicação, porém, pode ser que muitas ainda a enxergam como uma característica que só deve ser levada em conta durante o desenvolvimento se houver tempo e recurso, como se ela representasse um custo adicional.

Dependendo da frequência com que o software é empregado, os prejuízos para as empresas podem também ser expressivos, não só em decorrência do absenteísmo e da rotatividade do pessoal, mas também pela baixa produtividade, competitividade e menor retorno de investimento. Sistemas difíceis de usar implicam em erros e perda de tempo, fatores que se multiplicam com a frequência das tarefas e o número de usuários. A perda de dados e informações pode implicar na perda de clientes e de oportunidades. Acontecimentos deste tipo causam desde uma resistência ao uso do sistema até a sua subutilização e abandono completo, com o devido consentimento da empresa. O barato terá custado caro (CYBIS; BETIOL; FAUST, 2015).

Diante disso, o usuário que tiver uma experiência negativa na utilização do software, possivelmente desistirá de utilizá-lo, pois poderá encontrar dificuldades em realizar algumas tarefas devido à problemas de usabilidade.

Nielsen (1994) descreve que métodos de inspeção de usabilidade são meios eficientes para identificar problemas de usabilidade, por serem relativamente informais, fáceis e mais baratos de utilizar.

O *redesign*¹ de uma interface geralmente ocorre uma vez que ela foi testada com usuários ou foi exposta a especialistas para uma avaliação (FERREIRA; BARR; NOBLE, 2005).

De acordo com Sommerville (2011), existe uma necessidade constante e inevitável de evolução por parte de todo software que espera manter-se ativo por um longo período de tempo. A manutenção é um processo de melhoria ou correção de um software já desenvolvido ou que está sendo desenvolvido, sendo possível também corrigir erros que são encontrados durante a utilização do sistema pelo usuário ou por testes realizados.

1.1 Motivação

A interface com usuário é a parte visível de um sistema, modernizá-la pode melhorar a usabilidade e trazer benefícios aos usuários finais (WALLNAU; SEACORD; ROBERT, 2000). No caso da evolução em perspectiva de Interação Humano-Computador (IHC), a evolução comumente se trata de evolução de *design*² e, potencialmente envolve *redesign*.

Um processo de evolução em IHC completamente baseado em especialistas pode gerar um resultado final diferente de uma evolução em IHC com participação de usuários,

¹ Alterar a estética (parte visual) de um produto.

² Parte visual ou estética de um produto.

isto é, diferentes tipos de problemas e níveis de satisfação, por exemplo. Deste modo esse trabalho propõe justamente identificar se essas diferenças existem, detalhando os resultados de cada etapa do estudo e utilizando um ambiente real na indústria.

1.2 Objetivo

Nosso trabalho tem como principal objetivo comparar a utilização de técnicas de avaliação e *redesign* de interfaces baseadas na participação de especialistas e baseadas na participação de usuários não especialistas para a evolução de um aplicativo móvel.

Para que o objetivo principal seja alcançado, os objetivos específicos abaixo devem ser atingidos:

- Escolher técnicas de avaliação e *redesign* de interface com base em usuários especialistas e usuários não especialistas;
- Escolher um aplicativo para utilização evolução;
- Aplicar técnicas com base em usuários especialistas e usuários não especialistas paralelamente para a obtenção de dois aplicativos evoluídos;
- Comparar a aplicação das técnicas e as versões do aplicativo, respectivamente.

1.3 Organização deste trabalho

O trabalho está organizado em 5 principais capítulos, iniciando no Capítulo 1, no qual é abordada a introdução. O Capítulo 2, fundamentação teórica - explica os principais conceitos utilizados no trabalho, complementados por subseções em que técnicas são explicadas.

No Capítulo 3 são apresentados os trabalhos relacionados à este tema que contribuíram nas etapas do trabalho. A metodologia adotada está descrita no Capítulo 4, demonstrando o percurso dessa monografia. Por fim, no Capítulo 5 são apresentados os resultados do trabalho e no Capítulo 6 são resumidas as principais contribuições, as limitações e os trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Esse capítulo apresenta a fundamentação teórica, a qual explica inicialmente o que é IHC, Usabilidade e Design em IHC. Em seguida são apresentadas algumas técnicas de *design* e avaliação em IHC, como o percurso cognitivo, teste de usabilidade e avaliação heurística.

2.1 Interação Humano-Computador

Segundo Hewett et al. (1992), a Interação Humano-Computador (IHC) é a área do conhecimento interessada no projeto, implementação e avaliação de sistemas computacionais interativos para uso do ser humano, simultaneamente com os fenômenos associados a esse uso.

O estudo da IHC ainda envolve quais os impactos de uma determinada tecnologia na vida das pessoas e como a interação deve ser abordada para oferecer conforto, bem-estar e facilidade de uso (SHARP; ROGERS; PREECE, 2007).

2.2 Usabilidade

Sendo um dos principais pilares dentro da IHC, Sommerville (2011) descreve a usabilidade como uma propriedade que reflete a facilidade de se usar um sistema. Também, segundo Cybis, Betiol e Faust (2015), a usabilidade aborda a forma como o usuário se comunica com a máquina e como a tecnologia responde à interação do usuário, considerando as seguintes habilidades de acordo com a seção 11 da norma ISO-9241 (2018):

- Facilidade de aprendizado: a utilização do sistema requer pouco treinamento;
- Fácil de memorizar: o usuário deve lembrar como utilizar a interface depois de algum tempo;
- Maximizar a produtividade: a interface deve permitir que o usuário realize a tarefa de forma rápida e eficiente;
- Minimizar a taxa de erros: caso aconteçam erros, a interface deve avisar o usuário e permitir a correção de modo fácil;
- Maximizar a satisfação do usuário: a interface deve dar-lhe confiança e segurança.

2.2.1 *Design* em IHC

Em termos gerais, segundo Lawson (2006), pode-se estruturar a atividade de *design* como sendo:

- A análise da situação atual: estudar e interpretar a situação atual;

- A síntese de uma intervenção: planejar e executar uma intervenção na situação atual;
- A avaliação da nova situação: verificar o efeito da intervenção, comparando a situação analisada anteriormente com a nova situação, atingida após a intervenção.

Design pode ser definido também como o estudo de uma situação (seja ela a não existência de uma interface, o que seria um processo de *design*, ou seja ela uma interface já existente, um processo de *redesign*) a fim de identificar formas de melhorá-la.

2.2.1.1 Prototipação

A prototipação é um método amplamente usado para projetar, testar e refinar interfaces de usuário (SNYDER, 2004). É muito conhecido e utilizado em IHC e seu principal objetivo é tornar tangível e visual um projeto. Além disso um bom protótipo pode ser ricamente explicativo, visual e até interativo, possibilitando a demonstração para usuários finais ou partes interessadas.

Protótipos podem ser classificados em três categorias fundamentais (PROTOTIPAÇÃO..., 2018): Alta, média e baixa fidelidade. Protótipos de baixa fidelidade podem ser até mesmo um rascunho feito com um lápis em um papel, é onde surgem as primeiras ideias e o esforço é menor.

Protótipos de média fidelidade, ainda que possam ser desenhados no papel, já é possível considerar a utilização de algum software para elaborá-los. Nesse nível de protótipo também é preciso começar a se preocupar com a disposição dos elementos gráficos da interface, na organização estrutural e intuitiva da informação e pensar cuidadosamente em cada interatividade.

Protótipos de alta fidelidade possuem uma representação de alta proximidade com o produto final, podendo também ser funcionais. Tal protótipo, geralmente é concebido na fase final do projeto de interface. Pode ser desenvolvido em softwares com essa finalidade ou em softwares de edição de imagem.

2.2.2 Avaliação de IHC

Esta é uma atividade fundamental em qualquer processo de desenvolvimento que busque produzir um sistema interativo com alta qualidade de uso (BARBOSA; SILVA, 2010).

Existem métodos de avaliação com especialistas, e métodos de avaliação com não especialistas. Os métodos de avaliação com especialistas permitem ao avaliador examinar uma solução para tentar antever as possíveis consequências de certas decisões de *design*. Segundo Barbosa e Silva (2010), esses métodos não envolvem diretamente os usuários, portanto tratam de experiências de uso potenciais, e não reais. Por outro lado, a avaliação

com não especialistas busca trabalhar em um cenário mais concreto, envolvendo usuários reais na avaliação.

Ao inspecionar uma interface, os avaliadores tentam se colocar no lugar de um usuário com determinado perfil, com um certo conhecimento e experiência em algumas atividades, para então tentar identificar problemas que os usuários podem vir a ter quando interagirem com o sistema.

2.2.2.1 *System Usability Scale (SUS)*

O uso de questionários é uma das técnicas de obtenção de dados mais frequentemente utilizadas em que os participantes respondem a fim de fornecer os dados necessários em uma pesquisa, análise ou avaliação (BARBOSA; SILVA, 2010).

O método criado por Brooke (1986) pode ser usado para avaliar produtos, serviços, software, *websites*, aplicações e qualquer outro tipo de interface de forma barata, rápida e fácil. Os critérios que o SUS ajuda a avaliar são:

- Efetividade: os usuários conseguem completar seus objetivos?
- Eficiência: quanto esforço e recursos são necessários para isso?
- Satisfação: a experiência foi satisfatória?

O SUS é um questionário de satisfação aplicado depois do usuário cumprir tarefas especificadas por um avaliador para serem executadas em uma aplicação. Para a execução do SUS, Brooke (1986) definiu 10 perguntas que devem ser respondidas pelo usuário após a utilização da aplicação. O usuário responde a pergunta através de uma escala de 1 a 5, onde 1 significa discordar completamente e 5 significa concordar completamente.

1. Eu acho que gostaria de usar esse sistema com frequência.
2. Eu acho o sistema desnecessariamente complexo.
3. Eu achei o sistema fácil de usar.
4. Eu acho que precisaria de ajuda de uma pessoa com conhecimentos técnicos para usar o sistema.
5. Eu acho que as várias funções do sistema estão muito bem integradas.
6. Eu acho que o sistema apresenta muita inconsistência.
7. Eu imagino que as pessoas aprenderão como usar esse sistema rapidamente.
8. Eu achei o sistema atrapalhado de usar.
9. Eu me senti confiante ao usar o sistema.

10. Eu precisei aprender várias coisas novas antes de conseguir usar o sistema.

Para calcular a média final das respostas, é necessário seguir alguns passos:

- Para as perguntas ímpares (1, 3, 5, 7 e 9), subtrai-se 1 da pontuação que o usuário respondeu. Ou seja, se o usuário respondeu 5, contabiliza-se 4. Se o usuário respondeu 3, contabiliza-se 2.
- Para as perguntas pares (2, 4, 6, 8, 10), subtrai-se a resposta de 5. Ou seja, se o usuário respondeu 2, contabiliza-se 3. Se o usuário respondeu 4, contabiliza-se 1.
- Agora soma-se todos os valores das dez perguntas, e multiplica-se por 2.5.
- Tem-se a pontuação final, que pode ir de 0 a 100.

Uma pontuação final abaixo de 68 pontos é considerada abaixo da média, o que pode indicar que a aplicação possui problemas graves de usabilidade (USABILITY, 2018).

2.2.2.2 Avaliação Heurística

Nielsen e Molich (1990) a definem como um método informal de análise de usabilidade, onde vários avaliadores inspecionam uma interface de usuário e precisam comentar sobre o que é bom e o que é ruim nessa interface.

A avaliação heurística também pode ser definida como uma técnica baseada em especialistas que ajuda a identificar problemas de usabilidade em uma interface. Em outras palavras, ela pode ser definida como uma revisão de uma interface, levando em consideração os aspectos da experiência do usuário.

Uma avaliação heurística não deve substituir o teste de usabilidade. Embora as heurísticas referem-se a critérios que afetam a usabilidade, os problemas identificados em uma avaliação heurística são diferentes daqueles encontrados em um teste de usabilidade.

Ela fornece um *feedback* rápido, pode ser feita em diversas fases do projeto e também pode ser utilizada em conjunto com outras metodologias de avaliação. Porém para realizar uma avaliação heurística, é preciso que o aplicador tenha conhecimento para aplicar a avaliação de uma forma eficaz.

As heurísticas definidas por Nielsen e Molich (1990) são dez princípios gerais do *design* de interface do usuário. Elas podem ser utilizadas para a realização de uma avaliação heurística, durante ou após o desenvolvimento. Durante o desenvolvimento, as heurísticas são usadas como base para o bom desenho de interface focada em uma boa navegação, interação e experiência. Já após o desenvolvimento, elas são utilizadas para identificar problemas na aplicação de acordo com cada heurística.

As heurísticas de Nielsen e Molich (1990) são:

1. Visibilidade do estado do sistema: O sistema sempre deve informar ao usuário o que está acontecendo no momento da interação. Isso pode ser feito por meio de *feedbacks* instantâneos para orientá-lo.
2. Equivalência entre o sistema e o mundo real: Toda comunicação do produto deve falar a linguagem do usuário e não ser orientada ao sistema, ou seja, não deve-se usar linguagem técnica ou termos que são de conhecimento específico. Todas as nomenclaturas devem ser contextualizados e ser coerente com o modelo mental do usuário. Isso também é aplicado à ícones e imagens ilustrativas.
3. Liberdade e controle do usuário: Dar liberdade ao usuário das decisões e ações que podem ser tomadas. Facilitar as “saídas de emergência”, permitir desfazer ou refazer alguma ação no sistema e retornar ao ponto anterior quando estiver perdido ou em situações inesperadas.
4. Consistência e padrões: Manter a consistência visual e de linguagem. Manter padrões de interação em diferentes contextos.
5. Prevenção de erro: Ações drásticas devem ser bem sinalizadas. Além disso, deve-se ter sempre uma confirmação ou possibilidade de desfazer o que foi feito.
6. Reconhecer ao invés de relembrar: Deve-se evitar acionar a memória do usuário o tempo inteiro. Ter ajudas contextuais no sistema e fluxos de ações de acordo com o contexto que o usuário se encontra é de grande importância.
7. Flexibilidade e eficiência de uso: O sistema pode ser ágil para usuários avançados e ser fácil de utilizar pelos usuários leigos. Isso é o que se espera de um sistema flexível e eficiente.
8. Estética e *design* minimalista: Não deve-se usar desnecessariamente excessos de cores e elementos visuais que confundam o usuário. Dialogar de forma simples e direta, com um layout mais limpo, com diálogos naturais, de fácil entendimento e que apareçam em momentos necessários.
9. Auxiliar usuários a reconhecer, diagnosticar e recuperar ações erradas: Prevenir um erro é algo de máxima importância, mas tão importante quanto, é ajudar o usuário a identificar e resolver os problemas que acabam sendo inevitáveis.
10. Ajuda e documentação: Uma interface intuitiva e clara evita a solicitação de ajuda em algumas situações. Mesmo assim deve-se manter ao alcance do usuário, itens de auxílio para determinadas ações.

A execução de uma avaliação heurística deve ser dividida em três fases, o planejamento, a execução e a revisão. No planejamento, deve-se ter um objetivo claro do que se

espera com a avaliação e definir esses objetivos antes de iniciá-la. Nesta fase é importante entender exatamente o que precisa ser avaliado. Também é essencial que se saiba quem são os usuários, necessidades, motivações e comportamentos, e conduzir a avaliação heurística na perspectiva desses usuários. Na execução, define-se uma lista de heurísticas para utilizar na avaliação. Existem vários modelos que podem ser seguidos, entre eles, o mais conhecido são as 10 heurísticas de Nielsen e Molich (1990), citadas anteriormente. Na revisão, cada especialista gera um relatório apresentável da sua avaliação e compartilha com o restante da equipe.

Já a gravidade (severidade) do problema é calculada como uma combinação dos fatores. Frequência com que o problema ocorre, impacto do problema e persistência do problema. A gravidade é avaliada de acordo com a seguinte escala:

- 0 - Não concordo que isto seja um problema;
- 1 - Problema cosmético: não precisa ser consertado;
- 2 - Problema pequeno: o conserto é desejável;
- 3 - Problema grande: importante ser consertado;
- 4 - Catastrófico: é imperativo consertar este problema.

2.2.2.3 Percurso cognitivo

O Percurso Cognitivo, baseado na engenharia cognitiva¹, avalia, através da exploração da interface, a facilidade de aprendizado por exploração de um sistema. As tarefas dos usuários são decompostas em ações, sequência de passos necessários para realizar a tarefa. O avaliador analisa cada ação tentando se colocar no lugar do usuário, questionando se é possível realizar as ações e se elas conduzem ao cumprimento da tarefa com sucesso, anotando as características de usabilidade problemáticas. Caso ocorra problema na realização das ações, o avaliador levanta hipóteses sobre o problema e propõem soluções (BARBOSA; SILVA, 2010).

Barbosa e Silva (2010) propõem uma sequência de quatro atividades para aplicar essa técnica. A Tabela 1 detalha essas atividades.

De acordo com Barbosa e Silva (2010), para cada problema encontrado, deve-se definir a localização do problema e sua gravidade.

Com relação a localização, o problema pode estar em um único local da interface; em dois ou mais locais na interface, casualmente; na estrutura geral da interface, de forma sistemática; ou pode ser algo que não esteja visível e precise ser incluído na interface.

¹ A Engenharia Cognitiva foi proposta por Donald Norman, em 1986, na tentativa de utilizar conhecimentos da psicologia cognitiva, da ciência cognitiva e dos fatores humanos para entender os processos cognitivos humanos.

Tabela 1 – Atividades para aplicação do percurso cognitivo.

Atividade	Tarefa
Preparação	<ul style="list-style-type: none"> • identificar os perfis de usuários; • definir quais tarefas farão parte da avaliação; • descrever as ações necessárias para realizar cada tarefa; • obter uma representação da interface, executável ou não.
Coleta de dados e interpretação	<ul style="list-style-type: none"> • percorrer a interface de acordo com a sequência de ações necessárias para realizar cada tarefa; • para cada ação enumerada, analisar se o usuário executaria a ação corretamente, respondendo e justificando resposta às seguintes perguntas: <ul style="list-style-type: none"> - O usuário vai tentar atingir o efeito correto? (Vai formular a intenção correta?) - O usuário vai notar que a ação correta está disponível? - O usuário vai associar a ação correta com o efeito que está tentando atingir? - Se a ação for executada corretamente, o usuário vai perceber que está progredindo na direção de concluir a tarefa? • relatar uma história aceitável sobre o sucesso ou falha em realizar cada ação que compõe a tarefa.
Consolidação dos resultados	<ul style="list-style-type: none"> • sintetizar resultados sobre: <ul style="list-style-type: none"> - o que o usuário precisa saber a priori para realizar as tarefas; - o que o usuário deve aprender enquanto realiza as tarefas; - sugestões de correções para os problemas encontrados.
Relato dos resultados	<ul style="list-style-type: none"> • gerar um relatório consolidado com os problemas encontrados e sugestões de correção.

Fonte: Barbosa e Silva (2010)

2.2.2.4 Teste de usabilidade com usuário

Esta técnica avalia a usabilidade de sistemas interativos por meio da observação de usuários reais realizando tarefas típicas e pré-definidas, no intuito de encontrar problemas e assim poder melhorar a facilidade de uso do sistema (PRESSMAN, 2009).

Os testes de usabilidade realizam medições do desempenho dos usuários (números de erros e tempo de duração da tarefa) na experiência de uso do sistema, possuindo uma abordagem quantitativa. No entanto, Pressman (2009) e Barbosa e Silva (2010) afirmam que os resultados também possuem uma abordagem qualitativa, visto que apenas medir os resultados quantitativamente é uma abordagem simplista, sendo necessário o avaliador julgar e interpretar os resultados no intuito de identificar problemas e recomendar soluções.

Como complemento que pode ser utilizado em conjunto com o teste de usabilidade, no teste de pensar em voz alta, o mediador pede aos participantes do teste que usem o sistema enquanto continuamente pensam em voz alta, ou seja, simplesmente verbalizam seus pensamentos enquanto se movem pela interface do usuário (NNGROUP, 2012).

Para executar o teste, são precisas apenas três coisas: recrutar usuários representativos, dar tarefas representativas para executar e deixar que os usuários falem (NNGROUP, 2012).

Os benefícios trazidos na utilização do teste de pensar em voz alta são vários, mas como mais importante, ele serve para verificar realmente o que os usuários pensam sobre seu *design*. Escutar os problemas que os usuários enfrentam, geralmente faz com que surjam recomendações de *redesign*. Além disso, o protocolo tem um baixo custo, é robusto, flexível, convincente e fácil de utilizar (NNGROUP, 2012).

3 TRABALHOS RELACIONADOS

Esse capítulo apresenta o conjunto de trabalhos relacionados. Para realizar a busca, o primeiro passo foi identificar palavras-chave relacionadas ao tema para a elaboração da *string* de busca. Foram identificadas as seguintes palavras-chave: *redesign*, *interface*, *hci*, *human-computer interaction*, *legacy system*, *inspection*, *comparison*, *user-centered*, *techniques* e *methods*. O segundo passo foi a definição das bases de dados, e as escolhidas foram a *IEEE Xplore*¹ e a *ACM Digital Library*².

Como critério de inclusão, os artigos encontrados devem abordar os temas de *redesign* ou IHC, bem como técnicas de *redesign* em IHC. Como critérios de exclusão, foi definido que artigos que não abordem *redesign* ou IHC, que não possuam resumo, livros, editoriais, prefácios sejam excluídos dos resultados.

A busca retornou uma quantia de 403 trabalhos. Nestes, foi realizada a aplicação de uma filtragem para a área de IHC, restando 182 trabalhos. Após essa etapa, foi realizada uma leitura no título desses trabalhos, procurando identificar trabalhos relacionados ao tema, e destes 182 trabalhos, 10 foram selecionados para a leitura e interpretação. Destes 10, foram escolhidos 4 trabalhos que puderam contribuir com o tema.

3.1 An experimental comparison of usage-based and checklist-based reading

O trabalho de Thelin, Runeson e Wohlin (2003) teve como objetivo comparar, e consequentemente avaliar o quão bem funciona a leitura baseada em uso comparada a outros métodos. O estudo apresenta um experimento controlado, onde a leitura baseada em uso é comparada à uma leitura baseada em *checklist*.

Os autores definem a leitura baseada em *checklist* como uma técnica de leitura onde os revisores utilizam uma lista de questões que auxilia-os a saber que tipos de falhas procurar. Já a leitura baseada em uso tem como principal ideia concentrar o esforço de leitura na detecção das falhas mais críticas do objeto inspecionado.

Para a execução do experimento, foram selecionados como revisores do estudo 23 alunos do quarto ano de mestrado de Engenharia de Software, sendo que muitos desses estudantes possuem uma extensa experiência no desenvolvimento de software. Após o planejamento e execução do experimento, na análise dos dados coletados os autores identificaram que os resultados foram promissores pois mostraram que a leitura baseada em uso é significativamente melhor que a leitura baseada em *checklist*, em termos de eficiência e eficácia em encontrar as falhas que mais afetam os usuários.

¹ Disponível em: <<http://ieeexplore.ieee.org>>

² Disponível em: <<https://dl.acm.org>>

3.2 Project management system review and *redesign* using user-centred *design* methodology

O trabalho de Iqbal et al. (2011), discute o *redesign* de um sistema de gerenciamento utilizando uma metodologia de *design* centrada no usuário. Esse sistema era utilizado por uma ampla gama de usuários; funcionários acadêmicos, professores, estudantes, clientes externos e equipe administrativa. Porém o sistema original não estava atendendo às necessidades. Ele destaca que as vantagens do *redesign* incluem: melhor utilidade; melhor eficiência e produtividade; tempo de aprendizado reduzido; usabilidade melhorada; e maior aceitação dos usuários.

Iqbal et al. (2011) propuseram uma metodologia estendendo a metodologia tradicional do *design* centrado no usuário adicionando uma análise profunda das atividades atuais ou práticas de trabalho, a fim de avaliar se o sistema atual suporta essas atividades. As etapas da abordagem são: Análise das atividades atuais; Elicitação de requisitos; *Redesign*; Avaliação pelos usuários; e avaliação somativa³.

O primeiro passo para o *redesign* foi dado com a prototipação em um papel para ter uma melhor comunicação com os usuários. Usando um lápis e papel, os autores revelam que foi fácil evoluir rapidamente o *design*. Nesta fase, também foi fácil envolver os usuários com o novo *design*. Verificou-se que alguns usuários estavam particularmente interessados em compartilhar suas ideias de *design* e, usando desenhos de lápis e papel, poderiam ser incluídos no novo *design*, também conhecido como *design* participativo⁴.

Para avaliar o novo sistema de gerenciamento de projetos, os autores realizaram testes de usuário e observação. Os testes foram realizados nos sistemas antigo e novo, a fim de compará-los. Ambos os testes de usuário foram conduzidos em 13 participantes, 5 alunos do último ano do curso de graduação e 7 estudantes de pós-graduação. Para a quantificação dos resultados de cada tarefa na avaliação sumativa, a tarefa recebia 0 pontos se não fosse completada, 1 ponto se fosse completada com dificuldade e 2 pontos se fosse completada com facilidade.

Executando a metodologia proposta, no processo de avaliação, foram explorados muitos problemas de usabilidade no sistema original, e baseado nas análises foram propostas recomendações de *design* que foram utilizadas no *redesign* do sistema. Diante disso, os autores obtiveram como resultado que há melhorias significativas no desempenho das tarefas realizadas pelos usuários no sistema reprojeto, se comparado ao original.

³ Segundo Barbosa e Silva (2010), avaliação somativa é realizada após o término do produto com o objetivo de melhorar a qualidade do sistema ou verificar sua conformidade com um determinado padrão.

⁴ Metodologia democrática de projeto que enfatiza o envolvimento ativo de todas as pessoas que são afetadas pelo projeto, em especial, aquelas que são historicamente excluídas de processos decisórios.

3.3 A comparison of usability techniques for evaluating *design*

Doubleday et al. (1997) relataram uma série de experimentos projetados para comparar métodos de teste de usabilidade em uma interface. O objetivo principal foi investigar os problemas encontrados pelas pessoas durante a execução das tarefas de recuperação de informações e avaliar os métodos de avaliação, examinando o foco no problema, a qualidade dos resultados e a relação custo-eficácia de cada método.

Os autores utilizaram as heurísticas de Nielsen e teste com usuário final para identificar problemas de usabilidade. O propósito da pesquisa é ajudar usuários a decidir qual a melhor forma custo-eficácia para avaliar uma interface.

Embora ambas as avaliações compartilhem o mesmo objetivo, Doubleday et al. (1997) concluíram que os resultados reais produzidos por cada técnica são bastante diferentes. Por exemplo, em uma das tarefas realizadas, observar os usuários mostrou que eles frequentemente escolheram a opção de menu errada para mover do navegador para uma janela de consulta, mas não se sabe, apenas por observação, se isso ocorre por causa da má terminologia do menu, desorientação devido a afastar-se da janela de consulta e pesquisando o menu, ou por algum outro motivo. A avaliação heurística declarou explicitamente que a terminologia da opção de menu em questão não estava clara, observou que o menu estava muito longe da consulta e recomendou que um botão para essa tarefa estivesse disponível no navegador, além da opção de menu.

Isso sugere que uma diferença entre as técnicas é que o teste do usuário final pode indicar o sintoma do problema, enquanto o teste heurístico pode de alguma forma identificar a causa do problema.

3.4 A Survey of User-Centered Design Practice

O artigo de Vredenburg et al. (2002) relata os resultados de uma pesquisa com profissionais de *design* centrado no usuário. A pesquisa envolveu mais de cem (100) participantes. O objetivo principal foi investigar o uso de *design* centrado no usuário na prática dentro da indústria, evidenciando o que funciona e o que não funciona na prática.

Os autores identificaram os métodos e processos mais amplamente utilizados, os fatores-chave que predizem o sucesso e as práticas críticas que os praticantes devem fazer na aplicação de métodos e processos de *design* centrado no usuário.

Os principais métodos e processos mais comumente utilizados identificados foram: *design* iterativo, avaliação de usabilidade, análise de tarefas, grupos focais, avaliações heurísticas, entrevistas com usuários e prototipação sem participação do usuário final. A partir desses métodos e processos foram listados os benefícios e fraquezas de cada um.

Porém, os resultados da pesquisa foram inconclusivos quanto à economia no tempo de desenvolvimento e custos em todas as organizações. Os autores citam que talvez os entrevistados tenham se concentrado apenas no tempo e nos custos de desenvolvimento

de curto prazo, mas não em economias de longo prazo.

3.5 Lições aprendidas

Os trabalhos relacionados de maneira geral se destacam e contribuem para nosso trabalho por trazerem avaliações e comparações, entrevistas na indústria e até mesmo a evolução de um sistema que era utilizado e não atendia as necessidades a que ele se propunha a fazer.

O primeiro trabalho, de Thelin, Runeson e Wohlin (2003) contribuiu com o conceito da leitura baseada em uso. No trabalho, os autores definem todo o protocolo do experimento realizado e definem tarefas para os participantes realizarem. Semelhante ao teste de usabilidade que foi realizado em nosso trabalho.

O segundo trabalho, de Iqbal et al. (2011), teve como principal contribuição a evolução de um software que não atendia ao que foi proposto, utilizando *redesign*, uma metodologia definida, e comparando o sistema antigo com o novo.

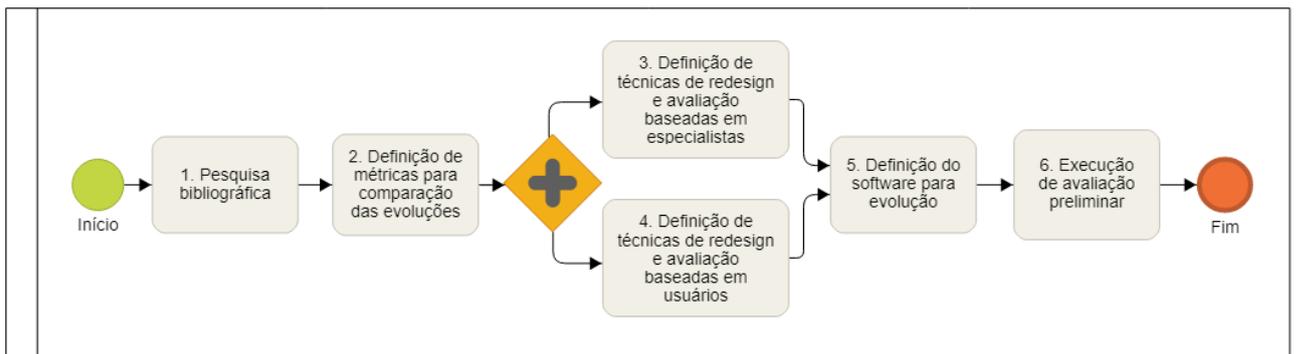
O terceiro trabalho, de Doubleday et al. (1997) tem como principal contribuição a comparação dos erros encontrados em testes de usabilidade com usuário final, e avaliação heurística. No trabalho, os autores também utilizam classificação de severidade nos erros.

Por fim, o trabalho de Vredenburg et al. (2002) apresenta os principais métodos e processos utilizados na indústrias, mostrando os pontos fortes e pontos fracos de cada um.

4 METODOLOGIA

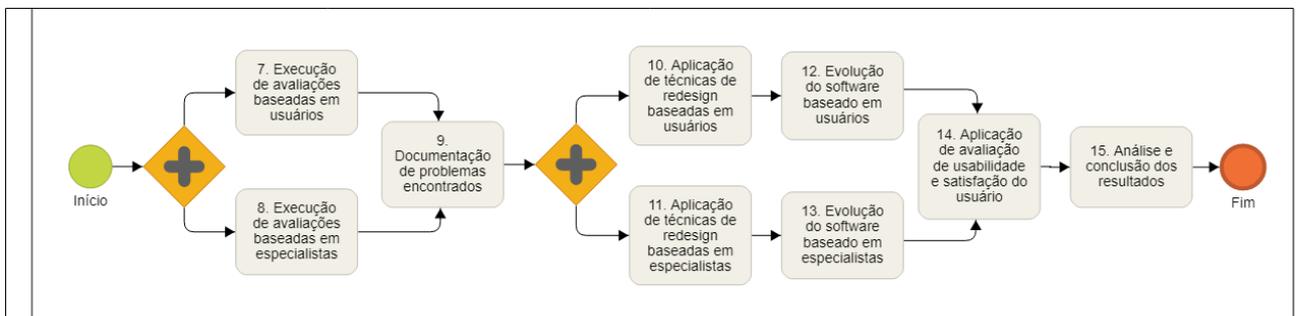
Neste capítulo é apresentado o desenho de pesquisa, representado na Figura 1 e Figura 2. A concepção foi dividida em duas partes, com a Figura 1 representando o Trabalho de Conclusão de Curso (TCC) 1 e a Figura 2 representando o Trabalho de Conclusão de Curso (TCC) 2. As figuras foram elaboradas utilizando BPMN¹, um padrão utilizado para modelagem de processos. O software utilizado foi o Heflo².

Figura 1 – Desenho e organização TCC 1



Fonte: o autor

Figura 2 – Desenho e organização TCC 2



Fonte: o autor

4.1 Pesquisa bibliográfica

O trabalho teve início na tarefa de pesquisa bibliográfica, onde foram definidas palavras chaves, critérios de inclusão e exclusão, e as bases de dados a serem utilizadas. O objetivo dessa etapa foi encontrar trabalhos relacionados que possam contribuir como materiais de apoio para esse trabalho. Os trabalhos encontrados estão localizados no Capítulo 3.

¹ Disponível em: <<https://www.heflo.com/pt-br/bpm/notacao-bpmn>>

² Disponível em: <<https://www.heflo.com>>

4.2 Definição de métricas para comparação das evoluções

Essa etapa tem como objetivo a definição de métricas para comparar os resultados do *redesign* que está previsto na execução do TCC 2. Serão comparadas as duas versões que serão geradas nas etapas 12 e 13 da Figura 2. As métricas de comparação definidas foram: quantidade de problemas encontrados; grau de severidade dos problemas e a relação de participantes que realizaram o *redesign* para um determinado problema ou não. Também foi realizado um questionário de satisfação do usuário em cada um dos produtos gerados com pessoas que não participaram das etapas anteriores. O questionário aplicado foi o SUS. Essas métricas foram escolhidas pois poderiam ser consideradas e trabalhadas de diferentes formas na análise final.

4.3 Definição de técnicas de avaliação e de *redesign*

Referente às etapas 3 e 4 da Figura 1, o objetivo principal é identificar e definir quais técnicas serão utilizadas para a avaliação e para o *redesign*, baseadas em especialistas e baseadas em usuário.

As técnicas de avaliação que foram utilizadas para identificar problemas de usabilidade foram o percurso cognitivo e o teste de usabilidade com usuário. Já para o *redesign* que foi realizado nas etapas 10 e 11 da Figura 2, definiu-se que especialistas e não especialistas iriam desenvolver protótipos de baixa fidelidade com lápis, caneta e papel, tendo a opção de utilizar lápis de cor, régua e outras ferramentas de personalização.

4.4 Definição do software para evolução

Foi definido através de um termo de consentimento de uso com a DevPampa³, que será utilizado o aplicativo para dispositivos móveis Efipan. Se somada a quantidade de *downloads* na *Play Store* e *App Store*, as instalações chegam a aproximadamente 5 mil, tendo assim uma importância relevante na região.

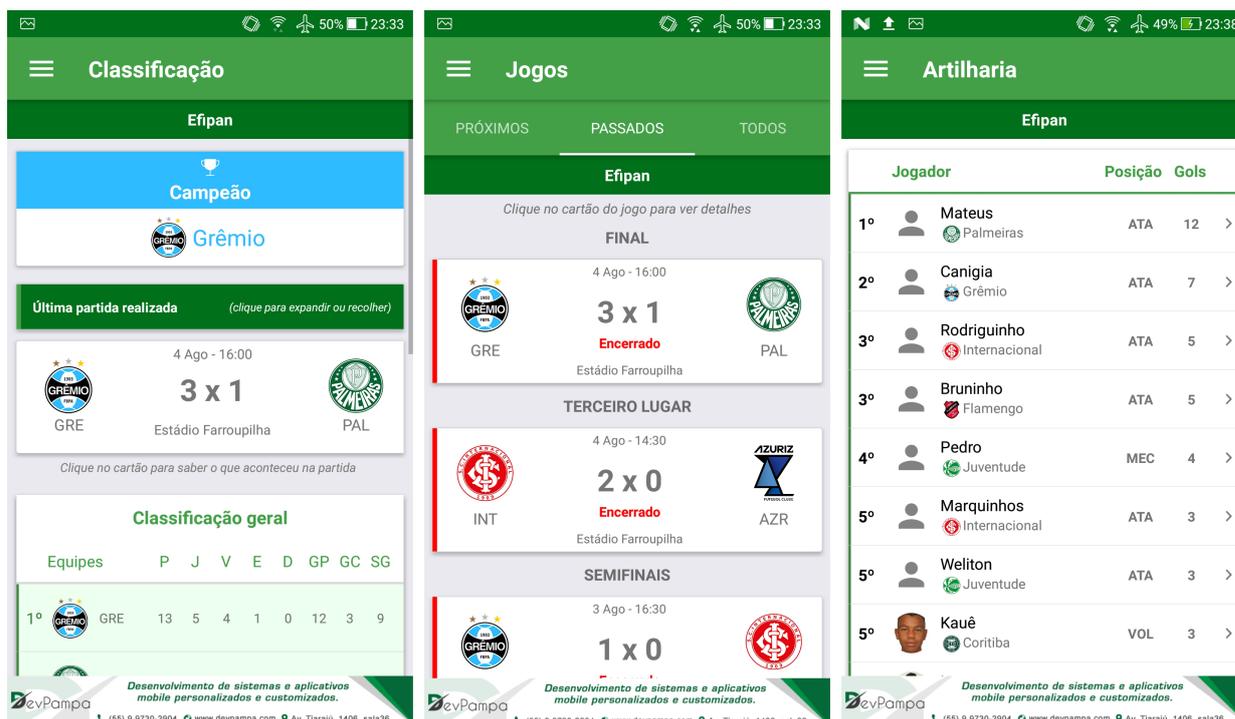
O termo de consentimento reconhecido pela DevPampa está anexado no Apêndice A.

4.5 Execução de avaliação preliminar

Com o objetivo de verificar se as tarefas definidas resultam em dados relevantes, foi realizada uma avaliação baseada em especialistas utilizando o percurso cognitivo, descrito na subseção 2.2.2.3, no aplicativo Efipan. O especialista que executou a avaliação é bacharel em Engenharia de Software e possui conhecimento na área de IHC.

³ DevPampa é uma empresa de software atuante no nicho esportivo.

Figura 3 – Aplicativo Efipan versão original.



Fonte: o autor

4.6 Execução das avaliações

Nessa etapa foram executadas as avaliações baseadas em não especialistas e as avaliações baseadas em especialistas com o objetivo de encontrar problemas que interfiram ou dificultam a satisfação do usuário ao utilizar a aplicação.

Para a realização dessa etapa, foi escolhido o percurso cognitivo, disponível na subseção 2.2.2.3, e o teste de usabilidade com usuário, aliado com a técnica do pensamento em voz alta, disponíveis na subseção 2.2.2.4.

Tanto o percurso cognitivo como o teste de usabilidade utilizaram como base uma lista de tarefas a serem executadas pelo avaliador. Para evitar diferenças, foi utilizada a mesma lista de tarefas para ambas as avaliações. A seguir são apresentadas as tarefas contendo um passo a passo que foram utilizados pelo especialista na execução do percurso cognitivo.

1. O usuário consegue encontrar quem foi o campeão da edição de 2018 do Efipan?
 - a) Acessar o aplicativo, o campeão deve aparecer na tela inicial
2. O horário da primeira partida da 3ª rodada?
 - a) Acessar o aplicativo
 - b) Deslizar para acessar o menu lateral

- c) Clicar em “Jogos”
 - d) Navegar até o label “3ª rodada”, o horário deve estar no *card*
 - e) Deslizar para acessar o menu lateral
3. O usuário consegue encontrar o resultado da última partida?
 - a) Acessar o aplicativo, a última rodada deve aparecer na tela inicial
 4. O usuário consegue encontrar o artilheiro da competição?
 - a) Acessar o aplicativo
 - b) Deslizar para acessar o menu lateral
 - c) Clicar em “Artilheiros”
 5. O usuário consegue encontrar a pontuação da equipe da Chapecoense?
 - a) Acessar o aplicativo
 - b) Verificar a tabela de classificação
 6. O usuário consegue encontrar a equipe com a defesa menos vazada?
 - a) Acessar o aplicativo
 - b) Deslizar para acessar o menu lateral
 - c) Clicar em “Defesa menos vazada”
 7. O usuário consegue encontrar quantos gols a equipe do Palmeiras sofreu?
 - a) Acessar o aplicativo
 - b) Clicar na equipe do Palmeiras
 - c) Ir para a aba “Estatísticas”
 8. O usuário consegue encontrar o nome do goleiro da equipe do Internacional?
 - a) Acessar o aplicativo
 - b) Clicar na equipe do Internacional
 - c) Procurar pelo goleiro
 9. O usuário consegue encontrar o nome do técnico da equipe do Atlético Paranaense?
 - a) Acessar o aplicativo
 - b) Clicar na equipe do Atlético Paranaense
 - c) Procurar pelo técnico

10. O usuário consegue encontrar o local da primeira partida da primeira rodada?
 - a) Acessar o aplicativo
 - b) Deslizar para acessar o menu lateral
 - c) Clicar em “Jogos”
 - d) Ir para a aba “Todos”
 - e) Verificar local descrito no *card*

11. O usuário consegue encontrar a escalação da partida final?
 - a) Acessar o aplicativo
 - b) Deslizar para acessar o menu lateral
 - c) Clicar em “Jogos”
 - d) Ir para a aba “Todos”
 - e) Procurar pela partida final
 - f) Clicar no *card* da partida final
 - g) Ir para a aba “Escalação”

12. O usuário consegue encontrar a comparação entre as duas equipes que se enfrentam na primeira partida da 2ª rodada?
 - a) Acessar o aplicativo
 - b) Deslizar para acessar o menu lateral
 - c) Clicar em “Jogos”
 - d) Ir para a aba “Todos”
 - e) Procurar pela partida
 - f) Clicar no *card* da partida
 - g) Ir para a aba “Comparação”

4.6.1 Percurso cognitivo

As avaliações foram realizadas em uma sala, com apenas o pesquisador e o avaliador. O dispositivo móvel utilizado foi o modelo Zenfone 3 Max, da marca Asus, com tela de 5.5 polegadas e resolução de 1080x1920 (Full HD). Após a execução de cada tarefa, o especialista respondeu cada uma das perguntas listadas na Tabela 1, referente às atividades do percurso cognitivo.

Após a execução de cinco avaliações individuais com cinco especialistas diferentes, os problemas encontrados foram documentados na etapa 9 da Figura 2. A execução da avaliação preliminar foi considerada nos resultados por não ter acarretado em mudanças no protocolo de execução.

4.6.2 Teste de usabilidade com usuário

As tarefas utilizadas no teste de usabilidade com usuário foram as mesmas realizadas na avaliação com percurso cognitivo, e estão disponíveis na seção 4.6. Porém, diferentemente do percurso cognitivo, não foi apresentado para os participantes o passo a passo para se realizar cada tarefa. O avaliador também poderia sugerir aos participantes que comentassem sobre a aplicação durante a realização das tarefas, podendo até mesmo ser uma sugestão ou algo que ele não gostou. Tudo que o participante relatasse foi considerado pelo avaliador.

Essa avaliação foi realizada individualmente com cinco usuários não especialistas (não estudantes de cursos de tecnologia) diferentes. O dispositivo utilizado foi o mesmo da etapa anterior, para evitar diferenças na resolução e proporção de tela.

4.7 Documentação dos problemas encontrados

Todo o material gerado durante as avaliações foi reunido e documentado nessa etapa. Os erros encontrados foram categorizados com uma descrição, tarefas associadas, grau de severidade e as heurísticas que foram violadas.

4.8 Aplicação das técnicas de *redesign*

Nessa etapa foi realizado o *redesign* com cinco usuários especialistas e cinco usuários não especialistas. Os especialistas utilizaram como base uma lista de 15 problemas identificados na avaliação de percurso cognitivo, enquanto os não especialistas utilizaram uma lista de 10 problemas identificados na avaliação de usabilidade com usuário.

A prototipação de ambos foi realizada em papel utilizando lápis, caneta, dando a possibilidade para os executores utilizarem outros materiais, como lápis de cores e régua.

4.9 Análise e escolha de soluções de *redesign*

Tendo o material gerado pela etapa de *redesign*, um juiz especialista da área de IHC cursando doutorado analisou as soluções junto do pesquisador para escolher a coleção de soluções compatíveis que poderiam ser implementadas para corrigir os problemas, levando em consideração alguns critérios, como por exemplo a quantidade de vezes que uma solução foi repetida para solucionar um mesmo problema (se foi repetida).

4.10 Evolução do software

Foi disponibilizado pela empresa DevPampa os códigos-fonte do aplicativo Efipan, o qual é desenvolvido com o *framework React Native*⁴. Utilizando o material gerado na

⁴ Documentação disponível em: <<https://facebook.github.io/react-native>>

etapa anterior, foram gerados dois novos aplicativos pelo pesquisador, um deles considerando o *redesign* por especialistas e o outro considerando o *redesign* por não especialistas. Alguns exemplos visuais das versões geradas estão disponíveis na seção 5.4.

4.11 Aplicação do questionário de avaliação

Foi aplicado um questionário de avaliação nos aplicativos evoluídos, com o objetivo de comparar a satisfação do usuário entre um aplicativo redesenhado por especialistas com um aplicativo redesenhado por não especialistas.

O questionário de avaliação escolhido foi o SUS, descrito na subseção 2.2.2.1. Ele foi aplicado com 20 participantes diferentes das etapas anteriores, executando as mesmas tarefas definidas na seção 4.6.

4.12 Análise e conclusão dos resultados

Tendo documentado todas as métricas anteriores de cada etapa de avaliação e *redesign*, levando em consideração o questionário de avaliação aplicado nos dois aplicativos finais, foram realizadas algumas análises estatísticas a fim de identificar as diferenças entre realizar avaliações e *redesign* com especialistas ou com não especialistas.

5 RESULTADOS

Esse capítulo apresenta nas seções a seguir os resultados obtidos durante a execução desse trabalho.

5.1 Aplicação das avaliações

O participante convidado foi direcionado ao ambiente escolhido para realização da avaliação, e então foi introduzido ao contexto do trabalho pelo pesquisador, que apresentou as intenções e as instruções seguindo o protocolo que foi criado.

Foram realizadas ao total 10 avaliações, sendo 5 avaliações com usuários especialistas e 5 avaliações com usuários não especialistas.

As avaliações realizadas com especialistas utilizaram a técnica do percurso cognitivo, e após a execução de cada uma das 12 tarefas, para cada tarefa o avaliador respondeu as quatro questões previstas na Tabela 1.

Na Tabela 2 estão os dados de duração da execução para cada tarefa, contando o tempo máximo, tempo mínimo e o tempo médio de todas as avaliações, sem considerar o tempo utilizado para responder as questões do protocolo do percurso cognitivo, disponíveis na Tabela 1. A média do tempo para realização das tarefas foi de 17,9 segundos.

Tabela 2 – Compilação de dados da execução das avaliações com especialistas

Tarefa	Não completada	Tempo máximo	Tempo mínimo	Tempo médio
1	-	7s	2s	4,4s
2	-	38s	24s	31,2s
3	-	36s	2s	16,6s
4	-	22s	9s	13s
5	-	10s	4s	6,6s
6	-	37s	5s	14,8s
7	-	35s	8s	20,6s
8	-	30s	8s	15,4s
9	-	55s	10s	22,2s
10	-	27s	10s	19,4s
11	-	32s	5s	18,2s
12	-	67s	18s	33,2s

Fonte: o autor

As avaliações realizadas com não especialistas foram realizadas com a técnica do teste de usabilidade com usuário, onde os avaliadores sabiam o escopo da tarefa, mas não sabiam o passo a passo para realizá-la.

Na Tabela 3 estão os dados de duração da execução para cada tarefa, contando o tempo máximo, tempo mínimo e o tempo médio de todas as avaliações. A tarefa 7 não foi completada em 3 oportunidades enquanto a tarefa 12 não foi completada 1 vez. A média do tempo máximo das tarefas realizadas pelos não especialistas foi de 44,1 segundos.

Tabela 3 – Compilação de dados da execução das avaliações com não especialistas

Tarefa	Não completada	Tempo máximo	Tempo mínimo	Tempo médio
1	-	21s	1s	5,6s
2	-	80s	16s	35,4s
3	-	47s	3s	12,8s
4	-	16s	4s	9,6s
5	-	22s	4s	11s
6	-	76s	6s	21,8s
7	3 vezes	75s	8s	39,4s
8	-	29s	4s	11s
9	-	25s	7s	14,4s
10	-	30s	13s	21s
11	-	49s	5s	18,4s
12	1 vez	60s	10s	24,8s

Fonte: o autor

Vale ressaltar que os especialistas tinham conhecimento sobre o passo a passo das tarefas para concluí-las, enquanto os não especialistas precisavam adotar uma forma mais explorativa para conseguir concluir as tarefas. Isso também explica o por que de nenhuma vez os especialistas não terem concluído alguma das tarefas, algo que aconteceu com os não especialistas em 4 oportunidades.

Na aplicação do teste Shapiro-Wilk¹ que avalia se uma amostra provém de uma população cuja distribuição é diferente da distribuição normal, percebe-se que os valores de significância (Sig.) dos tempos de especialistas e não especialistas estão menores que 0,05 (Figura 4), indicando que ambas amostras não têm distribuição normal com confiabilidade de 95%.

Figura 4 – Teste de normalidade de tempos para completar tarefas.

	Testes de Normalidade		
		Shapiro-Wilk	
	Estatística	gl	Sig.
Tempos de não especialistas	,777	59	,000
Tempos de especialistas	,906	59	,000

Fonte: o autor

Nesse caso, testes não paramétricos são mais adequados para medir diferenças entre os tempos de especialistas e não especialistas para completar as tarefas. Um teste

¹ Detalhes sobre o teste aplicado podem ser encontradas em <<http://www.real-statistics.com/tests-normality-and-symmetry/statistical-tests-normality-symmetry/shapiro-wilk-test/>>

Mann-Whitney² (uma alternativa não paramétrica ao teste-t de Student) foi então aplicado. Considerando o valor de significância (Sig.) superior a 0,05, não é possível afirmar que as medianas dos tempos para especialistas e não especialistas são significativamente diferentes, ou seja, não há diferença nos tempos para especialistas e não especialistas completarem as tarefas.

Figura 5 – Resultado do teste Mann-Whitney para tempos para completar tarefas.

Testes não paramétricos

Resumo de Teste de Hipótese				
	Hipótese nula	Teste	Sig.	Decisão
1	A distribuição de tempo é igual nas categorias de grupo.	Amostras Independentes de Teste U de Mann-Whitney	,357	Reter a hipótese nula.
São exibidas significâncias assintóticas. O nível de significância é ,050.				

Fonte: o autor

5.2 Compilação dos resultados das avaliações

Após a execução das avaliações, os problemas foram compilados com as tarefas associadas, o grau de severidade e as heurísticas que foram violadas. A atividade de identificar o grau de severidade e as heurísticas violadas foram inspiradas na técnica de inspeção heurística, disponível na subseção 2.2.2.2.

Nas avaliações com especialistas foram identificados 15 problemas, que destes, o mais alto grau de severidade foi 4. A Tabela 4 apresenta os dados de cada problema encontrado.

Nas avaliações com não especialistas foram identificados 10 problemas, que também foram compilados e categorizados com as tarefas associadas, o grau de severidade e as heurísticas que foram violadas pelos problemas. O mais alto grau de severidade também foi 4, conforme mostrado na Tabela 5.

A fim de avaliar diferenças entre as gravidades dos problemas encontrados por especialistas e não especialistas, foi aplicado um teste de mediana de Mood³. Esse teste não paramétrico é adequado para amostras pequenas de dados ordinais, mostrando-se suficiente nesse caso. O resultado do teste apresentado na Figura 6 indica que não há diferença observável nas amostras (Sig. > 0,05). Assim, não se pode dizer que uma técnica tenha encontrado problemas mais severos ou menos severos que a outra.

² Detalhes sobre o teste aplicado podem ser encontrados em <http://www.real-statistics.com/non-parametric-tests/mann-whitney-test>

³ Detalhes sobre o teste aplicado podem ser encontrados em <http://www.real-statistics.com/non-parametric-tests/moods-median-test-two-samples>.

Tabela 4 – Compilação dos problemas encontrados por usuários especialistas

ID	Descrição do problema	Tarefas associadas	Grau de severidade	Heurísticas violadas
1	Muitas informações acessáveis somente pelo menu lateral	2, 3 e 4	2	H8
2	Termo gols contra é confuso na tela de estatísticas do time	7	3	H2
3	Falta de affordance para demonstrar que a tabela de classificação é clicável	7, 8 e 9	3	H7
4	A inversão da ordem da listagem dos jogos nas abas Passados, Próximos, Todos pode gerar confusão	10, 11 e 12	1	H5
5	Falta de informação das fases pós-grupo na tela inicial	1	3	H8
6	Pouco destaque no texto que identifica a rodada na tela de jogos	2	1	H7
7	Na tela inicial, o card de última partida realizada está desconectado da informação textual que identifica que aquela é a última partida	3	1	H9
8	Diferenciar o primeiro colocado na defesa menos vazada	5	1	H7
9	Falta de legenda para identificar abreviaturas (P, J, SG) na tela inicial	6	3	H2, H10
10	Muita informação textual na tela de comparação entre os times que se enfrentam	12	1	H9
11	Dificuldade em voltar para a tela inicial	2	4	H3
12	Tamanho da fonte do horário é pequena	2	1	H7
13	Ordenação dos jogadores e comissão técnica está confusa	8 e 9	2	H2, H6
14	Não está explícito o ano em que ocorreu a competição	1	1	H7
15	Identificar a rodada da última partida que é mostrada na tela inicial	11	2	H7

Fonte: o autor

5.3 Aplicação do *redesign*

Tendo os problemas documentados, foram elaborados protocolos para execução do *redesign*. Enquanto os participantes especialistas foram apresentados aos 15 problemas provenientes das avaliações com especialistas, os participantes não especialistas foram apresentados aos 10 problemas provenientes das avaliações com não especialistas. Na execução, o pesquisador indagava o participante sobre como ele resolveria cada um dos

Tabela 5 – Compilação dos problemas encontrados por usuários não especialistas

ID	Descrição do problema	Tarefas associadas	Grau de severidade	Heurísticas violadas
1	Muitas informações acessáveis somente pelo menu lateral	2	2	H8
2	Termo gols contra é confuso na tela de estatísticas do time	7	3	H2
3	Falta de affordance para demonstrar que a tabela de classificação é clicável	8	3	H7
4	A inversão da ordem da listagem dos jogos nas abas Passados, Próximos, Todos pode gerar confusão	2, 10	1	H5
5	Falta de legenda para identificar abreviaturas (P, J, SG) na tela inicial	6	3	H2, H10
6	Dificuldade em voltar para a tela inicial	2	4	H3
7	Botão do menu está difícil de clicar	4	2	H6
8	Ordenação dos jogadores e comissão técnica está confusa	8, 9	2	H2, H6
9	Última partida realizada pode não estar muito visível na tela inicial	3, 11	2	H7
10	Comparativo entre os times pode não estar muito fácil de ser encontrado	12	2	H7

Fonte: o autor

Figura 6 – Teste de medianas sobre severidade dos problemas.

Resumo de Teste de Hipótese				
	Hipótese nula	Teste	Sig.	Decisão
1	As medianas de gravidade são iguais nas categorias de especialistas e não especialistas.	Amostras Independentes de Teste de Median	1,000 ^a	Retar a hipótese nula.

São exibidas significâncias assintóticas. O nível de significância é ,050.

a. A exata significância é exibida para este teste.

Fonte: o autor

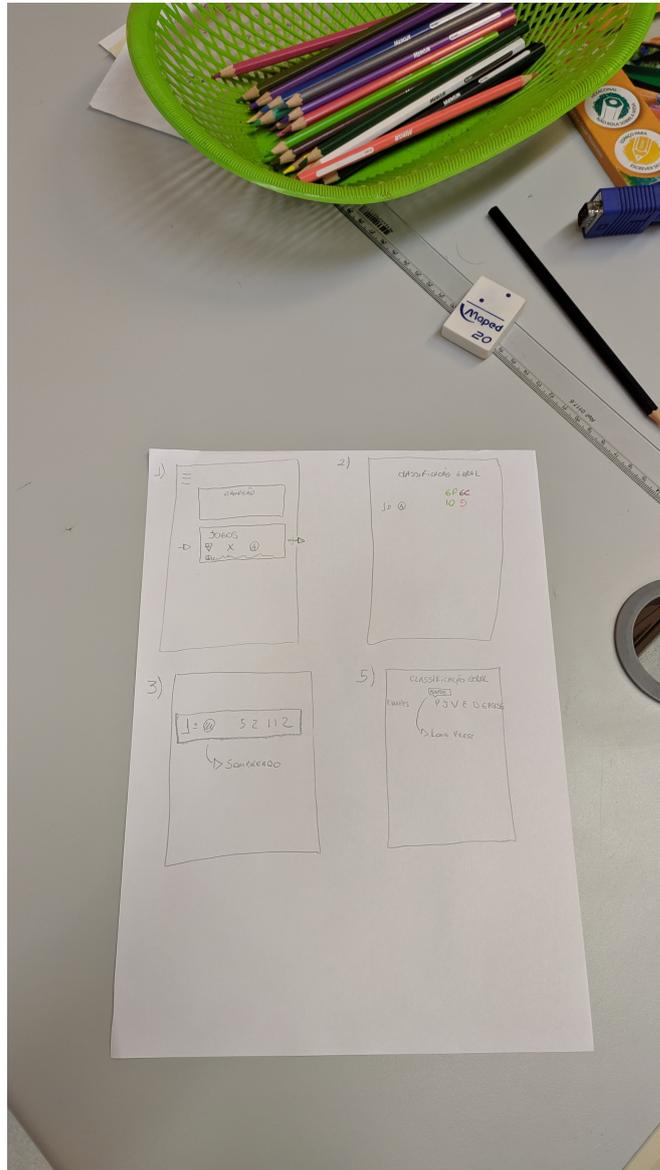
problemas citados, um de cada vez. Para representar sua resposta, era recomendado que o mesmo prototipasse a solução em um papel, tendo também a opção de não considerar um problema ou responder oralmente, caso considerasse a resposta óbvia.

Os participantes foram informados que poderiam não propor uma solução para um suposto problema se não o considerassem um. Desta forma, foi realizada uma relação entre problemas e soluções propostas, identificando que alguns problemas que foram relatados ou percebidos na etapa de avaliação, acabaram não se mostrando como problemas.

Os protótipos desenvolvidos estão anexados nos Apêndice B e Apêndice C

Na Figura 7 o cenário de execução de *redesign* é mostrado, onde estão disponíveis para o participante todos os materiais citados na seção 4.3.

Figura 7 – Material utilizado para execução do *redesign*.



Fonte: o autor

A Tabela 6 mostra o conjunto de decisões de cada participante especialista para cada problema, isto é, se ele propôs uma solução a relação estará preenchida com um “Sim”, mas caso não tenha sido proposta nenhuma solução, a relação estará preenchida com um “Não”. A mesma regra se aplica para a Tabela 5. Também, na última coluna das Tabelas 6 e 7 é mostrada a quantidade de soluções distintas que foram elaboradas para a resolução de cada problema.

Em seguida, foi realizada uma análise de concordância entre os participantes utili-

Tabela 6 – Relação da execução do *redesign* para cada problema por cada especialista.

Problemas	Especialistas					Soluções distintas
	A	B	C	D	E	
1	Sim	Sim	Sim	Sim	Sim	3
2	Sim	Sim	Sim	Sim	Não	1
3	Sim	Sim	Sim	Sim	Sim	4
4	Sim	Sim	Não	Sim	Não	2
5	Sim	Sim	Sim	Sim	Sim	4
6	Sim	Sim	Sim	Não	Sim	4
7	Sim	Sim	Sim	Sim	Sim	4
8	Não	Sim	Sim	Sim	Sim	4
9	Sim	Sim	Sim	Sim	Sim	5
10	Não	Não	Não	Não	Não	0
11	Sim	Sim	Sim	Sim	Sim	3
12	Sim	Sim	Não	Sim	Sim	1
13	Sim	Sim	Sim	Sim	Sim	1
14	Sim	Sim	Sim	Sim	Sim	3
15	Sim	Sim	Sim	Sim	Sim	4

Fonte: o autor

Tabela 7 – Relação da execução do *redesign* para cada problema por cada usuário não especialista.

Problemas	Não Especialistas					Soluções distintas
	A	B	C	D	E	
1	Sim	Sim	Sim	Sim	Sim	5
2	Sim	Sim	Sim	Sim	Não	2
3	Sim	Sim	Sim	Sim	Sim	5
4	Não	Não	Não	Não	Sim	1
5	Sim	Sim	Sim	Sim	Sim	4
6	Sim	Sim	Sim	Sim	Sim	3
7	Sim	Sim	Sim	Sim	Sim	2
8	Sim	Sim	Sim	Sim	Não	1
9	Sim	Sim	Sim	Sim	Não	2
10	Sim	Não	Não	Não	Sim	2

Fonte: o autor

zando o coeficiente Fleiss Kappa⁴. Os valores de kappa são -1 para discordância completa, 0 para concordância como obra do acaso (sorte), +1 para concordância total que não aconteceu ao acaso (há um fator que causa concordância).

Analisando a Figura 8, é possível perceber que o valor de kappa 0,414 representa concordância moderada entre os especialistas e é significativo (Sig. < 0,05).

Já o kappa com valor 0,313 observado na Figura 9 representa uma concordância

⁴ Detalhes sobre o teste aplicado podem ser encontrados em <<https://statistics.laerd.com/spss-tutorials/fleiss-kappa-in-spss-statistics.php>>

Figura 8 – Análise de concordância Fleiss Kappa entre especialistas.

➔ **Fleiss Multirater Kappa**

Acordo Geral ^{a,b}						
	Kappa	Assintótico			Intervalo de Confiança 95% Assintótico	
		Erro padrão	z	Sig.	Limite inferior	Limite superior
Acordo Geral	,414	,082	5,071	,000	,409	,419

a. Os dados de amostra contêm 15 assunto efetivo e 5 avaliadores.

b. Os valores de categoria de classificação fazem distinção entre maiúsculas e minúsculas

Fonte: o autor

marginal entre não especialistas, embora também significativa (Sig. < 0,05).

Figura 9 – Análise de concordância Fleiss Kappa entre não especialistas.

➔ **Fleiss Multirater Kappa**

Acordo Geral ^{a,b}						
	Kappa	Assintótico			Intervalo de Confiança 95% Assintótico	
		Erro padrão	z	Sig.	Limite inferior	Limite superior
Acordo Geral	,313	,100	3,125	,002	,306	,319

a. Os dados de amostra contêm 10 assunto efetivo e 5 avaliadores.

b. Os valores de categoria de classificação fazem distinção entre maiúsculas e minúsculas

Fonte: o autor

Esses dados apontam uma tendência maior dos avaliadores especialistas concordarem no processo de *redesign* ao terem que classificar potenciais problemas de usabilidade como problemas de fato (oferecendo soluções) ou como falsos problemas.

Analisando também a quantidade de soluções distintas no *redesign* realizado por especialistas e não especialistas (Tabelas 6 e 7), um teste Mann-Whitney foi aplicado para verificar diferenças. O resultado disponível na Figura 10 indica que não há diferença significativa entre a quantidade de soluções distintas de *redesign* com especialistas e não especialistas (Sig. > 0,05), ou seja, não é possível dizermos que uma gera soluções mais convergentes que outra.

5.4 Desenvolvimento das versões

Foi realizada uma análise por um juiz especialista na área em conjunto com o pesquisador para escolher soluções que seriam implementadas nas versões finais do produto.

Figura 10 – Resultado do teste Mann-Whitney comparando diferentes propostas de solução de *redesign*.

Resumo de Teste de Hipótese				
	Hipótese nula	Teste	Sig.	Decisão
→ 1	A distribuição de solucoes é igual nas categorias de categoria.	Amostras Independentes de Teste U de Mann-Whitney	,546 ^a	Reter a hipótese nula.
São exibidas significâncias assintóticas. O nível de significância é ,050.				
a. A exata significância é exibida para este teste.				

Fonte: o autor

Em alguns casos soluções compatíveis foram mescladas, em outros a solução adotada era a mais recorrente ou uma combinação das mais recorrentes.

Tendo as soluções escolhidas, foi desenvolvido pelo pesquisador uma versão da aplicação utilizando as soluções propostas por participantes especialistas e outra versão da aplicação utilizando soluções propostas por participantes não especialistas.

A Figura 11 mostra duas telas de exemplo do aplicativo gerado com base em soluções propostas por especialistas. Os problemas (Tabela 4) resolvidos nessas duas telas são os problemas 1, 5, 7, 8, 14 e 15.

Já na Figura 12 são mostradas duas telas de exemplo do aplicativo gerado com base em soluções propostas por não especialistas. Os problemas (Tabela 5) resolvidos nessas duas telas são os problemas 1, 2, 3, 5, 6 e 7.

Comparando as Figuras 11 e 12 é possível perceber que mesmo com alguns problemas se repetindo entre especialistas e não especialistas, as soluções propostas se diferem bastante uma das outras.

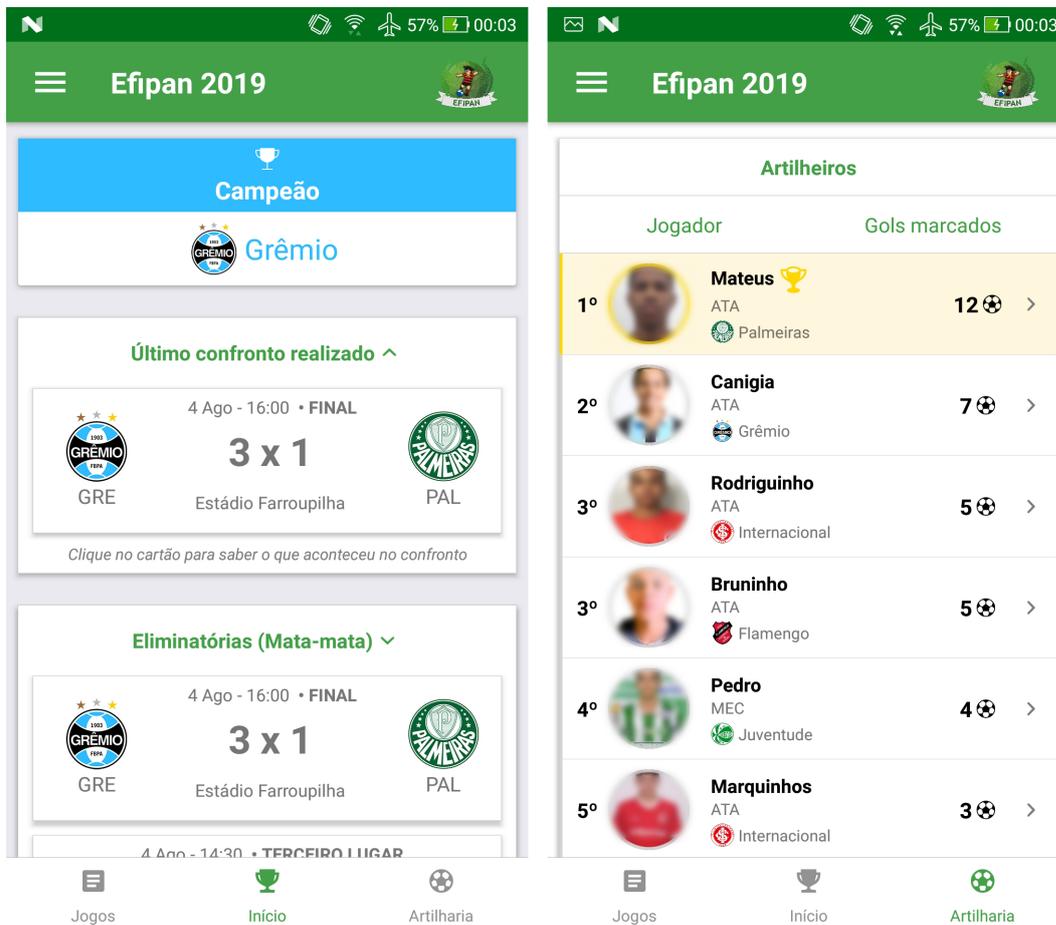
5.5 Avaliação de satisfação

Com as duas versões da aplicação já desenvolvidas, foi realizada a etapa da avaliação da satisfação, onde foram realizadas 20 avaliações, intercalando avaliações no aplicativo gerado por soluções de especialistas com avaliações no aplicativo gerado por soluções de não especialistas.

As avaliações de satisfação foram realizadas utilizando as mesmas tarefas da seção 4.6, e para isso foi necessário que os participantes dessa etapa não tivessem participado de etapas anteriores, para evitar influências nas decisões. Ao final da avaliação, o participante era convidado a responder um questionário SUS, disponível na subseção 2.2.2.1.

A aplicação gerada por especialistas acabou ficando com uma nota média de 89,50. A maior nota foi 100 e a menor nota foi 50. Já a aplicação gerada por não especialistas acabou ficando com uma nota média de 91,25, sendo 100 sua maior nota e 75 a menor nota.

Figura 11 – Exemplo de telas do aplicativo desenvolvido baseado no *redesign* com especialistas



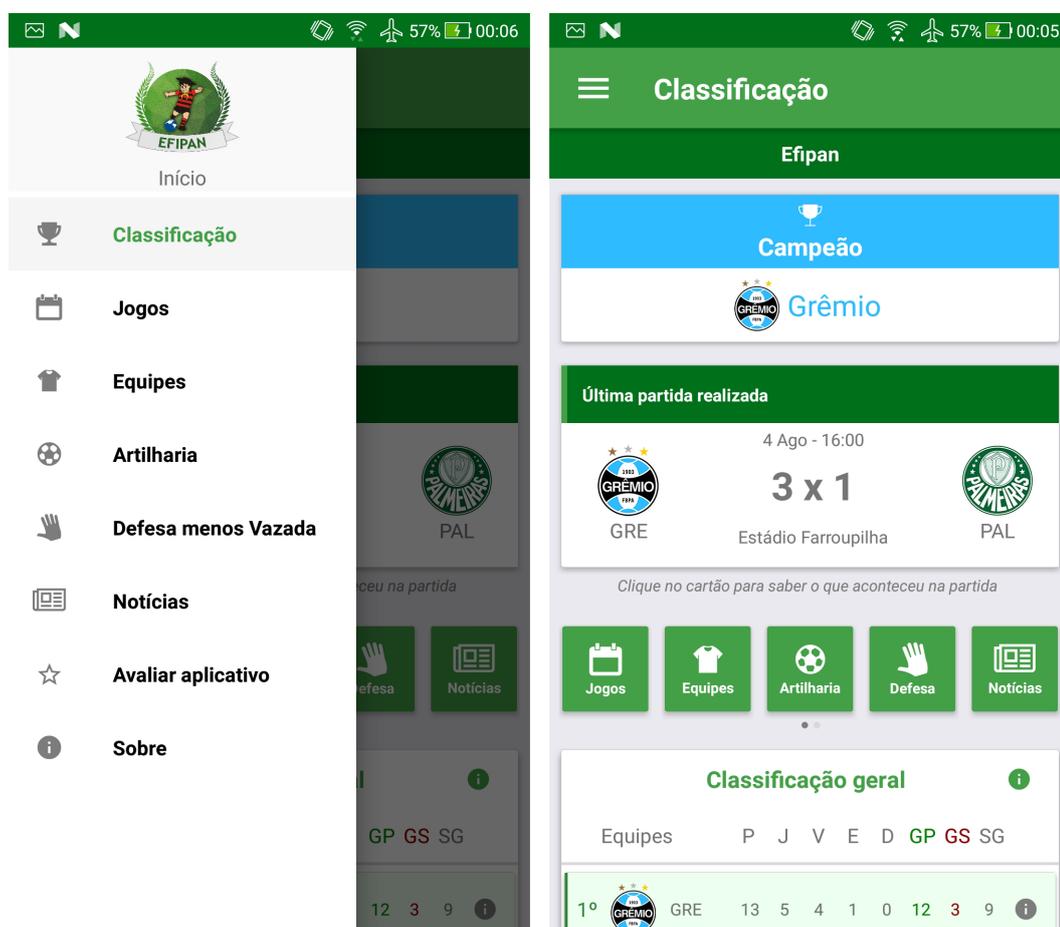
Fonte: o autor

Tabela 8 – Resultados da avaliação final SUS - Versão Especialistas

Avaliadores	Perguntas										Nota
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	
A1	5	1	5	1	5	1	5	1	5	1	100
A2	5	1	5	1	5	1	5	1	5	1	100
A3	5	1	5	1	5	1	5	1	5	1	100
A4	4	3	4	3	5	1	4	3	3	4	65
A5	5	2	4	1	5	1	4	1	4	2	87,5
A6	5	1	5	1	4	2	5	1	5	1	95
A7	2	3	3	4	5	2	5	4	3	5	50
A8	5	1	5	1	5	1	5	1	5	1	100
A9	5	2	5	1	5	1	5	1	5	1	97,5
A10	5	1	5	1	5	1	5	1	5	1	100
Média Final da Nota 89,50											

Fonte: o autor

Figura 12 – Exemplo de telas do aplicativo desenvolvido baseado no *redesign* com não especialistas



Fonte: o autor

Tabela 9 – Resultados da avaliação final SUS - Versão Não especialistas

Avaliadores	Perguntas										Nota
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	
A1	5	1	5	1	5	2	5	2	5	1	95
A2	4	2	4	1	5	1	5	1	4	1	90
A3	1	2	4	1	5	1	2	1	4	1	75
A4	5	1	5	1	5	1	5	1	5	1	100
A5	5	1	5	1	5	1	4	1	5	1	97,5
A6	5	1	4	2	5	1	5	1	5	3	90
A7	5	1	5	1	5	1	5	1	5	1	100
A8	4	1	4	1	5	1	5	1	5	2	92,5
A9	5	2	5	2	5	1	3	1	5	1	90
A10	4	3	5	2	4	1	5	2	4	1	82,5
Média Final da Nota 91,25											

Fonte: o autor

5.6 Análise e conclusão dos resultados

Observando os dados das Tabelas 8 e 9 referentes à avaliação final, aplicou-se um teste Shapiro-Wilk para verificar a distribuição dos dados (Figura 13). Como é possível observar, as distribuições não são as mesmas: os escores de não especialistas seguem uma distribuição normal (Sig. $> 0,05$) e os escores de especialistas não (Sig. $\leq 0,05$).

Figura 13 – Teste de normalidade de dados dos escores do SUS.

	Shapiro-Wilk		
	Estatística	gl	Sig.
Não especialistas	,909	10	,273
Especialistas	,676	10	,000

Fonte: o autor

Sob as condições indicadas, é necessário aplicar um teste não paramétrico. Assim, foi utilizado um teste Mann-Whitney conferir se há diferença entre as pontuações das avaliações SUS aplicadas nas aplicações geradas.

Pode ser observado na Figura 14 uma significância (Sig.) de 0,436, indicando que não é possível dizer que há diferença significativa entre as pontuações do SUS nos aplicativos gerados pelo *redesign* envolvendo especialistas e não especialistas (Sig. $> 0,05$).

Figura 14 – Resultado do teste Mann-Whitney para escores do SUS.

Resumo de Teste de Hipótese				
	Hipótese nula	Teste	Sig.	Decisão
1	A distribuição de escores do SUS é igual nas categorias de especialistas e não especialistas.	Amostras Independentes de Teste U de Mann-Whitney	,436 ^a	Retar a hipótese nula.

São exibidas significâncias assintóticas. O nível de significância é ,050.

a. A exata significância é exibida para este teste.

Fonte: o autor

6 CONSIDERAÇÕES FINAIS

O trabalho apresentado visa identificar diferenças entre a utilização de técnicas de avaliação e *redesign* executadas por especialistas em IHC com técnicas de avaliação e *redesign* executadas por não especialistas em IHC.

A execução desse estudo envolveu avaliações de usabilidade, *redesign*, desenvolvimento de software e avaliações de satisfação. Para cada uma dessas atividades foram coletados dados a partir de especialistas em IHC e também de não especialistas, a fim de obter uma comparação entre os diferentes conjuntos de técnicas.

Como contribuição importante, foi constatado que há similaridade nos tempos para completar as tarefas nas avaliações realizadas, fazendo com que não seja mais rápido aplicar um conjunto de técnicas específico. A diferença entre o grau de severidade dos problemas encontrados também não é significativa para podermos afirmar que especialistas tendem a encontrar mais problemas mais graves do que não especialistas. Verificou-se também que especialistas tendem a concordar mais na classificação de problemas de usabilidade em etapas de *redesign* em comparação aos não especialistas (algo que pode ser justificado pela formação técnica específica dos especialistas). Porém, não é possível dizer que há diferença significativa entre a quantidade de soluções propostas por especialistas e não especialistas.

Comparando os resultados das avaliações de satisfação, também não é possível afirmar que uma versão gerada por especialistas é melhor que uma versão gerada por não especialistas, ou vice-versa. Pela análise estatística realizada, foi constatado que a diferença entre os dois resultados não é significativa o suficiente para identificar uma abordagem que gere resultado de melhor qualidade.

Desta forma, os resultados apresentados demonstram que na prática, ambas as abordagens de avaliação e *redesign* se mostraram similarmente eficientes no contexto deste trabalho.

Foram identificadas algumas ameaças que podem ter afetado os resultados deste trabalho. A primeira delas, é a baixa quantidade de participantes (20) devido à falta de tempo na etapa de avaliação de satisfação utilizando o questionário SUS. Outra ameaça é a possibilidade de alguns participantes terem utilizado o aplicativo escolhido em sua versão original, podendo afetar a realização das tarefas nas etapas de avaliação por já terem experiência com o software e conhecerem sua estrutura. A utilização de um único juiz especialista para escolher as soluções que foram utilizadas no desenvolvimento das versões finais também pode introduzir viés no trabalho, e poderia ter sido substituída por uma etapa na qual os próprios participantes que propuseram soluções entrassem em consenso sobre as soluções implementadas. Essa etapa não foi realizada por falta de tempo e sua ausência é considerada uma ameaça aos resultados. Por fim, a última ameaça considerada foi que as duas versões de aplicativos foram desenvolvidas pelo próprio pesquisador.

Considera-se como trabalho futuro a execução das avaliações de satisfação com

um maior número de participantes, a fim de verificar se uma quantidade maior de participantes permite detectar diferença significativa na qualidade dos produtos. Também, como trabalho futuro, seria possível executar novamente as etapas de *redesign* e realizar um grupo focal entre os participantes para montar uma solução final unindo as soluções propostas, minimizando assim o viés de ter um juiz especialista atuando nos resultados.

Por fim, conclui-se que o objetivo do trabalho foi atingido, tendo em vista os resultados que mostram de forma sistemática e com rigor estatístico que não há diferenças significativas entre utilizar técnicas de avaliação e *redesign* de interfaces baseadas em especialistas e baseadas na participação de usuários não especialistas para a evolução de um aplicativo móvel.

REFERÊNCIAS

- BARBOSA, S.; SILVA, B. **Interação Humano-Computador**. Elsevier, 2010. Disponível em: <https://books.google.com.br/books?id=qk0skwr_cewC&printsec=frontcover&hl=pt-BR&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false>. Citado 5 vezes nas páginas 22, 23, 26, 27 e 30.
- BROOKE, J. Sus: A quick and dirty usability scale. **Usability Eval. Ind.**, v. 189, 11 1986. Citado na página 23.
- CYBIS, W. de A.; BETIOL, A.; FAUST, R. **Ergonomia e Usabilidade 3ª edição: Conhecimentos, Métodos e Aplicações**. NOVATEC, 2015. ISBN 9788575224595. Disponível em: <<https://books.google.com.br/books?id=AOa5CgAAQBAJ>>. Citado 2 vezes nas páginas 19 e 21.
- DOUBLEDAY, A. et al. A comparison of usability techniques for evaluating design. In: **Proceedings of the 2Nd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques**. New York, NY, USA: ACM, 1997. (DIS '97), p. 101–110. ISBN 0-89791-863-0. Disponível em: <<http://doi.acm.org/10.1145/263552.263583>>. Citado 2 vezes nas páginas 31 e 32.
- FERREIRA, J.; BARR, P.; NOBLE, J. The semiotics of user interface redesign. In: **Proceedings of the Sixth Australasian Conference on User Interface - Volume 40**. Darlinghurst, Australia, Australia: Australian Computer Society, Inc., 2005. (AUIC '05), p. 47–53. ISBN 1-920682-22-8. Disponível em: <<http://dl.acm.org/citation.cfm?id=1082243.1082250>>. Citado na página 19.
- HEWETT, T. T. et al. **ACM SIGCHI Curricula for Human-Computer Interaction**. New York, NY, USA, 1992. Citado na página 21.
- Iqbal, R. et al. Project management system review and redesign using user-centred design methodology. In: **Proceedings of the 2011 15th International Conference on Computer Supported Cooperative Work in Design (CSCWD)**. [S.l.: s.n.], 2011. p. 139–145. Citado 2 vezes nas páginas 30 e 32.
- ISO-9241. 2018. <<https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en>>. Acessado em: 23/04/2019. Citado na página 21.
- LAWSON, B. How designers think – the design process demystified. **University Press, Cambridge**, 01 2006. Citado na página 21.
- NIELSEN, J. Usability inspection methods. In: . [S.l.]: Conference Companion on Human Factors in Computing Systems, 1994. Citado na página 19.
- NIELSEN, J.; MOLICH, R. Heuristic evaluation of user interfaces. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: ACM, 1990. (CHI '90), p. 249–256. ISBN 0-201-50932-6. Disponível em: <<http://doi.acm.org/10.1145/97243.97281>>. Citado 2 vezes nas páginas 24 e 26.
- NNGROUP. 2012. <<https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/>>. Acessado em: 30/04/2019. Citado 2 vezes nas páginas 27 e 28.

PRESSMAN, R. **Engenharia de Software - 7.ed.** McGraw Hill Brasil, 2009. ISBN 9788580550443. Disponível em: <<https://books.google.com.br/books?id=y0rH9wuXe68C>>. Citado na página 27.

PROTOTIPAÇÃO, A MELHOR FORMA DE TESTAR E VALIDAR UM PROJETO. 2018. <<https://designcomcafe.com.br/prototipacao-a-melhor-forma-de-testar-e-validar-um-projeto/>>. Acessado em: 22/08/2019. Citado na página 22.

SHARP, H.; ROGERS, Y.; PREECE, J. **Interaction Design: Beyond Human Computer Interaction.** [S.l.]: John Wiley & Sons, 2007. ISBN 0470018666. Citado na página 21.

Acknowledgments. In: SNYDER, C. (Ed.). **Paper Prototyping.** Burlington: Morgan Kaufmann, 2004, (Interactive Technologies). p. xxiii – xxiv. ISBN 978-1-55860-870-2. Disponível em: <<http://www.sciencedirect.com/science/article/pii/B9781558608702500243>>. Citado na página 22.

SOMMERVILLE, I. **Engenharia de software.** PEARSON BRASIL, 2011. ISBN 9788579361081. Disponível em: <<https://books.google.com.br/books?id=H4u5ygAACAAJ>>. Citado 2 vezes nas páginas 19 e 21.

THELIN, T.; RUNESON, P.; WOHLIN, C. An experimental comparison of usage-based and checklist-based reading. **IEEE Transactions on Software Engineering**, v. 29, n. 8, p. 687–704, Aug 2003. ISSN 0098-5589. Citado 2 vezes nas páginas 29 e 32.

USABILITY. 2018. <<https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>>. Acessado em: 17/05/2018. Citado na página 24.

VREDENBURG, K. et al. A survey of user-centered design practice. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.** New York, NY, USA: ACM, 2002. (CHI '02), p. 471–478. ISBN 1-58113-453-3. Disponível em: <<http://doi.acm.org/10.1145/503376.503460>>. Citado 2 vezes nas páginas 31 e 32.

WALLNAU, K. C.; SEACORD, R. C.; ROBERT, J. A survey of legacy system modernization approaches. In: . [S.l.: s.n.], 2000. Citado na página 19.

Apêndices

APÊNDICE A – TERMO DE CONSENTIMENTO COM DEVPAMPA

TERMO DE CONSENTIMENTO

Eu, **Adriel Herter Seffrin**, de número de RG _____ e CPF _____, (doravante denominado **realizador do estudo**) solicito a disponibilização do código-fonte do produto "**Efipan**" (doravante denominado **produto**) para a aplicação de um estudo de caso dentro da **DevPampa**, de CNPJ **32.110.313/0001-99**, (doravante denominada **empresa**) e relacionado ao Trabalho de Conclusão de Curso de Bacharelado em Engenharia de Software para que sejam realizadas avaliações diversas para encontrar problemas de usabilidade e redesign do **produto**.

A **empresa** é detentora do **produto** original e de quaisquer subprodutos gerados nesse estudo de caso, sendo apenas a autoria dos subprodutos atribuída ao **realizador do estudo**.

Se houver alguma dúvida, qualquer representante da **empresa** poderá entrar em contato com o realizador do estudo pelo endereço de e-mail < adriel.herter@gmail.com > ou com seu orientador pelo endereço de e-mail < jeancheiran@unipampa.edu.br > para solicitar esclarecimentos.

A **empresa** não terá qualquer custo na disponibilização do **produto** para esse estudo. Ainda, a **empresa** não receberá qualquer vantagem pela disponibilização do **produto** nesse estudo.

Vale ressaltar que a realização desse estudo de caso poderá trazer benefícios à **empresa**, uma vez que as avaliações e os processos de redesign tendem a aumentar a qualidade do **produto** para seus usuários finais.

O **realizador do estudo** é o responsável pelo código-fonte do **produto**, é sua responsabilidade a não divulgação desse código-fonte em qualquer meio.

A sua colaboração é de extrema importância e acredito que trará benefício para as duas partes.

Sendo assim, havendo concordância entre as partes, esse termo serve como consentimento da colaboração da **empresa** e do **realizador do estudo** no estudo de caso que será realizado.

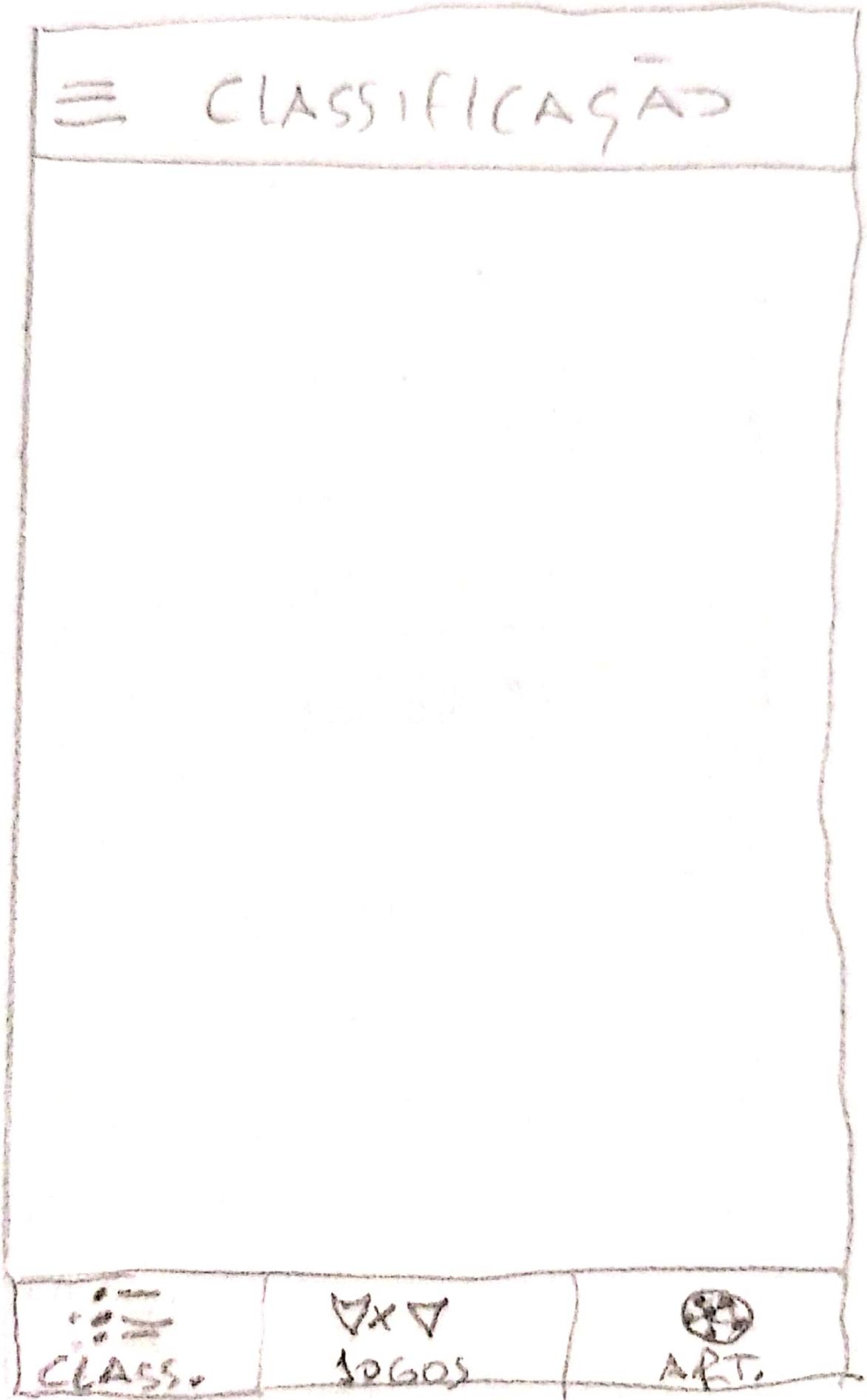
Assinatura dos responsáveis:

Adriel Herter Seffrin
(Realizador do estudo)

DevPampa
(Administrador da empresa)

APÊNDICE B – REDESIGN COM ESPECIALISTAS

1)



1)

≡ _____

CAMPEÃO

○ ~

Última partida

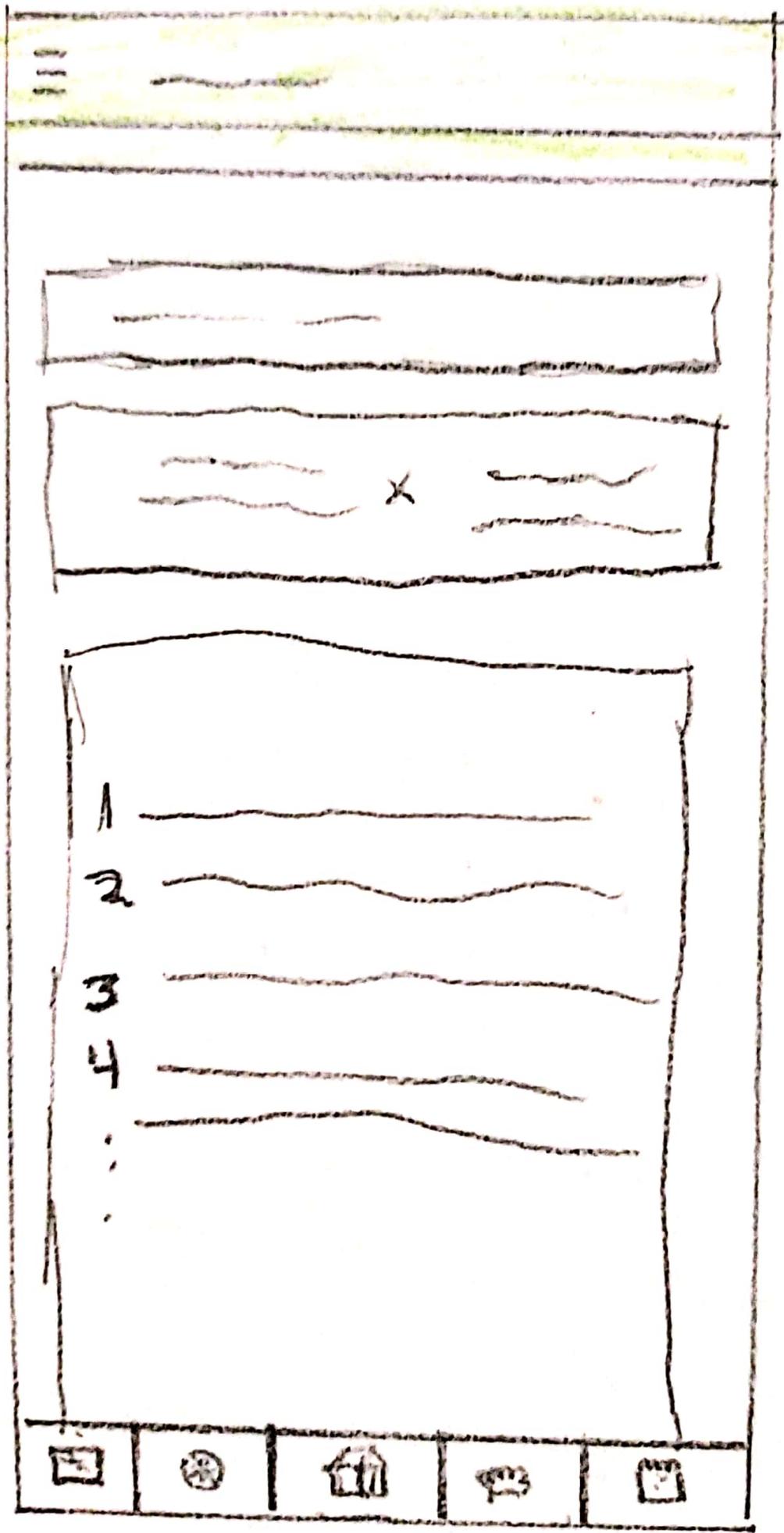
○ x ○

127 artilharia →

127 Jogos →

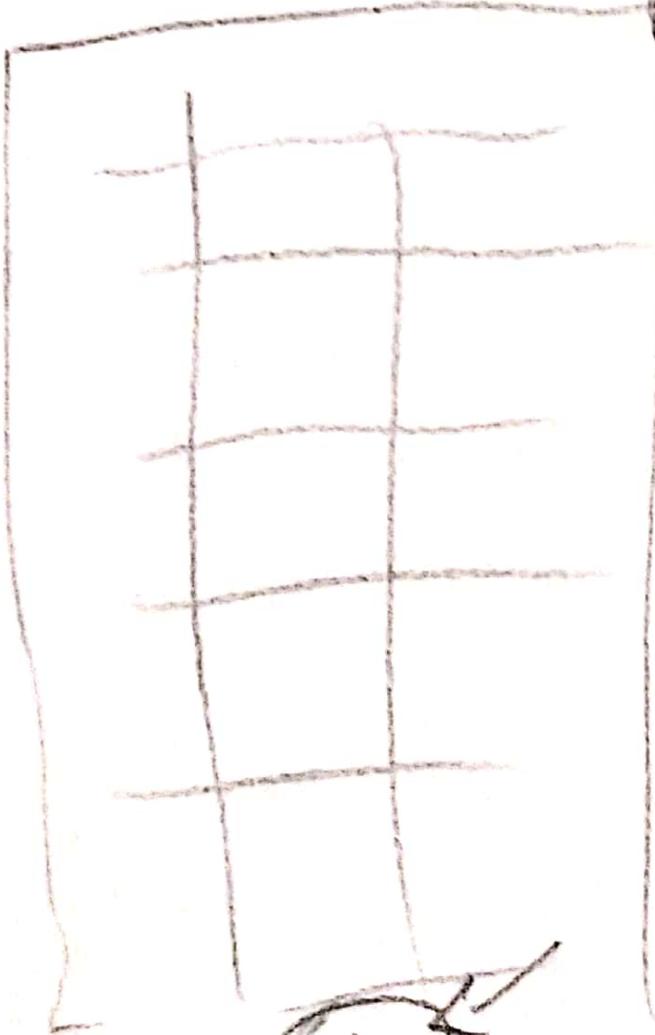
Classificação

1.

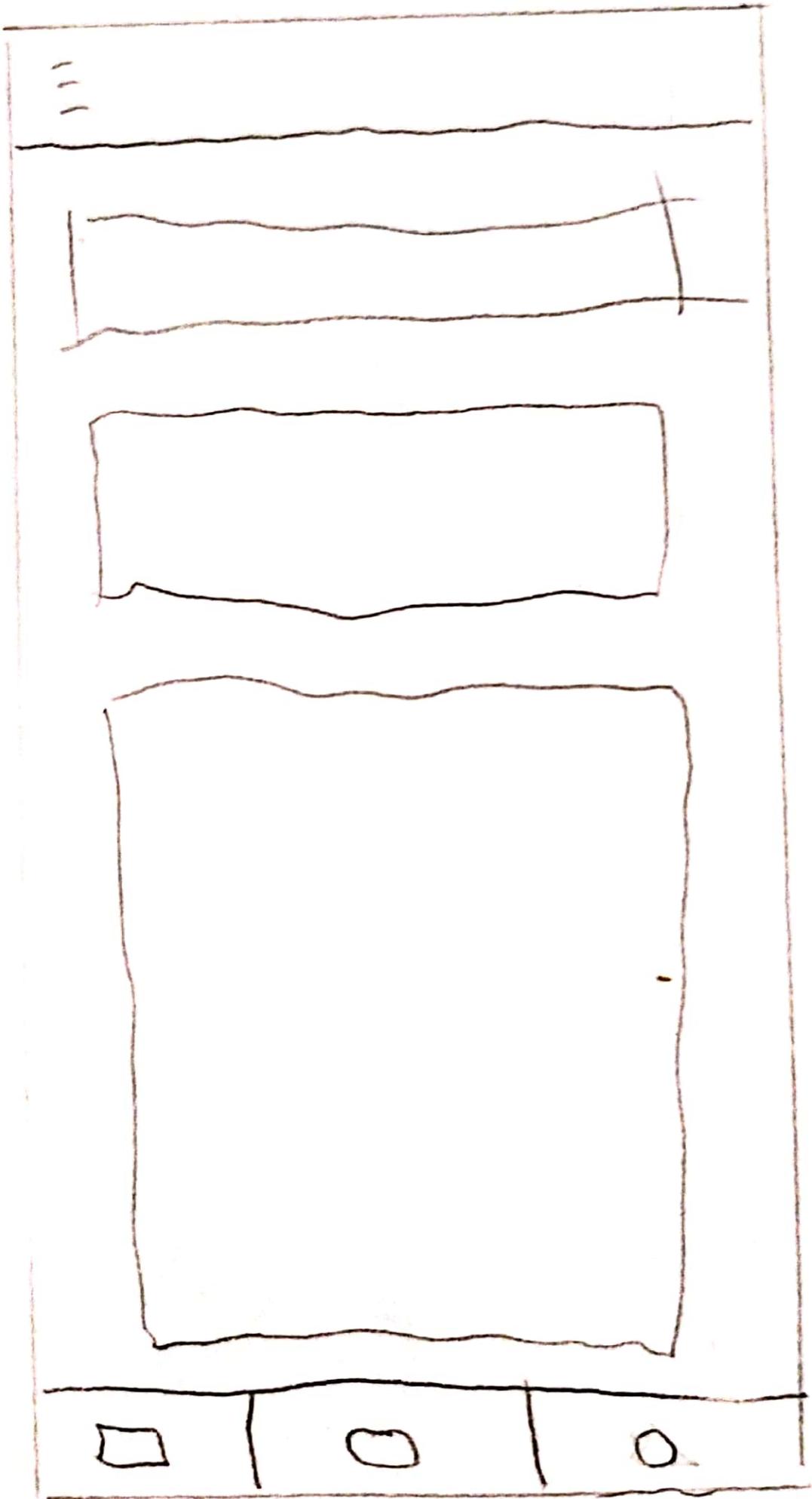


1

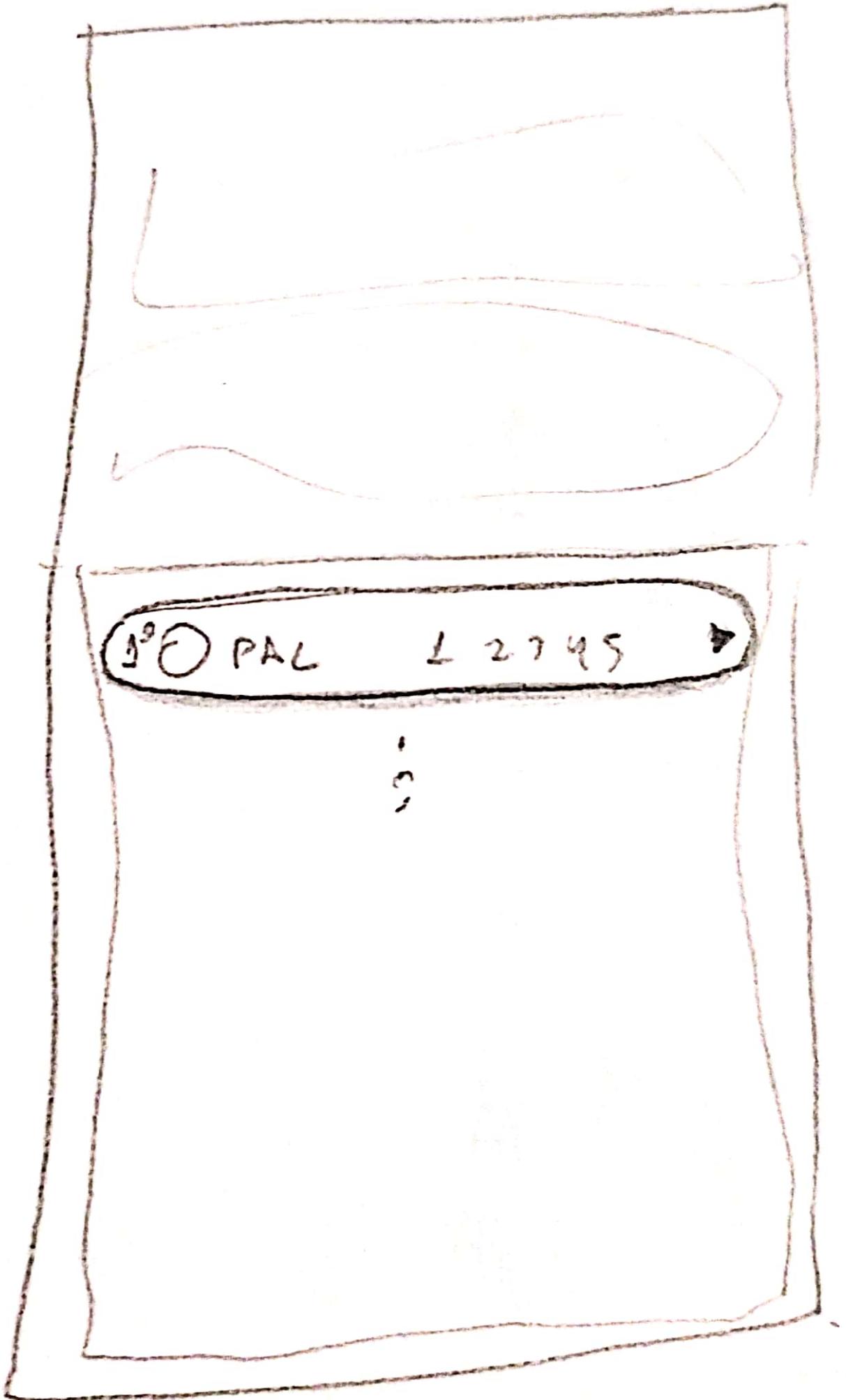
III



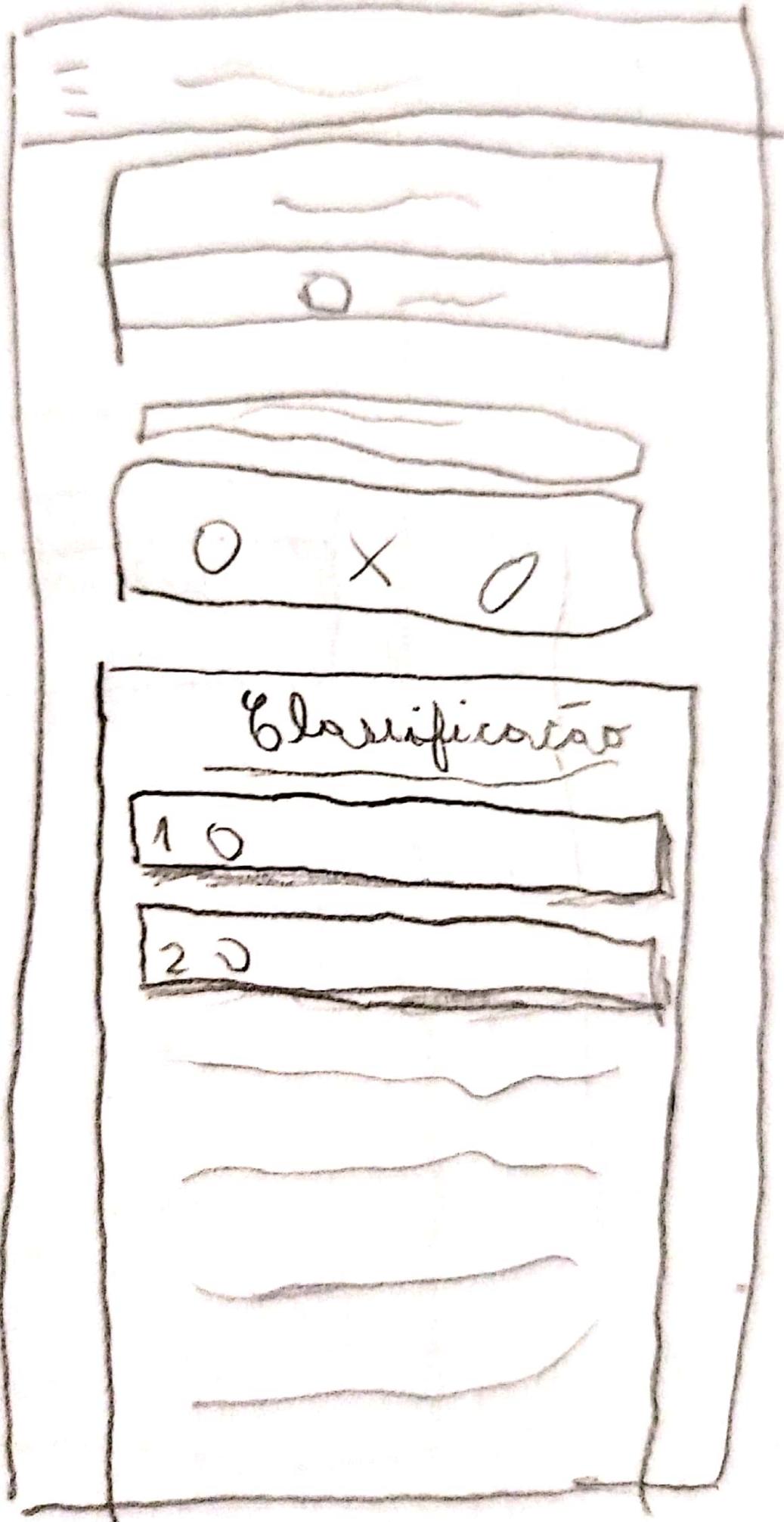
1 =



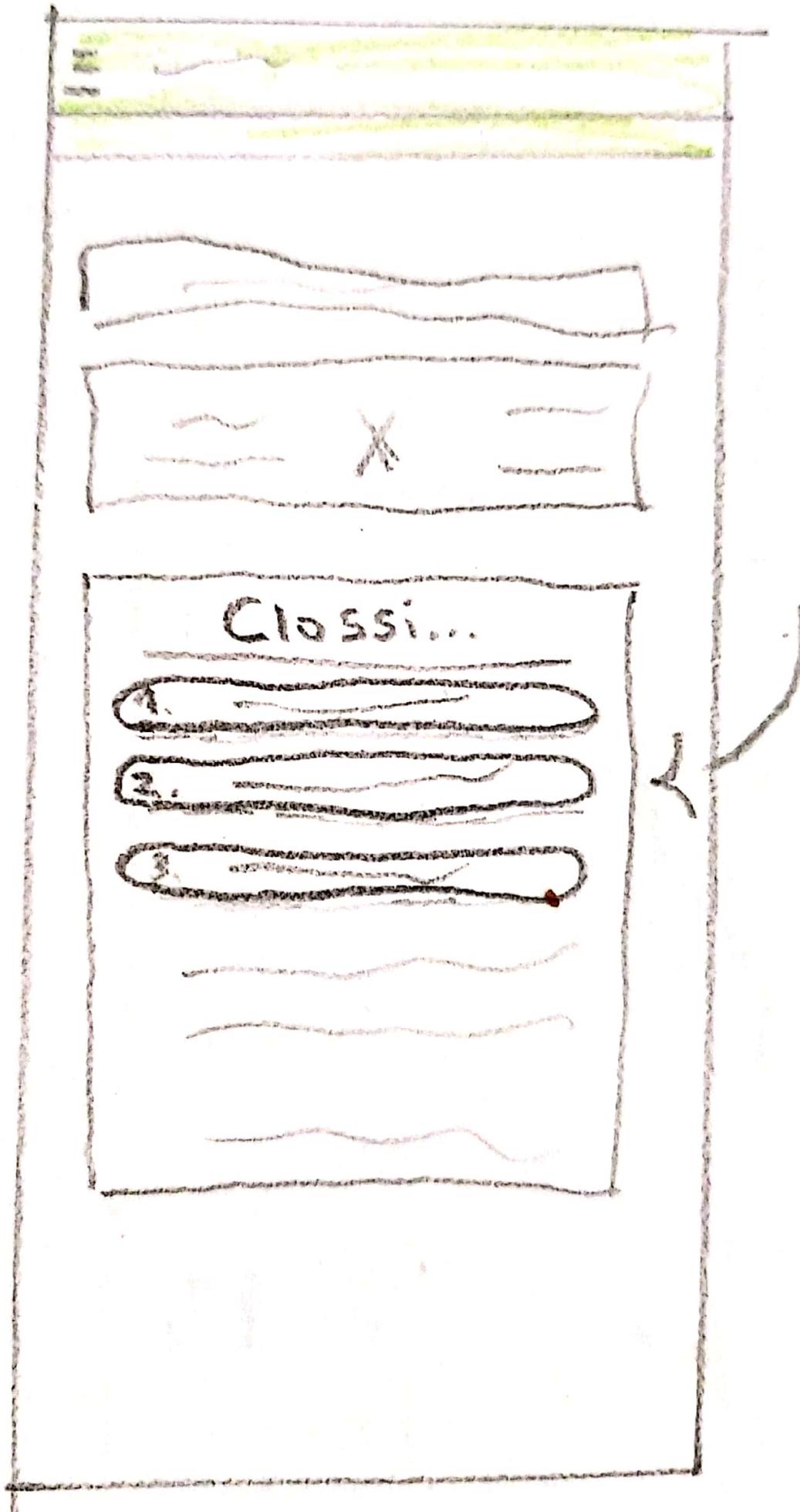
3)



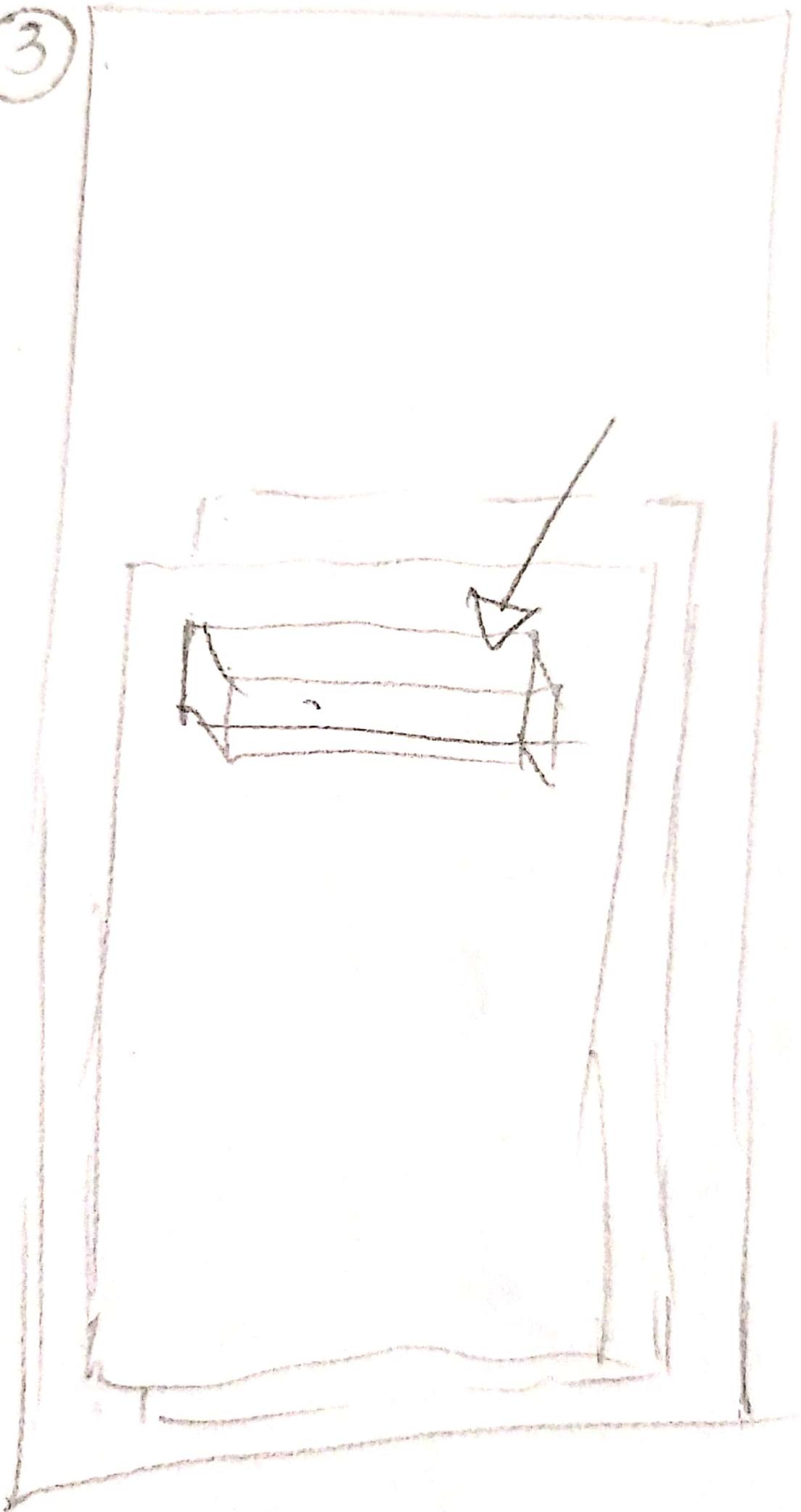
3)



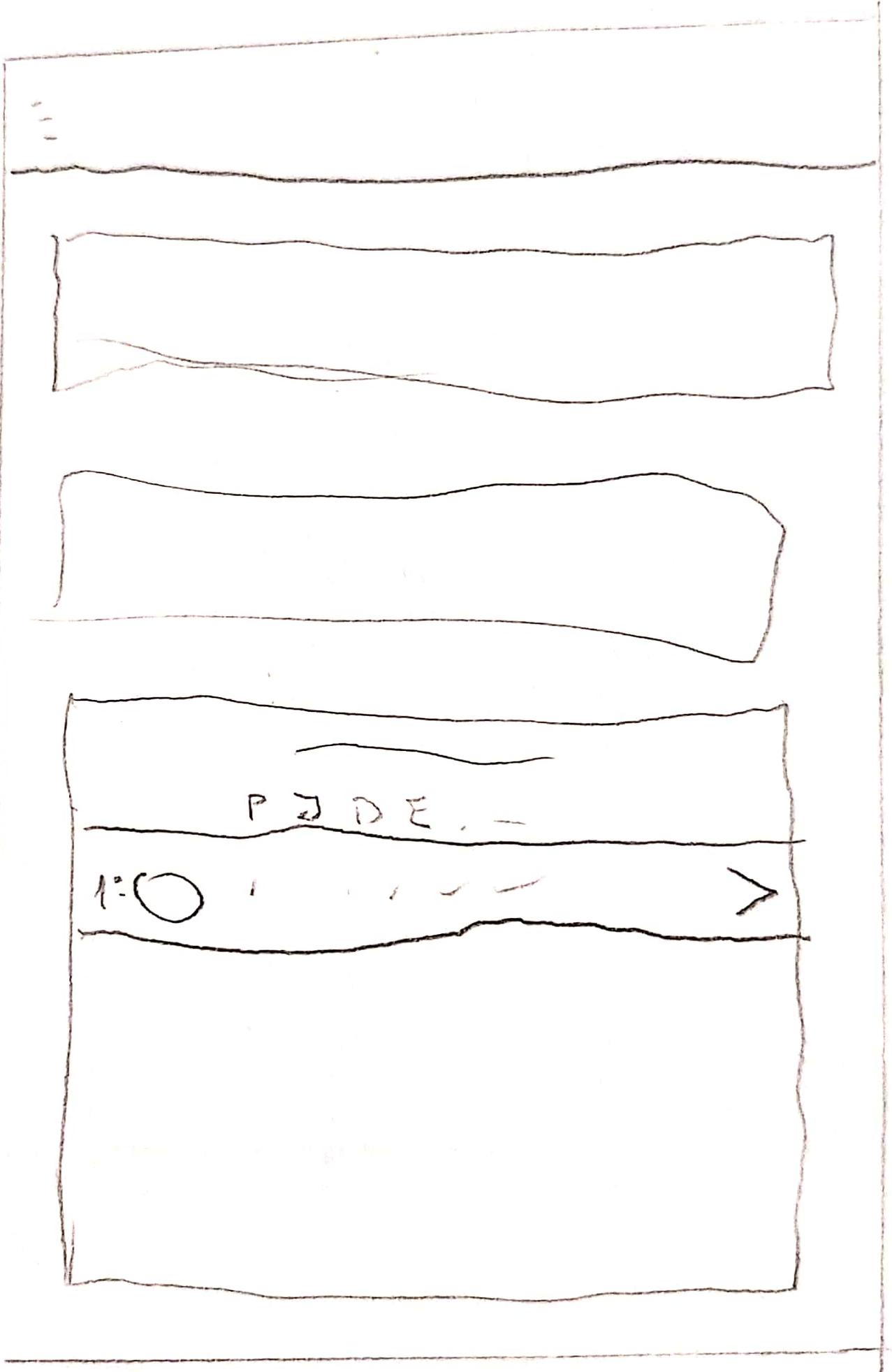
3.



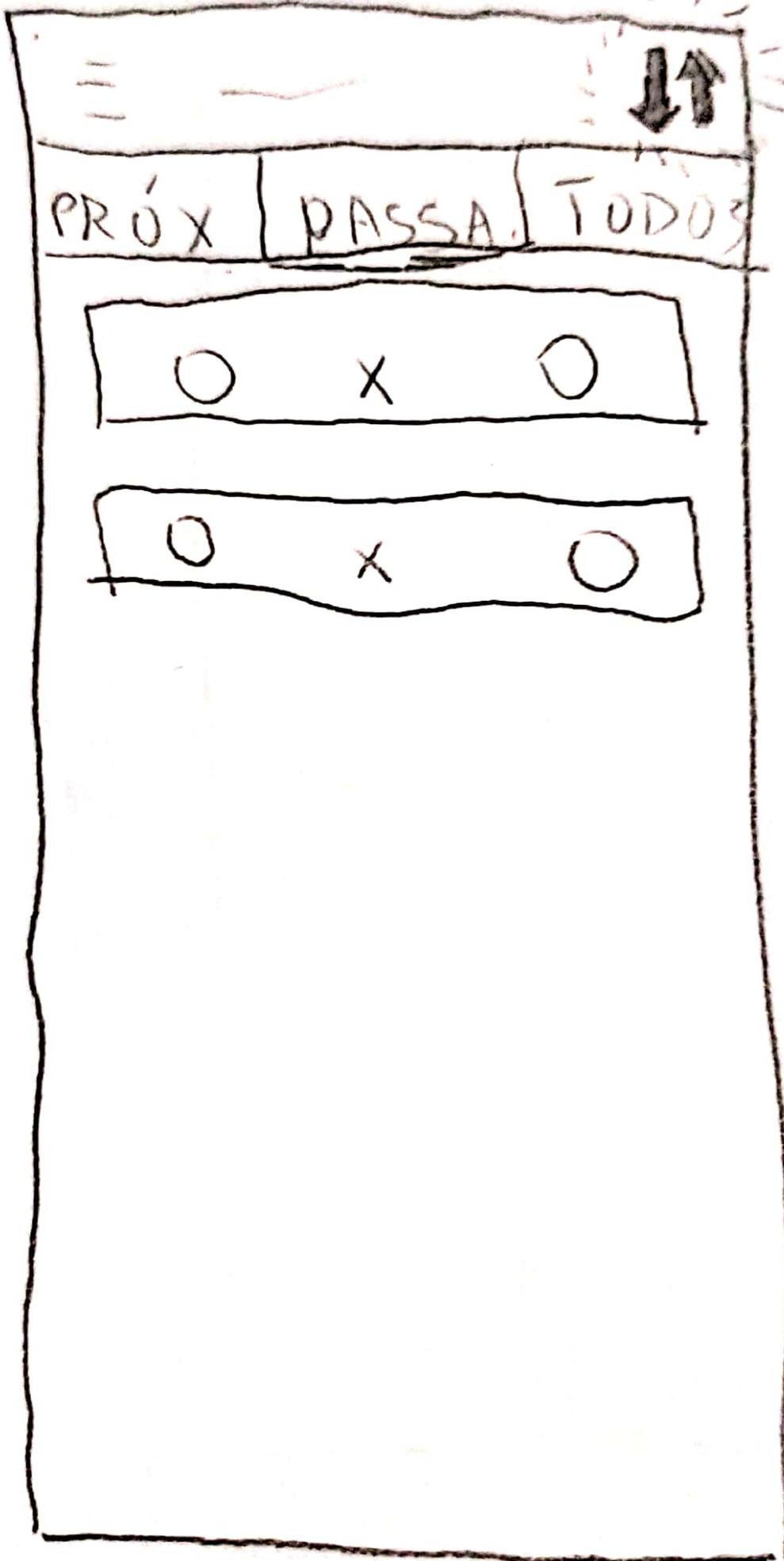
3



3-



4)

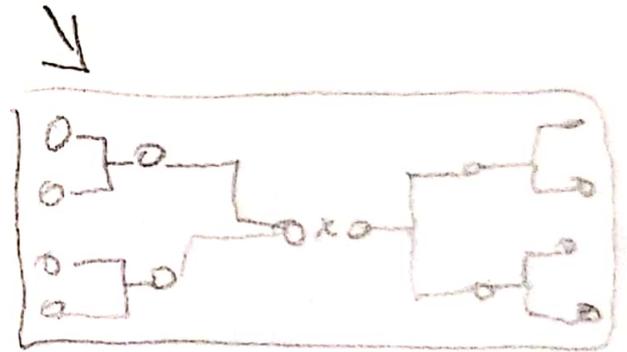
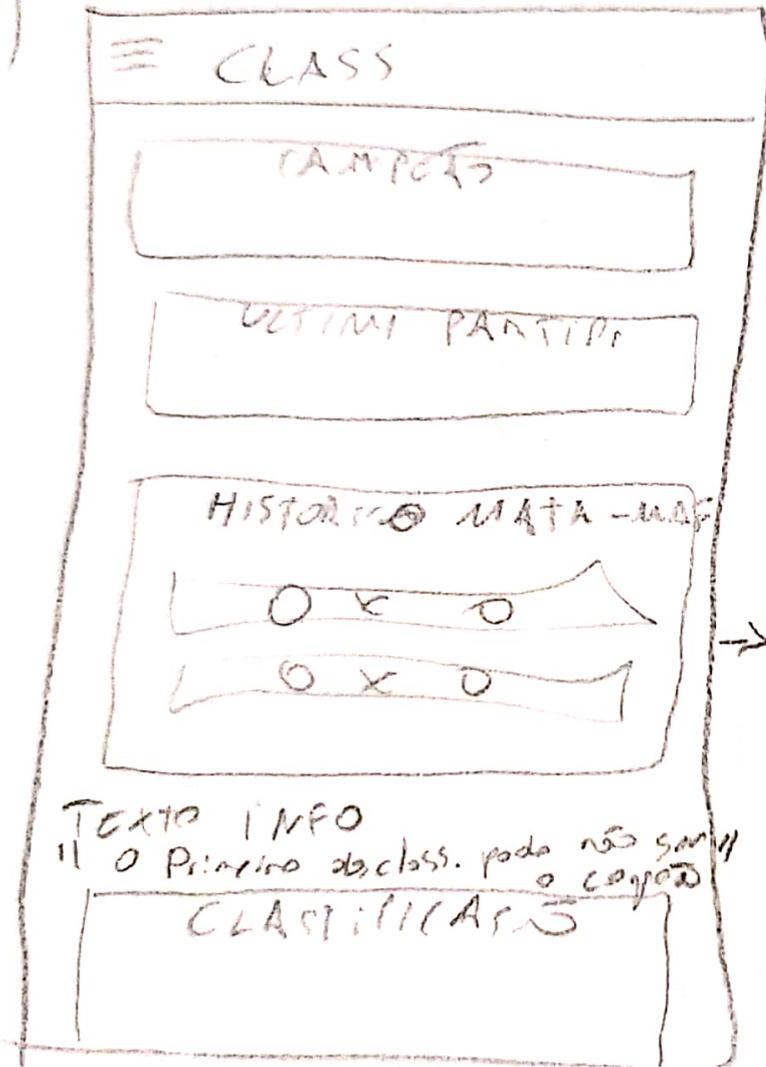


Order

4)

≡ JOGOS		
PROA	PASS.	TODOS
ORDENAÇÃO; CRESCENTE ▾		
CBASI	JOGOS	ART

5)

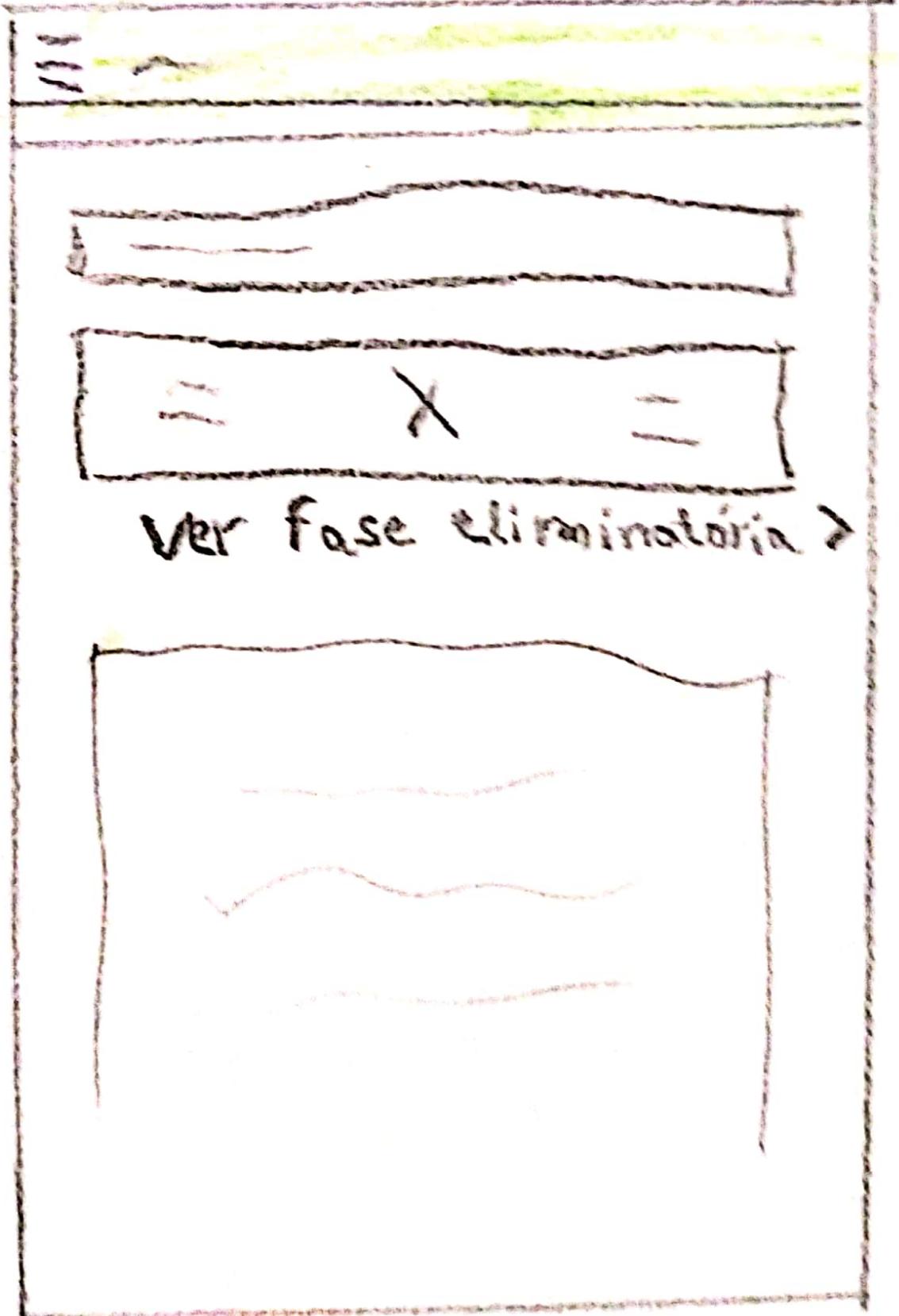


57)

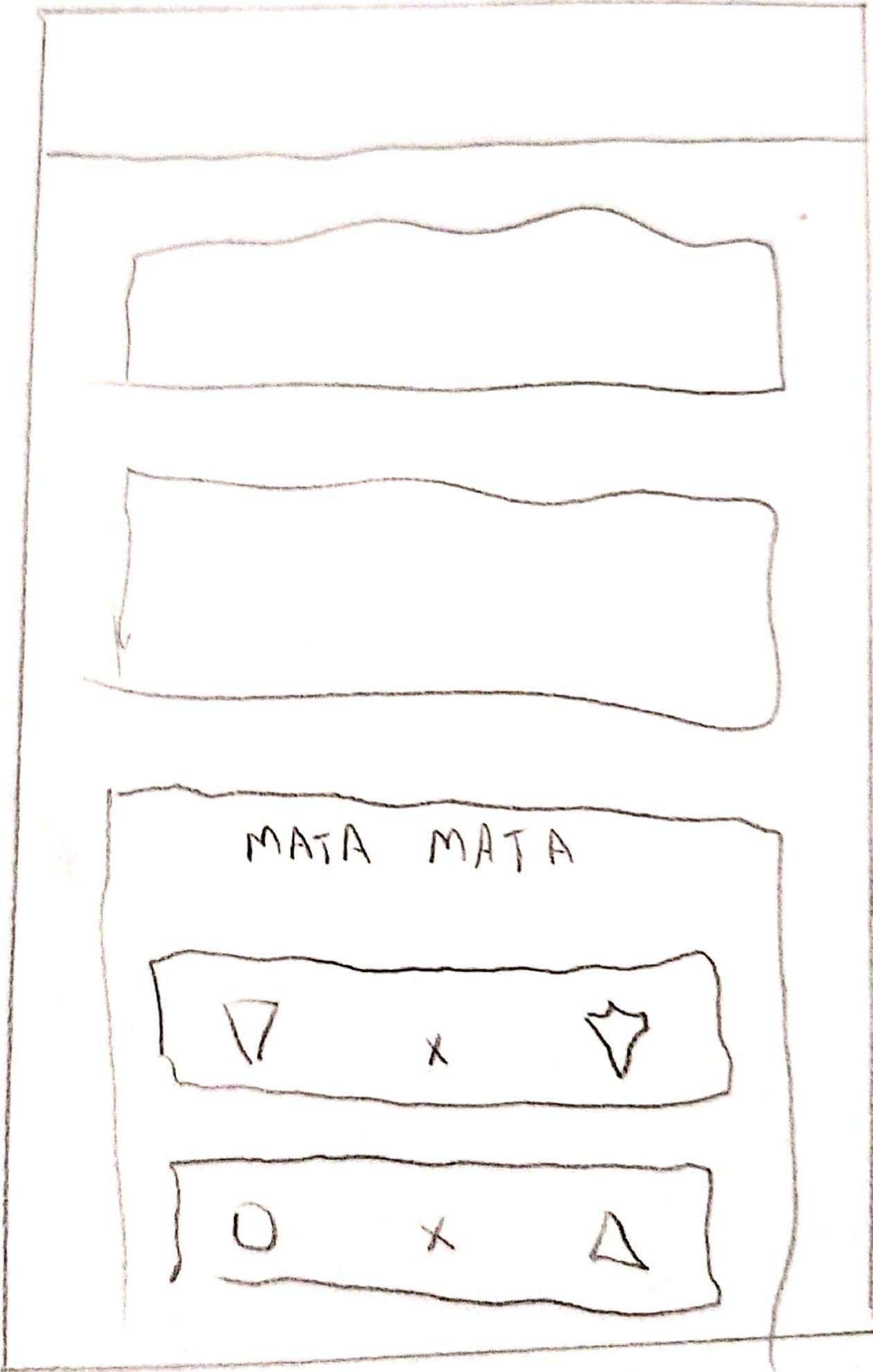


expandit

5.



51



6)

≡ J2602

4º RODADA

O X O

O X O

5º RODADA

O V O

O X O

6.

3 Jogos

7ª RODADA

— X —

— X —

FINAL

— X —

6



FINAL



TERCEIRO
LUGAR



6-

JOGOS

RODADA 1



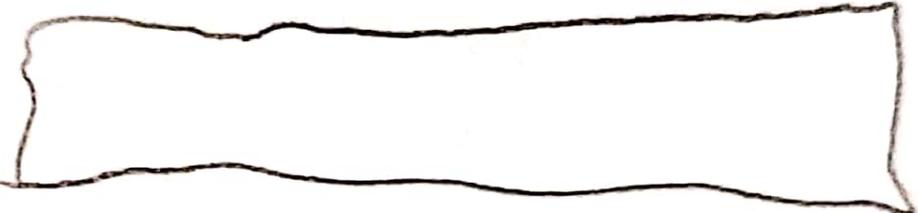
RODADA 2



7 -



ultima partida

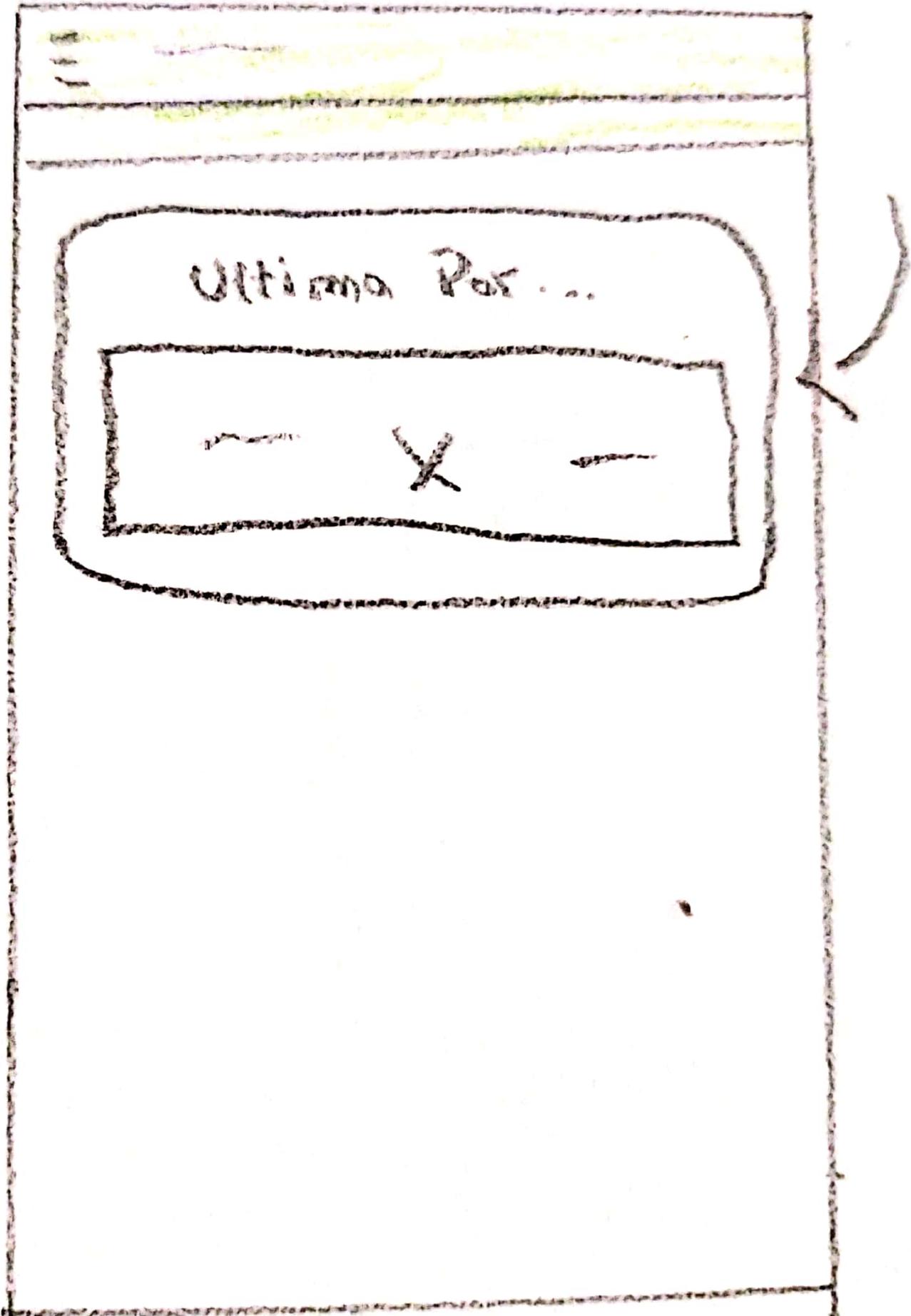


7

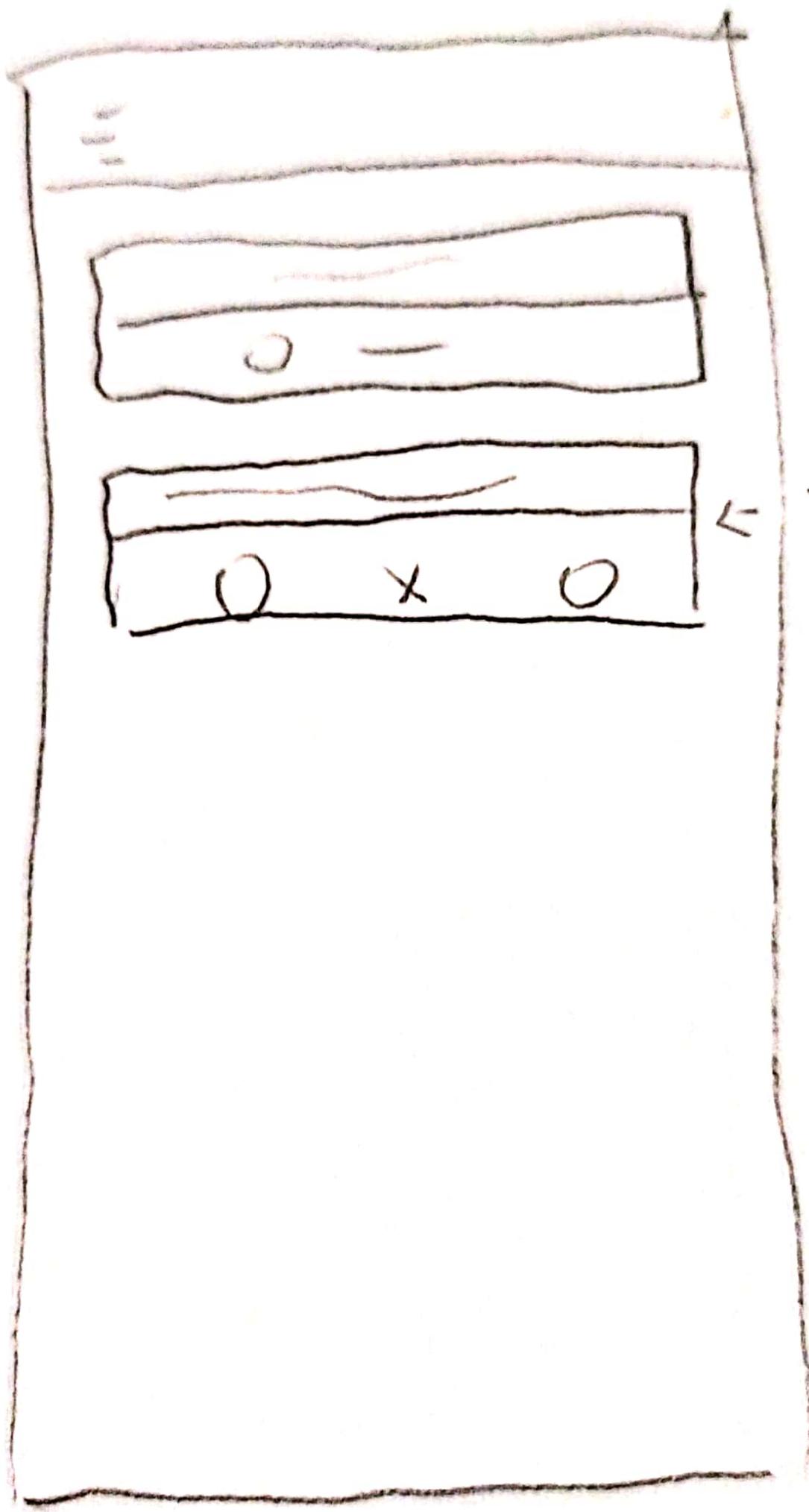


CARD

7

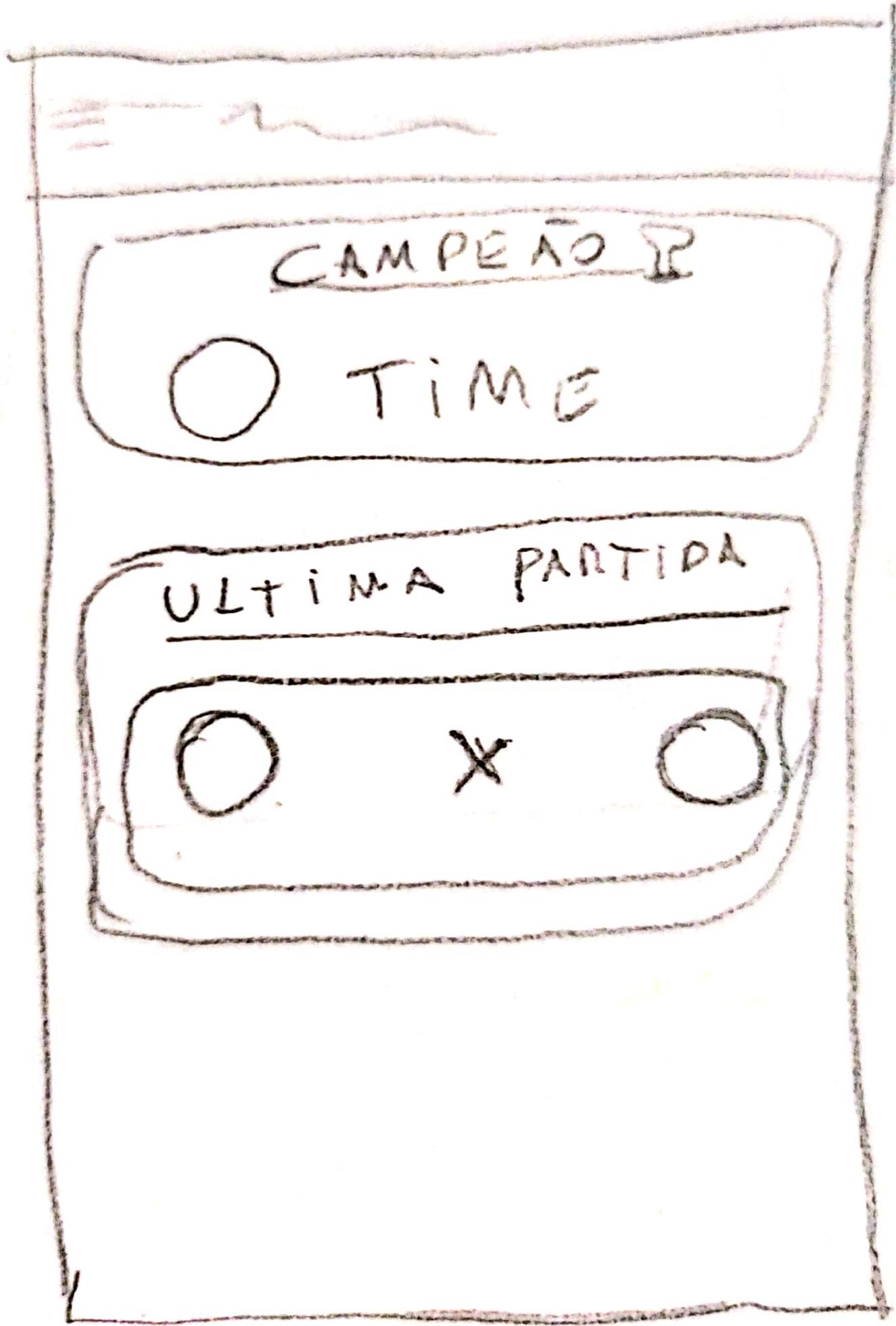


F)

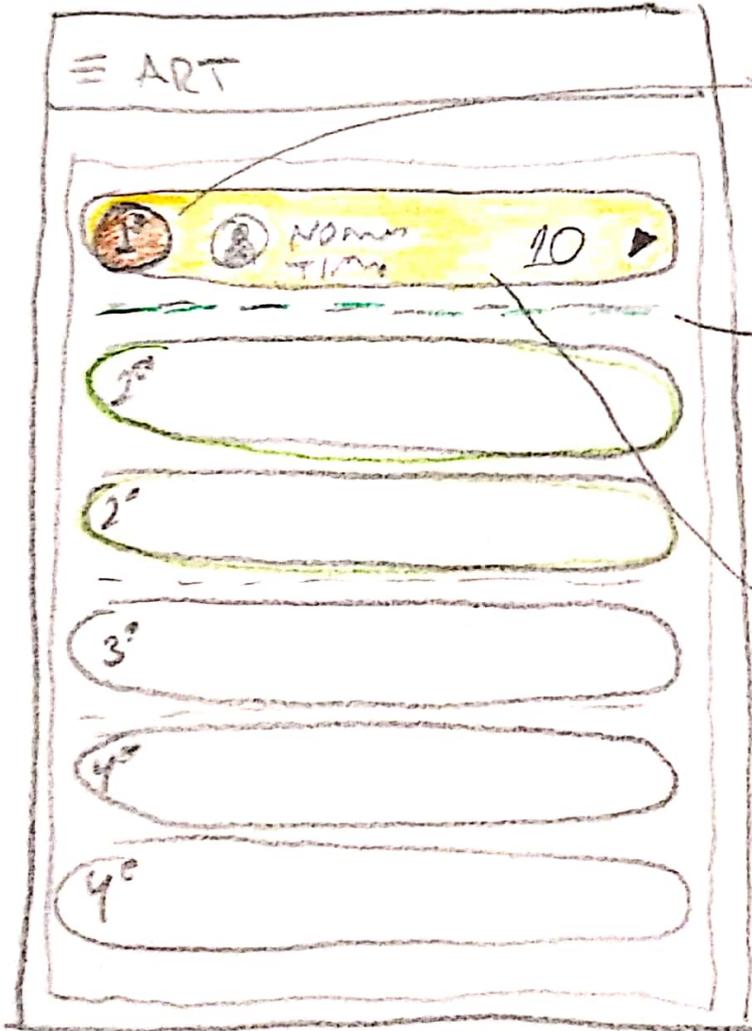


← Junta

7)



8)



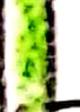
→ Texto: posição dentro de um recorte dourado (GLI BOLA)

→ Separador para posição do número; diferentes.

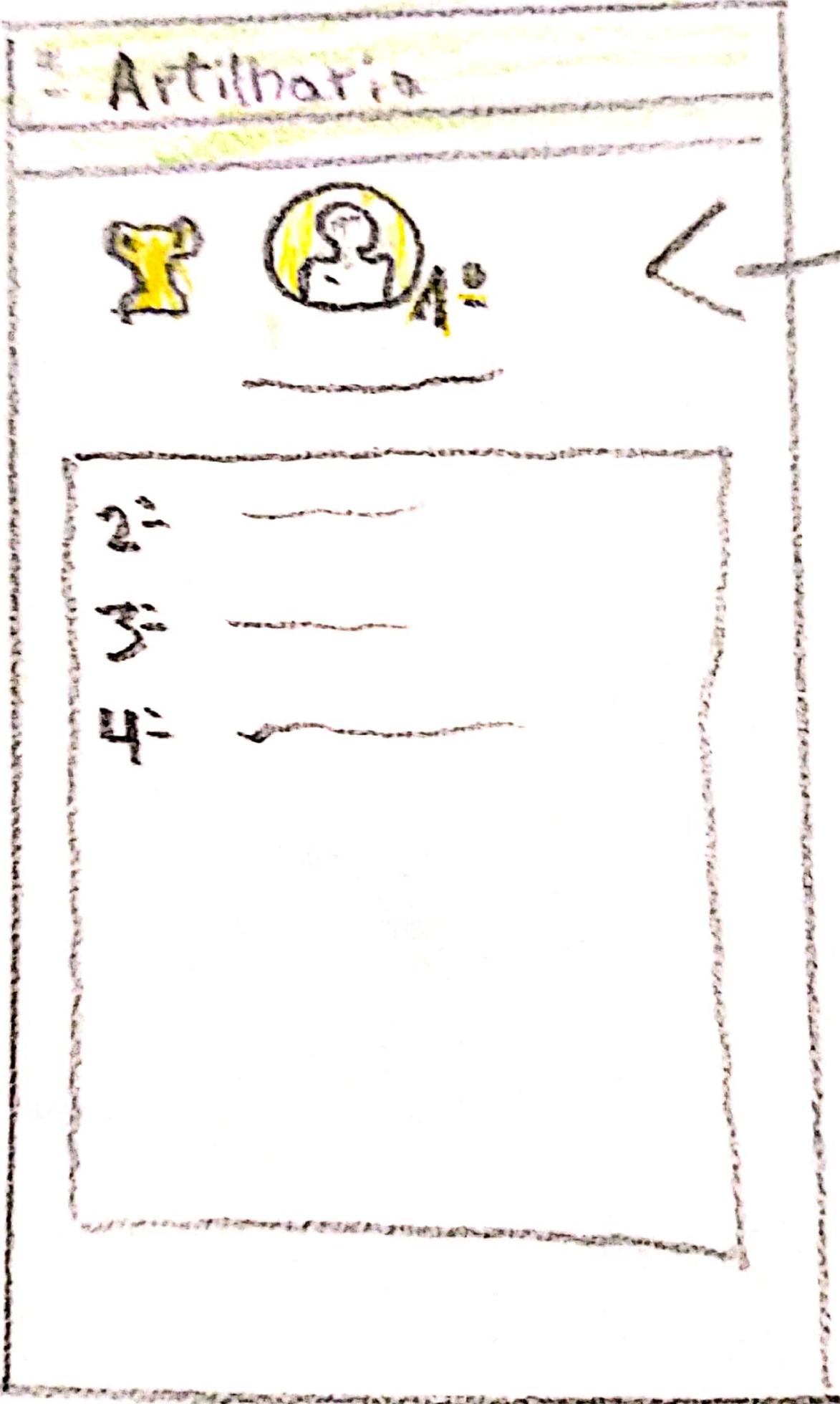
→ Cor diferentes

8)

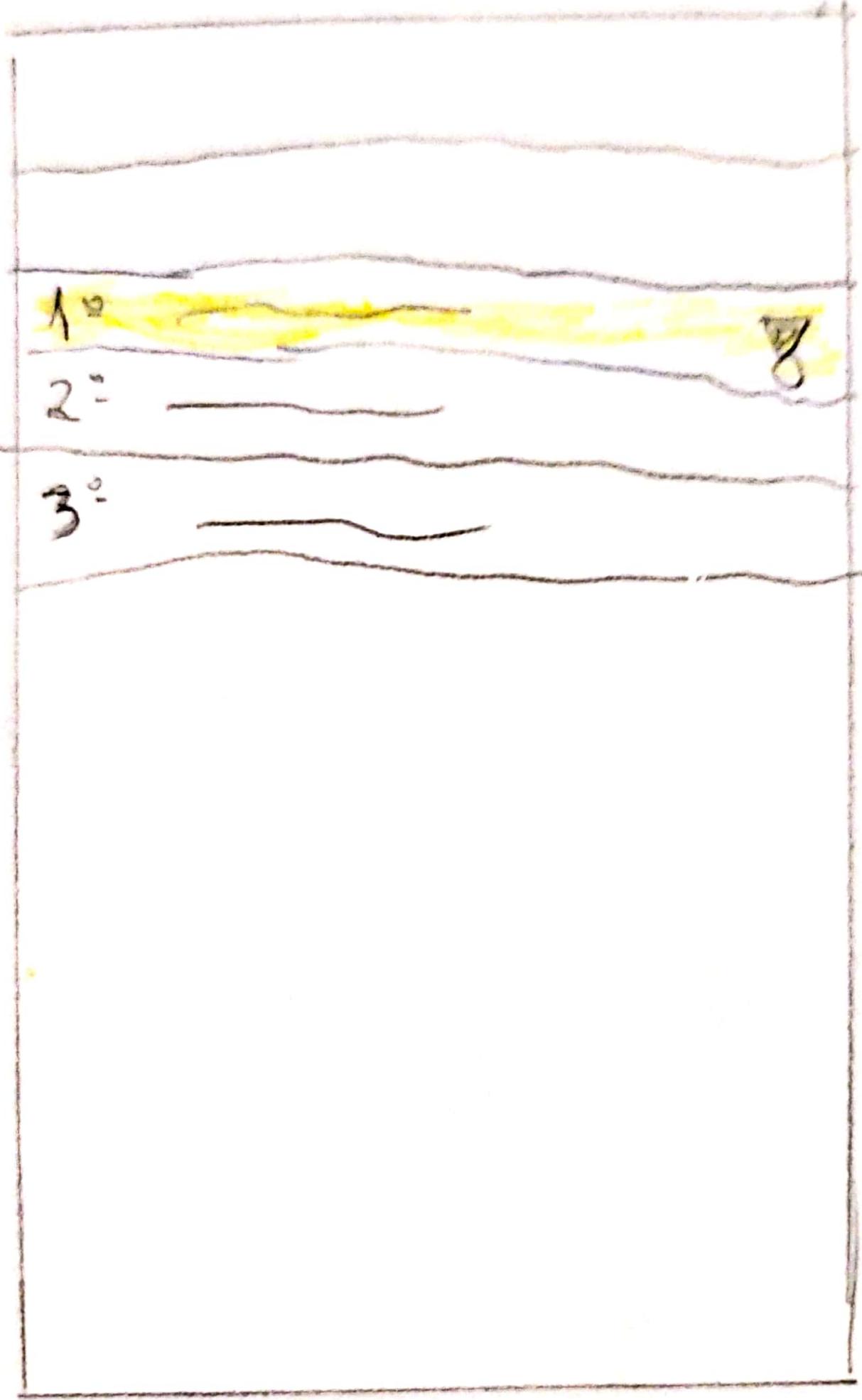
≡ Dates A

-  ○ ~~~~~
-  ○ ~~~~~
-  ○ ~~~~~
-  ○ ~~~~~
-  ○ ~~~~~
-  ○ ~~~~~
-  ○ ~~~~~
-  ○ ~~~~~
-  ○ ~~~~~

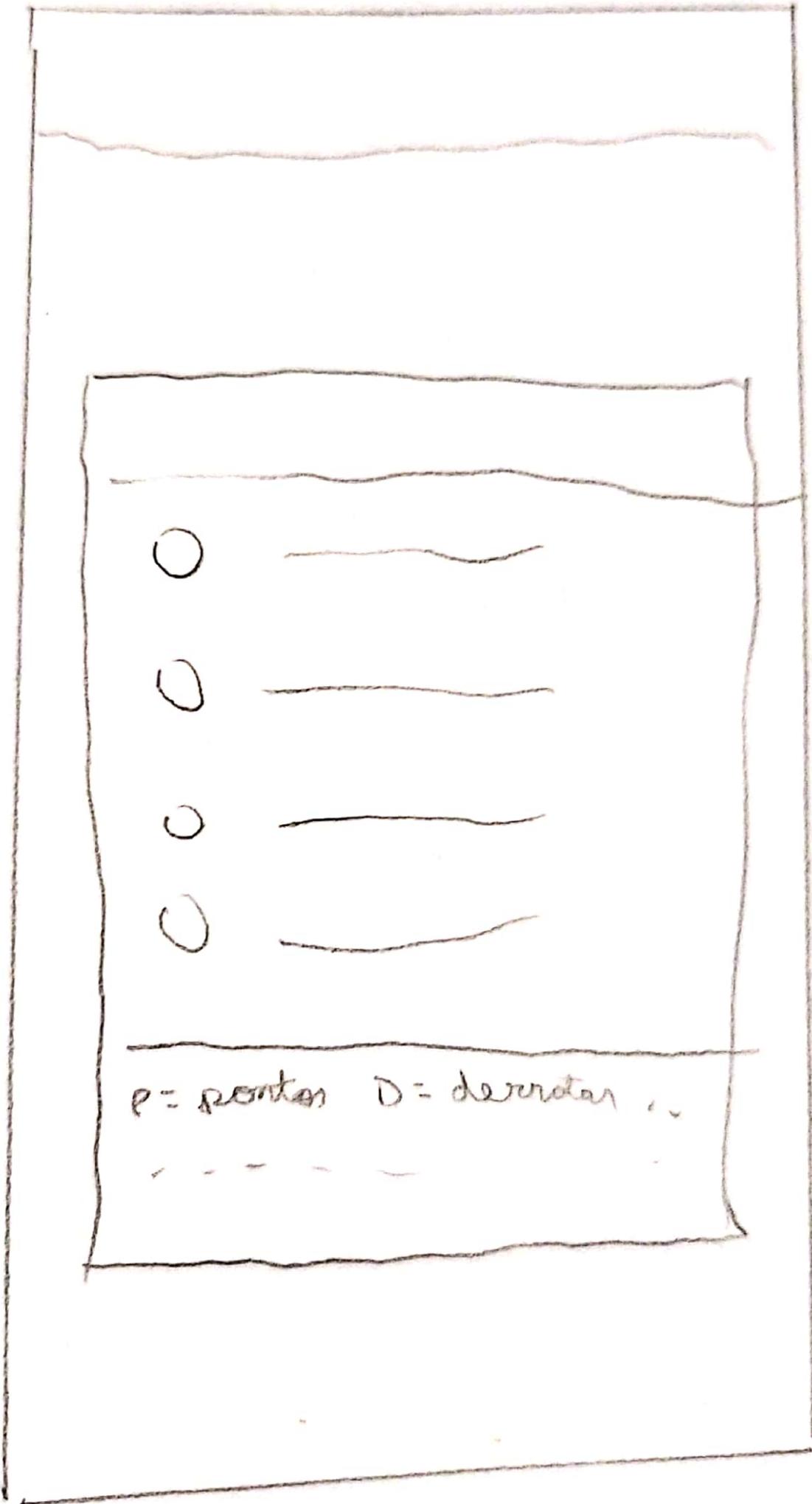
8.



8.

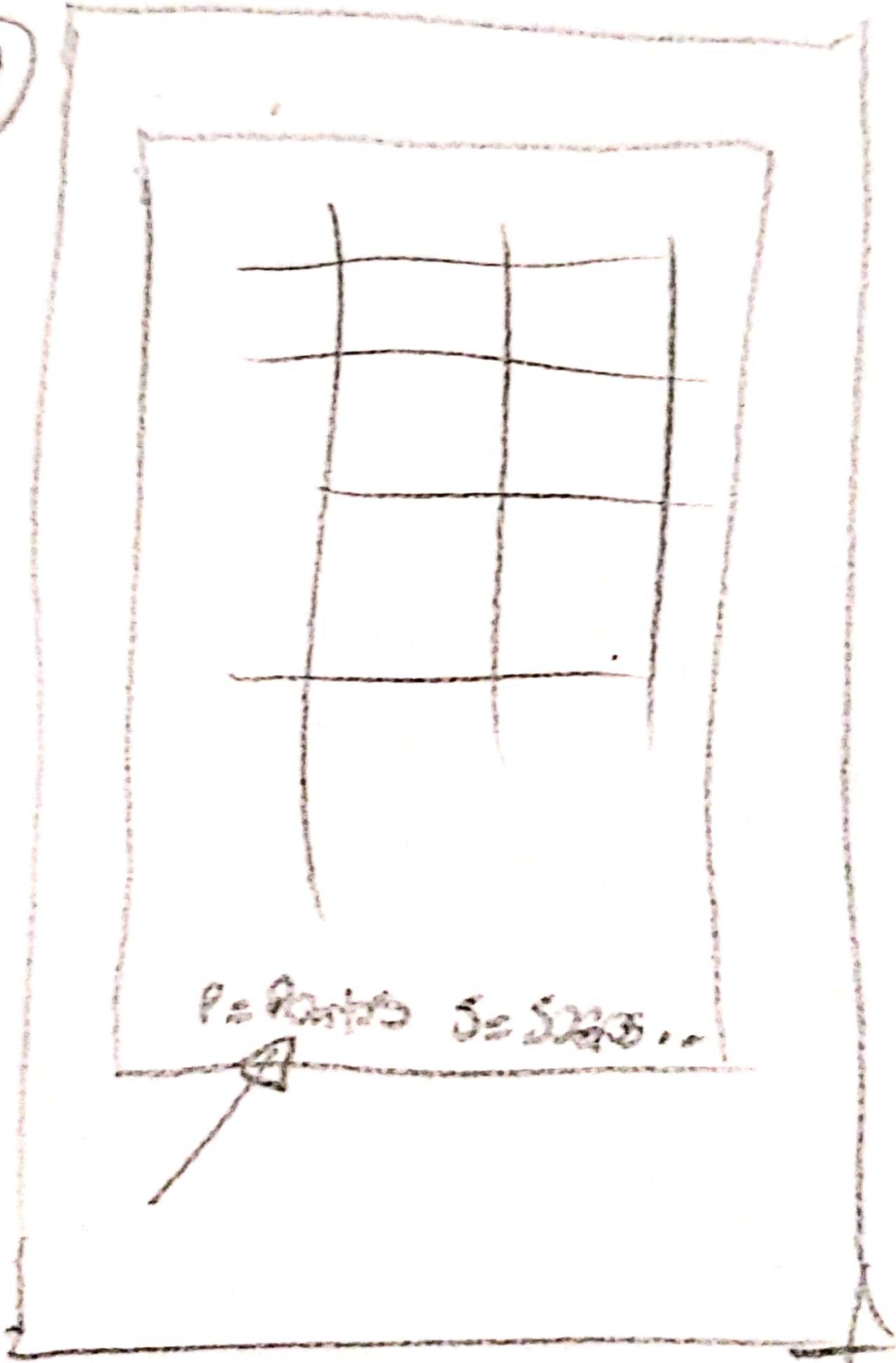


9 -

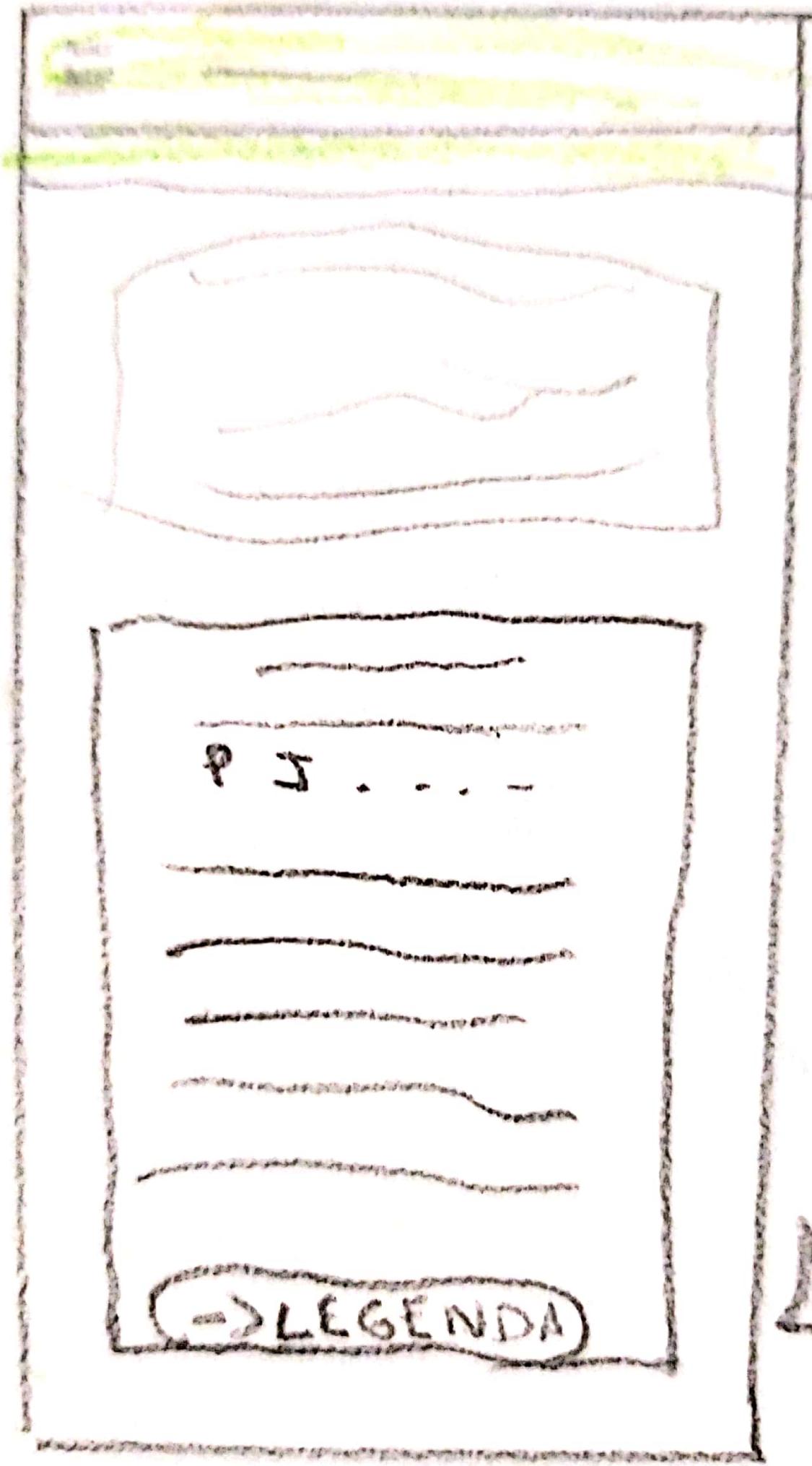


P = pointer D = derefer

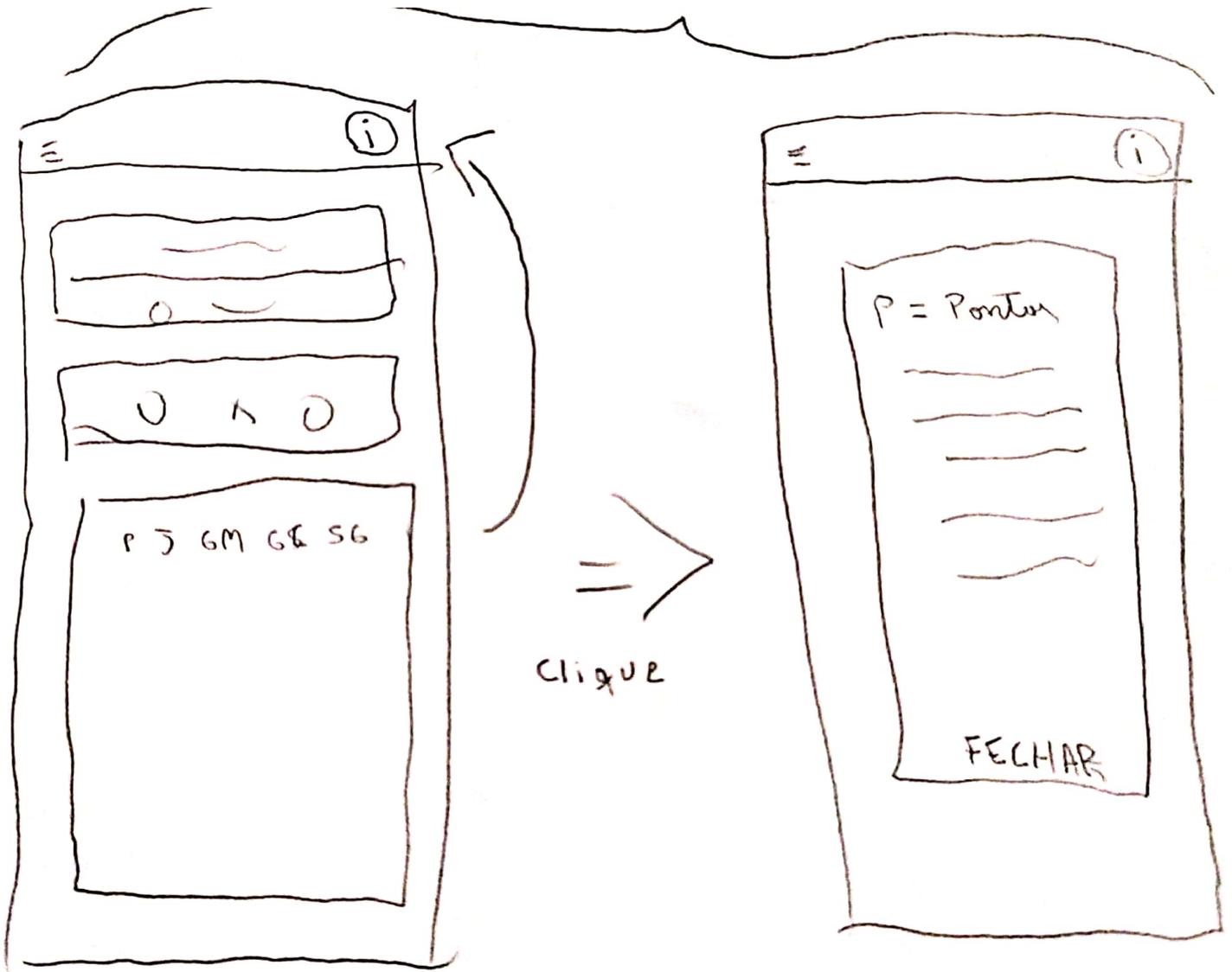
9



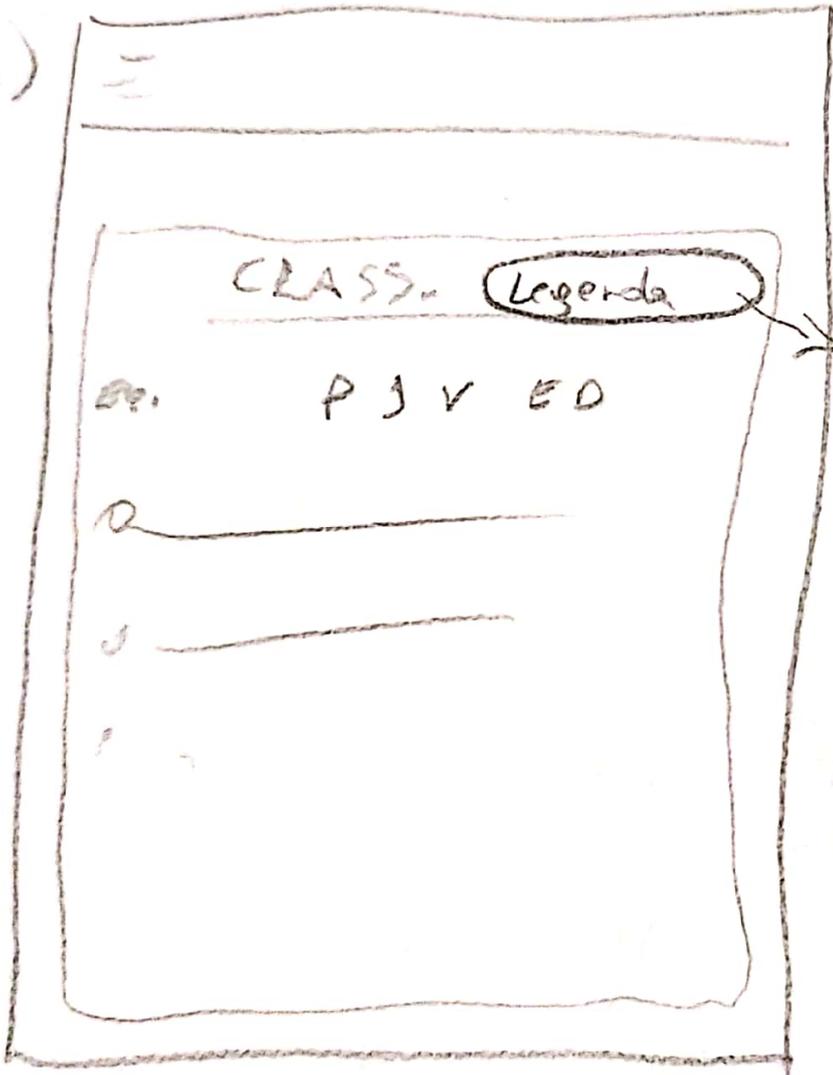
19.



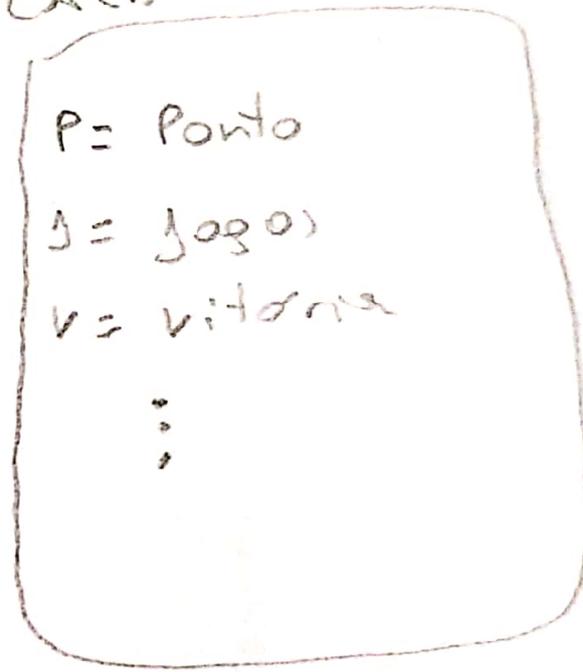
9)



9)



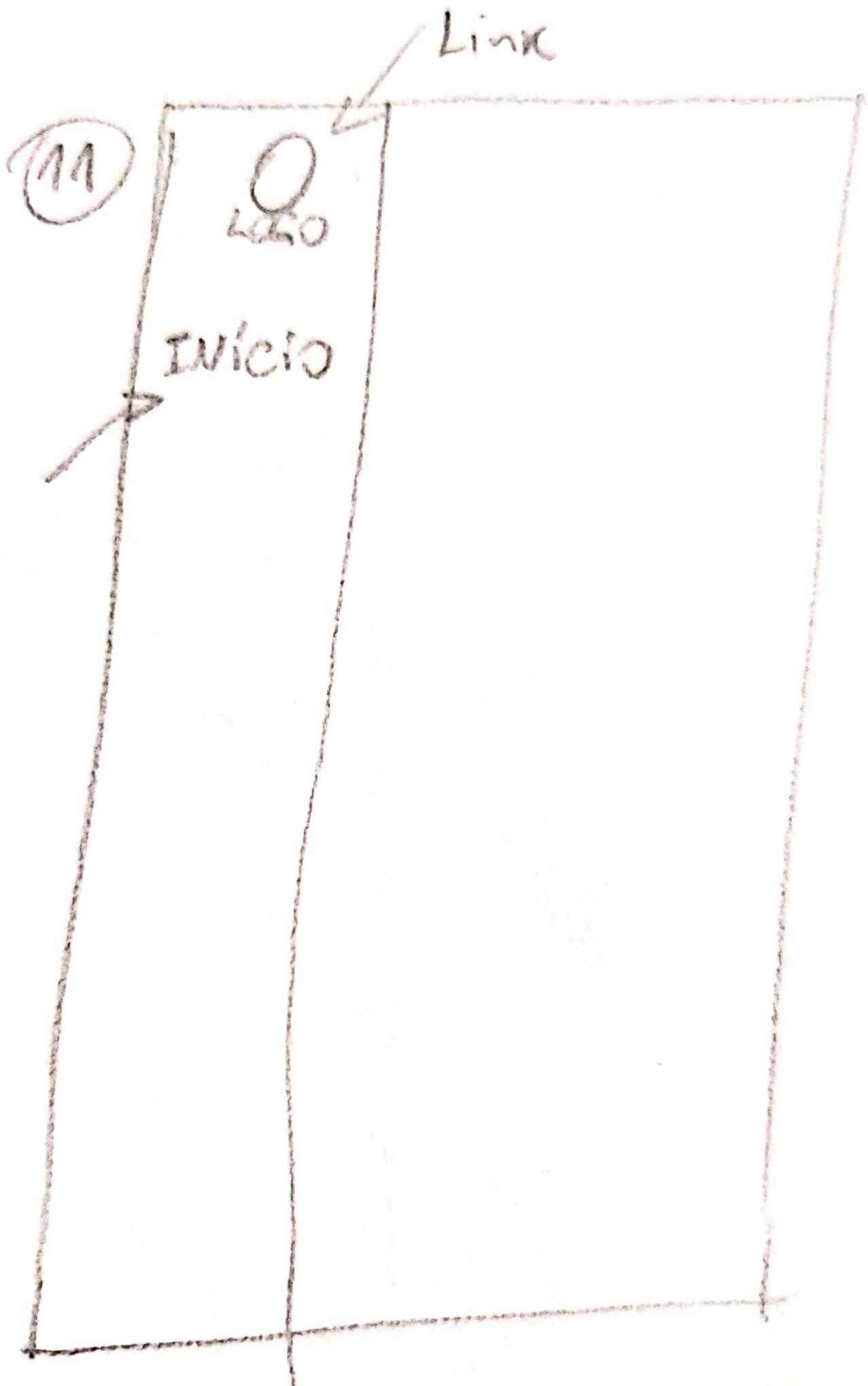
CARD



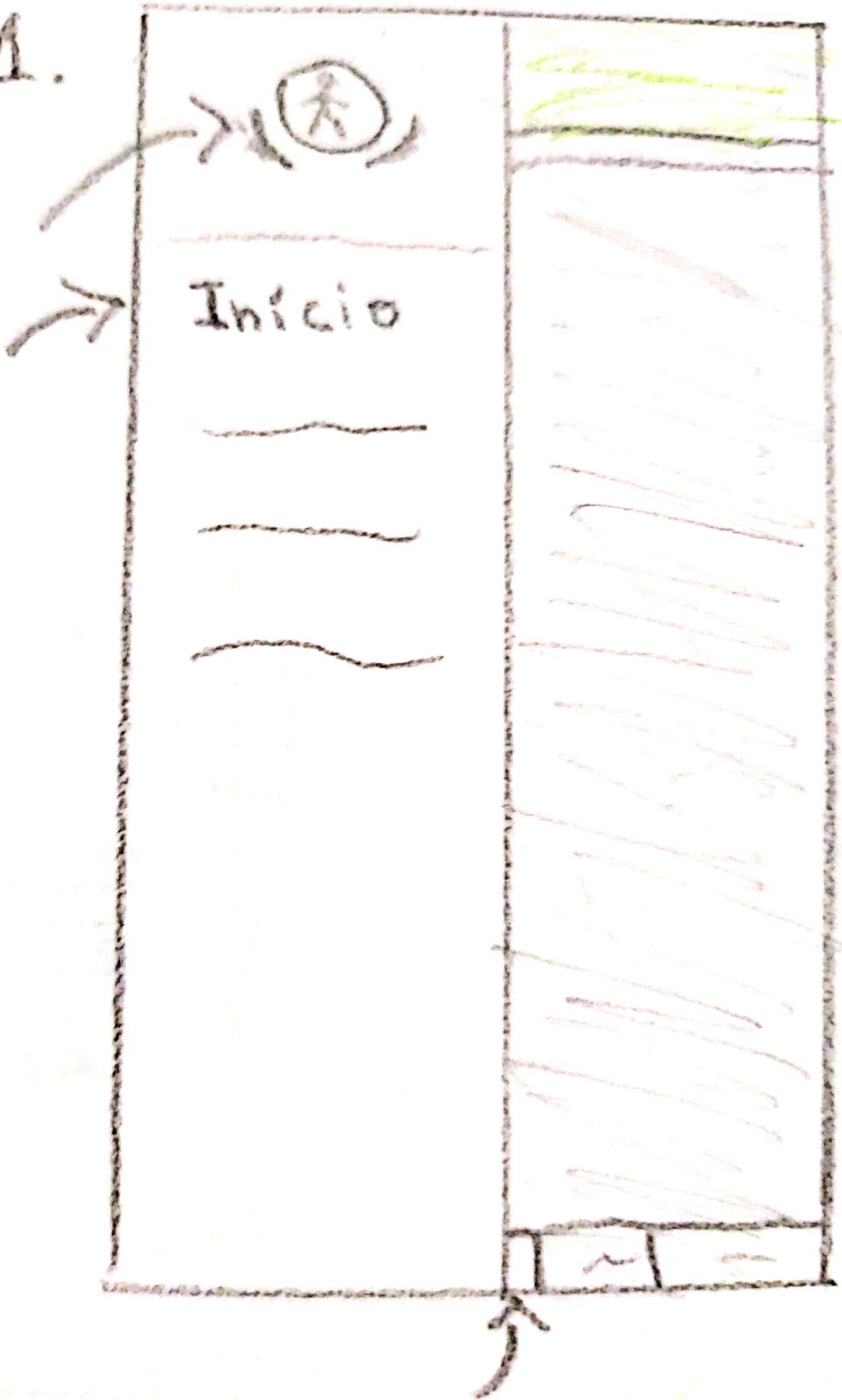
1 1 +

0
INICIO



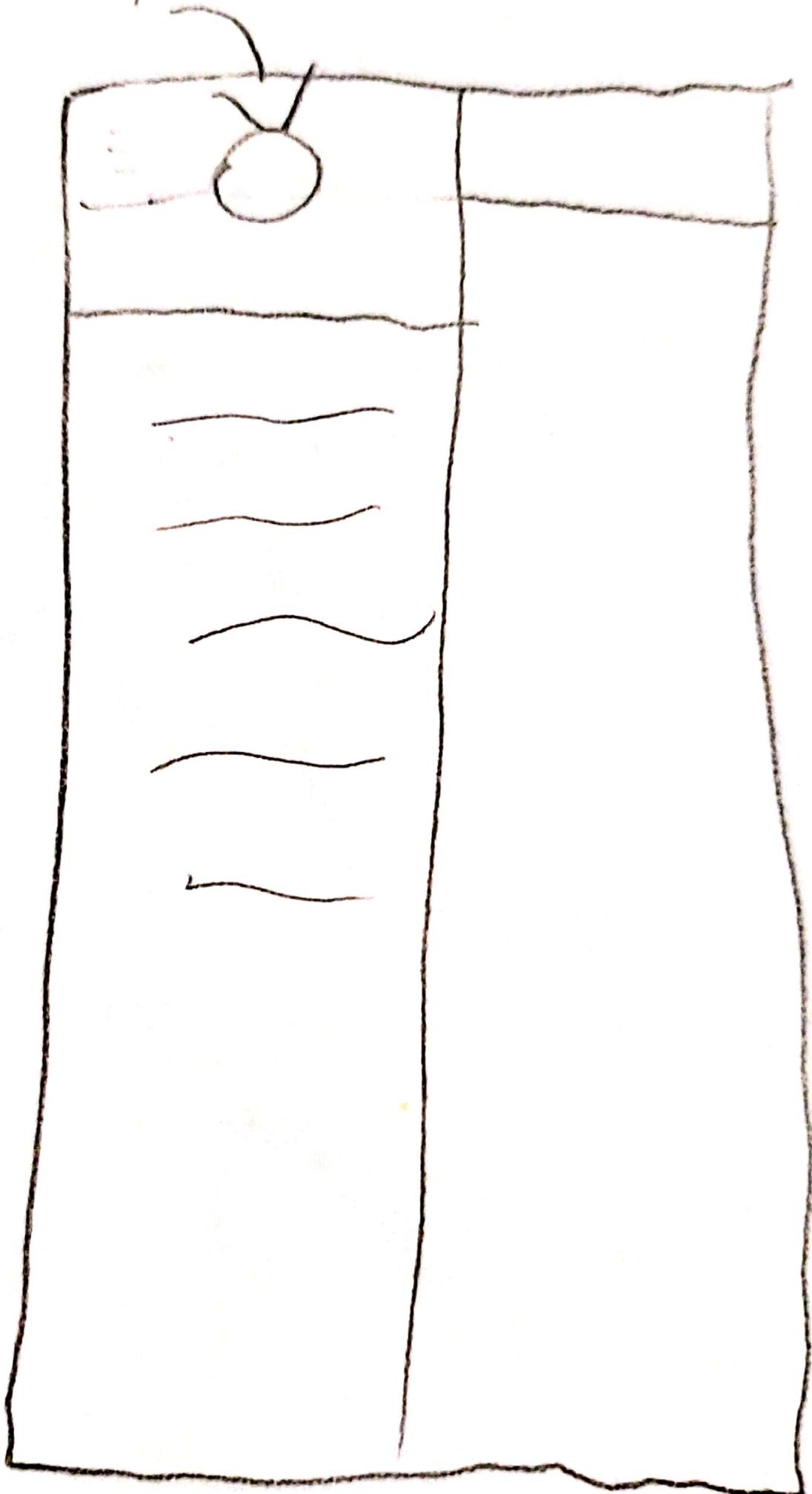


11.

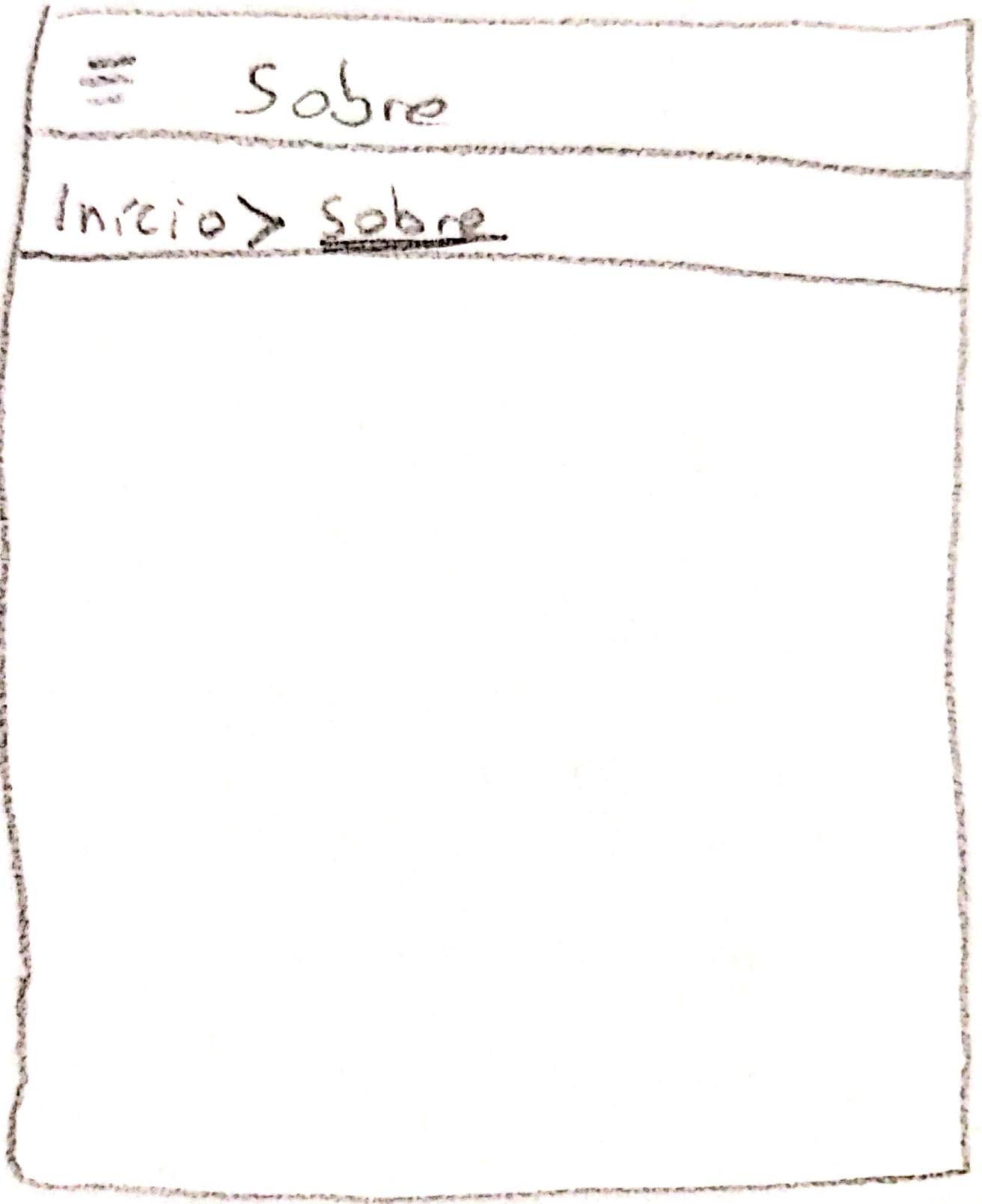


clique

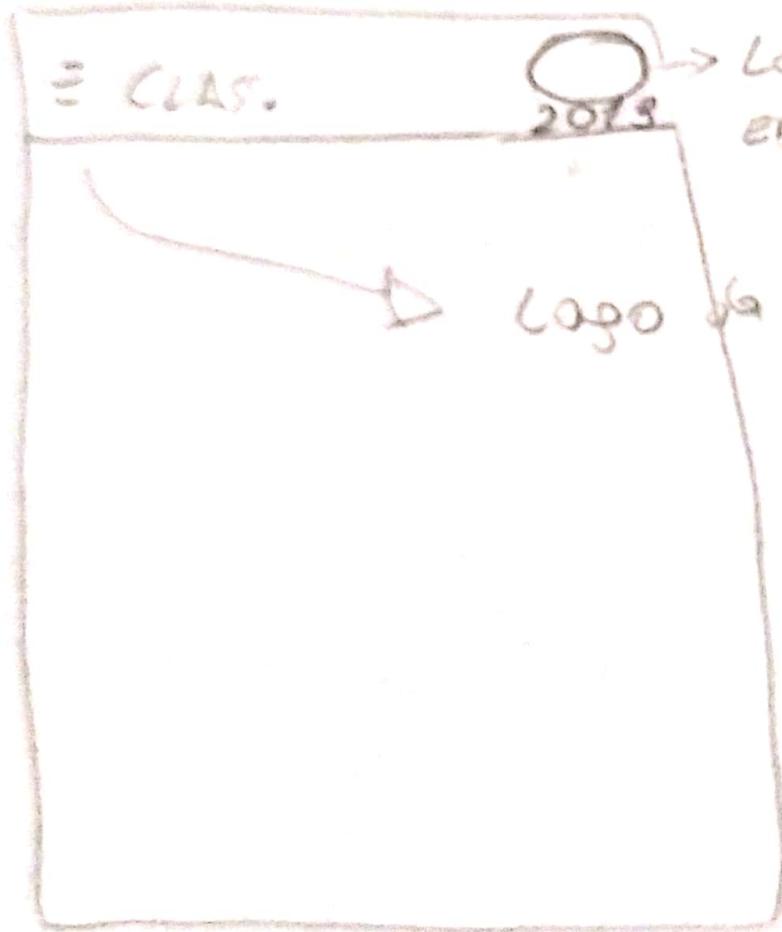
11)



11)



14)



2019

Logo da
edição + TEXTANO

Logo da edição no Menu

C

Tela Sobre / SPLASH

14

Gfipan 2018

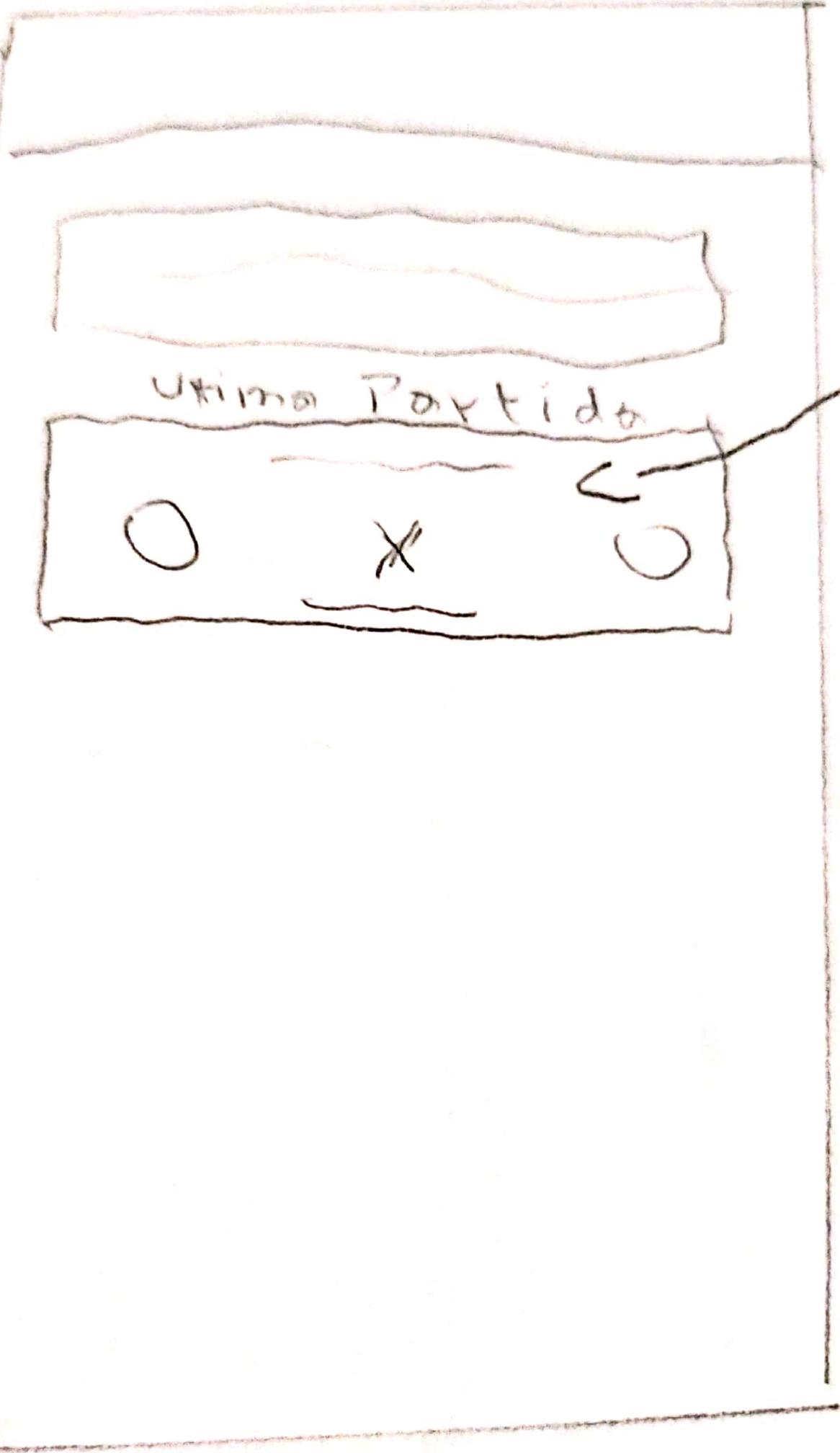


14 -

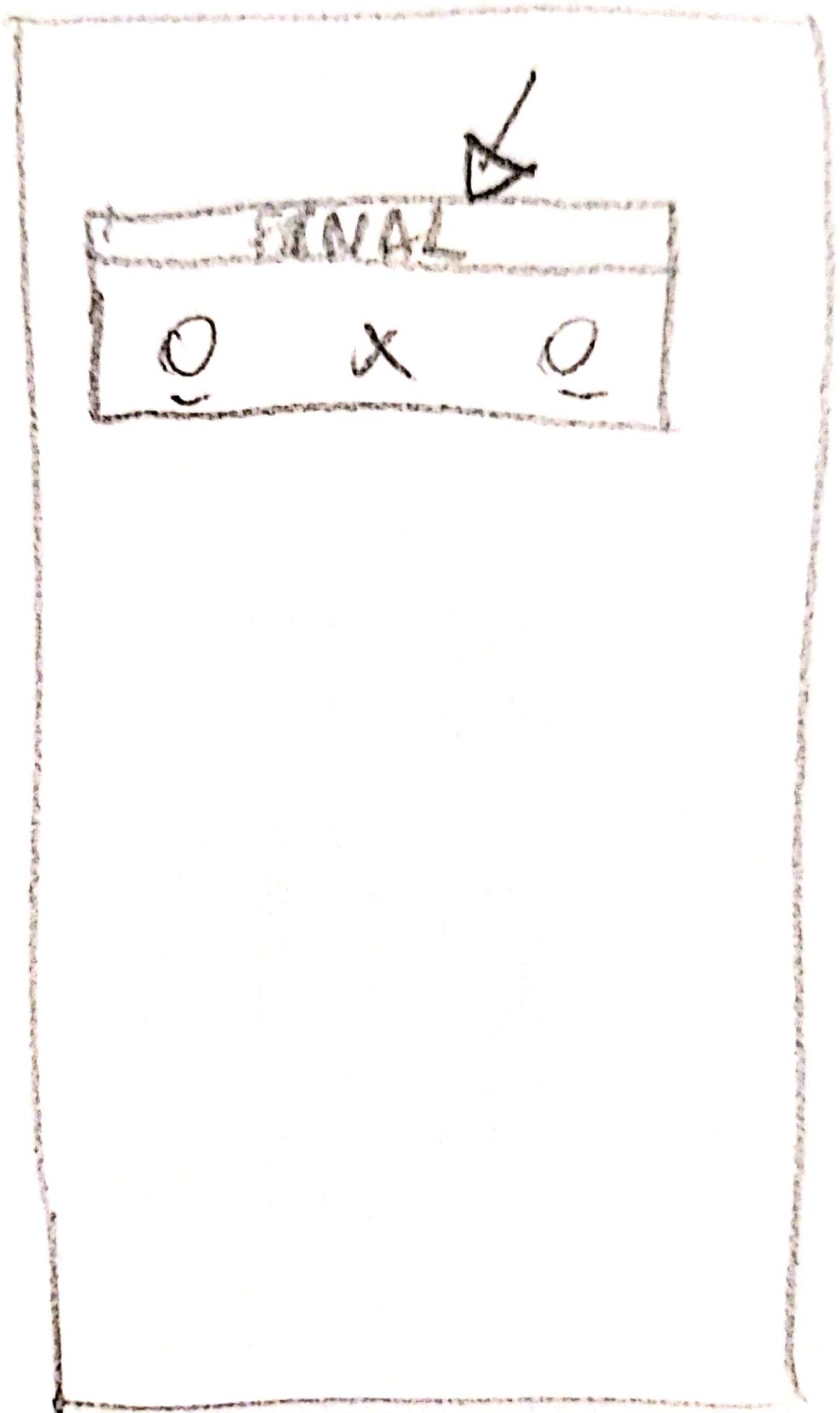
EFIPAN 2019



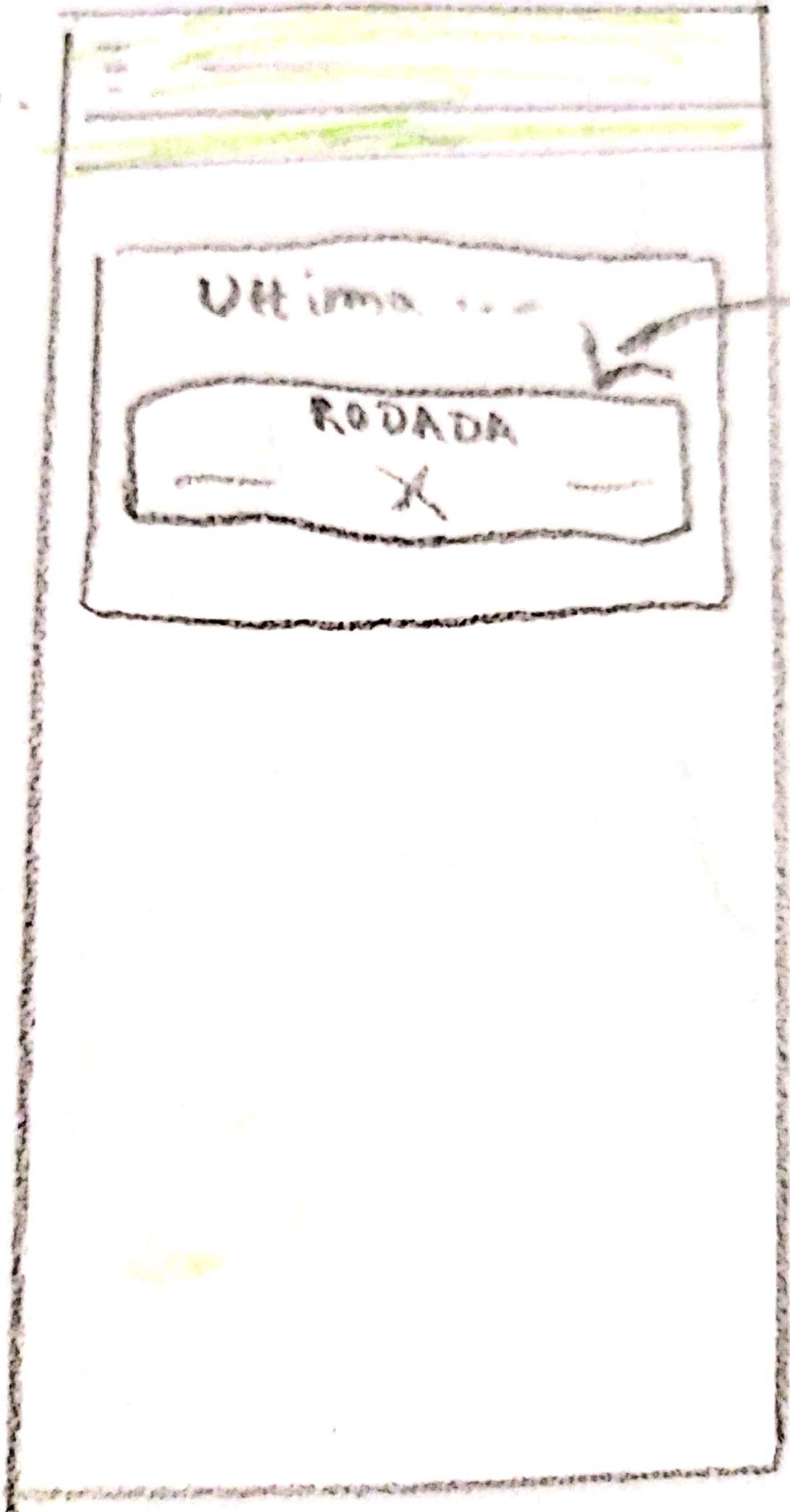
15 -



AS



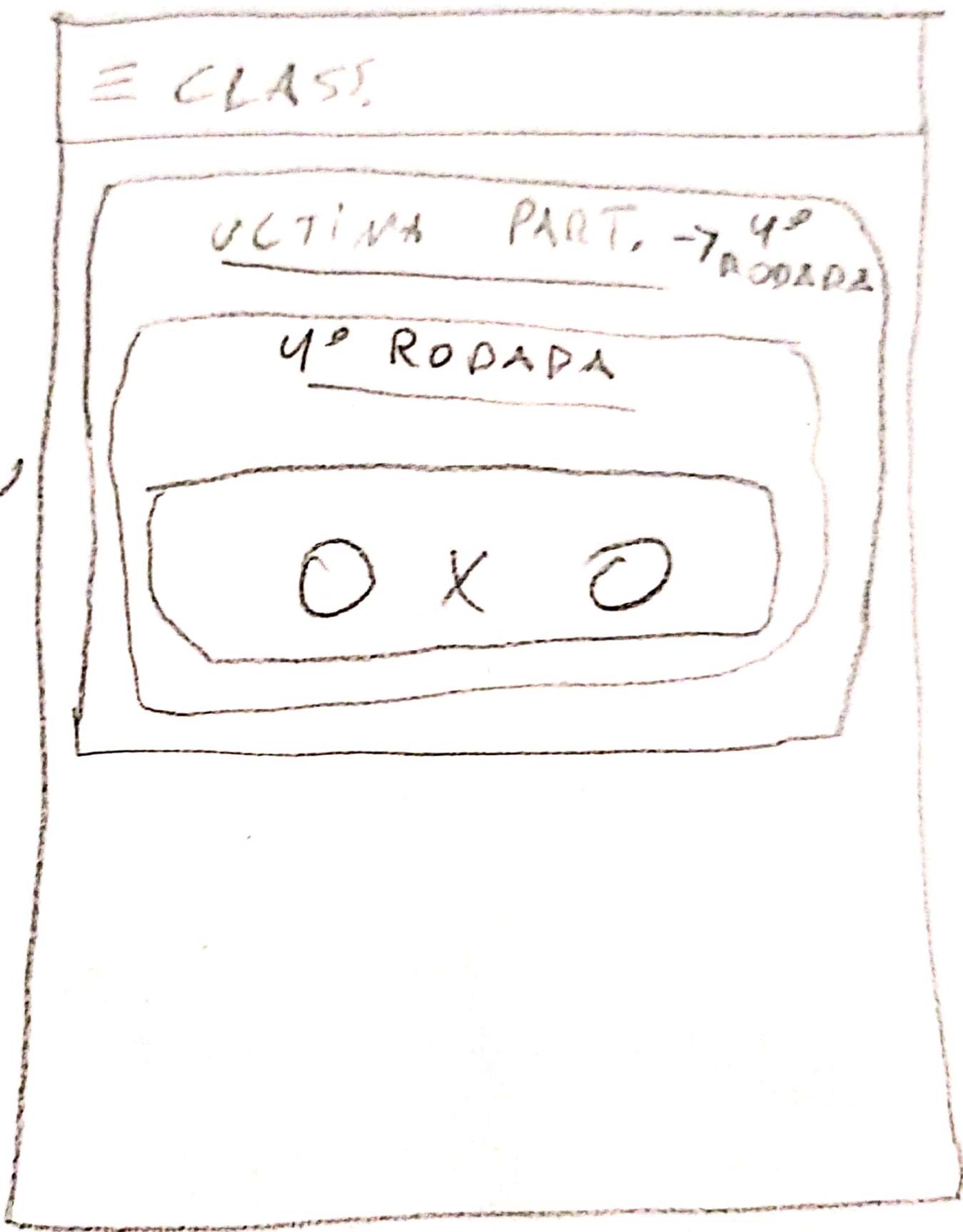
15.



15)



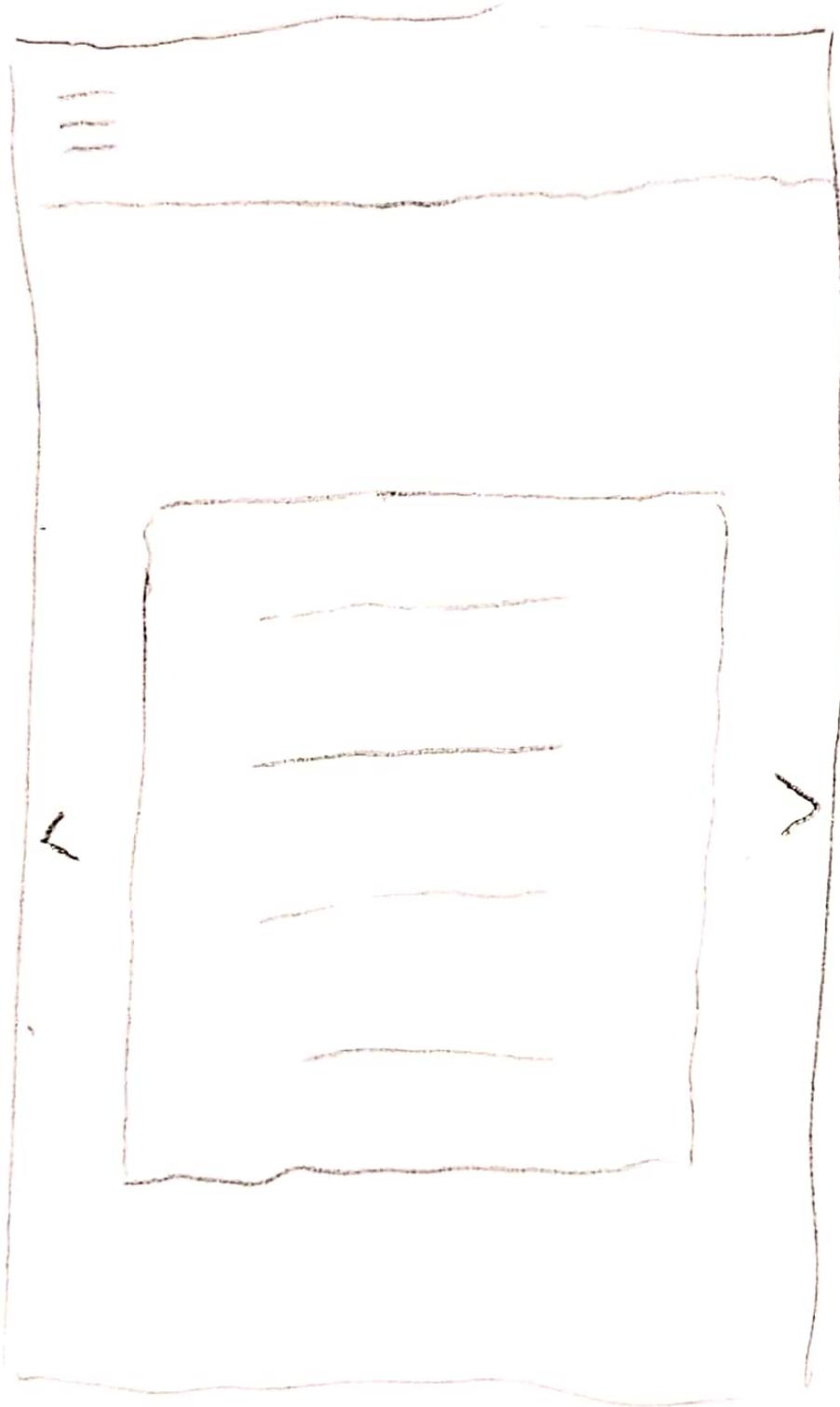
15)



a

APÊNDICE C – REDESIGN COM NÃO ESPECIALISTAS

1.



← *Carrossel*

1)



< ARTILHARIA JOGOS >

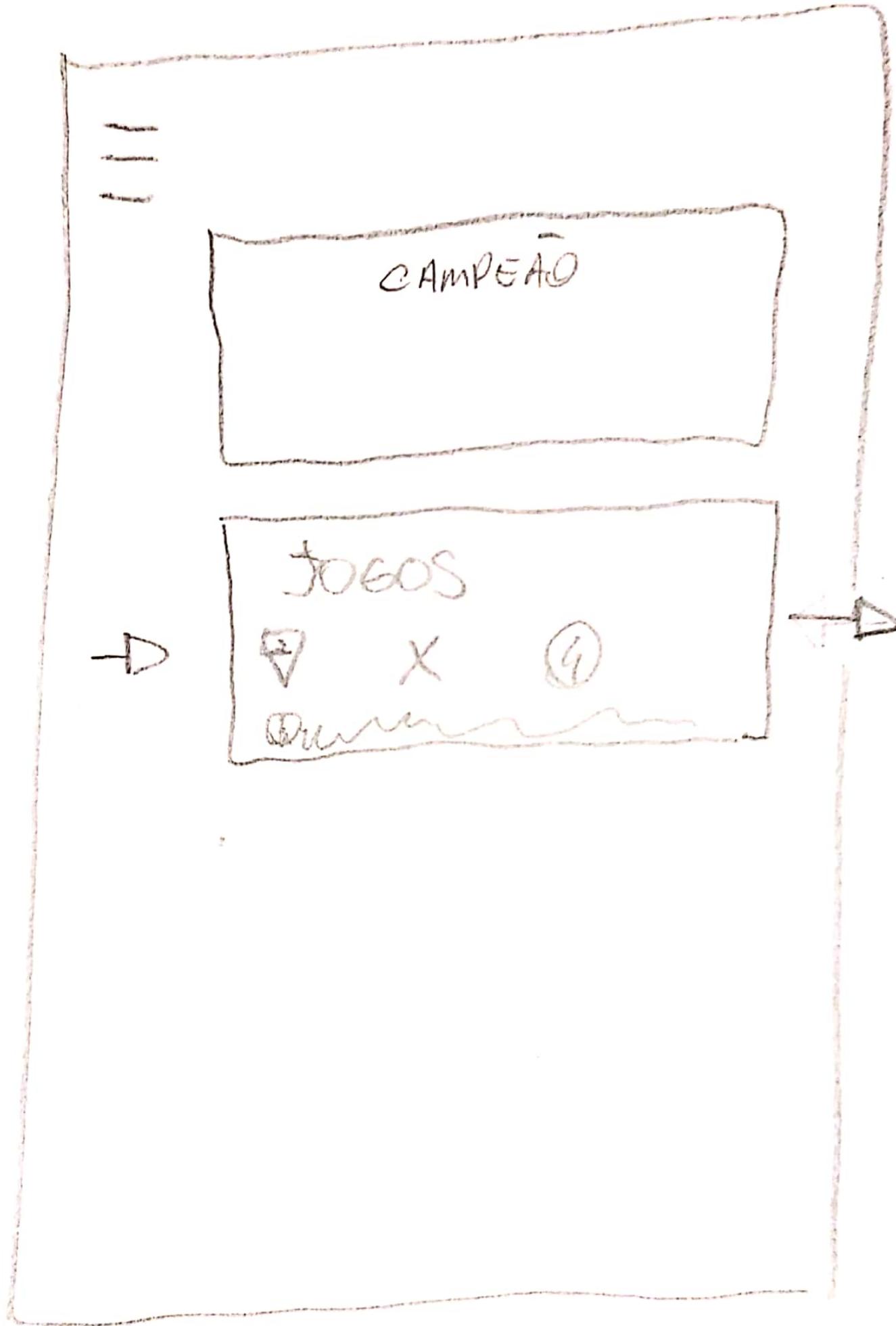
1



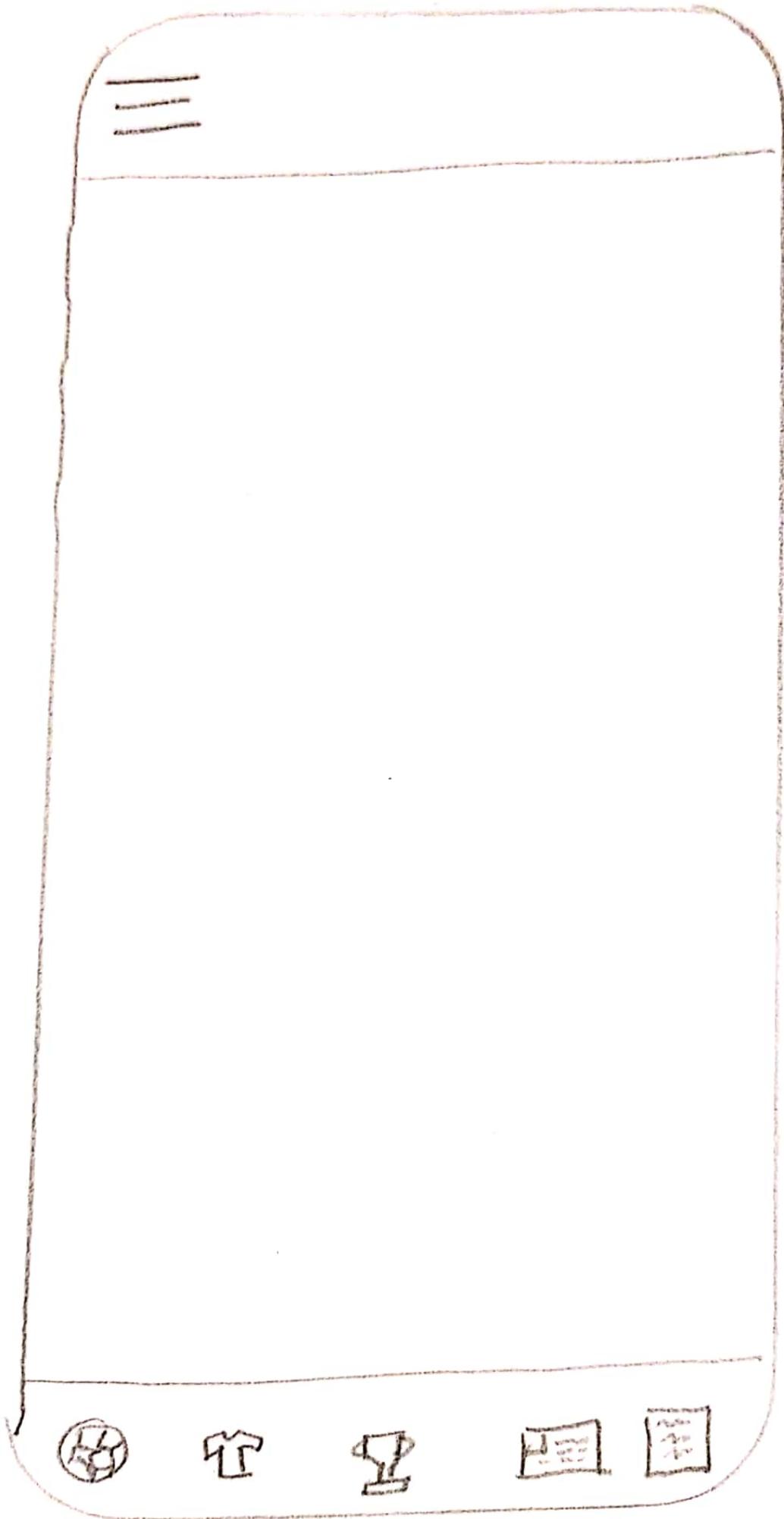
ÍNICIO OU EFIPAN

LOGO
EFIPAN





1-



2)

CLASSIFICAÇÃO GERAL

Jo (2)

6P 6C

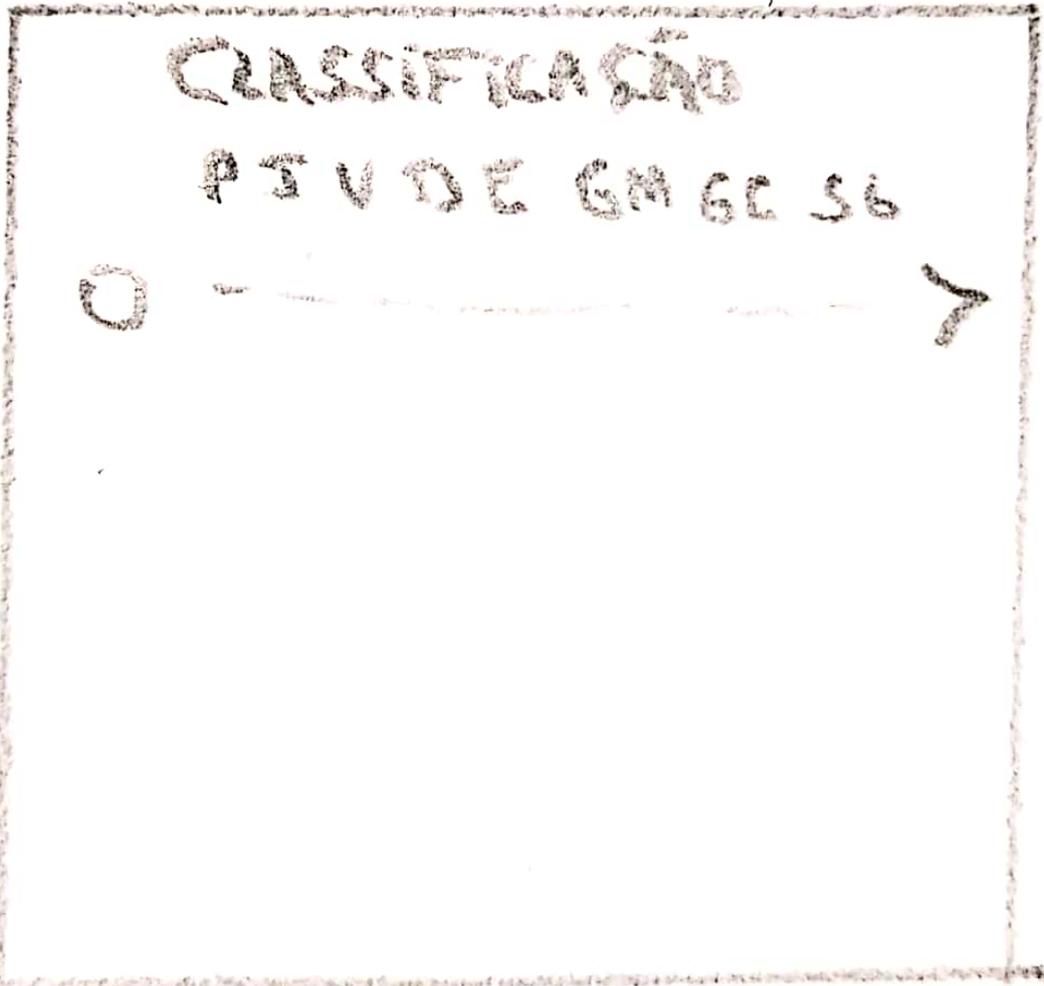
10 5

3)

└─ 10 (11) 3 2 1 1 2

└─> SOMPREADO

31



3-

EQUIPE

P J V E D G P G C S G

~~M/10~~

1^o ○ GRE

~~M/17~~

2^o ○ PAL



3



1°

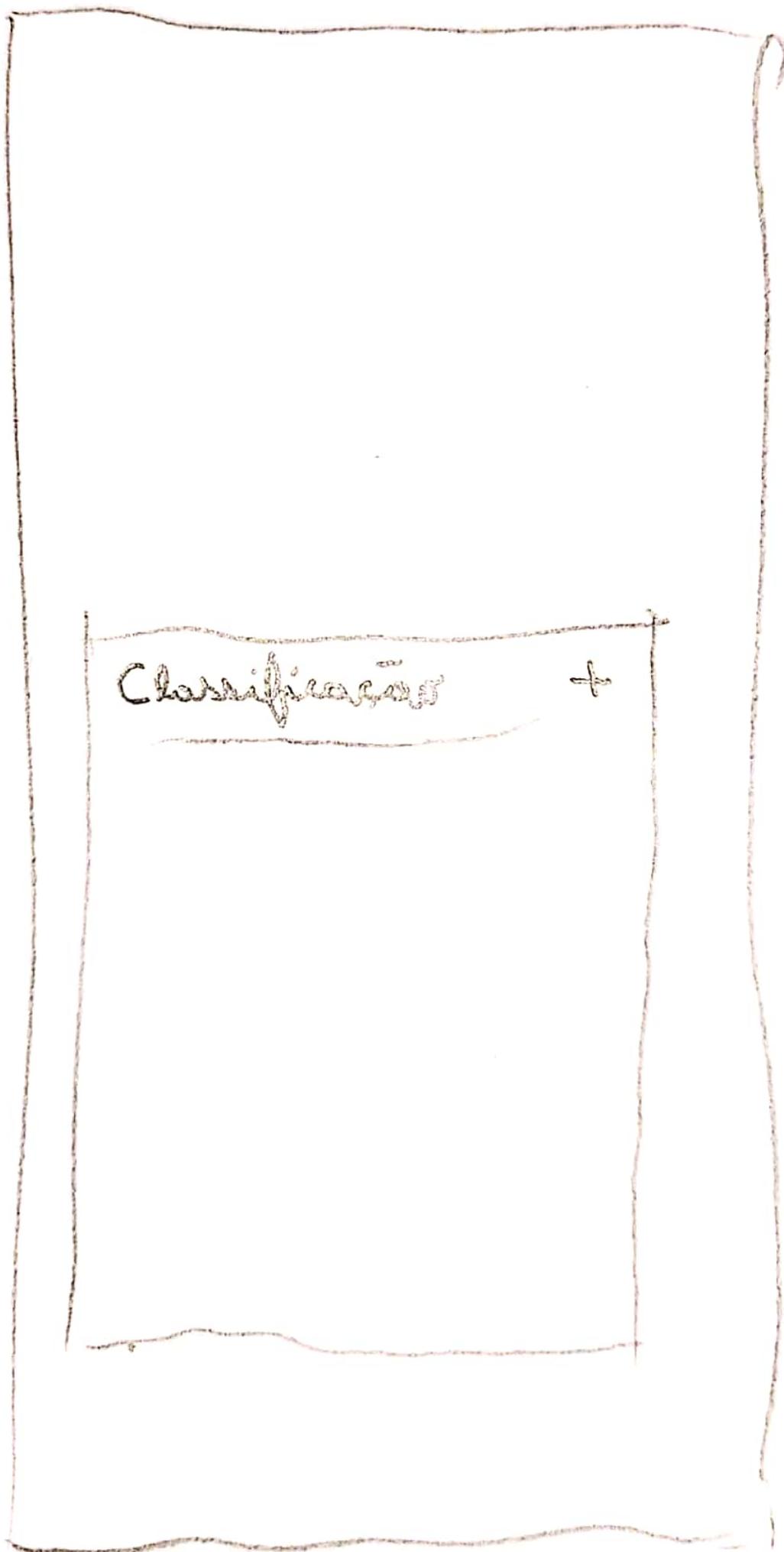


GRE

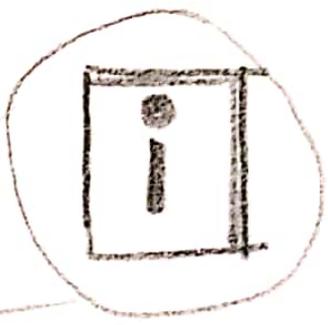
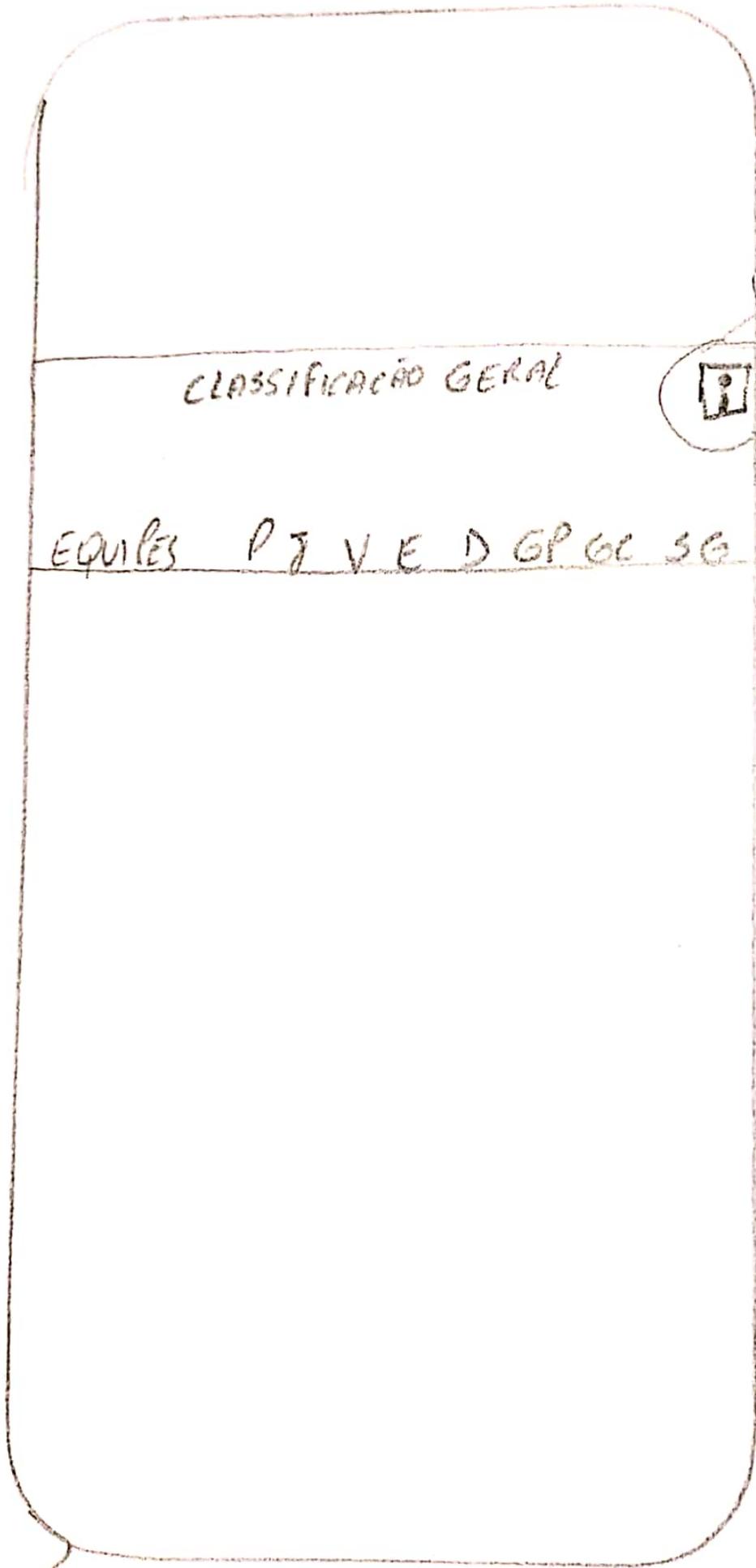
3.

<u>Classificação</u>	
Clique para Ver mais	
1 0	_____
2 0	_____
3 0	_____

5.



5.



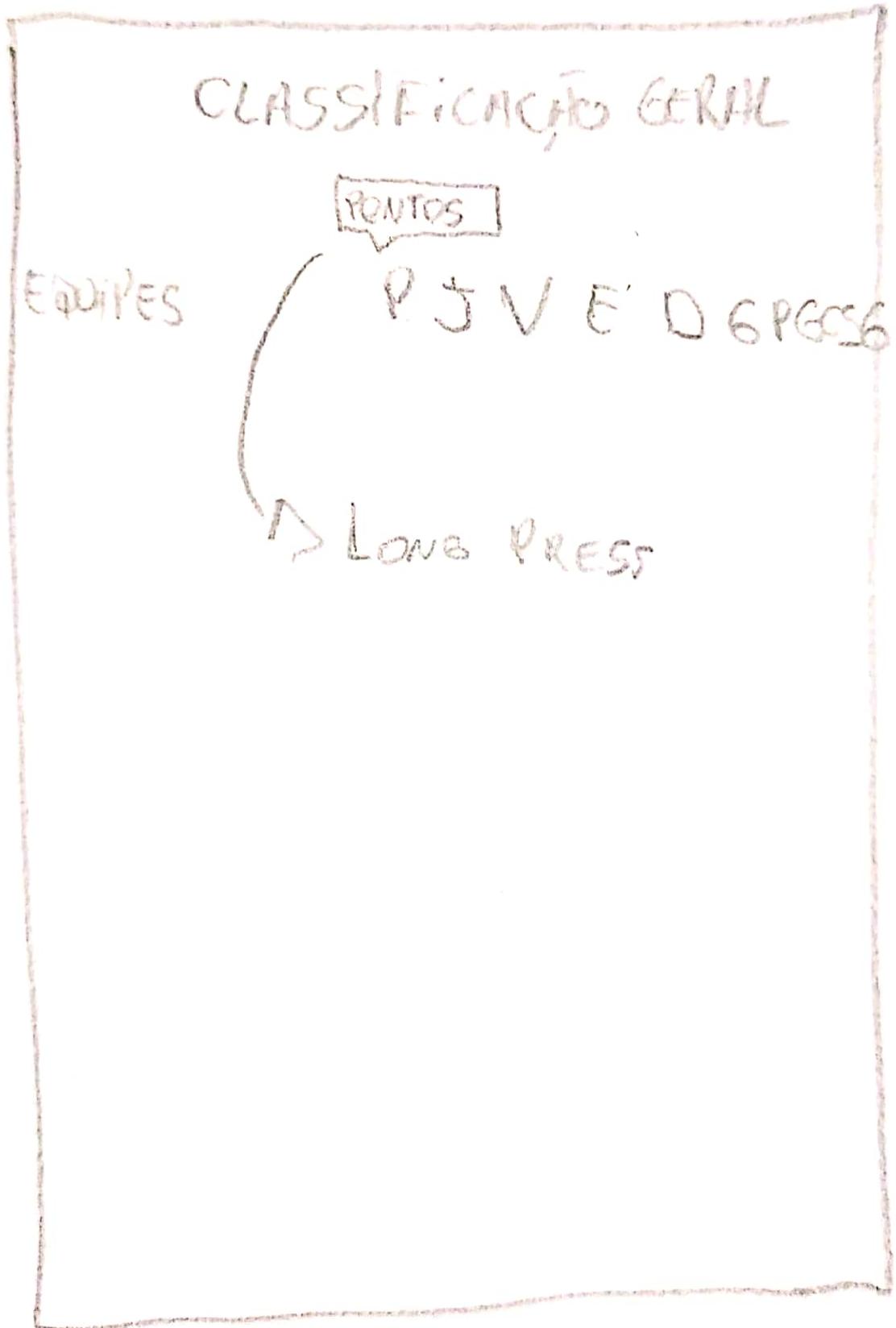
57

(our choice)

P J V D

JOGOS

5)



5

CLASSIFICAÇÃO GERAL

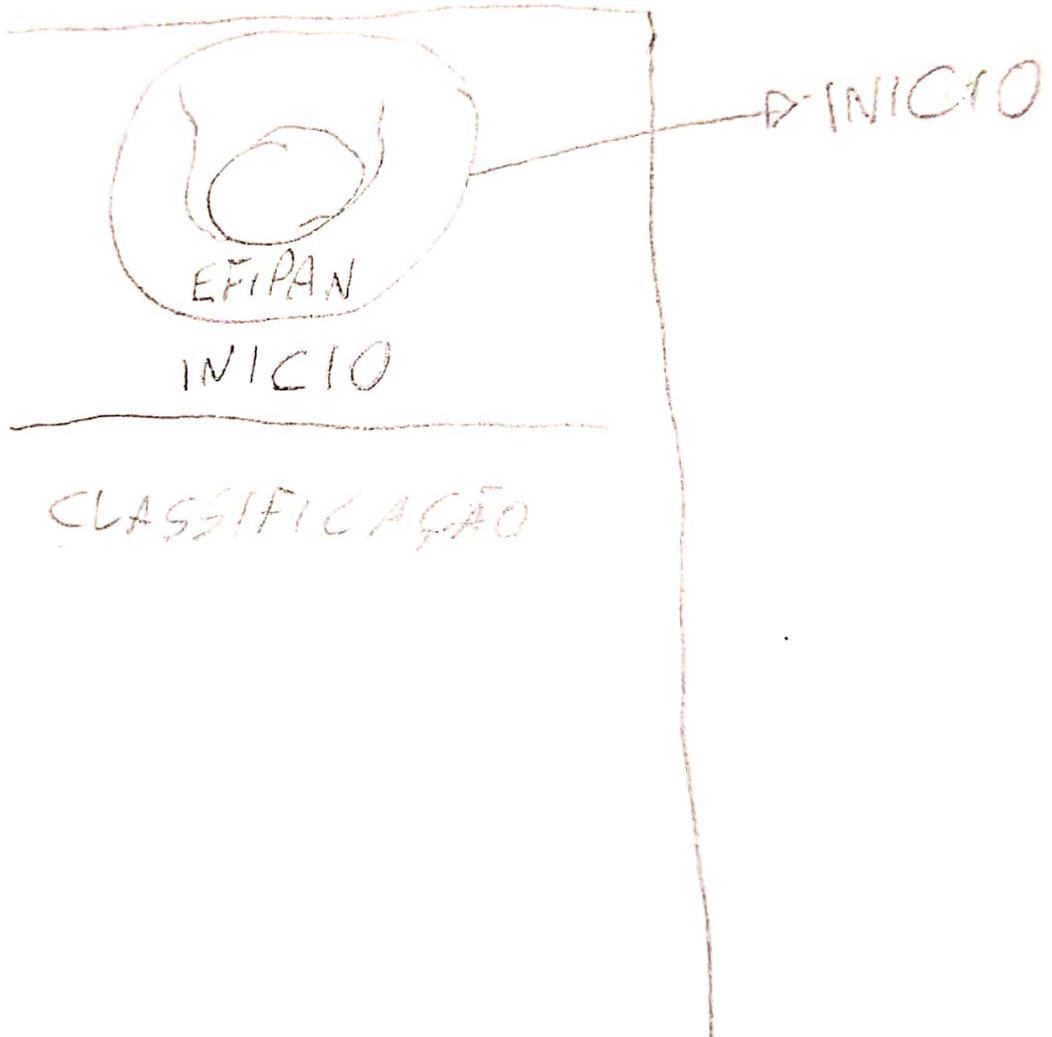
EQUIPES

P J V E D GP GC SG

(i)

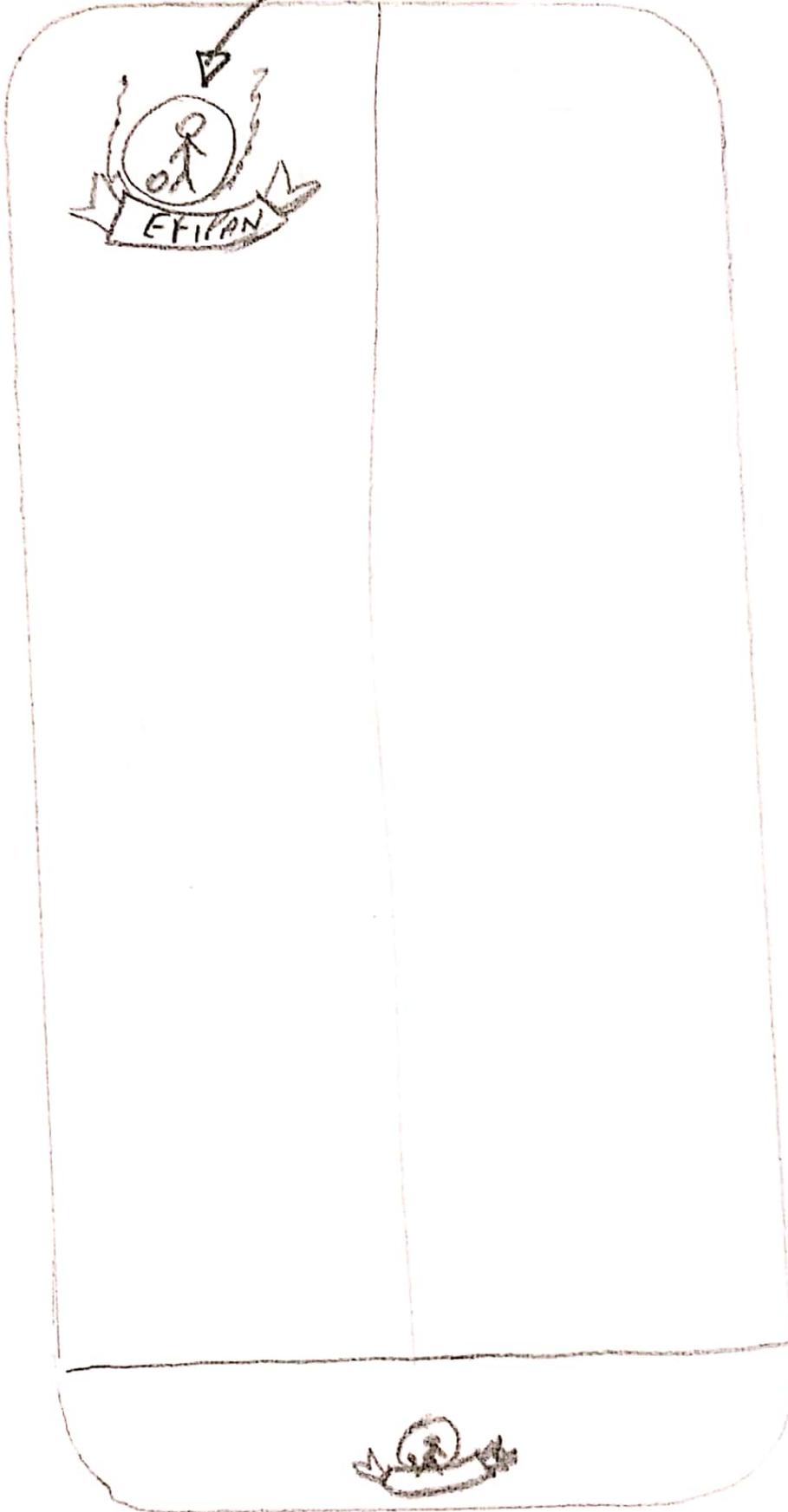
P = PARTIDAS X

6



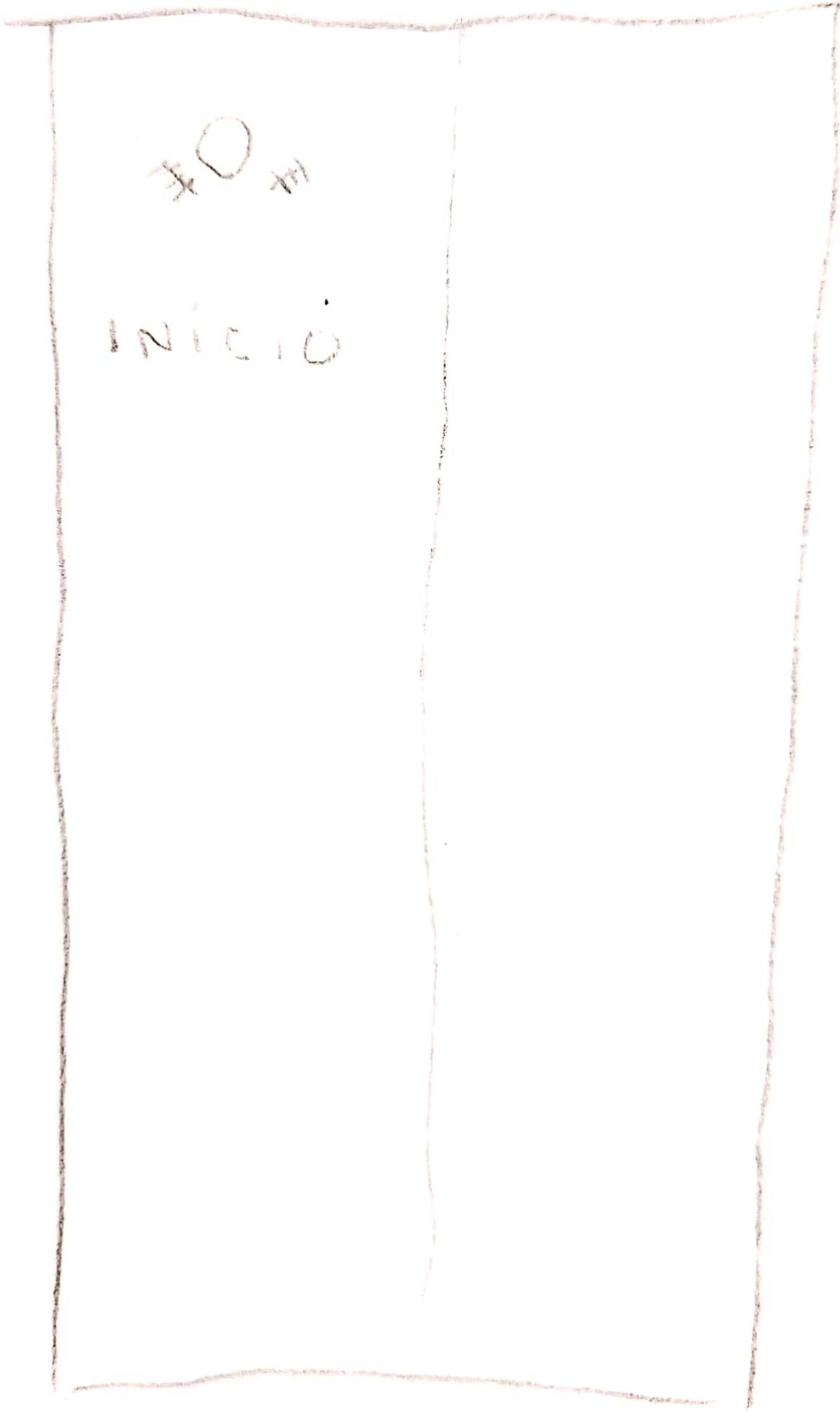
CLICAR AQUI PARA
VOLTAR AO INICIO.

6-



BOTÃO PARA VOLTAR
PARA O INICIO.

6.

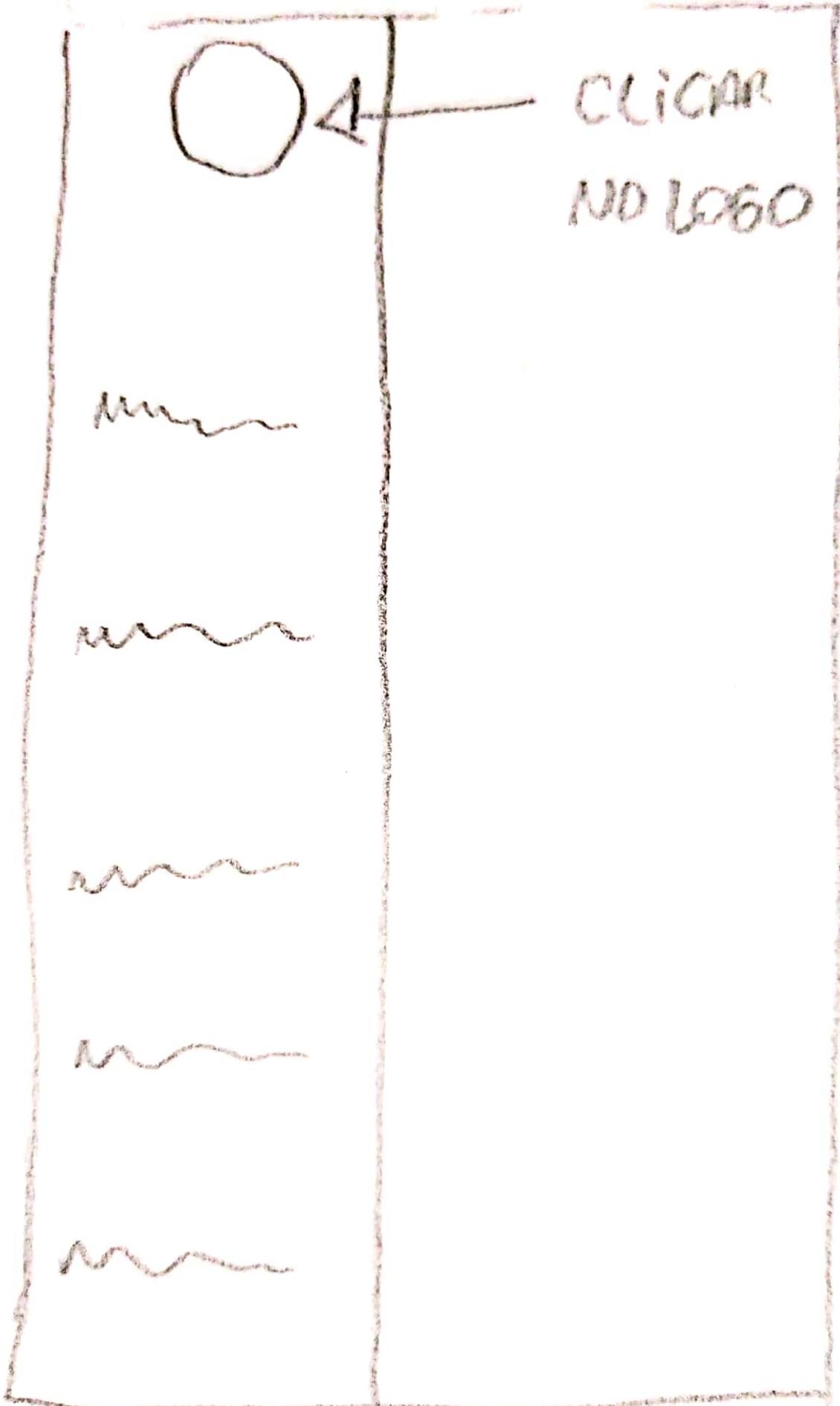


61

≡ JOGUS



6)



7



↳ BOTÃO INTEIRO

CLICAVEL

8-

← GRÊMIO

Equipe Jogo estat

GRÊMIO

==

JOGADORES

- 1 goleiro
- 2 lateral dir
- 3 zagueiro
- 4 zagueiro
- 5 volante

ORDENAR POR:

- POSIÇÃO
- NUMERO

==

9.



9-




CAMPEÃO

 GRÊMIO

ULTIMA PARTIDA REALIZADA

4 AGOS - 26:00



GRÊ

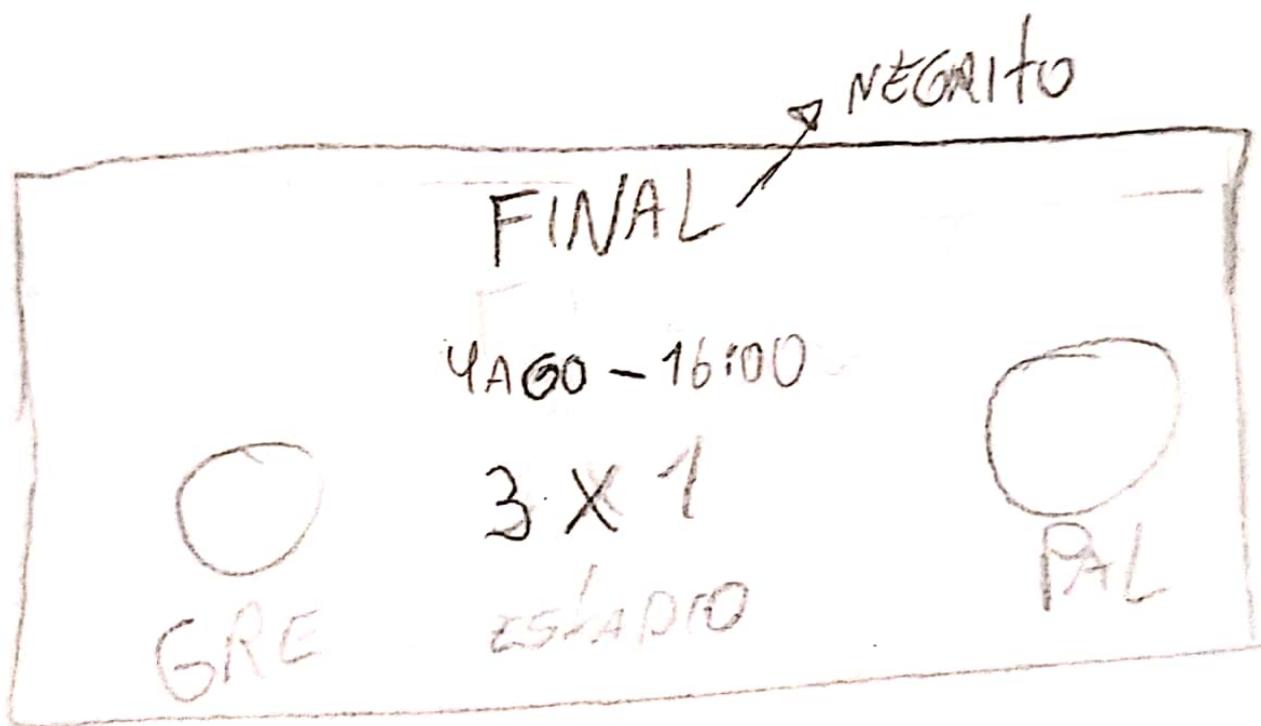
3X1

ESTADIO



PAL

9



10)

3 DETALHE JOBO



LANCES ✓

ESCALAÇÃO ✓

COMPARATIVO ✓

1	—————	0
2	—————	1
3	—————	0
0	—————	1

10)

≡ CLASSIFICAÇÃO

⊗ 3 x 1 ⊗

CLIQUE E VEJA O COMPARATIVO

ÍNDICE

IHC, 7, 17, 19, 21, 22, 29, 34, 38, 53

SUS, 11, 13, 17, 23, 34, 39, 49–53

TCC, 11, 33, 34