



Universidade de Brasília – UnB

Instituto de Psicologia

Programa de Pós-Graduação em Psicologia Clínica e Cultura

**Protótipo de um *game-based assessment* para avaliação de processos cognitivos em adultos: uma proposta fundamentada na abordagem *evidence-centered design***

Michelle Ferreira de Andrade

Brasília

Junho de 2026



Universidade de Brasília – UnB

Instituto de Psicologia

Programa de Pós-Graduação em Psicologia Clínica e Cultura

**Protótipo de um *game-based* assessment para avaliação de processos cognitivos em adultos: uma proposta fundamentada na abordagem *evidence-centered design* <sup>1</sup>**

Michelle Ferreira de Andrade

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Psicologia Clínica e Cultura do Instituto de Psicologia da Universidade de Brasília, como parte dos requisitos para obtenção do título de mestre em Psicologia Clínica e Cultura.

Orientadora: Prof.<sup>a</sup> Dra. Cristiane Faiad Moura

Coorientador: Prof. Dr. Alexandre L. O. Serpa

Brasília, junho de 2026

<sup>1</sup> Pesquisa apoiada pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior — CAPES.

Protótipo de um *game-based* assessment para avaliação de processos cognitivos em adultos: uma proposta fundamentada na abordagem *evidence-centered design*

Dissertação apresentada e avaliada por banca examinadora constituída por:

---

Prof.<sup>a</sup> Dra. Cristiane Faiad Moura (*Presidente - Orientadora*)

Universidade de Brasília - UnB

---

Prof. Dr. Alexandre Luiz de Oliveira Serpa

(*Coorientador*) Universidade Presbiteriana Mackenzie - SP

---

Prof.<sup>a</sup> Dra. Marcela Mansur-Alves (*Membro Externo*)

Universidade Federal de Minas Gerais (UFMG)

---

Prof. Dr. Carlos Manoel Lopes Rodrigues (*Membro Interno*)

Universidade de Brasília (UnB)

---

Prof. Dr. Sérgio Eduardo Silva de Oliveira (*Membro Suplente*)

Universidade de Brasília (UnB)

Brasília, junho de 2026

*“Que a arte nos aponte uma resposta, mesmo que ela não saiba.  
E que ninguém a tente complicar, porque é preciso simplicidade para fazê-la florescer.  
E que a minha loucura seja perdoada, porque metade de mim é amor  
E a outra metade também.”*

Oswaldo Montenegro - “Metade”

## **Agradecimentos**

Quero agradecer à minha família, minha base e meu alicerce em todas as etapas desta trajetória. Aos meus pais, agradeço pelos valores que me ensinaram e que continuam sustentando cada conquista da minha vida: o trabalho, a honestidade, a perseverança e a coragem de seguir mesmo diante dos desafios.

Ao meu marido, Jander Vasconcelos, meu companheiro de todas as horas, minha gratidão mais profunda. A caminhada foi difícil em muitos momentos, mas estar acompanhada tornou o percurso mais leve, possível e significativo. Sua paciência, seu apoio silencioso, sua serenidade e sua presença constante foram fundamentais nos dias mais exigentes. Caminhar ao seu lado me lembrou, muitas vezes, que algumas jornadas podem até ser individuais no esforço, mas se tornam mais fortes quando vividas com amor, parceria e cuidado.

À minha filha amada, Maria Luiza, minha maior inspiração, e aos meus enteados, filhos do meu coração, agradeço pelo carinho, pela convivência e por fazerem parte da construção da nossa família. A presença de vocês tornou esta caminhada mais afetiva, mais completa e cheia de sentido.

À minha tia Luzia Andrade, agradeço o cuidado generoso com a Malu nos períodos em que precisei me dedicar com mais intensidade aos estudos. Sua presença, disponibilidade e afeto foram essenciais para que eu pudesse caminhar com mais tranquilidade nesta etapa. Saber que a Malu estava acolhida e cuidada fez toda a diferença.

À minha querida cunhada, Luiza Elita Casado e à minha sogrinha querida, tenho profunda gratidão por tudo o que sempre fazem por mim e por nossa família. O apoio, a presença e a disponibilidade fazem parte desta conquista. Há ajudas que não aparecem nas páginas da dissertação, mas sustentam silenciosamente cada capítulo escrito.

À minha orientadora, Prof.<sup>a</sup> Dra. Cristiane Faiad, expresso minha profunda gratidão pela confiança, pela generosidade intelectual e pelo rigor acadêmico com que conduziu esta trajetória. Sua orientação foi essencial para transformar uma ideia desafiadora em um projeto possível e metodologicamente sustentado.

Ao meu coorientador, Dr. Alexandre Serpa, agradeço pela visão inovadora, pelas contribuições teóricas e metodológicas e pelo olhar atento à interface entre Psicologia, Psicometria e tecnologia. Suas provocações intelectuais ampliaram meus horizontes e contribuíram decisivamente para o amadurecimento deste trabalho.

Ao Programa de Pós-Graduação em Psicologia Clínica e Cultura da Universidade de Brasília (PPG-PsiCC/UnB), agradeço pela oportunidade de formação, pelo acolhimento institucional e pelo espaço de desenvolvimento acadêmico e científico. Estendo minha gratidão aos professores, em especial ao Prof. Dr. Sérgio Eduardo Silva de Oliveira e ao Prof. Dr. Carlos Manoel Lopes Rodrigues, com quem tive o privilégio de aprender, dialogar e amadurecer intelectualmente ao longo desta trajetória.

Ao grupo de pesquisa LabPAM-UnB, em especial à Milene Araújo e ao Erick Fontenele, pela parceria, disponibilidade e contribuições que tornaram este percurso mais colaborativo e enriquecedor.

Aos membros da banca, pelas valiosas contribuições que qualificaram este trabalho, e à CAPES, pelo apoio financeiro concedido durante esta etapa formativa, o qual contribuiu para a realização desta pesquisa.

Aos meus alunos e ao CEUB, que são fonte constante de inspiração e renovam, a cada encontro, o sentido da minha trajetória docente. Ensinar e aprender com vocês fortalece meu compromisso com uma Psicologia ética, crítica, sensível e transformadora.

A todos, minha sincera gratidão.

## Lista de Tabelas

### Manuscrito 1

#### Tabela 1

*Estudos identificados sobre avaliação cognitiva baseada em jogos, realidade virtual ou plataformas digitais.....49*

### Manuscrito 2

#### Tabela 1

*Definições operacionais dos processos cognitivos e sua participação nas tarefas do Neurotasks .....107*

#### Tabela 2

*Arquitetura inferencial preliminar do Neurotasks: tarefas, indicadores observáveis, processos examinados, decisões interpretativas e ameaças à interpretação.....119*

#### Tabela 3

*I-CVI dos juízes por item e critério de avaliação..... 122*

#### Tabela 4

*Proporção de respostas favoráveis dos usuários na avaliação de aparência e aceitabilidade do Neurotasks.....125*

## Lista de Figuras

### Manuscrito 1

#### Figura 1

*Fluxograma PRISMA 2020 adaptado do processo de seleção dos estudos incluídos na revisão de escopo.....45*

### Manuscrito 2

#### Figura 1

*Arquitetura inferencial do Neurotasks segundo o Evidence-Centered Design .....92*

#### Figura 2

*Fase 1: Exemplo do ambiente do quarto em que o jogador precisa procurar o celular.....97*

#### Figura 3

*Avaliação de juízes via web.....105*

#### Figura 4

*Mockup de tela do cenário do quarto da tarefa 1 .....110*

#### Figura 5

*Mockup de tela do cenário da cama da tarefa 1 .....111*

#### Figura 6

*Mockup de tela da lista de tarefas no celular da tarefa 2.....112*

#### Figura 7

*Mockup de tela da boca precisando de escovação da tarefa 3.....113*

#### Figura 8

*Mockup de tela da cozinha da tarefa 4 .....114*

#### Figura 9

*Mockup de tela do computador da tarefa 5 .....115*

#### Figura 10

*Mockup de tela da organização de documentos na tarefa 6.....116*

#### Figura 11

*Tela inicial do site Neurotasks .....117*

## Sumário

|                                                                                                                                                                     |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Lista de Tabelas.....                                                                                                                                               | 6   |
| Resumo Geral.....                                                                                                                                                   | 10  |
| Abstract.....                                                                                                                                                       | 11  |
| Apresentação.....                                                                                                                                                   | 12  |
| Referências.....                                                                                                                                                    | 24  |
| Manuscrito 1 - Avaliação digital de processos cognitivos em adultos: mapeamento de game-based assessments, tarefas computadorizadas e evidências psicométricas..... | 34  |
| Resumo.....                                                                                                                                                         | 34  |
| Abstract.....                                                                                                                                                       | 35  |
| Método.....                                                                                                                                                         | 39  |
| Discussão.....                                                                                                                                                      | 58  |
| Considerações Finais.....                                                                                                                                           | 63  |
| Referências.....                                                                                                                                                    | 66  |
| Manuscrito 2 – Neurotasks: desenvolvimento e evidências iniciais de validade baseadas no conteúdo segundo o Evidence-Centered Design.....                           | 72  |
| Resumo.....                                                                                                                                                         | 72  |
| Abstract.....                                                                                                                                                       | 73  |
| Método.....                                                                                                                                                         | 87  |
| Delineamento.....                                                                                                                                                   | 87  |
| Etapa 1 — Desenvolvimento conceitual e tecnológico do protótipo.....                                                                                                | 88  |
| Arquitetura inferencial do instrumento.....                                                                                                                         | 88  |
| Participantes.....                                                                                                                                                  | 92  |
| Instrumentos.....                                                                                                                                                   | 94  |
| Protótipo do Neurotasks.....                                                                                                                                        | 94  |
| Protocolo de avaliação por juízes.....                                                                                                                              | 97  |
| Protocolo de avaliação por usuários.....                                                                                                                            | 97  |
| Procedimentos.....                                                                                                                                                  | 98  |
| Etapa 1 — Desenvolvimento do protótipo.....                                                                                                                         | 98  |
| Etapa 2 — Documentação do jogo.....                                                                                                                                 | 99  |
| Etapa 3 — Avaliação por juízes.....                                                                                                                                 | 99  |
| Etapa 4 — Avaliação por usuários.....                                                                                                                               | 100 |
| Análise de dados.....                                                                                                                                               | 100 |
| Considerações éticas.....                                                                                                                                           | 104 |
| Resultados.....                                                                                                                                                     | 104 |
| Arquitetura geral do protótipo.....                                                                                                                                 | 105 |
| Tarefas, processos-alvo e métricas observáveis.....                                                                                                                 | 108 |
| Evidências iniciais de validade de conteúdo, obtidas por juízes especialistas.....                                                                                  | 118 |
| Resultados relativos aos processos de resposta, à aparência e à aceitabilidade entre                                                                                |     |

|                                                                        |     |
|------------------------------------------------------------------------|-----|
| usuários.....                                                          | 122 |
| Síntese integrada dos achados.....                                     | 127 |
| Discussão.....                                                         | 128 |
| Equidade e fairness.....                                               | 128 |
| Níveis progressivos e tarefas adaptativa.....                          | 129 |
| Considerações Finais.....                                              | 141 |
| Referências.....                                                       | 146 |
| Apêndices.....                                                         | 152 |
| Apêndice A.....                                                        | 152 |
| Termo de Consentimento Livre e Esclarecido (TCLE).....                 | 152 |
| Apêndice B.....                                                        | 157 |
| Protocolo de Avaliação Remota baseado no AvaliaJS.....                 | 157 |
| Apêndice C.....                                                        | 178 |
| Ficha de validade de conteúdo: avaliação por juízes especialistas..... | 178 |
| Parte I - Avaliação dos itens.....                                     | 179 |
| Parte II - Avaliação semântica.....                                    | 182 |
| Parte III - Avaliação global.....                                      | 182 |
| Apêndice D.....                                                        | 183 |
| Instruções ao jogador.....                                             | 183 |

## Resumo Geral

A avaliação psicológica de processos cognitivos em adultos demanda instrumentos psicometricamente fundamentados, ecologicamente pertinentes e capazes de captar indicadores do desempenho. Diante das limitações de alguns formatos tradicionais, como tarefas em lápis e papel, e da menor aproximação das demandas funcionais do cotidiano, observa-se a necessidade de ferramentas digitais que registrem métricas comportamentais e favoreçam a construção progressiva de evidências de validade. Nesse contexto, a *game-based assessment* (GBA) configura-se como uma abordagem que utiliza atividades baseadas em jogos como tarefas avaliativas, permitindo coleta automatizada de dados de desempenho e interação. Este estudo teve como objetivo desenvolver e analisar um protótipo digital de GBA, denominado Neurotasks, voltado à investigação de indicadores potenciais de funcionamento cognitivo em adultos. A construção foi orientada pela abordagem *Evidence-Centered Design* (ECD), utilizada para articular modelo de proficiência, modelo de evidências e modelo de tarefas. A dissertação foi estruturada em dois manuscritos: uma revisão de escopo sobre a construção e evidências de validade de GBAs aplicadas à avaliação de processos cognitivos em adultos; e um estudo de desenvolvimento do Neurotasks, com evidências iniciais de validade baseadas no conteúdo, obtidas por meio da avaliação de juízes especialistas, e evidências preliminares relacionadas aos processos de resposta, à aparência e à aceitabilidade, obtidas por meio da avaliação dos usuários. O Neurotasks (Protótipo 1; versão 2025) foi desenvolvido em plataforma web, no formato *point-and-click*, com registros automatizados das interações do usuário. O jogo é composto por quatro ambientes virtuais — quarto, banheiro, cozinha e escritório — que simulam situações funcionais do cotidiano, como busca de objetos, organização de informações, execução sequencial de ações e resolução de problemas práticos. As tarefas foram planejadas com hipóteses inferenciais relacionadas à atenção, memória de trabalho, memória de curto prazo, evocação, planejamento, organização, monitoramento da ação e controle executivo, sendo documentadas em um *Game Design Document* (GDD). Para a obtenção de evidências iniciais de validade baseadas no conteúdo, o protótipo foi avaliado por sete juízes especialistas, obtendo um Índice de Validade de Conteúdo geral de 0,82. Os usuários adultos avaliaram aspectos relacionados aos processos de resposta, à aparência, à aceitabilidade e à experiência de interação, por meio de instrumento adaptado do Protocolo AvaliaJS. Foram registrados 48 cadastros, 44 execuções do jogo e 23 avaliações completas, com proporção geral de respostas favoráveis de 0,76. Os resultados indicam adequação inicial do protótipo quanto à coerência das tarefas, relevância dos indicadores propostos e potencial de aplicação em avaliação psicológica informatizada, embora apontem necessidade de refinamentos na interface, nas instruções, nos feedbacks e nas tarefas. Conclui-se que o Neurotasks ainda não constitui um instrumento psicológico validado, mas representa uma etapa inicial de desenvolvimento de um GBA orientado pelo ECD. A pesquisa contribui ao propor um percurso metodológico para a construção de GBAs, articulando métricas processuais, experiência do usuário e inferências sobre o funcionamento cognitivo. Estudos futuros deverão investigar evidências adicionais de validade, precisão, validade convergente, aproximação funcional com situações cotidianas e aplicabilidade clínica.

**Palavras-chave:** avaliação psicológica informatizada; *game-based assessment*; processos cognitivos; *evidence-centered design*; validade de conteúdo; métricas de processo.

## Abstract

The psychological assessment of cognitive processes in adults requires psychometrically grounded and ecologically relevant instruments capable of capturing both performance outcomes and process indicators. Given the limitations of some traditional assessment formats, such as paper-and-pencil tasks and their lower approximation to functional daily-life demands, there is a need for digital tools that record behavioral metrics and support the progressive development of validity evidence. In this context, game-based assessment (GBA) emerges as an approach that uses game-based activities as assessment tasks, enabling the automated collection of performance and interaction data. This study aimed to develop and analyze a digital GBA prototype, named Neurotasks, designed to investigate potential indicators of cognitive functioning in adults. Its development was guided by the Evidence-Centered Design (ECD) approach, which articulates the proficiency model, evidence model, and task model. The dissertation was structured into two manuscripts: a scoping review on the construction and validity evidence of GBAs applied to the assessment of cognitive processes in adults; and a development study of Neurotasks, presenting a development study of Neurotasks, presenting initial validity evidence based on content and preliminary findings related to response processes, face-related aspects, acceptability, and user interaction. Neurotasks (Prototype 1; 2025 version) was developed as a web-based, *point-and-click* platform with automated recording of user interactions. The game consists of four virtual environments—bedroom, bathroom, kitchen, and office—that simulate functional daily-life situations, such as object search, information organization, sequential execution of actions, and practical problem solving. The tasks were designed with inferential hypotheses related to attention, working memory, short-term memory, recall, planning, organization, action monitoring, and executive control, and were documented in a *Game Design Document* (GDD). For content validity, the prototype was evaluated by seven expert judges, obtaining an overall content validity index of 0.82. A total of 48 users were registered, 44 game sessions were recorded, and 23 users completed the full evaluation, yielding an overall favorable-response proportion of 0.76. The findings indicate initial adequacy of the prototype regarding task coherence, relevance of the proposed indicators, and potential application in computerized psychological assessment, although refinements are needed in interface, instructions, feedback, and task design. It is concluded that Neurotasks is not yet a validated psychological instrument but represents an initial stage in the development of an ECD-guided GBA. This research contributes by proposing a methodological pathway for the construction of GBAs, articulating process metrics, user experience, and inferences about cognitive functioning. Future studies should investigate additional evidence of validity, reliability, convergent validity, functional approximation to everyday situations, and clinical applicability.

**Keywords:** computerized psychological assessment; game-based assessment; cognitive processes; evidence-centered design; content validity; process metrics.



## **Apresentação**

A avaliação de processos cognitivos em adultos constitui uma área central da avaliação psicológica, por permitir investigar funções associadas à adaptação do indivíduo às demandas cotidianas, acadêmicas, laborais e sociais. Esse campo se apoia em contribuições clássicas da neuropsicologia e da psicologia cognitiva, especialmente nos estudos sobre funções corticais superiores, atenção seletiva, memória operacional, interferência cognitiva e funções executivas (Baddeley & Hitch, 1974; Broadbent, 1958; Lezak, 1982; Luria, 1966, 1973; Stroop, 1935). Modelos posteriores ampliaram essa compreensão ao destacar o papel do controle atencional e executivo em processos como planejamento, monitoramento da ação, flexibilidade cognitiva e resolução de problemas, compreendidos como componentes interdependentes do funcionamento cognitivo orientado a objetivos (Dias & Seabra, 2014; Malloy-Diniz et al., 2010, 2018; Miyake et al., 2000; Miotto *et al.*, 2018; Norman & Shallice, 1986; Posner & Petersen, 1990; Seabra, 2012, 2014).

Entretanto, os processos cognitivos constituem construtos latentes e, portanto, não são diretamente observáveis. Sua avaliação depende da inferência realizada a partir de comportamentos produzidos em situações estruturadas, como acertos, erros, tempos de execução, tentativas, omissões, correções e sequências de ações. Para que esses indicadores sustentem interpretações psicologicamente defensáveis, a construção do instrumento deve articular a delimitação teórica do construto, os comportamentos esperados, as regras de pontuação e as evidências empíricas que justificam as interpretações propostas (Anastasi & Urbina, 2000; Hutz et al., 2015, 2018; Pasquali, 2007, 2010; Primi et al., 2019; Reppold et al., 2017). Assim, o registro de que um participante clicou, errou, demorou ou corrigiu uma resposta somente adquire significado avaliativo quando sua relação com o funcionamento cognitivo é teoricamente explicitada e empiricamente investigada.

A interpretação torna-se particularmente complexa em tarefas contextualizadas, porque atividades como localizar um objeto, organizar documentos ou executar uma sequência de ações podem mobilizar simultaneamente busca visual, atenção seletiva, memória operacional, planejamento, controle inibitório, tomada de decisão, monitoramento de erros e flexibilidade cognitiva (Malloy-Diniz et al., 2018; Miotto et al., 2018; Seabra, 2012). Desse modo, um mesmo comportamento pode admitir hipóteses cognitivas concorrentes e também sofrer influência de fatores externos ao construto, como compreensão das instruções, familiaridade tecnológica e demandas motoras. A interpretação dos indicadores deve, portanto, considerar tanto os processos-alvo quanto explicações alternativas para o desempenho observado (Kane, 2013; Messick, 1994).

Historicamente, a avaliação de processos cognitivos foi desenvolvida por meio de instrumentos padronizados, em grande parte administrados em formato lápis-papel ou em tarefas presenciais. Esses instrumentos permanecem essenciais para a prática psicológica e neuropsicológica, pois oferecem parâmetros normativos, padronização e evidências de validade e precisão (Hutz et al., 2015, 2018; Malloy-Diniz et al., 2018). No entanto, parte dessas medidas traz como produto final do desempenho o número de acertos, erros, tempo total de execução ou pontuações agregadas. Embora esses indicadores sejam clinicamente úteis, eles podem ser insuficientes para compreender o percurso cognitivo realizado pelo avaliado durante a execução da tarefa (Laurence et al., 2018; Serpa et al., 2024).

Dois indivíduos podem alcançar o mesmo resultado final por caminhos cognitivos distintos. Um pode apresentar desempenho lento, cauteloso e monitorado; outro pode responder de modo rápido, impulsivo e com sucessivas autocorrekções. Nesses casos, o escore final pode ocultar diferenças relevantes no processo de resposta. Por essa razão, cresce o interesse por modelos avaliativos capazes de registrar não apenas o resultado, mas também o

modo como o desempenho foi produzido (Shute, 2011; Shute & Ventura, 2013; von Davier et al., 2019). Esse deslocamento do produto para o processo amplia as possibilidades interpretativas da avaliação, desde que os indicadores observados sejam teoricamente justificados e submetidos a estudos psicométricos adequados, com especial atenção ao desenvolvimento de evidências que fundamentem a construção de um argumento de validade baseada no processo de respostas (American Educational Research Association et al., 2014).

Na literatura contemporânea, essas modalidades podem ser compreendidas no campo mais amplo das Technology-Based Assessments (TBAs), que empregam recursos computacionais na construção, apresentação, aplicação, registro, pontuação ou comunicação dos resultados. Embora esses recursos ampliem as possibilidades de elaboração de tarefas interativas e de análise dos processos de resposta, permanecem submetidos às mesmas exigências de validade, precisão e fairness aplicáveis às avaliações realizadas por meios tradicionais. Além disso, formatos inovadores demandam atenção específica à comparabilidade dos escores, às interações entre características da pessoa e da tarefa e à possível introdução de variância irrelevante ao construto (Bennett et al., 2025).

Tarefas computadorizadas, plataformas online, ambientes virtuais, sistemas adaptativos e *game-based assessments* permitem registrar automaticamente indicadores temporais, sequenciais e de precisão, como latência, cliques, tentativas, erros, correções e padrões de navegação (Alves & Serpa, 2019; Bennett et al., 2025; Mansur-Alves & Serpa, 2018; Muner et al., 2024; Pasquali, 2007; Serpa et al., 2024). Esses registros oferecem uma descrição mais detalhada do processo de resolução, mas sua utilidade depende da correspondência teórica e empírica entre o comportamento captado e o processo cognitivo investigado (Shute et al., 2015; von Davier et al., 2019).

Nesse contexto, Mansur-Alves e Serpa (2018) destacam que as novas tecnologias de informação e comunicação modificaram diferentes etapas da avaliação psicológica, incluindo aplicação, correção, armazenamento, interpretação e devolução de resultados. Os autores discutem que a testagem computadorizada, a avaliação online, as plataformas digitais e modelos como o *Computerized Adaptive Testing* podem aumentar a eficiência da avaliação e favorecer formas mais dinâmicas de apresentação dos estímulos e registro de respostas. No entanto, também alertam que a passagem do formato tradicional para o digital não garante, por si só, equivalência psicométrica ou validade interpretativa. Aspectos como familiaridade com computador, uso de mouse, tamanho da tela, velocidade de conexão, características da interface, ansiedade tecnológica, usabilidade e condições de aplicação podem influenciar o desempenho e introduzir variância irrelevante ao construto (Huff & Sireci, 2001; Mansur-Alves & Serpa, 2018; Naglieri et al., 2004; Russell et al., 2003).

Diretrizes recentes específicas para avaliações baseadas em tecnologia reforçam que o desenvolvimento e o uso de TBAs devem contemplar, de maneira integrada, validade, fairness, acessibilidade, segurança, privacidade, funcionamento da plataforma e consequências do uso dos resultados. Essas dimensões não devem ser tratadas como aspectos acessórios, pois podem alterar as condições de acesso à tarefa e a interpretação dos escores produzidos (International Test Commission & Association of Test Publishers [ITC/ATP], 2025).

Dessa forma, a avaliação psicológica digital exige uma postura metodológica simultaneamente inovadora e crítica. O uso de tecnologia não torna uma medida automaticamente superior, mais precisa ou mais válida. Pelo contrário, a sofisticação tecnológica pode aumentar a complexidade inferencial da avaliação. Ambientes digitais podem gerar grande volume de dados, mas nem todo dado comportamental constitui

evidência psicológica válida. O simples registro de cliques, tempos, trajetórias ou tentativas não assegura que esses indicadores representem atenção, memória operacional, planejamento ou qualquer outro processo cognitivo (American Educational Research Association et al., 2014 ; Kane, 2013; Messick, 1994). Para que esses dados tenham valor avaliativo, é necessário explicitar a cadeia que conecta tarefa, comportamento observado, processo cognitivo e interpretação psicológica (Cunha et al., 2023; Faiad et al., 2024; Serpa et al., 2024).

Destaca-se a relevância dos *game-based assessments*, definidos como avaliações que utilizam atividades baseadas em jogos para produzir evidências sobre conhecimentos, habilidades, competências ou outros atributos psicológicos a partir do comportamento do jogador (Landers, 2015; Mislevy et al., 2015; Shute & Kim, 2014). Diferentemente do uso de jogos para entretenimento, ensino ou intervenção, a GBA possui finalidade avaliativa. Isso significa que seus elementos lúdicos, narrativos e interativos devem ser planejados para eliciar comportamentos relevantes aos construtos de interesse (DiCerbo & Mislevy, 2017; Kim & Shute, 2015; Salen & Zimmerman, 2004). Nessa perspectiva, o jogo não deve ser compreendido como mero recurso motivacional ou como uma interface lúdica sobreposta a uma tarefa avaliativa tradicional. Ao contrário, ele compõe a própria estrutura metodológica de eliciação, registro e organização das evidências, uma vez que suas regras, objetivos, demandas interativas, condições de resposta e métricas de desempenho são planejados para produzir dados observáveis alinhados às inferências psicológicas pretendidas.

As GBAs apresentam potencial para ampliar o engajamento, reduzir a dependência de autorrelato, registrar trajetórias comportamentais e aproximar a avaliação de situações mais funcionais e ecologicamente plausíveis (Kleitman *et al.*, 2022; Shute, 2011; Weiner & Sanchez, 2020). Além disso, permitem coletar dados processuais que dificilmente seriam

observados com a mesma precisão em instrumentos tradicionais, como sequência de ações, padrões de exploração, tempo entre decisões, tentativas sucessivas, erros de rota e estratégias de resolução (DiCerbo & Mislevy, 2017; Serpa *et al.*, 2024). Contudo, esses mesmos recursos aumentam a necessidade de controle metodológico, uma vez que tarefas gamificadas podem envolver múltiplas fontes de variação no desempenho, incluindo experiência prévia com jogos, familiaridade digital, coordenação visomotora, compreensão da interface, motivação, tolerância à ambiguidade e responsividade aos feedbacks (Chamorro-Premuzic *et al.*, 2016; Leutner *et al.*, 2023; Mansur-Alves & Serpa, 2018).

Essa preocupação foi reforçada por Markovitch *et al.* (2025), que compararam jogadores frequentes e pouco frequentes em jogos destinados à avaliação cognitiva. Os participantes com menor frequência de jogo relataram experiência menos favorável e apresentaram menor precisão dos dados em medidas cognitivas mais complexas, com redução da validade das mensurações. A experiência prévia com jogos deve, portanto, ser considerada como potencial fonte de variância irrelevante ao construto e como variável relevante em estudos de fairness.

Entretanto, resultados mais recentes indicam que a experiência produzida pelo formato gamificado não é necessariamente positiva para todos os participantes. Ohlms *et al.* (2024), ao compararem uma GBA de habilidade cognitiva com uma avaliação em formato tradicional, encontraram relação positiva entre os escores, mas observaram avaliações menos favoráveis da experiência gamificada, especialmente entre mulheres e pessoas com menor experiência prévia com jogos. Esses achados indicam que atratividade, experiência do usuário e validade devem ser investigadas separadamente, evitando pressupor que a gamificação aumenta automaticamente a aceitação ou a qualidade da medida.

No caso da avaliação de processos cognitivos, essa questão é especialmente sensível. Uma tarefa aparentemente voltada ao planejamento pode ser influenciada pela clareza das instruções, pelo tamanho dos estímulos, pela forma de navegação ou pela familiaridade do usuário com ambientes *point-and-click*. Do mesmo modo, uma tarefa de busca visual pode refletir atenção seletiva, mas também pode ser afetada por contraste visual, ambiguidade dos objetos clicáveis ou expectativa do participante sobre a lógica do jogo. Essas fontes de influência configuram possíveis ameaças à interpretação dos resultados e precisam ser consideradas desde o desenvolvimento do instrumento (Cunha et al., 2023; Huff & Sireci, 2001; Muner et al., 2024). Assim, a inovação tecnológica deve estar articulada a uma arquitetura psicométrica rigorosa, capaz de explicitar o construto-alvo, os comportamentos observáveis esperados, as métricas extraídas da interação do usuário e as evidências empíricas necessárias para sustentar as inferências pretendidas. Em instrumentos digitais, a qualidade da avaliação não depende apenas da atratividade da interface ou da sofisticação do recurso tecnológico, mas da clareza sobre o que se pretende avaliar, de que modo os dados são produzidos e quais evidências sustentam a interpretação dos resultados, a fim de produzir medidas válidas, fidedignas e equânimes.

Entre as abordagens metodológicas capazes de orientar esse tipo de construção, destaca-se o *Evidence-Centered Design (ECD)*. O *ECD* propõe que uma avaliação seja desenvolvida a partir de uma lógica de evidências, na qual se define, inicialmente, o que se pretende inferir sobre o avaliado; em seguida, quais comportamentos observáveis podem servir como evidência dessas inferências; e, por fim, quais tarefas devem ser construídas para eliciar esses comportamentos (Almond et al., 2002; Mislevy et al., 2003, 2004, 2006; Mislevy, 2018). Essa abordagem organiza o processo avaliativo por meio da articulação entre modelo de proficiência, modelo de evidência e modelo de tarefa, favorecendo maior

coerência entre teoria, design, coleta de dados e interpretação dos resultados (Arieli-Attali et al., 2019; DiCerbo & Mislevy, 2017; von Davier & Mislevy, 2013). O modelo de proficiência define os atributos ou processos cognitivos que se pretende inferir. O modelo de evidências especifica os comportamentos observáveis e as métricas processuais capazes de subsidiar essas inferências. Por sua vez, o modelo de tarefas descreve as condições, os estímulos, as regras e as atividades necessárias para eliciar os comportamentos previstos. Em *game-based assessments* (GBAs), essa articulação é particularmente relevante, pois permite relacionar tarefas digitais, indicadores processuais e interpretações psicológicas, reduzindo o risco de inferências intuitivas ou pouco fundamentadas sobre o desempenho do avaliado (DiCerbo & Mislevy, 2017; Kim & Shute, 2015; Shute & Ventura, 2013; von Davier & Mislevy, 2013).

A aplicação do ECD é particularmente pertinente para avaliações digitais e gamificadas, pois impede que os dados gerados pelo sistema sejam interpretados de maneira direta ou intuitiva, sem fundamentação teórica suficiente. Em instrumentos baseados em jogos, a quantidade de dados pode produzir uma falsa sensação de precisão. Entretanto, do ponto de vista psicométrico, o valor desses dados depende de sua relação com o construto e com as hipóteses interpretativas do instrumento (Mislevy et al., 2015; Serpa et al., 2024). Dessa forma, o ECD contribui para transformar dados digitais em evidências psicológicas potencialmente interpretáveis, desde que essas relações sejam posteriormente testadas em estudos empíricos de validade, precisão, estrutura interna e relações com variáveis externas (AERA et al., 2014; Kane, 2013).

Foi a partir dessa problemática que se desenvolveu a presente dissertação. O estudo parte da lacuna existente entre o potencial das tecnologias digitais para registrar dados processuais e a necessidade de modelos metodológicos capazes de sustentar inferências válidas sobre processos cognitivos em adultos. No contexto brasileiro, ainda são escassos os

estudos que descrevem sistematicamente o desenvolvimento de *game-based assessments* voltados à avaliação psicológica de processos cognitivos, especialmente com explicitação da relação entre tarefa, indicador observável, processo cognitivo e interpretação psicológica (Alves & Serpa, 2019; Muner et al., 2024; Serpa et al., 2024). Assim, o presente trabalho pretende contribuir para essa agenda ao propor um protótipo digital fundamentado no ECD e orientado por evidências iniciais de validade.

Nesse cenário, foi concebido o Neurotasks, um protótipo digital em ambiente web, desenvolvido como *game-based assessment* para avaliação de processos cognitivos em adultos. O instrumento apresenta uma rotina funcional em que o participante executa tarefas contextualizadas em ambientes cotidianos, como quarto, banheiro, cozinha e escritório. As tarefas envolvem busca de objetos, organização de informações, execução sequencial de ações e resolução de problemas práticos, com interação predominantemente no formato *point-and-click*. O protótipo foi planejado para registrar indicadores temporais, sequenciais e de precisão, armazenados como logs comportamentais produzidos durante a execução das tarefas. Nesta etapa, esses registros são compreendidos como evidências potenciais, cuja relação com os processos cognitivos propostos deverá ser investigada empiricamente em estudos posteriores.

Diante da necessidade de desenvolver instrumentos digitais que articulem inovação tecnológica, fundamentação teórica e rigor psicométrico, esta dissertação foi organizada em dois manuscritos complementares. O primeiro apresenta uma revisão de escopo sobre *game-based assessment* e avaliação digital de processos cognitivos, com o objetivo de mapear a produção científica e identificar fundamentos relevantes para a construção de instrumentos digitais. O segundo descreve o desenvolvimento do protótipo Neurotasks, fundamentado no *Evidence-Centered Design*, e apresenta evidências iniciais de validade baseadas no conteúdo,

provenientes da avaliação dos juízes especialistas, além de evidências preliminares relacionadas aos processos de resposta, à aparência, à aceitabilidade e à experiência de interação dos usuários.

O primeiro manuscrito, intitulado “Avaliação digital de processos cognitivos em adultos: mapeamento de *game-based assessments*, tarefas computadorizadas e evidências psicométricas”, apresenta revisão de escopo sobre construção e evidências de validade de GBAs na avaliação de processos cognitivos em adultos. Seu objetivo é mapear tipos de tarefas, métricas observáveis, evidências psicométricas e desafios inferenciais reportados na literatura. Esse manuscrito contribui para situar o Neurotasks no campo da avaliação psicológica digital e para identificar lacunas relacionadas à construção de instrumentos gamificados, especialmente no que se refere à explicitação da cadeia inferencial entre tarefa, indicador comportamental, processo cognitivo e interpretação psicológica.

O segundo manuscrito, intitulado “Neurotasks: desenvolvimento e evidências iniciais de validade baseadas no conteúdo segundo o Evidence-Centered Design”, descreve o desenvolvimento do protótipo e apresenta evidências iniciais de validade baseadas no conteúdo e resultados preliminares relativos aos processos de resposta, à aparência, à aceitabilidade e à experiência de interação dos usuários. O objetivo é explicitar a arquitetura inferencial do instrumento, descrever suas tarefas, processos-alvo e métricas observáveis, bem como apresentar os resultados da apreciação por juízes especialistas e usuários. Nesse manuscrito, são discutidas as evidências preliminares de adequação conceitual e aceitabilidade do protótipo, além de fragilidades operacionais relacionadas à interface, instruções, feedbacks, acessibilidade e possíveis fontes de variância irrelevante ao construto.

Assim, a presente dissertação parte da premissa de que instrumentos digitais e gamificados podem contribuir para a avaliação psicológica contemporânea, desde que sejam

construídos a partir de modelos teóricos claros, tarefas intencionalmente planejadas, métricas interpretáveis e estudos progressivos de validade. A contribuição esperada não está apenas na criação de um jogo aplicado à Psicologia, mas na proposição de uma arquitetura avaliativa capaz de transformar comportamento digital em evidência psicológica válida, interpretável e cientificamente defensável.

### **Objetivo geral**

Esta dissertação tem por objetivo desenvolver um protótipo de game-based assessment para avaliação de processos cognitivos em adultos, fundamentado na abordagem Evidence-Centered Design, e apresentar evidências iniciais de validade baseadas no conteúdo, além de resultados preliminares relativos aos processos de resposta, à aparência e à aceitabilidade entre usuários.

### **Objetivos específicos**

Para atingir o objetivo geral, foram desenvolvidos dois manuscritos, que atendem a diferentes objetivos específicos:

**Manuscrito 1:** mapear, por meio de uma revisão de escopo, as abordagens digitais utilizadas na avaliação de processos cognitivos em adultos, com ênfase em game-based assessments, tarefas computadorizadas, adaptações para plataformas digitais, ambientes virtuais, métricas processuais, evidências psicométricas e desafios inferenciais reportados na literatura.

**Manuscrito 2:** descrever o desenvolvimento do Neurotasks, um protótipo de game-based assessment fundamentado no Evidence-Centered Design, explicitando sua arquitetura inferencial, tarefas, processos-alvo e métricas observáveis, bem como apresentar evidências iniciais de validade baseadas no conteúdo e resultados preliminares referentes aos processos de resposta, à aparência, à aceitabilidade e à experiência de interação dos usuários.

Como Produtos Técnico-Tecnológicos (PTTs) oriundos da presente dissertação, foram produzidos: o Neurotasks, um sistema funcional automatizado hospedado em um site para interação com os usuários, bem como um Game Design Document (GDD), que descreve o planejamento, as mecânicas e o desenho do jogo. Os produtos podem ser acessados Os produtos e os respectivos materiais técnicos podem ser acessados por meio do repositório indicado na página 110 desta dissertação.

## Referências

- Almond, R. G., Steinberg, L. S., & Mislevy, R. J. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *Journal of Technology, Learning, and Assessment, 1*(5), 1–63.
- Alves, M. M., & Serpa, A. L. O. (2019). Avaliação psicológica por meio das novas tecnologias de informação e comunicação. In M. N. Baptista, M. Muniz, C. T. Reppold, C. H. S. S. Nunes, L. F. Carvalho, R. Primi, A. P. P. Noronha, A. G. Seabra, S. M. Wechsler, C. S. Hutz, & L. Pasquali (Orgs.), *Compêndio de avaliação psicológica* (pp. 160–172). Vozes.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- American Psychological Association. (1986). *Guidelines for computer-based tests and interpretations*. American Psychological Association.
- Anastasi, A., & Urbina, S. (2000). *Testagem psicológica*. Artmed.
- Arieli-Attali, M., Wylie, E. C., & Bauer, M. I. (2019). The expanded evidence-centered design (e-ECD) for learning and assessment systems: A conceptual framework for incorporating learning goals and processes within assessment design. *Frontiers in Psychology, 10*, Article 853. <https://doi.org/10.3389/fpsyg.2019.00853>
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). Academic Press.
- Barros, P. M., & Hazin, I. (2013). Avaliação das funções executivas na infância: Revisão dos conceitos e instrumentos. *Psicologia em Pesquisa, 7*(1), 13–22.

- Bennett, R. E., LaMar, M. M., & Mazzeo, J. (2025). Technology-based assessment: Validity, modeling, and analysis issues. In L. L. Cook & M. J. Pitoniak (Eds.), *Educational measurement* (5th ed., pp. 581–654). Oxford University Press.  
<https://doi.org/10.1093/oso/9780197654965.003.0009>
- Broadbent, D. E. (1958). *Perception and communication*. Pergamon Press.
- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 3(3), 296–322.
- Buckley, P., Noonan, S., Geary, C., MacKessy, T., & Nagle, E. (2019). An empirical study of gamification frameworks. *Journal of Organizational and End User Computing*, 31(1), 22–38. <https://doi.org/10.4018/JOEUC.2019010102>
- Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A., & Hogan, R. (2016). New talent signals: Shiny new objects or a brave new world? *Industrial and Organizational Psychology*, 9(3), 621–640. <https://doi.org/10.1017/iop.2016.6>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- Cunha, R. D., Faiad, C., Baptista, M. N., & Cardoso, H. F. (2023). The data generation mechanism: Relationship between constructs and their indicators. *Psicologia: Teoria e Pesquisa*, 39, Article e39nspe08.  
<https://doi.org/10.1590/0102.3772e39nspe08.en>
- Das, S., & Misra, S. (2014). Gamification of learning and instruction: Game-based methods and strategies for training and education. *International Journal of Information and Learning Technology*, 31(2), 134–136.
- Dias, N. M. (Org.). (2014). *Avaliação neuropsicológica cognitiva* (Vol. 4). Memnon.

- DiCerbo, K. E., & Mislevy, R. J. (2017). Evidence-centered design and the future of assessment. In M. von Davier, C. Zhu, & P. C. Kyllonen (Eds.), *Innovative assessment of collaboration* (pp. 19–37). Springer.
- Drasgow, F. (Ed.). (2015). *Technology and testing: Improving educational and psychological measurement*. Routledge. <https://doi.org/10.4324/9781315871493>
- Faiad, C., Schelini, P. W., & Argimon, I. I. L. (2024). Avaliação psicológica, tecnologias sociais e inovação: Editorial. *Recital*, 5, 8–10.
- Georgiou, K., & Das, R. (2016). The role of game-based assessment in personnel selection. *International Journal of Selection and Assessment*, 24(4), 327–331.
- Georgiou, K., Gouras, A., & Nikolaou, I. (2017). Gamification in employee selection: The development of a gamified assessment. *International Journal of Selection and Assessment*, 27(2), 91–103.
- Gomez, M. J., Ruipérez-Valiente, J. A., Martínez, P. A., & Kim, Y. J. (2022). Applying learning analytics to game-based assessment. *Computers in Human Behavior*, 130, Article 107171.
- Gomez, M. J., Ruipérez-Valiente, J. A., & García-Clemente, F. J. (2023). A systematic literature review of game-based assessment studies: Trends and challenges. *IEEE Transactions on Learning Technologies*, 16(4), 500–515.  
<https://doi.org/10.1109/TLT.2022.3226661>
- Huff, K. L., & Sireci, S. G. (2001). Validity issues in computer-based testing. *Educational Measurement: Issues and Practice*, 20(3), 16–25.  
<https://doi.org/10.1111/j.1745-3992.2001.tb00066.x>
- Hutz, C. S., Bandeira, D. R., & Trentini, C. M. (Orgs.). (2015). *Psicometria*. Artmed.

- Hutz, C. S., Bandeira, D. R., & Trentini, C. M. (Orgs.). (2018). *Avaliação psicológica da inteligência e da personalidade*. Artmed.
- International Test Commission, & Association of Test Publishers. (2025). *ITC/ATP guidelines for technology-based assessment*. International Test Commission.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Kim, Y. J., & Shute, V. J. (2015). The interplay of game elements with psychometric qualities, learning, and enjoyment in game-based assessment. *Computers & Education*, 87, 340–356. <https://doi.org/10.1016/j.compedu.2015.07.009>
- Kim, Y. J., Almond, R. G., & Shute, V. J. (2016). Applying evidence-centered design for the development of game-based assessments in Physics Playground. *International Journal of Testing*, 16(2), 142–163.  
<https://doi.org/10.1080/15305058.2015.1108322>
- Kimura, T. (2017). The impacts of computer adaptive testing from a variety of perspectives. *Journal of Educational Evaluation for Health Professions*, 14, Article 12.  
<https://doi.org/10.3352/jeehp.2017.14.12>
- Kleitman, S., Huang, P.-S., & Adjerid, I. (2022). Applying evidence-centered design to game-based assessment of psychological resilience. *Frontiers in Psychology*, 12, Article 717568. <https://doi.org/10.3389/fpsyg.2021.717568>
- Krkovic, K., Pásztor-Kovács, A., Molnár, G., & Greiff, S. (2014). New technologies in psychological assessment: The example of computer-based collaborative problem solving assessment. *International Journal of e-Assessment*, 1, 1–13.
- Landers, R. N. (2015). An introduction to game-based assessment: Frameworks for the measurement of knowledge, skills, abilities and other human characteristics using

- behaviors observed within videogames. *International Journal of Gaming and Computer-Mediated Simulations*, 7(4), iv–vii.
- Landers, R. N., & Sanchez, D. R. (2022). Game-based, gamified, and gamefully designed assessments for employee selection: Definitions, distinctions, design, and validation. *International Journal of Selection and Assessment*, 30(1), 1–13.  
<https://doi.org/10.1111/ijsa.12376>
- Laurence, P. G., Mecca, T. P., Serpa, A., Martin, R., & Macedo, E. C. (2018). Eye movements and cognitive strategy in a fluid intelligence test: Item type analysis. *Frontiers in Psychology*, 9, Article 380.
- Leutner, F., Codreanu, S.-C., Brink, S., & Bitsakis, T. (2023). Game-based assessments of cognitive ability in recruitment: Validity, fairness and test-taking experience. *Frontiers in Psychology*, 13, Article 942662.  
<https://doi.org/10.3389/fpsyg.2022.942662>
- Lezak, M. D. (1982). The problem of assessing executive functions. *International Journal of Psychology*, 17(1–4), 281–297.
- Malloy-Diniz, L. F., Fuentes, D., Mattos, P., & Abreu, N. (Orgs.). (2018). *Avaliação neuropsicológica* (2ª ed.). Artmed.
- Mansur-Alves, M., & Serpa, A. (2018). Avaliação psicológica por meio das novas tecnologias de informação e comunicação. In C. S. Hutz, D. R. Bandeira, C. M. Trentini, & J. S. Krug (Orgs.), *Psicodiagnóstico* (pp. 161–176). Artmed.
- Markovitch, B., Kamps, J. C. C., Markopoulos, P., & Birk, M. V. (2025). Do cognitive assessment games leave infrequent video game players behind? Evaluating frequent and infrequent players' gaming experience and data quality. *Computers in Human Behavior*, 172, Article 108720. <https://doi.org/10.1016/j.chb.2025.108720>

- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23.
- Miotto, E. C., Lucia, M. C. S., & Scaff, M. (Orgs.). (2018). *Manual de avaliação neuropsicológica* (Vol. 1). Vetor.
- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design* (ETS Research Report No. RR-03-16). Educational Testing Service. <https://doi.org/10.1002/j.2333-8504.2003.tb01908.x>
- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2004). *A brief introduction to evidence-centered design* (CSE Report No. 632). National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles.
- Mislevy, R. J., Corrigan, S., Oranje, A., DiCerbo, K., Bauer, M. I., von Davier, A., & John, M. (2015). Psychometrics and game-based assessment. In F. Drasgow (Ed.), *Technology and testing: Improving educational and psychological measurement* (pp. 23–48). Routledge. <https://doi.org/10.4324/9781315871493-2>
- Mislevy, R. J., Steinberg, L. S., Almond, R. G., & Lukas, J. F. (2006). Concepts, terminology, and basic models of evidence-centered design. In D. M. Williamson, I. I. Bejar, & R. J. Mislevy (Eds.), *Automated scoring of complex tasks in computer-based testing* (pp. 15–47). Lawrence Erlbaum Associates.
- Mislevy, R. J. (2018). *Sociocognitive foundations of educational measurement*. Routledge.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100.

- Muner, L. C., Serpa, A. L. O., Martins, G. D., Rodrigues, J. C., & Becker, N. (2024). De analógico a digital: Desafios e estratégias na adaptação de testes e tarefas neuropsicológicas para plataformas digitais. In L. R. R. Carreiro, S. M. B. Assis, & E. H. M. Brum (Orgs.), *Ciências do desenvolvimento humano: Produção de conhecimento e impactos de pesquisas em psicologia, saúde e educação* (pp. 172–186). Editora CESMAC.
- Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., & Velasquez, R. (2004). Psychological testing on the Internet: New problems, old issues. *American Psychologist*, 59(3), 150–162.  
<https://doi.org/10.1037/0003-066X.59.3.150>
- Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol. 4, pp. 1–18). Plenum Press.
- Ohlms, M. L., Melchers, K. G., & Kanning, U. P. (2024). Can we playfully measure cognitive ability? Construct-related validity and applicant reactions. *International Journal of Selection and Assessment*, 32, 91–107. <https://doi.org/10.1111/ijsa.12450>
- Pasquali, L. (2007). *Teoria de Resposta ao Item*. Laboratório de Pesquisa em Avaliação e Medida—LabPAM/UnB.
- Pasquali, L. (2010). *Instrumentação psicológica: Fundamentos e práticas*. Artmed.
- Pedersen, M. K., Castaño Díaz, C. M., Wang, Q. J., Alba-Marrugo, M. A., Amidi, A., Basaiawmoit, R. V., Bergenholtz, C., Christiansen, M. H., Gajdacz, M., Hertwig, R., Ishkhanyan, B., Klyver, K., Ladegaard, N., Mathiasen, K., Parsons, C., Rafner, J., Ryom Villadsen, A., Wallentin, M., Zana, B., & Sherson, J. F. (2023). Measuring

cognitive abilities in the wild: Validating a population-scale game-based cognitive assessment. *Cognitive Science*, 47(6), Article e13308.

<https://doi.org/10.1111/cogs.13308>

Primi, R., Muniz, M., & Nunes, C. H. S. S. (2009). Definições contemporâneas de validade de testes psicológicos. In C. S. Hutz (Org.), *Avanços e polêmicas em avaliação psicológica* (pp. 243–265). Casa do Psicólogo.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, 13, 25–42.

Reppold, C. T., Serafini, A. J., Ramires, D. A., & Gurgel, L. G. (2017). Análise dos manuais psicológicos aprovados pelo SATEPSI para avaliação de crianças e adolescentes no Brasil. *Avaliação Psicológica*, 16(1), 19–28.

<https://doi.org/10.15689/ap.2017.1601.03>

Salen, K., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. MIT Press.

Santana, A. N. (2019). Funções executivas e avaliação neuropsicológica. In A. G. Seabra & N. M. Dias (Orgs.), *Avaliação neuropsicológica cognitiva*. Memnon.

Santana, A. N., Melo, M. R. A., & Minervino, C. A. S. M. (2019). Instrumentos de avaliação das funções executivas: Revisão sistemática dos últimos cinco anos. *Avaliação Psicológica*, 18(1), 96–107. <https://doi.org/10.15689/ap.2019.1801.14668.11>

Seabra, A. G. (Org.). (2012). *Avaliação neuropsicológica cognitiva: Atenção e funções executivas* (Vol. 1). Memnon.

Seabra, A. G. (Org.). (2014). *Avaliação neuropsicológica cognitiva: Linguagem oral* (Vol. 2). Memnon.

Serpa, A. L. O., Rodrigues, M. M., & Maruyama, J. M. (2024). Psicometria computacional: Fundamentos e intersecções entre psicologia, estatística e ciência da computação. In

- S. M. B. Assis, L. R. R. Carreiro, & E. H. M. Brum (Orgs.), *Ciências do desenvolvimento humano: Produção de conhecimento e impactos de pesquisas em psicologia, saúde e educação* (pp. 559–574). Editora CESMAC.
- Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503–524). Information Age Publishing.
- Shute, V. J., & Kim, Y. J. (2014). Formative and stealth assessment. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 311–321). Springer.
- Shute, V. J., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games*. MIT Press.
- Shute, V. J., Ventura, M., & Kim, Y. J. (2015). Assessment and learning of qualitative physics in Newton's Playground. *The Journal of Educational Research*, 106(6), 423–430.
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3(3), 271–295.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643–662.
- von Davier, M., & Mislevy, R. J. (2013). Design and psychometric considerations for game-based assessment. *Educational Testing Service Research Report Series*.
- von Davier, M., Zhu, C., & Kyllonen, P. C. (Eds.). (2019). *Innovative assessment of collaboration*. Springer.
- Wang, C., Chang, H.-H., & Huebner, A. (2011). Restrictive stochastic item selection methods in cognitive diagnostic computerized adaptive testing. *Journal of Educational Measurement*, 48(3), 255–273. <https://doi.org/10.1111/j.1745-3984.2011.00145.x>

Zanini, D. S., Reppold, C. T., & Faiad, C. (2021). Do lápis e papel à modalidade remota: Considerações sobre a avaliação psicológica em tempos de pandemia. In K. L. Oliveira, P. W. Schelini, & S. M. Barroso (Orgs.), *Formação e estratégias de ensino em avaliação psicológica* (pp. 37–51). Vozes.

**Manuscrito 1 – Avaliação digital de processos cognitivos em adultos:  
mapeamento de *game-based assessments*, tarefas computadorizadas e evidências  
psicométricas**

**Resumo**

A avaliação de processos cognitivos tem sido transformada pela expansão de tecnologias digitais, tarefas computadorizadas, ambientes virtuais, jogos sérios e *game-based assessments* (GBAs), os quais possibilitam a obtenção de indicadores comportamentais e processuais para além dos escores finais de desempenho. Este estudo teve como objetivo mapear, por meio de uma revisão de escopo, as abordagens digitais utilizadas na avaliação de processos cognitivos em adultos, com ênfase em *game-based assessments*, tarefas computadorizadas, adaptações para plataformas digitais e ambientes virtuais, descrevendo os tipos de tarefas, as métricas processuais, as evidências psicométricas e os desafios inferenciais reportados na literatura. A revisão foi delineada e relatada conforme o *Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews* (PRISMA-ScR), e estruturada com base na estratégia PCC, contemplando: População, adultos; Conceito, avaliação de processos cognitivos, com ênfase em GBAs; e Contexto, avaliação psicológica, neuropsicológica, clínica, educacional, ocupacional e em saúde. As buscas foram conduzidas nas bases PubMed/MEDLINE, Scopus e SciELO. Foram identificados 1.392 registros, dos quais quatro estudos compuseram a síntese qualitativa final. Os resultados indicaram um campo promissor, porém heterogêneo, com diversidade de tecnologias, construtos e métricas. As principais lacunas referem-se à baixa presença de GBAs voltados à avaliação cognitiva de adultos, à limitada padronização das métricas e à fragilidade na explicitação da cadeia inferencial entre tarefa, indicador comportamental, processo cognitivo e interpretação psicológica. Conclui-se que instrumentos digitais podem contribuir para a avaliação contemporânea, desde que sustentados por modelos teóricos, evidências de validade e arquitetura inferencial explícita.

**Palavras-chave:** avaliação psicológica digital; processos cognitivos; *game-based assessment*; psicomетria computacional; revisão de escopo.

## Abstract

The assessment of cognitive processes has been transformed by the expansion of digital technologies, computerized tasks, virtual environments, serious games, and game-based assessments (GBAs), which allow the recording of process indicators beyond final performance outcomes. This study aimed to map, through a scoping review, digital approaches used to assess cognitive processes in adults, with an emphasis on GBAs, computerized tasks, virtual reality, process metrics, psychometric evidence, and inferential challenges. The review was reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews. The review was structured according to the PCC strategy: Population, adults; Concept, digital, computerized, or gamified assessment of cognitive processes; and Context, psychological, neuropsychological, clinical, educational, occupational, and health assessment. Searches were conducted in PubMed/MEDLINE, Scopus, and SciELO. A total of 1,392 records were identified, of which four studies were included in the final qualitative synthesis. The results indicated a promising but heterogeneous field, marked by diverse technologies, constructs, and metrics. The main gaps refer to the limited presence of GBAs aimed at adult cognitive assessment, low metric standardization, and insufficient clarification of the inferential chain between task, behavioral indicator, cognitive process, and psychological interpretation. It is concluded that digital instruments may contribute to contemporary assessment, provided that their development is grounded in theoretical models, validity evidence, and an explicit inferential architecture.

**Keywords:** digital psychological assessment; cognitive processes; game-based assessment; computational psychometrics; scoping review.

A avaliação psicológica de processos cognitivos enfrenta, atualmente, um desafio central: avançar da simples mensuração do resultado para a compreensão do caminho percorrido pelo indivíduo até chegar à resposta. Em modelos tradicionais de avaliação psicológica, o desempenho costuma ser descrito por indicadores como número de acertos, erros, tempo total de execução e pontuações compostas. Esses parâmetros são fundamentais, especialmente por sua utilidade clínica, comparabilidade normativa e contribuição para a tomada de decisão. Nesse cenário, a incorporação de métricas de processo — como latência, sequência de ações, número de tentativas, correções, padrões de navegação e estratégias de interação — amplia as possibilidades de análise do desempenho, permitindo uma leitura mais refinada e dinâmica do funcionamento cognitivo observado. Entretanto, evidências recentes indicam que a relação entre indicadores extraídos de jogos e medidas tradicionais de habilidade cognitiva é apenas moderada e pode variar conforme o tipo e a amplitude das métricas utilizadas, reforçando a necessidade de investigar empiricamente o significado psicológico dos dados produzidos durante a interação (Anastasi & Urbina, 2000; Bipp et al., 2024; Hutz et al., 2015; Laurence et al., 2018; Mansur-Alves & Serpa, 2018; Shute, 2011; von Davier et al., 2019).

Nos últimos anos, a expansão das tecnologias digitais no campo da avaliação psicológica tem modificado de forma significativa os modos de aplicação, registro, correção e interpretação dos instrumentos (Mansur-Alves & Serpa, 2018; Naglieri et al., 2004; Pasquali, 2007; Russell et al., 2003). Instrumentos informatizados, tarefas computadorizadas, plataformas remotas, simulações interativas, ambientes de realidade virtual e *game-based*

*assessments* passaram a oferecer novas possibilidades de registro comportamental, incluindo tempos de reação, latências, número de tentativas, revisões de rota, erros de sequência, navegação espacial, uso de pistas e outros indicadores processuais (Alves & Serpa, 2019; Huff & Sireci, 2001; Muner et al., 2024; Serpa et al., 2024; Shute, 2011; von Davier *et al.*, 2019). Esse movimento amplia o nível de detalhamento dos dados e aproxima a avaliação de contextos mais dinâmicos e ecologicamente relevantes. Ao mesmo tempo, recoloca uma questão central da psicometria: a sofisticação tecnológica do formato não substitui a necessidade de evidências de validade. A qualidade de um instrumento depende da coerência estabelecida entre o construto-alvo, as demandas da tarefa, as evidências observáveis produzidas pelo desempenho e a interpretação psicológica sustentada a partir desses dados. (American Educational Research Association et al., 2014; ; Cunha et al., 2023; Faiad et al., 2024; Hutz et al., 2015; Kane, 2013; Messick, 1994).

Em avaliações digitais, cliques, tempos de resposta, latências, rotas de navegação e sequências de ações somente adquirem valor psicométrico quando podem ser interpretados como evidências relevantes do construto, e não apenas como registros operacionais do sistema (DiCerbo & Mislevy, 2017; Mislevy et al., 2015; Serpa et al., 2024; Shute, 2011). Essa questão ganha especial importância em instrumentos digitais complexos, nos quais diferentes fontes de variância podem influenciar o desempenho, tais como tipo de estímulo, demanda cognitiva da tarefa, qualidade da interface, familiaridade com tecnologias digitais, experiência prévia com jogos, coordenação visomotora, compreensão das instruções, tamanho da tela, usabilidade, acessibilidade e tolerância à ambiguidade (Alves & Serpa, 2019; Huff & Sireci, 2001; Mansur-Alves & Serpa, 2018; Muner et al., 2024; Naglieri et al., 2004; Russell et al., 2003). Portanto, em avaliações computadorizadas, *web-based* ou gamificadas, torna-se fundamental distinguir a variância atribuível ao construto de interesse daquela decorrente das

exigências acessórias do ambiente digital, evitando que limitações de interface ou de familiaridade tecnológica sejam equivocadamente interpretadas como diferenças cognitivas reais.

Em contextos digitais, uma mesma resposta pode refletir tanto o processo cognitivo-alvo quanto demandas acessórias da tarefa. Uma atividade aparentemente voltada ao planejamento, por exemplo, pode ser influenciada pela experiência prévia com jogos, habilidade de navegação, uso do mouse, compreensão da interface, adaptação ao ambiente digital ou ansiedade tecnológica. Evidências recentes indicam que jogadores frequentes e pouco frequentes podem interagir com avaliações cognitivas gamificadas de maneira distinta, especialmente quanto à percepção de domínio, frustração e experiência de uso. Essas diferenças podem afetar a qualidade e a interpretação dos dados, mesmo quando os participantes apresentam esforço e envolvimento semelhantes durante a tarefa. A literatura sobre avaliação psicológica informatizada destaca, portanto, que o uso de plataformas digitais não elimina os desafios psicométricos tradicionais; ao contrário, introduz potenciais novas fontes de erro que precisam ser consideradas na construção e validação dos instrumentos (Huff & Sireci, 2001; Mansur-Alves & Serpa, 2018; Markovitch et al., 2025; Miguel, 2019; Naglieri et al., 2004; Russell et al., 2003). Na lógica da construção de instrumentos, esses fatores correspondem ao risco de variância irrelevante ao construto, isto é, à influência de elementos que afetam o desempenho, mas não pertencem diretamente ao atributo psicológico que se pretende avaliar (Cunha et al., 2023; Serpa et al., 2024).

Além disso, fairness não deve ser reduzida apenas à ausência de diferenças médias entre grupos. Em avaliações digitais e gamificadas, sua investigação deve considerar acessibilidade, oportunidades equivalentes de interação, possíveis efeitos diferenciais da interface, precisão dos escores em subgrupos e adequação das interpretações para populações

com diferentes níveis de experiência tecnológica (Leutner et al., 2023; Markovitch et al., 2025; ITC/ATP, 2025).

Nesse contexto, cresce o interesse por abordagens digitais capazes de registrar dados processuais e aproximar a avaliação psicológica de situações funcionais e ecologicamente mais válidas. Entre essas abordagens, destacam-se as *game-based assessments* (GBAs), que combinam mecânicas interativas, cenários contextualizados e coleta automatizada de indicadores comportamentais ao longo da tarefa, como tempo de resposta, latência, erros, tentativas e padrões de navegação (DiCerbo & Behrens, 2012; Landers, 2015; Mislevy et al., 2014; Rupp et al., 2010; Shute & Ventura, 2013). Esses formatos podem favorecer maior engajamento, reduzir a dependência de autorrelato e ampliar a observação de processos de resposta em contextos mais próximos das demandas cotidianas (Chaytor & Schmitter-Edgecombe, 2003; Parsons, 2015; Serpa & Macedo, 2021; Takieddine & Serpa, 2024).

Diante desse cenário, torna-se relevante mapear como a literatura recente tem tratado a avaliação digital de processos cognitivos em adultos, quais formatos vêm sendo utilizados, que tipos de métricas observáveis têm sido coletadas e quais evidências psicométricas e inferenciais têm sido reportadas.

Assim, este estudo teve como objetivo mapear as abordagens digitais utilizadas na avaliação de processos cognitivos em adultos e descrever seus formatos, tarefas, processos cognitivos investigados, métricas observáveis, evidências psicométricas e desafios inferenciais reportados na literatura. Foram consideradas, com ênfase, *game-based assessments*, tarefas computadorizadas, adaptações para plataformas digitais de instrumentos originalmente desenvolvidos em formatos analógicos e ambientes virtuais.

## **Método**

### **Delineamento do estudo**

Este estudo foi delineado como uma revisão de escopo, com o objetivo de mapear abordagens digitais utilizadas na avaliação de processos cognitivos em adultos e descrever os formatos tecnológicos, as características das tarefas, os processos cognitivos investigados, as métricas observáveis, as evidências psicométricas e os desafios inferenciais reportados nos estudos incluídos. Foram consideradas, com ênfase, game-based assessments, tarefas computadorizadas, adaptações para plataformas digitais de instrumentos originalmente desenvolvidos em formatos analógicos e ambientes virtuais. A escolha por esse delineamento justifica-se pela heterogeneidade do campo investigado, que reúne estudos com diferentes tecnologias, construtos, populações, métricas comportamentais e níveis de evidência psicométrica.

A revisão não teve como finalidade estimar efeitos, comparar intervenções ou produzir síntese quantitativa, mas mapear os formatos de avaliação digital, as características metodológicas dos estudos, os tipos de tarefas, os processos cognitivos investigados, as métricas registradas, as evidências psicométricas e os desafios relacionados à interpretação dos indicadores comportamentais.

A revisão de escopo mostrou-se adequada por permitir a identificação de tendências, categorias metodológicas e lacunas em uma área ainda em consolidação. Nesse sentido, a presente revisão buscou não apenas localizar estudos sobre avaliação cognitiva digital, mas examinar de que modo tais estudos descrevem instrumentos, tarefas, métricas, evidências de validade e relações inferenciais entre comportamento observado e processos cognitivos avaliados.

### **Formulação da pergunta de revisão**

O planejamento e o relato da revisão seguiram as recomendações do Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR), diretriz específica para a apresentação transparente de revisões de escopo (Tricco et al., 2018).

A questão norteadora foi estruturada com base na estratégia PCC, composta por População, Conceito e Contexto. A população foi definida como adultos. O conceito foi definido como a avaliação digital, computadorizada ou gamificada de processos cognitivos, contemplando estudos sobre game-based assessments, tarefas computadorizadas, adaptações para plataformas digitais, ambientes virtuais, métricas processuais e diferentes evidências psicométricas relacionadas à construção, validação e interpretação dos instrumentos. O contexto contemplou os campos psicológico, neuropsicológico, clínico, educacional, ocupacional e da saúde. Com base nessa estrutura, formulou-se a seguinte pergunta de revisão: *Quais abordagens digitais têm sido utilizadas na avaliação de processos cognitivos em adultos e como os estudos descrevem seus formatos, tarefas, processos cognitivos investigados, métricas observáveis, evidências psicométricas e desafios inferenciais?*

### **Fontes de informação e estratégia de busca**

A busca bibliográfica foi conduzida nas bases PubMed, Scopus e SciELO, por sua relevância para as áreas de Psicologia, Neuropsicologia, Saúde, Psicometria e tecnologias digitais aplicadas à avaliação. Foram utilizados descritores e combinações de termos em português e inglês relacionados à população adulta, avaliação cognitiva, avaliação digital, jogos sérios, GBA, processos cognitivos e evidências psicométricas.

As buscas foram realizadas em março de 2026 e atualizadas pela última vez em maio de 2026. Foram considerados todos os registros indexados nas bases até a data da última busca, sem delimitação prévia quanto ao ano de publicação.

Entre os termos utilizados estavam: *adult, adults, cognitive assessment, cognitive processes, digital assessment, computerized assessment, game-based assessment, psychometric evidence, validity evidence, content validity*, avaliação cognitiva, avaliação digital, avaliação baseada em jogos e evidências de validade. As combinações foram adaptadas às especificidades de cada base, utilizando operadores booleanos AND e OR. A busca foi complementada pela identificação manual de um estudo adicional considerado relevante para o escopo da revisão, em razão de sua proximidade conceitual com a construção e validação de avaliação cognitiva baseada em jogo.

### **Critérios de elegibilidade**

Foram incluídos estudos que atenderam simultaneamente aos seguintes critérios: (a) pesquisas empíricas envolvendo participantes adultos, com idade igual ou superior a 18 anos, ou amostras mistas que apresentassem resultados separáveis para adultos; (b) estudos com finalidade avaliativa, voltados à mensuração de processos cognitivos, habilidades cognitivas ou desempenho neuropsicológico; (c) utilização de recursos digitais, computadorizados, gamificados ou imersivos, incluindo game-based assessments, tarefas computadorizadas, adaptações para plataformas digitais, jogos digitais, ambientes de realidade virtual e plataformas remotas de avaliação; (d) apresentação de informações sobre a construção, o funcionamento ou a avaliação psicométrica do instrumento ou tarefa; e (e) relato de pelo menos um dos seguintes componentes: métricas comportamentais ou processuais, evidências de validade, estimativas de precisão, relações com medidas externas, comparação com instrumentos de referência ou discussão inferencial sobre os indicadores produzidos.

Foram excluídos: (a) estudos em que jogos ou recursos digitais eram utilizados exclusivamente para treinamento, reabilitação, intervenção terapêutica ou ensino, sem

finalidade de avaliação ou mensuração; (b) estudos centrados apenas em experiência do usuário, engajamento, satisfação ou usabilidade, quando não havia análise de processos cognitivos ou de propriedades avaliativas do instrumento; (c) estudos sobre personalidade, atitudes, saúde mental, desempenho acadêmico, comportamento organizacional ou outros construtos sem componente cognitivo claramente definido; (d) estudos que empregavam tecnologia apenas para aplicação, registro ou correção de instrumentos tradicionais, sem modificação relevante da tarefa, coleta de métricas digitais ou discussão sobre o formato informatizado; (e) protocolos, editoriais, cartas, resumos sem dados suficientes, revisões teóricas e publicações que não apresentavam resultados empíricos de um instrumento ou tarefa avaliativa; (f) estudos exclusivamente com crianças ou adolescentes, quando os dados de adultos não podiam ser analisados separadamente; e (g) estudos cujo texto completo não permitia identificar com clareza a população, a finalidade avaliativa, os processos cognitivos investigados ou as características do recurso digital.

A distinção entre avaliação e intervenção foi considerada central. Estudos que utilizaram jogos ou tecnologias digitais para promover melhora cognitiva foram incluídos apenas quando apresentaram uma tarefa avaliativa claramente delimitada e resultados relacionados à mensuração de processos cognitivos. Da mesma forma, estudos sobre usabilidade ou experiência de interação foram considerados somente quando essas dimensões estavam vinculadas à qualidade da medida, aos processos de resposta ou à interpretação dos indicadores comportamentais.

### **Seleção dos estudos**

O processo de seleção ocorreu em etapas sucessivas. Inicialmente, foram identificados os registros nas bases consultadas. Em seguida, os estudos foram triados conforme critérios

previamente definidos. Após essa triagem, os textos com maior aderência temática foram avaliados quanto à elegibilidade. A seleção considerou a aderência dos estudos à avaliação cognitiva digital e a existência de dados ou argumentos capazes de sustentar inferências sobre processos cognitivos.

Na triagem por título e resumo, foram verificados principalmente a população, a finalidade avaliativa e a presença de recurso digital relacionado à mensuração cognitiva. Na leitura do texto completo, foram examinados o tipo de tarefa, os processos cognitivos investigados, as métricas registradas, as evidências psicométricas apresentadas e a possibilidade de interpretar os dados produzidos como indicadores de desempenho cognitivo.

A Figura 1 apresenta o fluxograma do processo de identificação, triagem, elegibilidade e inclusão dos estudos. O fluxo de identificação, triagem, elegibilidade e inclusão dos estudos foi organizado conforme as recomendações do PRISMA-ScR e apresentado por meio de fluxograma adaptado do modelo PRISMA 2020 (Page et al., 2021; Tricco et al., 2018).

### **Extração e organização dos dados**

Para cada estudo incluído na síntese qualitativa, foram extraídas as seguintes informações: autoria, ano de publicação, objetivo principal, contexto ou amostra, tecnologia ou formato avaliativo, processos cognitivos investigados, evidências ou achados principais e contribuição do estudo para o presente manuscrito. A extração dos dados buscou identificar de que modo cada estudo contribui para a compreensão da avaliação cognitiva baseada em jogos, realidade virtual ou plataformas digitais. Os dados foram organizados em uma matriz de síntese qualitativa.

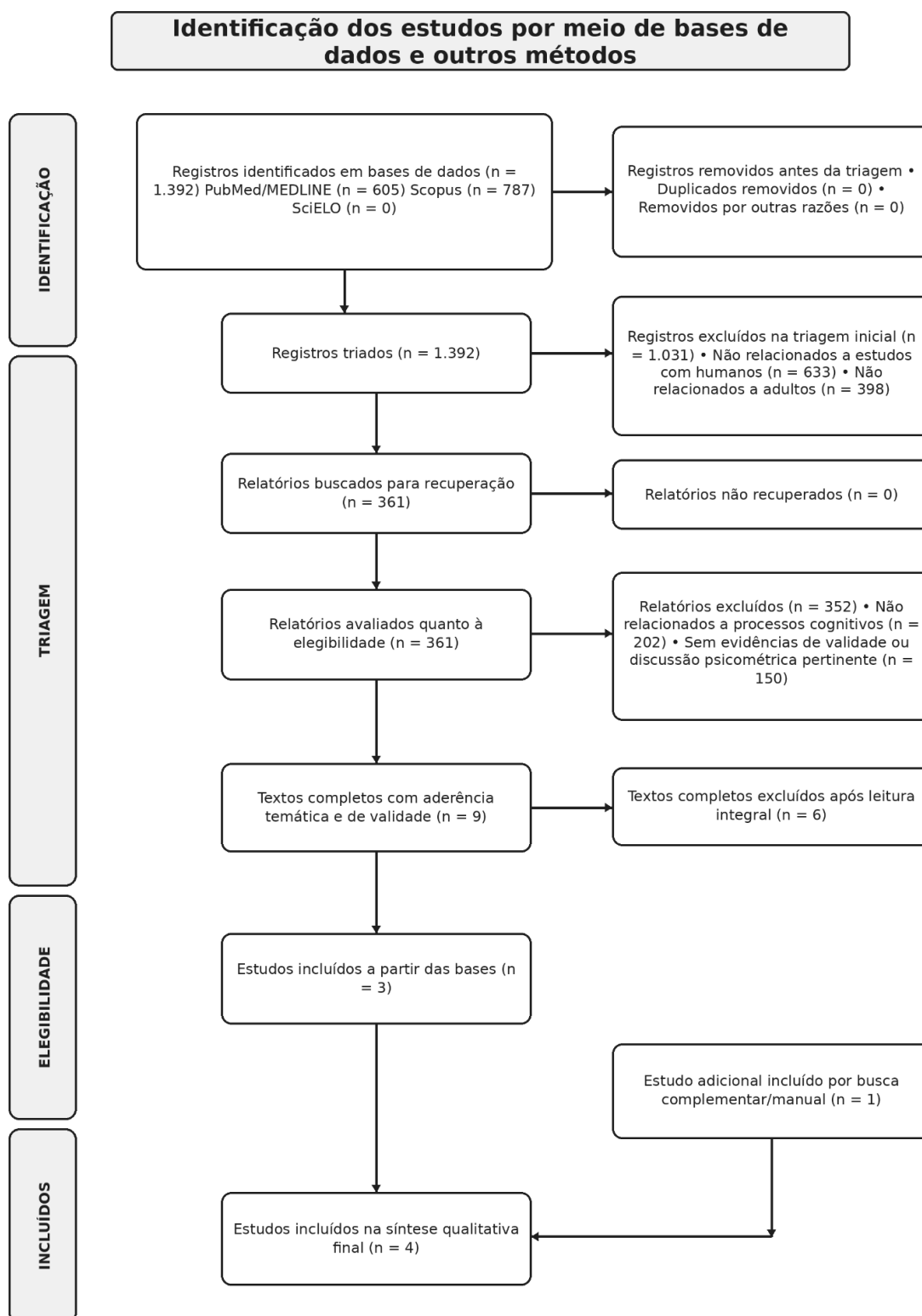
A revisão não utilizou formalmente os critérios do Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) para avaliar a qualidade

metodológica dos estudos ou das propriedades de medida reportadas. O referencial COSMIN foi considerado apenas como orientação conceitual para a discussão sobre evidências psicométricas, sem aplicação de seus checklists, critérios de risco de viés ou sistemas de classificação. Essa decisão decorreu do caráter exploratório da revisão de escopo, da heterogeneidade dos formatos tecnológicos e da diversidade das propriedades psicométricas apresentadas nos estudos incluídos.

As evidências psicométricas foram identificadas e descritas conforme reportadas pelos estudos incluídos. Foram examinados o tipo de evidência apresentada, o procedimento utilizado para obtê-la, a medida de comparação empregada e a interpretação proposta pelos autores. Entretanto, como não foi realizada avaliação formal da qualidade metodológica ou do risco de viés, os achados não foram classificados como evidências fortes, moderadas ou fracas. Nesta revisão, os termos “preliminar”, “parcial” e “promissora” foram utilizados para indicar evidências iniciais que ainda requerem replicação e complementação por outras fontes de validade e precisão.

### **Figura 1**

*Fluxograma PRISMA 2020 adaptado do processo de seleção dos estudos incluídos na revisão de escopo.*



*Nota. O fluxograma apresenta as etapas de identificação, triagem, elegibilidade e inclusão dos estudos. O relato foi orientado pelo PRISMA-ScR (Tricco et al., 2018), e a representação*

*gráfica foi adaptada do modelo PRISMA 2020 (Page et al., 2021). Fonte: elaboração própria.*

Os registros localizados nas bases PubMed, Scopus e SciELO foram submetidos a filtros sucessivos de seleção: estudos com humanos, adultos, processos cognitivos e evidências de validade. Além dos estudos recuperados diretamente nas bases, um artigo adicional foi incluído na busca complementar/manual. Fonte: elaboração própria, adaptada de Page et al. (2021).

## **Resultados**

### **Distribuição dos registros identificados**

Os resultados foram organizados de acordo com os componentes definidos na pergunta de revisão: formatos de avaliação digital, características das tarefas, processos cognitivos investigados, métricas observáveis, evidências psicométricas e desafios inferenciais.

A busca inicial identificou 1.392 registros nas bases consultadas, sendo 605 na PubMed, 787 na Scopus e nenhum registro na Scielo. Após a aplicação dos filtros iniciais, 759 registros estavam relacionados a estudos com humanos e 361 contemplavam a população adulta. Em etapa posterior, 159 registros apresentaram relação com processos cognitivos e nove estudos apresentaram evidências de validade ou discussão psicométrica pertinente. Ao final, três estudos foram incluídos a partir das bases de dados e um estudo adicional foi incorporado por busca complementar/manual, totalizando quatro estudos na síntese qualitativa final.

Após a descrição do processo de seleção, são apresentados individualmente os quatro estudos incluídos. Em seguida, os achados são sintetizados transversalmente quanto aos

formatos tecnológicos, processos cognitivos investigados, tipos de tarefas, métricas observáveis, evidências psicométricas e desafios inferenciais.

### **Estudos incluídos na revisão**

Foram incluídos quatro estudos na síntese qualitativa: Peters et al. (2021), Weiner e Sanchez (2020), Leutner et al. (2023) e Bhargava et al. (2024). Os estudos utilizaram diferentes formatos tecnológicos e contextos de aplicação, incluindo jogo digital em ambiente Minecraft, jogos em realidade virtual, avaliações gamificadas em contexto de recrutamento e jogos tridimensionais para dispositivos móveis.

Peters et al. (2021) desenvolveram e investigaram uma avaliação de inteligência baseada em jogo no ambiente Minecraft. O estudo foi o que apresentou maior proximidade com a concepção de uma game-based assessment cognitiva, por integrar tarefas de jogo e interpretação de desempenho psicológico.

Weiner e Sanchez (2020) examinaram evidências de validade para escores obtidos em jogos de realidade virtual como indicadores de habilidades cognitivas específicas, contribuindo para a análise da relação entre desempenho em ambientes imersivos e medidas cognitivas.

Leutner et al. (2023) investigaram validade, fairness e experiência de uso em avaliações baseadas em jogos aplicadas ao recrutamento. O estudo destacou a necessidade de considerar possíveis diferenças relacionadas à familiaridade com jogos e à experiência do participante durante a avaliação.

Bhargava et al. (2024) compararam jogos em realidade virtual e jogos tridimensionais para dispositivos móveis com o Addenbrooke's Cognitive Examination III. O estudo

contribuiu principalmente para a discussão sobre relações com instrumentos cognitivos de referência e aplicabilidade de jogos digitais na avaliação cognitiva.

As principais características desses estudos são apresentadas na Tabela 1.

**Tabela 1**

*Estudos identificados sobre avaliação cognitiva baseada em jogos, realidade virtual ou plataformas digitais.*

| <b>Estudo</b>        | <b>Objetivo principal</b>                                           | <b>Tecnologia/ Formato</b>          | <b>Processos Cognitivos</b>                                                                                                                                                                                                                                                                                | <b>Evidências psicométricas ou achados reportados</b>                                                       | <b>Como as evidências foram avaliadas</b>                            | <b>Contribuição para esta revisão</b>                                                            |
|----------------------|---------------------------------------------------------------------|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| Peters et al. (2021) | Desenvolver e validar uma avaliação de inteligência baseada em jogo | Jogo digital em ambiente Minecraft. | Peters et al. (2021) desenvolveram e investigaram uma avaliação de inteligência baseada em jogo no ambiente Minecraft. O processo cognitivo central investigado foi a inteligência geral ou habilidade cognitiva geral, operacionalizada por meio do desempenho em tarefas integradas ao ambiente do jogo. | Apresenta proposta de avaliação cognitiva em ambiente gamificado e discute evidências iniciais de validade. | Evidências iniciais e promissoras, mas não suficientes isoladamente. | Aproxima-se do escopo de GBAs cognitivos e contribui para discutir validade em ambiente de jogo. |

|                         |                                                                                                                         |                                                        |                                                                                                                                                                                                                                                                                                                                                                          |                                                                               |                                                                                                               |                                                                                            |
|-------------------------|-------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| Weiner e Sanchez (2020) | Investigar evidências de validade para escores de jogos em realidade virtual como indicadores de habilidades cognitivas | Jogos em realidade virtual                             | Weiner e Sanchez (2020) examinaram evidências de validade para escores obtidos em jogos de realidade virtual como indicadores de habilidades cognitivas específicas. O estudo investigou diferentes habilidades cognitivas por meio do desempenho em tarefas imersivas; entretanto, o texto atual da revisão não discrimina individualmente todos os domínios avaliados. | Explora relações entre desempenho em VR e habilidades cognitivas específicas. | Evidências convergentes preliminares, dependentes da tarefa e do indicador analisado.                         | Contribui para a discussão sobre métricas digitais e validade em ambientes imersivos.      |
| Leutner et al. (2023)   | Examinar validade, justiça e experiência de uso em avaliações baseadas em jogos para habilidade cognitiva.              | Avaliação baseada em jogos no contexto de recrutamento | Leutner et al. (2023) investigaram validade, fairness e experiência de uso em avaliações baseadas em jogos aplicadas ao recrutamento. O construto cognitivo central foi a habilidade cognitiva geral, examinada em associação com o desempenho                                                                                                                           | Discute validade, fairness e experiência do usuário em avaliação gamificada.  | Evidências relevantes, porém específicas ao contexto ocupacional e insuficientes para generalizaçã o clínica. | Amplia a discussão sobre uso aplicado de GBAs e controle de vieses no processo avaliativo. |

|                        |                                                                                           |                                      |                                                                                                                                                                                                                                                                                                                                                        |                                                                               |                                                                                                    |                                                                                                         |
|------------------------|-------------------------------------------------------------------------------------------|--------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|
|                        |                                                                                           |                                      | nas tarefas gamificadas e com possíveis diferenças relacionadas à experiência prévia com jogos.                                                                                                                                                                                                                                                        |                                                                               |                                                                                                    |                                                                                                         |
| Bhargava et al. (2024) | Validar e comparar jogos em realidade virtual e jogos 3D mobile para avaliação cognitiva. | Realidade virtual e jogos 3D mobile. | Bhargava et al. (2024) compararam jogos em realidade virtual e jogos tridimensionais para dispositivos móveis com o Addenbrooke's Cognitive Examination III. Foram contemplados domínios cognitivos amplos avaliados pelo instrumento de referência, incluindo atenção e orientação, memória, fluência verbal, linguagem e habilidades visuoespaciais. | Compara desempenho em jogos digitais com instrumento cognitivo de referência. | Evidência convergente promissora, mas preliminar e limitada pela amostra e pelo desenho do estudo. | Contribui para discutir validade convergente e aplicabilidade de jogos digitais na avaliação cognitiva. |

---

**Nota.** ACE-III = Addenbrooke's Cognitive Examination III; GBA = Game-Based Assessment; VR = Virtual Reality. Os estudos foram selecionados por apresentarem proximidade conceitual ou metodológica com a avaliação cognitiva baseada em jogos, realidade virtual ou plataformas digitais. Fonte: elaboração própria com base em Peters et al. (2021), Weiner e Sanchez (2020), Leutner *et al.* (2023) e Bhargava *et al.* (2024).

### Panorama geral do campo

Os estudos identificados utilizaram diferentes formatos de avaliação digital, incluindo tarefas computadorizadas, jogos digitais, realidade virtual, jogos tridimensionais para dispositivos móveis e avaliações gamificadas em contextos organizacionais ou clínicos. Em vez de um conjunto homogêneo de instrumentos, observa-se um ecossistema composto por diferentes famílias de abordagem: tarefas neuropsicológicas computadorizadas; adaptações para plataformas digitais de testes e tarefas originalmente desenvolvidas em formatos analógicos; plataformas remotas de avaliação; tarefas criadas originalmente em ambiente digital com registro automatizado de desempenho; ambientes de realidade virtual; jogos 3D; e, em menor proporção, propostas de GBA (Alves & Serpa, 2019; Muner et al., 2024; Serpa et al., 2024).

Essa heterogeneidade confirma a pertinência da revisão de escopo como estratégia metodológica, uma vez que o campo ainda não apresenta homogeneidade suficiente para sínteses quantitativas mais robustas. Os estudos localizados indicam que jogos digitais, ambientes de realidade virtual e plataformas gamificadas podem ser utilizados para avaliação de habilidades cognitivas, especialmente quando articulam tarefas estruturadas, coleta de dados processuais e comparação com medidas externas.

### **Síntese temática dos estudos incluídos**

Em conjunto, os quatro estudos indicam que a avaliação digital de processos cognitivos em adultos constitui um campo ainda reduzido e metodologicamente heterogêneo. As diferenças entre os estudos concentram-se nos formatos tecnológicos, nos processos cognitivos investigados, nas características das tarefas, nas métricas produzidas e nas estratégias utilizadas para examinar as propriedades psicométricas. A síntese a seguir organiza esses achados por eixos temáticos, evitando a análise isolada de cada estudo.

Esses achados dialogam com Serpa et al. (2024), que compreendem a psicometria computacional como um campo de integração entre psicologia, estatística e ciência da computação, voltado à modelagem de dados comportamentais complexos. Essa perspectiva é especialmente relevante em GBAs, nas quais os logs de interação podem gerar grande quantidade de variáveis, exigindo procedimentos de seleção, modelagem e validação que preservem a relação entre os indicadores digitais e o construto. Estudos recentes mostram que modelos computacionais e algoritmos de aprendizado de máquina podem melhorar a predição de habilidades e explorar dados processuais, mas não eliminam a necessidade de transparência, replicação, análise de precisão e avaliação da generalização dos resultados entre grupos e contextos (Bipp et al., 2024; Leutner et al., 2023; Markovitch et al., 2025).

A principal lacuna identificada diz respeito à explicitação da cadeia inferencial entre tarefa, indicador comportamental, processo cognitivo e interpretação psicológica. Em muitos estudos, a inovação tecnológica é mais desenvolvida do que o argumento psicométrico que sustenta a interpretação dos dados. Assim, embora métricas como tempo de resposta, erros, tentativas, rotas, cliques e desempenho em tarefas digitais ampliem a granularidade da avaliação, seu valor depende da demonstração de que tais indicadores representam adequadamente os processos cognitivos pretendidos.

Nesse sentido, os achados reforçam a relevância do desenvolvimento de game-based assessments fundamentados em arquiteturas inferenciais explícitas, especialmente quando articulam tarefas contextualizadas, indicadores processuais e estudos progressivos de validade. A proposta apresenta potencial de contribuição ao campo por buscar articular tarefas contextualizadas, indicadores processuais, validade de conteúdo e fundamentação psicométrica. O diferencial metodológico não reside apenas no uso de tecnologia ou de jogos,

mas na capacidade de transformar dados comportamentais digitais em evidências psicologicamente interpretáveis.

Considerando os quatro estudos incluídos, foram investigados processos e habilidades cognitivas em diferentes níveis de abrangência. Peters et al. (2021) e Leutner et al. (2023) concentraram-se predominantemente na inteligência ou habilidade cognitiva geral. Weiner e Sanchez (2020) investigaram habilidades cognitivas específicas por meio de tarefas em realidade virtual, embora os domínios não tenham sido discriminados individualmente nesta revisão. Bhargava et al. (2024) contemplaram atenção e orientação, memória, fluência verbal, linguagem e habilidades visuoespaciais, tomando o ACE-III como medida cognitiva de referência.

### **Processos cognitivos mais frequentemente avaliados**

Nos estudos incluídos, os processos cognitivos foram investigados em diferentes níveis de especificidade. Dois estudos focalizaram predominantemente a inteligência ou habilidade cognitiva geral. Um estudo examinou habilidades cognitivas específicas em ambiente de realidade virtual, sem discriminação detalhada dos domínios na presente síntese. Outro contemplou atenção e orientação, memória, fluência verbal, linguagem e habilidades visuoespaciais, tomando o ACE-III como instrumento cognitivo de referência.

Todavia, a recorrência desses construtos não significa que sua mensuração esteja sempre claramente demonstrada. Em tarefas digitais complexas, uma mesma ação pode depender de atenção visual, compreensão da instrução, estratégia exploratória, familiaridade com ambiente digital e tomada de decisão. Por isso, a inferência sobre processos cognitivos deve ser formulada com cautela, evitando interpretações lineares entre um único indicador e um único construto.

### **Tipos de tarefas digitais e métricas observáveis**

As tarefas variaram de atividades estruturadas, voltadas a processos cognitivos delimitados, a jogos e ambientes virtuais mais contextualizados. As principais métricas reportadas foram acurácia, tempo total, tempo de reação, latência, número de erros, tentativas, sequência de ações, padrões de navegação e desempenho em etapas específicas das tarefas. Em alguns estudos, esses indicadores foram combinados com escores finais ou comparados a medidas cognitivas externas.

Embora as métricas processuais ampliem a descrição do desempenho, sua disponibilidade tecnológica não garante relevância psicológica. O valor interpretativo desses registros depende da explicitação da relação entre as demandas da tarefa, o comportamento observado e o processo cognitivo hipotetizado.

### **Evidências psicométricas e de fidedignidade**

Os estudos incluídos apresentaram diferentes formas de avaliação psicométrica. Foram identificadas análises de relações com instrumentos cognitivos de referência, evidências convergentes, investigações sobre a validade dos escores, estimativas de precisão e análises relacionadas à fairness e à experiência de realização das tarefas. Essas evidências foram avaliadas pelos autores dos estudos por meio de comparações entre medidas, associações estatísticas, análises da estrutura dos escores ou exames de possíveis diferenças entre grupos e condições de uso.

Em conjunto, as evidências podem ser consideradas promissoras, mas ainda parciais e heterogêneas. Elas indicam que o desempenho em jogos digitais pode se relacionar a habilidades cognitivas e a medidas tradicionais de referência, mas não demonstram, de forma suficiente, que os jogos sejam equivalentes a instrumentos cognitivos consolidados ou que

seus escores possam ser utilizados isoladamente para decisões psicológicas, clínicas ou diagnósticas.

Para que o conjunto de evidências fosse considerado mais suficiente, seria necessário acumular resultados provenientes de diferentes fontes, incluindo evidências baseadas no conteúdo, nos processos de resposta, na estrutura interna, nas relações com variáveis externas, na precisão, na estabilidade temporal, na invariância entre grupos e nas consequências da utilização dos escores. Também seriam necessários estudos independentes, amostras mais amplas e diversificadas e replicações em diferentes contextos de aplicação.

Uma síntese meta-analítica recente mostrou que indicadores derivados de avaliações relacionadas a jogos apresentam associação moderada com medidas tradicionais de habilidade cognitiva, mas que essa associação varia em função da forma como o construto e os indicadores digitais são operacionalizados (Bipp et al., 2024). Esse resultado sugere que o desempenho em ambientes gamificados pode compartilhar variância com habilidades cognitivas consolidadas, sem necessariamente constituir uma medida equivalente aos testes convencionais. Assim, estudos de validade convergente devem ser interpretados em conjunto com evidências baseadas no conteúdo, nos processos de resposta, na estrutura interna e na análise das possíveis demandas acessórias introduzidas pelo ambiente de jogo.

A distribuição e o nível de detalhamento das evidências foram desiguais entre os estudos. Alguns apresentaram relações com medidas externas ou análises de validade dos escores, enquanto outros concentraram maior atenção no formato tecnológico, na experiência do usuário ou na aplicabilidade da tarefa. Assim, a revisão permite identificar quais evidências foram reportadas, mas não permite concluir de maneira padronizada sobre sua qualidade metodológica, suficiência ou risco de viés.

### **Lacunas identificadas**

A revisão permitiu identificar três lacunas principais no campo da avaliação digital de processos cognitivos em adultos.

A primeira refere-se ao número ainda reduzido de game-based assessments desenvolvidos especificamente para avaliação cognitiva de adultos. Embora tenham sido encontrados estudos com jogos digitais, realidade virtual e avaliações gamificadas, poucas propostas apresentaram uma arquitetura avaliativa claramente estruturada, com processos cognitivos previamente definidos, tarefas construídas para eliciar comportamentos relevantes e indicadores processuais vinculados a hipóteses interpretativas. Essa lacuna fundamentou o desenvolvimento do Neurotasks como um protótipo de GBA direcionado à avaliação de processos cognitivos em adultos.

A segunda lacuna diz respeito à heterogeneidade dos formatos tecnológicos, tarefas, processos cognitivos e métricas utilizadas. Os estudos empregaram diferentes formas de registro e interpretação do desempenho, sem padronização suficiente para comparar diretamente indicadores como tempo, erros, tentativas, navegação e sequência de ações. Em resposta a essa lacuna, o Neurotasks foi estruturado por meio do Evidence-Centered Design, com organização explícita entre modelo de proficiência, modelo de evidência e modelo de tarefa.

A terceira lacuna refere-se à explicitação insuficiente da cadeia inferencial entre tarefa, indicador comportamental, processo cognitivo e interpretação psicológica. Em parte dos estudos, os registros digitais foram descritos com maior detalhamento do que a justificativa teórica para interpretá-los como evidências dos processos cognitivos pretendidos. Essa lacuna orientou, no segundo manuscrito, a construção de uma arquitetura inferencial preliminar para o Neurotasks, na qual cada tarefa foi relacionada aos processos-alvo, aos

indicadores observáveis, às hipóteses interpretativas e às possíveis ameaças decorrentes de variância irrelevante ao construto.

Também foi observada produção ainda limitada de evidências psicométricas integradas. Os estudos apresentaram evidências promissoras, porém heterogêneas e parciais, com diferenças quanto às relações com medidas externas, estrutura dos escores, precisão, fairness e experiência dos usuários. Diante disso, o segundo estudo concentrou-se em uma etapa inicial do processo de validação, examinando evidências baseadas no conteúdo por meio de juízes especialistas e resultados preliminares relativos aos processos de resposta, à aparência, à aceitabilidade e à experiência de interação dos usuários.

Assim, o desenvolvimento do Neurotasks não decorreu apenas da ausência de um jogo cognitivo, mas da necessidade de responder metodologicamente às lacunas identificadas na revisão: escassez de GBAs cognitivos para adultos, ausência de arquiteturas inferenciais explícitas, heterogeneidade das métricas, insuficiência de evidências psicométricas e atenção limitada às possíveis fontes de variância irrelevante ao construto

## **Discussão**

A formulação dessa pergunta dialoga com a discussão apresentada por Serpa et al. (2024), segundo a qual a psicometria computacional possibilita a integração entre psicologia, estatística e ciência da computação, mas exige clareza sobre a relação entre construto psicológico, indicadores registrados e modelos utilizados para interpretação dos dados. Essa preocupação também é compatível com a discussão de Muner et al. (2024), que destacam os desafios envolvidos na adaptação de testes e tarefas neuropsicológicas para plataformas digitais.

A revisão de escopo sugere que a avaliação digital de processos cognitivos em adultos ocupa uma posição de transição entre duas tradições. De um lado, herda da psicometria e da neuropsicologia clássica a exigência de validade, fidedignidade, operacionalização clara do construto e controle das condições de medida (Hutz et al., 2015, 2018; Malloy-Diniz et al., 2010, 2018). De outro, incorpora possibilidades trazidas por tecnologias digitais, como registro contínuo do desempenho, simulações, cenários interativos e uso de dados processuais (Alves & Serpa, 2019; Muner et al., 2024; Serpa et al., 2024). O encontro entre essas tradições pode favorecer a inovação na avaliação psicológica no Brasil, mas ainda demanda amadurecimento metodológico.

O principal achado desta revisão é que a inovação tecnológica parece avançar mais rapidamente do que o refinamento inferencial. Esse resultado é coerente com a literatura de construção de testes, segundo a qual um novo instrumento não se justifica apenas por seu formato, mas pela necessidade de medir uma variável de modo teoricamente consistente, com vantagens claras em relação a medidas existentes e com critérios definidos para administração, pontuação e interpretação (Hutz et al., 2015). Estudos recentes também mostram que a associação entre indicadores produzidos em jogos e medidas tradicionais de habilidade cognitiva não é uniforme, podendo variar conforme a seleção das métricas, o formato da tarefa e a abrangência do construto avaliado (Bipp et al., 2024). A revisão da literatura e a operacionalização do traço latente em comportamentos observáveis continuam sendo, portanto, exigências centrais, independentemente de o instrumento estar em papel, em tela ou em ambiente gamificado.

Portanto, as evidências encontradas não devem ser interpretadas como confirmação definitiva da validade das avaliações cognitivas baseadas em jogos. Elas oferecem suporte inicial à hipótese de que indicadores produzidos em jogos podem compartilhar variância com habilidades cognitivas, mas a magnitude e o significado dessas relações dependem da tarefa, das métricas selecionadas, da população e do instrumento externo utilizado como referência.

Nesse sentido, o ponto crítico do campo não é a ausência de tecnologia, mas a frequência com que a tecnologia é tratada como se, por si só, garantisse maior precisão psicológica. O simples fato de uma tarefa registrar cliques, tempo, trajetória ou número de tentativas não assegura que tais indicadores representem o construto avaliado. A passagem de dado bruto para evidência psicológica requer teoria, definição operacional e demonstração empírica (Cunha et al., 2023). Quando essa etapa é pouco explicitada, a avaliação corre o risco de acumular dados ricos e interpretações pobres — ou, em linguagem menos elegante, grande volume de dados não necessariamente acompanhado de sustentação inferencial equivalente.

A heterogeneidade observada também merece destaque. O campo reúne desde tarefas computadorizadas bastante delimitadas, nas quais a inferência tende a ser mais clara, até propostas mais imersivas e ecológicas, cujas métricas são abundantes, porém mais difíceis de interpretar. Tarefas altamente controladas tendem a sacrificar contexto funcional em favor de maior pureza inferencial; já tarefas ecológicas e GBAs ampliam a naturalidade da situação e a complexidade do comportamento, mas tornam mais difícil isolar o que exatamente está sendo medido. Não há solução simples para equilibrar essas exigências concorrentes. O ponto-chave é explicitar a arquitetura inferencial adotada.

Embora os princípios do COSMIN sejam pertinentes à análise da qualidade das propriedades de medida, seus critérios não foram aplicados sistematicamente nesta revisão. A heterogeneidade dos estudos incluídos — envolvendo jogos digitais, realidade virtual, plataformas informatizadas e diferentes formas de operacionalização dos construtos — dificultou a utilização uniforme de uma ferramenta originalmente voltada à avaliação estruturada de propriedades de instrumentos de medida. Assim, as evidências psicométricas foram descritas conforme reportadas pelos estudos, sem classificação formal de sua qualidade metodológica.

Outra limitação decorre do fato de que a revisão descreveu as evidências conforme reportadas pelos estudos, sem realizar reanálise estatística ou classificação formal de sua qualidade. Dessa forma, expressões como “promissora” ou “preliminar” indicam a presença de resultados favoráveis iniciais, mas não equivalem a uma avaliação sistemática de evidência forte ou suficiente.

Ao organizar tarefas em cenários cotidianos e registrar indicadores como tempo de execução, sequência de ações, erros, correções, omissões, uso de pistas e navegação, o instrumento se alinha à tendência contemporânea de ampliar a análise processual do desempenho. Entretanto, sua contribuição mais robusta dependerá de quanto conseguir transformar essa coleta em inferência psicologicamente defensável. Em termos metodológicos, isso requer vincular cada métrica a hipóteses claras sobre o construto, discutir variância irrelevante ao construto e acumular evidências progressivas de validade.

As lacunas identificadas oferecem uma base direta para o desenvolvimento apresentado no segundo manuscrito. A reduzida presença de GBAs cognitivos voltados a adultos justificou a proposição do Neurotasks. A heterogeneidade das tarefas e métricas indicou a necessidade de uma estrutura metodológica que organizasse previamente os

processos-alvo, os comportamentos observáveis e as condições de eliciação. A fragilidade da cadeia inferencial motivou a adoção do Evidence-Centered Design, de modo a articular modelo de proficiência, modelo de evidência e modelo de tarefa. Por fim, a insuficiência de evidências psicométricas acumuladas orientou a realização de uma etapa inicial de análise da validade baseada no conteúdo e de investigação preliminar dos processos de resposta, da aparência, da aceitabilidade e da experiência de interação.

Em síntese, a escassez de GBAs cognitivos para adultos levou ao desenvolvimento do protótipo; a ausência de arquiteturas inferenciais explícitas fundamentou a adoção do ECD; a relação pouco clara entre tarefa, indicador e construto orientou a elaboração da matriz inferencial; a heterogeneidade das métricas motivou sua organização sistemática; as possíveis fontes de variância irrelevante foram incorporadas como ameaças à interpretação; e a insuficiência das evidências psicométricas justificou o início do processo de validação por meio da análise de conteúdo e da avaliação preliminar dos usuários.

Esses achados também indicam que o desenvolvimento de instrumentos digitais deve ser acompanhado por estratégias metodológicas capazes de sustentar interpretações. Entre elas, destacam-se: validação de conteúdo por juízes especialistas; estudos sobre a percepção dos usuários quanto à usabilidade, à experiência de interação, à acessibilidade percebida, à aparência e à aceitabilidade do protótipo, bem como sobre possíveis interferências nos processos de resposta; estudos de validade convergente com instrumentos consolidados; avaliação de fidedignidade; investigação de invariância, funcionamento diferencial, precisão dos escores em diferentes subgrupos e fairness da medida, incluindo possíveis diferenças associadas a idade, gênero, escolaridade, deficiência, familiaridade digital e experiência prévia com jogos; análise de variância irrelevante ao construto; e modelagens que integrem escores finais e indicadores processuais.

A fairness deverá ser examinada como parte do argumento de validade e não apenas como comparação descritiva entre grupos. Isso implica verificar se as tarefas oferecem condições adequadas e acessíveis de interação, se os indicadores apresentam significado comparável entre participantes e se os modelos de pontuação mantêm precisão e interpretação equivalentes nos diferentes subgrupos avaliados (ITC/ATP, 2025; Leutner et al., 2023; Markovitch et al., 2025).

Portanto, a avaliação digital de processos cognitivos em adultos já dispõe de uma base suficiente para justificar novas propostas, mas ainda não de um consenso consolidado sobre como integrar, de modo rigoroso, ecologia, interatividade e validade. O avanço do campo dependerá menos da promessa tecnológica em si e mais da qualidade da engenharia psicométrica que organiza a relação entre tarefa, evidência e interpretação.

### **Considerações finais**

Os resultados desta revisão apresentam implicações diretas para o segundo estudo desta dissertação. A escassez de game-based assessments voltados especificamente à avaliação cognitiva de adultos, a heterogeneidade das métricas digitais, a limitada explicitação da cadeia inferencial e a insuficiência de evidências psicométricas integradas fundamentaram o desenvolvimento do Neurotasks. Dessa forma, o segundo manuscrito foi planejado como uma resposta metodológica inicial às lacunas identificadas, e não apenas como aplicação tecnológica dos resultados da revisão.

A revisão indica que a aproximação funcional com situações cotidianas deve ser tratada como uma potencialidade, e não como garantia automática de validade ecológica ou de aplicabilidade clínica. A utilização de cenários cotidianos pode aproximar as tarefas de demandas funcionais, mas também amplia a quantidade de variáveis potencialmente

associadas ao desempenho. Por essa razão, fatores como familiaridade com jogos, compreensão da interface, experiência digital, habilidades perceptomotoras e condições de acessibilidade devem ser considerados no planejamento das tarefas e na interpretação dos resultados.

Em terceiro lugar, os achados sugerem que os indicadores processuais devem ser organizados em hipóteses inferenciais passíveis de investigação empírica. Maior latência inicial, por exemplo, pode ser hipotetizada como indicador de maior demanda de orientação atencional ou planejamento; maior número de interações irrelevantes pode sugerir exploração pouco eficiente; erros de sequência podem estar associados a dificuldades de planejamento ou monitoramento; e correções espontâneas podem representar estratégias de autorregulação. Essas relações, contudo, não devem ser interpretadas como conclusões diretas sobre o funcionamento cognitivo, mas como hipóteses que exigem sustentação teórica e empírica.

Por fim, a construção de um *game-based assessment* poderá contribuir de modo mais consistente para a avaliação de processos cognitivos se articular desenvolvimento tecnológico, fundamentação teórica e um plano progressivo de obtenção de evidências de validade. Esse plano deve contemplar, inicialmente, evidências baseadas no conteúdo e análises preliminares dos processos de resposta, da aparência, da aceitabilidade, da usabilidade e da acessibilidade; em seguida, estudos-piloto com análise das métricas processuais; e, posteriormente, investigações sobre estrutura interna, precisão, relações com variáveis externas, invariância e sensibilidade a diferenças individuais ou entre grupos.

A revisão de escopo indica que a avaliação digital de processos cognitivos em adultos constitui uma área promissora, caracterizada por expansão tecnológica, diversidade metodológica e crescente valorização de dados processuais. Tarefas computadorizadas,

adaptações para plataformas digitais de instrumentos originalmente desenvolvidos em formatos analógicos, ambientes de realidade virtual e game-based assessments ampliam as possibilidades de observação do desempenho e de análise dos processos de resposta, oferecendo caminhos relevantes para a avaliação psicológica e neuropsicológica contemporânea.

Entretanto, os resultados também sugerem que o avanço tecnológico do campo não foi acompanhado, na mesma proporção, pela explicitação das cadeias inferenciais e pelo acúmulo de evidências psicométricas. Persistem lacunas na definição dos construtos, na tradução de processos latentes em indicadores observáveis e na demonstração empírica de que as métricas processuais representam adequadamente os processos cognitivos que se pretende avaliar.

Para pesquisas futuras, os achados reforçam a importância de estudos que articulem teoria cognitiva, desenvolvimento de tarefas, modelagem de evidências e investigação de propriedades psicométricas. Nesse cenário, futuras propostas de *game-based assessment* poderão contribuir para o avanço da avaliação cognitiva digital, desde que sejam fundamentadas em modelos explícitos de relação entre construto, evidência e tarefa e submetidas a estudos empíricos destinados a examinar a sustentação de seus argumentos interpretativos.

De modo geral, as evidências encontradas podem ser caracterizadas como preliminares e promissoras, porém ainda insuficientes para sustentar o uso amplo ou independente das avaliações baseadas em jogos em decisões psicológicas. Os estudos indicam potencial de convergência com habilidades cognitivas e instrumentos de referência, mas o campo ainda demanda replicações, amostras maiores, análises de precisão, estrutura interna, invariância, fairness e investigação das possíveis fontes de variância irrelevante ao construto.

O desenvolvimento inicial do Neurotasks buscou responder a essas lacunas ao propor: (a) um GBA direcionado à avaliação de processos cognitivos em adultos; (b) tarefas contextualizadas organizadas segundo o Evidence-Centered Design; (c) explicitação dos processos-alvo, indicadores observáveis e hipóteses interpretativas; (d) identificação de possíveis fontes de variância irrelevante ao construto; e (e) obtenção inicial de evidências de validade baseadas no conteúdo e de informações preliminares sobre processos de resposta, aparência, aceitabilidade e experiência de interação. Essas respostas são ainda iniciais e não eliminam as lacunas do campo, mas estabelecem um percurso estruturado para sua investigação progressiva.

Conclui-se que o principal desafio da avaliação cognitiva digital não consiste apenas em criar tarefas mais atrativas ou tecnologicamente sofisticadas, mas em desenvolver instrumentos capazes de transformar comportamentos digitais em evidências psicológicas teoricamente justificadas, empiricamente sustentadas e interpretativamente defensáveis. Essa articulação representa uma das principais oportunidades científicas para propostas avaliativas orientadas pela psicometria, pela psicologia cognitiva, pela neuropsicologia e pelo design centrado em evidências.

### **Referências**

Alves, M. M., & Serpa, A. L. O. (2019). Avaliação psicológica por meio das novas tecnologias de informação e comunicação. In M. N. Baptista, M. Muniz, C. T. Reppold, C. H. S. S. Nunes, L. F. Carvalho, R. Primi, A. P. P. Noronha, A. G. Seabra, S. M. Wechsler, C. S. Hutz, & L. Pasquali (Orgs.), *Compêndio de avaliação psicológica* (pp. 160–172). Vozes.

- Bipp, T., Wee, S., Walczok, M., & Hansal, L. (2024). The relationship between game-related assessment and traditional measures of cognitive ability—A meta-analysis. *Journal of Intelligence*, 12(12), Article 129. <https://doi.org/10.3390/jintelligence12120129>
- Bhargava, Y., Kottapalli, A., & Bhat, V. (2024). Validation and comparison of virtual reality and 3D mobile games for cognitive assessment against ACE-III in 82 young participants. *Scientific Reports*, 14, Article 23918. <https://doi.org/10.1038/s41598-024-75065-1>
- Cunha, R. D., Faiad, C., Baptista, M. N., & Cardoso, H. F. (2023). The data generation mechanism: Relationship between constructs and their indicators. *Psicologia: Teoria e Pesquisa*, 39, Article e39nspe08. <https://doi.org/10.1590/0102.3772e39nspe08.en>
- Dias, N. M. (Org.). (2014). *Avaliação neuropsicológica cognitiva* (Vol. 4). Memnon.
- Faiad, C., Baptista, M. N., & Primi, R. (Orgs.). (2021). *Tutoriais em análise de dados aplicados à psicometria*. Vozes.
- Faiad, C., Schelini, P. W., & Argimon, I. I. L. (2024). Avaliação psicológica, tecnologias sociais e inovação. *Recital – Revista de Educação, Ciência e Tecnologia de Almenara/MG*, 5(3), 8–10. <https://doi.org/10.46636/recital.v5i3.520>
- Hutz, C. S., Bandeira, D. R., & Trentini, C. M. (Orgs.). (2015). *Psicometria*. Artmed.
- Hutz, C. S., Bandeira, D. R., & Trentini, C. M. (Orgs.). (2018). *Avaliação psicológica da inteligência e da personalidade*. Artmed.
- Landers, R. N. (2015). An introduction to game-based assessment: Frameworks for the measurement of knowledge, skills, abilities, and other human characteristics using

- behaviors observed within videogames. *International Journal of Gaming and Computer-Mediated Simulations*, 7(4), iv–viii.
- Laurence, P. G., Mecca, T. P., Serpa, A., Martin, R., & Macedo, E. C. (2018). Eye movements and cognitive strategy in a fluid intelligence test: Item type analysis. *Frontiers in Psychology*, 9, Article 380. <https://doi.org/10.3389/fpsyg.2018.00380>
- Leutner, F., Codreanu, S.-C., Brink, S., & Bitsakis, T. (2023). Game-based assessments of cognitive ability in recruitment: Validity, fairness and test-taking experience. *Frontiers in Psychology*, 13, Article 942662. <https://doi.org/10.3389/fpsyg.2022.942662>
- Malloy-Diniz, L. F., Fuentes, D., Leite, W. B., Correa, H., & Bechara, A. (Orgs.). (2010). *Avaliação neuropsicológica*. Artmed.
- Malloy-Diniz, L. F., Fuentes, D., Mattos, P., & Abreu, N. (Orgs.). (2018). *Avaliação neuropsicológica* (2ª ed.). Artmed.
- Miotto, E. C., Lucia, M. C. S., & Scaff, M. (Orgs.). (2018). *Manual de avaliação neuropsicológica* (Vol. 1). Vetor.
- Muner, L. C., Serpa, A. L. O., Martins, G. D., Rodrigues, J. C., & Becker, N. (2024). De analógico a digital: Desafios e estratégias na adaptação de testes e tarefas neuropsicológicas para plataformas digitais. In S. M. B. Assis, L. R. R. Carreiro, & E. H. M. Brum (Orgs.), *Ciências do desenvolvimento humano: Produção de conhecimento e impactos de pesquisas em psicologia, saúde e educação* (pp. 172–186). Editora CESMAC.

- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, Article n71.  
<https://doi.org/10.1136/bmj.n71>
- Peters, H., Kyngdon, A., & Stillwell, D. (2021). Construction and validation of a game-based intelligence assessment in Minecraft. *Computers in Human Behavior*, 119, Article 106701. <https://doi.org/10.1016/j.chb.2021.106701>
- Seabra, A. G. (Org.). (2012). *Avaliação neuropsicológica cognitiva: Atenção e funções executivas* (Vol. 1). Memnon.
- Seabra, A. G. (Org.). (2014a). *Avaliação neuropsicológica cognitiva: Leitura, escrita e aritmética* (Vol. 3). Memnon.
- Seabra, A. G. (Org.). (2014b). *Avaliação neuropsicológica cognitiva: Linguagem oral* (Vol. 2). Memnon.
- Serpa, A. L. O., & Macedo, E. C. (2021). Game-based assessment and intervention for cognitive deficits associated with procrastination behaviour. In *Annals of the ITC 12th Conference on Tests and Testing*.
- Serpa, A. L. O., Rodrigues, M. M., & Maruyama, J. M. (2024). Psicometria computacional: Fundamentos e intersecções entre psicologia, estatística e ciência da computação. In S. M. B. Assis, L. R. R. Carreiro, & E. H. M. Brum (Orgs.), *Ciências do desenvolvimento*

*humano: Produção de conhecimento e impactos de pesquisas em psicologia, saúde e educação* (pp. 559–574). Editora CESMAC.

Serpa, A. L. O., Timoteo, A. P. P., Querino, E. H. G., & Malloy-Diniz, L. F. (2019). Desenvolvimento do teste de planejamento Torre de Londres – versão brasileira (TOL-BR). *Revista Debates em Psiquiatria*, 9(4), 10–19.  
<https://doi.org/10.25118/2763-9037.2019.v9.45>

Souza, V. V., & Faiad, C. (2024). Os testes situacionais de inteligência emocional como um teste adaptativo computadorizado. *Recital – Revista de Educação, Ciência e Tecnologia de Almenara/MG*, 5(3), 105–126. <https://doi.org/10.46636/recital.v5i3.478>

Takieddine, J., & Serpa, A. L. O. (2024). Validação de game-based assessment para avaliação de atenção sustentada em indivíduos com e sem TDAH. In *Anais do III Curso de Inverno em Neurodesenvolvimento e Transtornos Associados e IX Mostra do Programa de Pós-Graduação em Ciências do Desenvolvimento Humano* (p. 40).

Tricco, A. C., Lillie, E., Zarin, W., O’Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., . . . Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–473.  
<https://doi.org/10.7326/M18-0850>

Weiner, E. J., & Sanchez, D. R. (2020). Cognitive ability in virtual reality: Validity evidence for VR game-based assessments. *International Journal of Selection and Assessment*, 28(3), 215–235. <https://doi.org/10.1111/ijsa.12295>

Zanini, D. S., Reppold, C. T., & Faiad, C. (2021). Do lápis e papel à modalidade remota:

Considerações sobre a avaliação psicológica em tempos de pandemia. In K. L. Oliveira,

P. W. Schelini, & S. M. Barroso (Orgs.), *Formação e estratégias de ensino em avaliação psicológica* (pp. 37–51). Vozes.

## Manuscrito 2 – Neurotasks: desenvolvimento e evidências iniciais de validade baseadas no conteúdo segundo o Evidence-Centered Design

### Resumo

A avaliação psicológica de processos cognitivos em adultos demanda instrumentos capazes de sustentar inferências válidas a partir de comportamentos observáveis, sobretudo quando o objetivo é examinar não apenas o produto final do desempenho, mas também indicadores processuais produzidos durante a execução de tarefas. Avaliações digitais e *game-based assessments* ampliam a possibilidade de registrar métricas como tempo de resposta, latência, sequência de ações, número de cliques, tentativas, erros, omissões e padrões de navegação. Contudo, a inovação tecnológica não assegura validade psicométrica, sendo necessário explicitar a relação entre construto, tarefa, evidência e interpretação. Nesse sentido, o presente estudo teve como objetivo descrever o desenvolvimento do Neurotasks, um protótipo de *game-based assessment* voltado à avaliação de processos cognitivos em adultos, fundamentado no *Evidence-Centered Design*, bem como apresentar evidências iniciais de validade baseadas no conteúdo e resultados preliminares relativos aos processos de resposta, à aparência, à aceitabilidade e à experiência de interação dos usuários. Trata-se de um estudo metodológico, exploratório, descritivo e transversal, organizado em quatro etapas: planejamento conceitual e inferencial, desenvolvimento do protótipo digital e apreciação preliminar por juízes especialistas e usuários. O Neurotasks foi estruturado em uma rotina funcional, com tarefas contextualizadas em ambientes cotidianos, como quarto, banheiro, cozinha e escritório. A arquitetura do instrumento articulou modelo de proficiência, modelo de evidência e modelo de tarefa. Participaram sete juízes especialistas e 23 usuários adultos com avaliações completas. Com base no protocolo AvaliaJS, os dados indicaram, entre os juízes, um Índice de Validade de Conteúdo (IVC) geral de 0,82, considerando 147 respostas; um Índice de Validade de Conteúdo da escala calculado pela média dos itens (*Scale-Level Content Validity Index/Average* — S-CVI/Ave) de 0,816 para relevância, 0,810 para clareza e 0,830 para representatividade. Entre os usuários, a proporção geral de respostas favoráveis foi de 0,76, considerando 690 respostas. Esse resultado foi interpretado como indicador preliminar de aparência, aceitabilidade e adequação da experiência de interação, além de fornecer informações sobre possíveis interferências nos processos de resposta. Os resultados indicaram adequação conceitual inicial do protótipo, sobretudo quanto à aproximação funcional com situações cotidianas, legibilidade, tempo de resposta, neutralidade da linguagem e pertinência do jogo para avaliação cognitiva. Foram identificadas fragilidades relacionadas à intuitividade da interface, clareza de mensagens de erro, feedback, acessibilidade, navegação e representação da tela de resultados. O Neurotasks apresenta potencial como instrumento digital em desenvolvimento, mas requer refinamentos operacionais e estudos psicométricos posteriores para fortalecer seu argumento interpretativo.

**Palavras-chave:** avaliação baseada em jogos; avaliação psicológica digital; processos cognitivos; *Evidence-Centered Design*; desenvolvimento de instrumento.

## Abstract

The psychological assessment of cognitive processes in adults requires instruments capable of supporting valid inferences from observable behaviors, especially when the goal is to examine not only final performance outcomes but also process indicators produced during task execution. Digital assessments and game-based assessments expand the possibility of recording metrics such as response time, latency, sequence of actions, number of clicks, attempts, errors, omissions, and navigation patterns. However, technological innovation alone does not ensure psychometric validity; it is necessary to make explicit the relationship among construct, task, evidence, and interpretation. This study aimed to describe the development of Neurotasks, a game-based assessment prototype designed to assess cognitive processes in adults, grounded in Evidence-Centered Design, and to present initial validity evidence based on content, as well as preliminary findings regarding response processes, face-related aspects, acceptability, and user interaction. This is a methodological, exploratory, descriptive, and cross-sectional study organized into four stages: conceptual and inferential planning, digital prototype development, and preliminary appraisal by expert judges and users. Neurotasks was structured as a functional routine with tasks contextualized in everyday environments, such as a bedroom, bathroom, kitchen, and office. The instrument architecture articulated a proficiency model, evidence model, and task model. The 23 complete user evaluations yielded 690 valid responses. Seven expert judges and 23 adult users with complete assessments participated. Among expert judges, the overall Content Validity Index (CVI) was 0.82, based on 147 responses, with S-CVI/Ave values of 0.816 for relevance, 0.810 for clarity, and 0.830 for representativeness. Among users, the overall proportion of favorable responses was 0.76 and was interpreted as a preliminary indicator of acceptability, face-related aspects, interaction adequacy, and potential influences on response processes. Results indicated initial conceptual adequacy of the prototype, particularly regarding perceived functional similarity to everyday situations, readability, response time, neutral language, and perceived suitability for cognitive assessment. However, weaknesses were identified regarding interface intuitiveness, clarity of error messages, feedback, accessibility, navigation, and representation of the results screen. It is concluded that Neurotasks shows potential as a digital instrument under development but requires operational refinements and further psychometric studies to strengthen its interpretive argument.

**Keywords:** game-based assessment; digital psychological assessment; cognitive processes; Evidence-Centered Design; instrument development.

A avaliação psicológica tem passado por profundas transformações em decorrência da incorporação de tecnologias digitais, que possibilitam o registro automatizado de respostas, a coleta de dados processuais e a criação de ambientes interativos de avaliação. Nesse contexto, instrumentos baseados em jogos, simulações e plataformas online vêm sendo progressivamente utilizados para mensurar construtos complexos e investigar desempenho funcional, ampliando as possibilidades tradicionais de mensuração e aplicação remota (Bennett et al., 2025; Ginige et al., 2025; Kleitman et al., 2022). Apesar dessas potencialidades, o desenvolvimento tecnológico deve ser acompanhado por avaliação sistemática da experiência do usuário, da acessibilidade e das propriedades psicométricas dos indicadores registrados.

Uma revisão sistemática recente sobre avaliações cognitivas baseadas em jogos para pessoas idosas mostrou que validade psicométrica, embora essencial, não é suficiente para garantir utilidade prática. Os estudos analisados apresentaram resultados geralmente favoráveis de usabilidade, mas participantes mais velhos e pessoas com comprometimento cognitivo tenderam a perceber algumas ferramentas como menos utilizáveis. A adequação do nível de desafio, a definição precisa da população-alvo e a participação dos usuários no desenvolvimento são, portanto, aspectos relevantes para a aceitabilidade e para a qualidade da interação (Mantell et al., 2025).

A vida adulta abrange diferentes momentos do desenvolvimento, desde a adultez jovem até a velhice, marcados por mudanças nas demandas acadêmicas, ocupacionais, familiares e funcionais. Essas diferenças também se relacionam ao contato com tecnologias digitais. Adultos mais jovens tendem a apresentar maior exposição a jogos, interfaces e dispositivos, enquanto participantes de gerações anteriores podem apresentar trajetórias mais heterogêneas de familiaridade digital. Assim, idade cronológica não deve ser interpretada isoladamente, mas em conjunto com escolaridade, experiência tecnológica, uso prévio de jogos, condições sensoriais e motoras e características neurocognitivas. Em avaliações digitais, essas variáveis podem influenciar a compreensão da interface e os processos de resposta, constituindo possíveis fontes de variância irrelevante ao construto (International Test Commission & Association of Test Publishers [ITC/ATP], 2025; Mantell et al., 2025; Markovitch et al., 2025).

O desenvolvimento de instrumentos para avaliação de processos cognitivos em adultos requer modelos metodológicos capazes de sustentar inferências psicométricas a partir de comportamentos observáveis, assegurando rigor teórico, padronização procedimental e precisão interpretativa. Por se tratarem de construtos latentes, os processos cognitivos não são diretamente observáveis e devem ser inferidos a partir de amostras de comportamento, desempenho ou performance em situações estruturadas de avaliação (Cronbach & Meehl, 1955; Pasquali, 2010). A inferência psicológica tem se apoiado predominantemente em escores finais de desempenho, derivados de tarefas fixas, frequentemente no formato lápis e papel (Cronbach & Meehl, 1955), o que limita a compreensão da dinâmica cognitiva subjacente à execução das tarefas e reduz a sensibilidade da avaliação a aspectos processuais relevantes (Shute & Ventura, 2013; von Davier et al., 2019). Com o rápido desenvolvimento

da tecnologia da informação, grandes mudanças ocorreram no trabalho e na vida das pessoas, apresentando problemas mais complexos de serem aferidos (Griffin et al., 2012).

Na neuropsicologia contemporânea, esses processos cognitivos são compreendidos como um conjunto integrado de operações mentais responsáveis por direcionar, regular e monitorar o comportamento em função de objetivos e demandas contextuais, incluindo atenção, memória de trabalho, planejamento, flexibilidade cognitiva, controle inibitório e monitoramento de ações (Barros & Hazin, 2013; Malloy-Diniz et al., 2018; Santana, 2019). A literatura destaca a interdependência funcional entre esses processos e sua relevância para o desempenho adaptativo em adultos, evidenciando as limitações de abordagens avaliativas centradas em medidas isoladas ou exclusivamente em escores finais. Esses achados reforçam a necessidade de estratégias de avaliação capazes de captar múltiplos domínios cognitivos de forma integrada, sensíveis aos processos subjacentes ao desempenho e coerentes com os avanços metodológicos da avaliação psicológica (Santana et al., 2019; Shute et al., 2015; von Davier et al., 2019).

Nesse cenário, o avanço das tecnologias digitais aplicadas à avaliação psicológica tem ampliado significativamente as possibilidades de mensuração, ao permitir o registro automatizado e contínuo de dados comportamentais e de processo, como tempo de execução, padrões de interação, sequência de ações, erros e estratégias utilizadas durante a realização das tarefas (Shute et al., 2015; von Davier et al., 2019). Esses dados possibilitam deslocar o foco avaliativo do produto final para os processos cognitivos mobilizados ao longo da atividade, favorecendo interpretações mais refinadas, dinâmicas e teoricamente informadas do desempenho observado e fortalecendo a pertinência dos jogos digitais como suporte metodológico para avaliações cognitivas (Kim & Shute, 2015; Shute, 2011; von Davier et al., 2019).

A complexidade inerente a esses ambientes exige modelos metodológicos que tornem explícita a lógica inferencial que sustenta a interpretação dos dados gerados. Nesse sentido, a *game-based assessment* (GBA) tem grande potencial para avaliar construtos que são difíceis de medir por meio de tarefas tradicionais em papel (Buckley et al., 2019; Kapp, 2012; Drasgow, 2015; Georgiou et al., 2017; Mislevy et al., 2015; Wang et al., 2011).

As game-based assessments (GBAs) são avaliações em que tarefas baseadas em jogos são intencionalmente planejadas para eliciar e registrar comportamentos capazes de sustentar inferências sobre conhecimentos, habilidades, competências ou atributos psicológicos. Diferenciam-se de jogos destinados exclusivamente ao entretenimento, ensino, treinamento ou intervenção porque apresentam finalidade avaliativa explícita, critérios de registro e pontuação e uma arquitetura que relaciona construto, tarefa, comportamento observado e interpretação. Sua qualidade depende da produção progressiva de evidências de validade, precisão e fairness (Landers, 2015; Landers & Sanchez, 2022; Mislevy et al., 2015).

As GBAs têm sido utilizadas em contextos educacionais, organizacionais, clínicos, neuropsicológicos e de pesquisa. Podem apoiar a avaliação de aprendizagem, inteligência e habilidade cognitiva geral, atenção, memória, funções executivas, resolução de problemas, competências profissionais e outros atributos. Sua principal contribuição é permitir o registro de dados processuais, como tempo, latência, erros, tentativas, escolhas, sequências de ações e padrões de navegação, complementando os escores finais de desempenho.

Entretanto, a utilização de jogos não assegura automaticamente validade ou equidade. O desempenho pode ser influenciado pela familiaridade com jogos, letramento digital, compreensão das regras, acessibilidade, demandas perceptomotoras e características da interface. Além disso, muitas propostas ainda apresentam dificuldade de calibrar tarefas,

definir regras de pontuação e produzir evidências baseadas na estrutura interna e nas relações com variáveis externas. Por isso, os indicadores gerados devem ser tratados como evidências potenciais até que sua relação com os construtos seja empiricamente demonstrada.

As avaliações baseadas em tecnologia, ou Technology-Based Assessments (TBAs), utilizam recursos computacionais em uma ou mais etapas da construção, apresentação, administração, registro, pontuação, análise ou comunicação dos resultados. Essa categoria abrange testes computadorizados, avaliações adaptativas, simulações, tarefas interativas, avaliações remotas e instrumentos baseados em jogos. As TBAs podem ampliar o registro dos processos de resposta e permitir tarefas que não seriam viáveis em formatos tradicionais. Contudo, introduzem desafios relacionados à comparabilidade entre dispositivos, estabilidade da plataforma, segurança e privacidade dos dados, acessibilidade, funcionamento da interface, pontuação automatizada, transparência e fairness. Por isso, seus escores devem atender às mesmas exigências de validade, precisão e justiça aplicáveis às demais modalidades de avaliação, acrescidas das evidências necessárias para demonstrar que o componente tecnológico não altera indevidamente o construto ou os processos de resposta (Bennett et al., 2025; ITC/ATP, 2025).

Na avaliação psicológica digital, as GBAs podem compor uma agenda de pesquisa voltada a: desenvolvimento de tarefas fundamentadas em modelos cognitivos; análise dos processos de resposta; calibração da dificuldade; investigação da estrutura interna e da precisão; relações com instrumentos consolidados e funcionamento cotidiano; estudos de invariância, acessibilidade e fairness; controle da familiaridade tecnológica; e análise da utilidade clínica e das consequências dos usos dos resultados. Essa agenda exige integração

entre psicologia cognitiva, psicomетria, neuropsicologia, ciência da computação e design de interação.

Para que uma proposta seja tecnicamente caracterizada como game-based assessment, não basta que utilize um jogo, uma narrativa lúdica ou elementos de gamificação. A GBA deve apresentar finalidade explicitamente avaliativa, isto é, ser desenvolvida para produzir informações destinadas à inferência sobre conhecimentos, habilidades, competências, processos cognitivos ou outros atributos psicológicos. O comportamento do participante durante a interação com o jogo constitui a principal fonte de dados, podendo incluir respostas finais, escolhas, sequências de ações, tempos, erros, tentativas, correções, estratégias e padrões de navegação. Esses registros, contudo, somente adquirem significado psicométrico quando são vinculados a hipóteses previamente formuladas sobre os construtos que se pretende avaliar (Landers, 2015; Landers & Sanchez, 2022; Mislevy et al., 2015).

Outro critério técnico refere-se ao desenho intencional da tarefa. Em uma GBA, as regras, os objetivos, os estímulos, as mecânicas, os níveis de dificuldade, as possibilidades de ação e as condições de feedback devem ser planejados para eliciar comportamentos relevantes ao construto. Dessa forma, o jogo não funciona apenas como uma camada estética ou motivacional acrescentada a um teste convencional, mas como parte do próprio mecanismo de geração de evidências. A tarefa deve criar oportunidades controladas para que os processos de interesse se manifestem, ao mesmo tempo em que limita, identifica ou monitora possíveis fontes de variância irrelevante ao construto (DiCerbo & Mislevy, 2017; Kim et al., 2016; Mislevy et al., 2015).

Revisões contemporâneas mostram que o campo das GBAs reúne propostas muito distintas quanto à finalidade, aos construtos, aos contextos de aplicação, às técnicas computacionais e aos modelos de pontuação. Entre os desafios recorrentes estão amostras

reduzidas, ausência de replicação, acesso limitado aos instrumentos e algoritmos, diversidade de métricas e insuficiência de estudos de validação independentes (Gomez et al., 2023).

Nesse contexto, a game-based assessment pode ser definida tecnicamente a partir de cinco componentes articulados: finalidade avaliativa, registro sistemático do comportamento, inferência sobre construtos, desenho intencional das tarefas e produção progressiva de evidências psicométricas. A finalidade avaliativa distingue a GBA de jogos utilizados exclusivamente para entretenimento, ensino, treinamento ou intervenção. O registro comportamental corresponde à coleta estruturada de dados produzidos durante a interação. A inferência sobre construtos exige que esses dados sejam relacionados a atributos psicológicos previamente definidos. O desenho intencional da tarefa implica planejar as mecânicas e condições do jogo para eliciar comportamentos relevantes. Por fim, as interpretações derivadas dos indicadores devem ser sustentadas por evidências de validade, estimativas de precisão e análises das possíveis fontes de erro e de variância irrelevante ao construto (Shute & Kim, 2014; Quiroga et al., 2015).

Esses componentes são interdependentes. Uma tarefa pode registrar grande quantidade de dados sem constituir uma GBA psicometricamente defensável, caso não haja clareza sobre o construto, sobre a função avaliativa das ações exigidas ou sobre a interpretação dos indicadores. Da mesma forma, a presença de narrativa, recompensas, pontuação, avatares ou desafios não transforma automaticamente uma atividade em avaliação baseada em jogo. A caracterização como GBA depende da existência de uma arquitetura avaliativa que conecte aquilo que se pretende inferir, os comportamentos que podem funcionar como evidência e as situações construídas para produzi-los.

Assim, uma GBA deve ser compreendida como um sistema de avaliação e não apenas como um produto tecnológico. Sua qualidade depende da coerência entre construto, tarefa,

comportamento registrado, regra de pontuação e interpretação proposta. Essa coerência precisa ser demonstrada empiricamente por meio de estudos progressivos de validade baseados no conteúdo, nos processos de resposta, na estrutura interna e nas relações com variáveis externas, além de estimativas de precisão, análises de fairness e investigação das consequências dos usos dos resultados (AERA et al., 2014; Kane, 2013; Landers & Sanchez, 2022).

Os avanços tecnológicos possibilitam a construção de ambientes avaliativos mais sensíveis à complexidade do desempenho humano, ao permitir a coleta de evidências não apenas sobre o resultado final da tarefa, mas também sobre os processos mobilizados durante sua execução. (Mislevy et al., 2003). Para a avaliação de construtos complexos, não basta elaborar tarefas formalmente adequadas. É necessário que elas apresentem demandas compatíveis com o construto investigado, de modo a evocar os processos psicológicos relevantes e produzir evidências observáveis alinhadas às inferências avaliativas pretendidas (Messick, 1994).

Projetar instrumentos avaliativos a partir dessa estrutura favorece a coerência entre os objetivos da avaliação, os construtos investigados, as evidências coletadas e os critérios de interpretação dos resultados. Desse modo, a coleta e a análise das evidências tornam-se teoricamente orientadas e metodologicamente alinhadas às inferências que a avaliação pretende sustentar (Almond et al., 2002).

Nesse cenário, as abordagens da GBA permitem integrar elementos lúdicos, narrativos e interativos aos processos avaliativos, favorecendo o engajamento dos participantes e a coleta de indicadores comportamentais e cognitivos. Além disso, possibilitam a extração de dados de processo, como tempo de resposta, padrões de decisão e estratégias utilizadas, aspectos dificilmente capturados por instrumentos tradicionais (DiCerbo, 2017; Shute, 2011).

Entretanto, a expansão das GBAs tem sido acompanhada por questionamentos relacionados à definição dos construtos, à transparência dos modelos de pontuação, à replicabilidade, à influência da experiência prévia com jogos, à experiência do participante e à suficiência das evidências psicométricas. Resultados recentes indicam que relações com testes cognitivos tradicionais tendem a ser positivas, porém variáveis, e que a experiência com jogos pode influenciar tanto a avaliação subjetiva quanto a precisão dos dados produzidos. Por essa razão, a validade de uma GBA deve ser sustentada por múltiplas fontes de evidência, análise dos processos de resposta e investigação de fairness, e não apenas por correlações com medidas externas (Bipp et al., 2024; Markovitch et al., 2025; Ohlms et al., 2024).

A operacionalização desses critérios demanda um modelo metodológico capaz de organizar previamente as inferências pretendidas, os indicadores comportamentais e as características das tarefas. No presente estudo, essa função foi atribuída ao Evidence-Centered Design, por permitir estruturar a GBA como um sistema de argumentação baseada em evidências, e não apenas como uma sequência de atividades digitais.

Apesar de seu potencial, as GBAs apresentam desafios psicométricos específicos. A grande quantidade de dados produzida pelo jogo não assegura que os indicadores representem o construto investigado. Tempos, erros, cliques e sequências podem refletir simultaneamente processos cognitivos, experiência prévia com jogos, familiaridade tecnológica, compreensão das regras e características da interface. Ruídos relacionados à legibilidade, navegação, feedback e estabilidade da plataforma podem introduzir variância irrelevante ao construto. Também é difícil calibrar tarefas que combinam demandas cognitivas, motoras e lúdicas, evitando efeitos de piso, teto, frustração ou abandono. Além disso, muitos estudos apresentam evidências externas limitadas e não investigam suficientemente precisão, invariância e

fairness entre grupos. Desse modo, a validade de uma GBA depende da explicitação da cadeia entre construto, tarefa, comportamento e interpretação, bem como da obtenção progressiva de evidências baseadas no conteúdo, nos processos de resposta, na estrutura interna e nas relações com variáveis externas (AERA et al., 2014; Bipp et al., 2024; ITC/ATP, 2025; Leutner et al., 2023; Markovitch et al., 2025).

Nesse cenário, o Evidence-Centered Design oferece uma estrutura para organizar essa agenda ao explicitar o que se pretende inferir, quais comportamentos podem funcionar como evidências e quais tarefas devem ser construídas para eliciá-los.

O *Evidence-Centered Design* (ECD) configura-se como um dos principais marcos teóricos para o desenvolvimento de avaliações baseadas em evidências, estruturando o processo de construção a partir de argumentos inferenciais claros e integrados (Mislevy et al., 2003, 2004). O ECD organiza o desenvolvimento em modelos interdependentes, que articulam competências latentes, evidências observáveis e tarefas avaliativas.

O *Evidence-Centered Design* (ECD) é uma abordagem de desenvolvimento de avaliações orientada por uma lógica inferencial, na qual a construção do instrumento parte da definição do que se pretende inferir sobre o participante, das evidências observáveis necessárias para sustentar essas inferências e das tarefas capazes de produzir tais evidências. Essa organização favorece a coerência entre construto, atividade avaliativa, dados registrados, critérios de pontuação e interpretação dos resultados, reduzindo o risco de que métricas digitais sejam analisadas de forma direta ou intuitiva, sem sustentação teórica suficiente (Mislevy et al., 2003, 2004, 2006; Mislevy, 2018).

No âmbito do *Conceptual Assessment Framework*, o ECD articula três componentes centrais: o modelo de proficiência, o modelo de evidência e o modelo de tarefa. O modelo de

proficiência delimita os atributos ou processos que se pretende avaliar, como atenção, memória operacional, planejamento, organização, monitoramento da ação e controle executivo. O modelo de evidência especifica os comportamentos observáveis e as métricas que podem subsidiar inferências sobre esses processos, incluindo acurácia, tempo de execução, latência, número de cliques, tentativas, erros, omissões e sequências de ações. Por sua vez, o modelo de tarefa descreve as condições, os estímulos, as regras, as ações possíveis e os níveis de demanda necessários para eliciar os comportamentos previstos.

Em *game-based assessments*, essa articulação é especialmente relevante, pois o ambiente digital pode produzir grande volume de dados comportamentais, sem que todos apresentem significado psicológico imediato. A aplicação do ECD permite planejar as tarefas de modo intencional e explicitar previamente a relação entre processos cognitivos, indicadores observáveis e interpretações propostas. Ainda assim, os registros produzidos pelo sistema devem ser compreendidos como evidências potenciais até que sua relação com os construtos seja examinada empiricamente em estudos de validade, precisão, estrutura interna, relações com variáveis externas e *fairness* (Arieli-Attali et al., 2019; DiCerbo & Mislevy, 2017; von Davier & Mislevy, 2013).

No presente estudo, o ECD foi adotado como estrutura metodológica para orientar o desenvolvimento do Neurotasks, organizando a definição dos processos-alvo, a seleção das métricas comportamentais e a construção das tarefas digitais. Essa escolha buscou assegurar maior transparência à arquitetura avaliativa do protótipo e fundamentar a cadeia inferencial entre construto, tarefa, comportamento registrado e interpretação psicológica.

No contexto brasileiro, ainda são escassos estudos que descrevem, de forma sistemática, como referenciais já consolidados na literatura psicométrica, como o

*Evidence-Centered Design* (ECD) e os *Standards for Educational and Psychological Testing*, podem ser operacionalizados no desenvolvimento de *game-based assessments* (GBAs) voltados à avaliação de processos cognitivos em adultos. Assim, a questão central não se refere à ausência de modelos teóricos para orientar a construção avaliativa, mas à limitada explicitação de como esses modelos são aplicados na prática, especialmente na articulação entre construtos cognitivos, tarefas digitais, indicadores comportamentais e interpretações psicológicas. Em GBAs, essa operacionalização exige atenção adicional às características do ambiente digital, uma vez que aspectos como interface, usabilidade, acessibilidade, clareza das instruções e estabilidade técnica podem interferir na interação do usuário com a tarefa e, conseqüentemente, nos dados observados. Dessa forma, tais elementos não constituem apenas barreiras técnicas, mas condições relevantes para a qualidade das evidências geradas e para a defensabilidade das inferências sobre o funcionamento cognitivo.

Diante dessas lacunas, foi desenvolvido o Neurotasks, um protótipo de GBA fundamentado no Evidence-Centered Design e voltado à investigação de indicadores potenciais de processos cognitivos em adultos. Sua construção buscou articular tarefas contextualizadas, registros comportamentais e hipóteses inferenciais, conforme descrito nas seções metodológicas.

Nesse sentido, o presente estudo teve como objetivo descrever o desenvolvimento do Neurotasks, um protótipo de game-based assessment voltado à avaliação de processos cognitivos em adultos e apresentar evidências iniciais de validade baseadas no conteúdo e resultados preliminares relacionados aos processos de resposta, à aparência, à aceitabilidade e à experiência de interação dos usuários. Nesta etapa, o Neurotasks atende de forma preliminar aos critérios de finalidade avaliativa, desenho intencional das tarefas, registro comportamental e explicitação das hipóteses sobre os construtos. Entretanto, as relações entre

os indicadores processuais e os processos cognitivos ainda não foram testadas empiricamente de modo suficiente. Por essa razão, os registros do sistema são tratados como evidências potenciais, e não como medidas cognitivas validadas ou bases para decisões individuais.

A interpretação das métricas digitais requer cautela, pois os registros produzidos pelo sistema podem refletir tanto os processos cognitivos hipotetizados quanto fatores como familiaridade tecnológica, experiência prévia com jogos, compreensão da interface, habilidades de navegação e condições de acessibilidade. Nesta etapa, esses fatores foram tratados como possíveis ameaças à interpretação e como aspectos a serem considerados em estudos posteriores (ITC/ATP, 2025; Leutner et al., 2023; Markovitch et al., 2025).

Especificamente, buscou-se: (a) explicitar a arquitetura inferencial do instrumento; (b) descrever suas tarefas, processos-alvo e métricas observáveis; (c) analisar a apreciação de juízes especialistas quanto à relevância, clareza e representatividade dos elementos do protótipo; (d) examinar a percepção dos usuários sobre usabilidade, acessibilidade percebida, aparência, aceitabilidade e experiência de interação; (e) identificar possíveis ameaças à interpretação dos indicadores processuais, especialmente aquelas relacionadas à interface, à familiaridade tecnológica e à variância irrelevante ao construto e (f) identificar, de maneira exploratória, fatores individuais e contextuais potencialmente relacionados à interação com o protótipo, especialmente idade, escolaridade, familiaridade com tecnologias digitais, experiência prévia com jogos e condições de neurodivergência, quando essas informações estiverem disponíveis.

O Neurotasks deve ser compreendido como um protótipo em processo de construção. Nesta etapa, foram examinadas a coerência conceitual das tarefas e dos indicadores por meio da apreciação de juízes especialistas e características preliminares da interação com os

usuários. As evidências obtidas são iniciais e baseadas predominantemente no conteúdo; não demonstram ainda estrutura interna, precisão, relações com variáveis externas ou capacidade de sustentar decisões psicológicas individuais.

## **Método**

### **Delineamento**

Trata-se de um estudo metodológico de desenvolvimento de instrumento, com delineamento exploratório, descritivo e transversal. O processo foi organizado em quatro etapas sequenciais: (1) desenvolvimento conceitual e tecnológico do protótipo; (2) documentação das decisões de design e da arquitetura avaliativa; (3) avaliação do protótipo por juízes especialistas; e (4) avaliação do protótipo por usuários adultos. Nas duas primeiras etapas, o Neurotasks constituiu o resultado do processo de construção. Nas etapas seguintes, a versão funcional produzida passou a ser utilizada como instrumento ou objeto de avaliação pelos juízes e usuários. O estudo examinou evidências iniciais de validade baseadas no conteúdo e resultados preliminares referentes aos processos de resposta, à aparência, à aceitabilidade, à usabilidade e à experiência de interação.

O estudo foi conduzido em quatro etapas. A primeira correspondeu ao desenvolvimento do protótipo, incluindo a definição dos processos cognitivos de interesse, dos cenários, das tarefas, das mecânicas de interação, da linguagem, dos estímulos e das métricas comportamentais a serem registradas. A segunda etapa envolveu a documentação do jogo, por meio da elaboração do Game Design Document (GDD) e da organização da arquitetura inferencial fundamentada no Evidence-Centered Design. A terceira etapa consistiu na avaliação da versão funcional por juízes especialistas, que examinaram aspectos técnicos, psicométricos, avaliativos, de design, jogabilidade, acessibilidade e instruções. A quarta etapa correspondeu à avaliação por usuários adultos, que executaram o protótipo e avaliaram

aspectos relacionados à aparência, à aceitabilidade, à usabilidade, à acessibilidade percebida e à experiência de interação.

## **Etapas 1 — Desenvolvimento conceitual e tecnológico do protótipo**

### **Arquitetura inferencial do instrumento**

O desenvolvimento do Neurotasks foi orientado pelo ECD, articulando modelo de proficiência, modelo de evidência e modelo de tarefa. No modelo de proficiência, foram definidos como processos-alvo atenção, memória operacional, planejamento, organização, monitoramento da ação, controle inibitório e resolução de problemas em contexto cotidiano. Esses processos foram selecionados por sua relevância para o funcionamento cognitivo adulto e por sua participação em atividades que exigem manutenção de objetivos, tomada de decisão, sequenciamento, inspeção visual, adaptação e correção de erros.

Os processos cognitivos foram definidos operacionalmente para orientar a construção das tarefas e a formulação das hipóteses inferenciais. A atenção seletiva corresponde à capacidade de direcionar recursos atencionais aos estímulos relevantes, inibindo ou reduzindo a interferência de elementos concorrentes. A atenção sustentada refere-se à manutenção do foco e da prontidão de resposta durante uma atividade ao longo do tempo. A memória operacional corresponde à capacidade de manter e manipular temporariamente informações necessárias à execução de uma tarefa orientada a objetivos.

O planejamento envolve a formulação antecipada de etapas, a seleção de estratégias e a organização de ações necessárias para alcançar uma meta. A organização refere-se à estruturação de informações, objetos ou ações segundo critérios funcionais, temporais ou conceituais. O monitoramento da ação corresponde ao acompanhamento contínuo do próprio

desempenho, incluindo identificação de erros, verificação do progresso e ajuste da estratégia. O controle inibitório envolve a supressão de respostas impulsivas, automáticas ou irrelevantes diante das demandas da tarefa.

A categorização corresponde ao agrupamento de estímulos ou informações com base em atributos compartilhados ou regras de classificação. A resolução de problemas envolve identificar uma discrepância entre o estado atual e a meta, selecionar uma estratégia e implementar ações para reduzir essa discrepância. O raciocínio quantitativo simples refere-se à utilização de relações numéricas básicas em situações estruturadas. A busca visual corresponde à exploração sistemática de um campo de estímulos para localizar um alvo, enquanto a inspeção visual envolve examinar características relevantes dos estímulos para orientar decisões e ações.

Esses processos não são considerados independentes. Em tarefas contextualizadas e funcionalmente complexas, diferentes operações cognitivas podem ser mobilizadas simultaneamente. Por essa razão, foi definido, para cada tarefa, um processo primário, correspondente à demanda central intencionalmente eliciada, e processos secundários, que contribuem para a execução, mas não constituem o foco inferencial principal.

A classificação como processo primário ou secundário foi estabelecida com base na análise cognitiva da tarefa. O processo primário corresponde à operação cuja manifestação foi intencionalmente priorizada pelo desenho da atividade e para a qual os principais indicadores foram definidos. Os processos secundários correspondem às demais operações necessárias à execução, mas que não devem ser interpretadas isoladamente a partir do desempenho. Essa distinção não implica pureza cognitiva das tarefas, mas explicita a hierarquia das hipóteses formuladas durante o desenvolvimento.

No modelo de evidência, foram definidos indicadores observáveis com potencial para subsidiar interpretações futuras: tempo total de execução, latência para início da ação, latência até a primeira ação relevante, número de cliques, número de tentativas, erros de sequência, omissões, correções, consultas a listas ou pistas, acurácia, desistência, aderência a sequências esperadas e padrões de navegação. Esses indicadores foram organizados como métricas temporais, métricas de contagem, métricas proporcionais e índices compostos a serem testados em estudos posteriores.

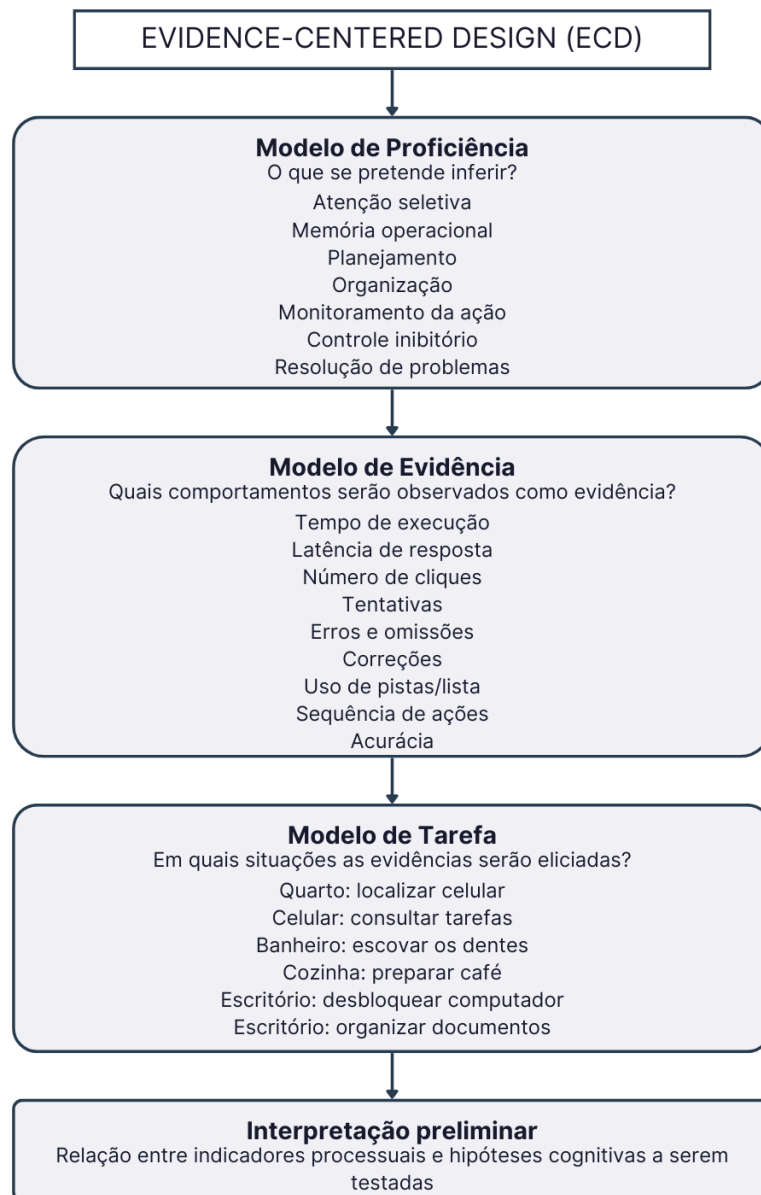
A seleção dos indicadores foi orientada pela literatura sobre avaliações digitais, game-based assessments e análise dos processos de resposta, que destaca a utilidade de métricas temporais, sequenciais, de precisão e de revisão da resposta. O tempo e a latência foram selecionados por permitirem examinar a eficiência temporal, a prontidão para iniciar uma ação e o intervalo necessário para orientação e tomada de decisão. O número de cliques, as tentativas e os erros foram incluídos como medidas de precisão, exploração e custo comportamental para alcançar a meta.

As sequências de ações foram escolhidas por permitirem examinar como o participante organiza o percurso até a solução, em vez de considerar apenas o resultado final. As omissões indicam etapas esperadas que não foram realizadas; as correções permitem observar detecção e reparação de erros; as consultas a listas ou pistas informam sobre uso de apoio externo; e a desistência identifica interrupções que podem decorrer de dificuldade, incompreensão, frustração ou problemas de interface. Esses indicadores foram escolhidos por possibilitarem uma descrição mais granular do desempenho, mas ainda requerem investigação empírica antes de serem interpretados como evidências dos processos cognitivos propostos.

No modelo de tarefa, os processos e indicadores foram operacionalizados em atividades contextualizadas em quatro ambientes cotidianos: quarto, banheiro, cozinha e escritório. A narrativa do jogo apresenta um adulto que acorda atrasado e precisa organizar sua rotina para sair de casa. Essa estrutura foi adotada para integrar as tarefas em uma sequência funcional, evitando que pareçam exercícios isolados. A lógica de construção é apresentada na Figura 1.

**Figura 1**

*Arquitetura inferencial do Neurotasks segundo o Evidence-Centered Design*



*Nota.* A figura representa a lógica de construção do Neurotasks segundo o ECD. Nesta etapa, os indicadores comportamentais são compreendidos como evidências potenciais, cuja relação com os processos cognitivos deverá ser testada empiricamente em estudos posteriores.

### Participantes

Participaram do estudo dois grupos: juízes especialistas e usuários finais. Participaram sete juízes especialistas, cada um responsável pela avaliação dos 21 itens do protocolo, totalizando 147 respostas válidas. O banco de dados utilizado nesta etapa não continha

informações sociodemográficas ou profissionais padronizadas referentes à idade, titulação, tempo de experiência, conhecimento prévio sobre *game-based assessment*, experiência com jogos digitais ou familiaridade tecnológica. As avaliações dos juízes foram registradas entre 29 de janeiro e 19 de abril de 2026. Dessa forma, não foi possível calcular estatísticas descritivas para essas variáveis, o que limita a caracterização detalhada do perfil dos especialistas. Os critérios de elegibilidade incluíram formação superior e experiência relacionada a pelo menos uma das dimensões examinadas no protótipo: construção de instrumentos, avaliação psicológica, processos cognitivos, design de sistemas digitais, usabilidade ou desenvolvimento tecnológico. Após acesso, os juízes puderam executar as tarefas digitais e responder aos critérios avaliativos no próprio ambiente do sistema. Os sete juízes foram previamente cadastrados no sistema e realizaram a avaliação diretamente na plataforma Neurotasks. A plataforma registrava cada sessão de uso como uma execução independente. Por essa razão, foram identificadas 19 execuções associadas aos sete participantes, indicando que alguns juízes acessaram ou executaram o protótipo mais de uma vez. Para as análises de validade de conteúdo, a unidade de análise foi a resposta avaliativa atribuída por cada juiz aos itens do protocolo, totalizando 147 respostas válidas. As 19 execuções não foram tratadas como 19 participantes independentes.

O grupo de usuários foi composto por adultos recrutados por conveniência. Foram adotados como critérios de inclusão: idade igual ou superior a 18 anos, capacidade de compreender as instruções apresentadas em português e disponibilidade de dispositivo com acesso à internet e recursos compatíveis com a plataforma. A familiaridade tecnológica não foi utilizada como critério rígido de exclusão, pois constituiu uma variável relevante para a análise exploratória da interação com o protótipo. Os participantes foram cadastrados na plataforma Neurotasks e receberam acesso ao protótipo em ambiente digital. Após a execução

das tarefas, foram convidados a responder ao instrumento de avaliação de aparência e aceitabilidade, adaptado do Protocolo AvaliaJS. A plataforma registrou automaticamente cada sessão de uso do jogo, de modo que a unidade “execução” correspondeu ao acesso e à interação com o protótipo, enquanto a unidade “avaliação” correspondeu ao preenchimento completo do questionário. Assim, embora 48 usuários tenham sido cadastrados e 44 execuções tenham sido registradas, foram consideradas para análise apenas as 23 avaliações completas, que totalizaram 690 respostas válidas para o cálculo das proporções de respostas favoráveis relativas à aparência e à aceitabilidade.

Embora o cadastro previsse informações sobre idade, escolaridade, familiaridade tecnológica, experiência prévia com jogos e condições neurodivergentes, os registros disponíveis não permitiram produzir estatísticas descritivas completas dessas variáveis. Essa limitação reduz a possibilidade de examinar diferenças individuais relacionadas à interação com o protótipo.

## **Instrumentos**

### ***Protótipo do Neurotasks***

O Neurotasks — Protótipo 1, versão 2025 — foi desenvolvido como uma aplicação web destinada à execução em navegador de internet. A interação foi estruturada predominantemente no formato point-and-click, com uso de mouse, de modo a padronizar a resposta e reduzir demandas motoras relacionadas a controles mais complexos. O protótipo foi organizado em quatro ambientes cotidianos — quarto, banheiro, cozinha e escritório — integrados por uma narrativa na qual um adulto acorda atrasado e precisa concluir tarefas antes de sair para o trabalho.

A aplicação foi programada para apresentar instruções, controlar o fluxo entre as tarefas, registrar eventos produzidos durante a interação e armazenar os dados associados a cada sessão. Os registros planejados incluíram identificação da sessão, horário de início e término, tempo de execução, latência, número de cliques, tentativas, sequência de ações, erros, omissões, consultas, acertos e conclusão das tarefas. Esses dados foram armazenados como logs comportamentais vinculados à execução do participante. Nesta etapa, os logs foram utilizados para documentar o funcionamento da plataforma e o percurso de uso, mas não foram empregados para produzir escores cognitivos individuais ou interpretações diagnósticas.

Como a aplicação foi remota e realizada em dispositivos próprios, não houve padronização integral das condições tecnológicas de execução. Características como tamanho e resolução da tela, escala de exibição do navegador, brilho, contraste, tipo de dispositivo, uso de mouse ou touchpad, sensibilidade do cursor, qualidade do áudio, sistema operacional, navegador e velocidade da conexão podem ter alterado a visibilidade dos estímulos, a precisão dos cliques, a latência registrada e a experiência de navegação. Essas variáveis foram consideradas possíveis fontes de variância irrelevante ao construto e deverão ser registradas e, quando possível, controladas em estudos posteriores.

## **Figura 2**

*Fase 1: Exemplo do ambiente do quarto, em que o jogador precisa procurar o celular.*



### ***Documentação do jogo***

As decisões conceituais, funcionais e tecnológicas foram documentadas em um Game Design Document (GDD). O documento organizou a finalidade do jogo, o público-alvo, a narrativa, os ambientes, as tarefas, os objetivos de cada atividade, as regras, as mecânicas, os estímulos, os objetos interativos, os feedbacks, as transições entre telas e os indicadores comportamentais planejados. Também foram produzidos mockups das telas e descrições dos fluxos de interação, utilizados como referência para a comunicação entre a equipe de pesquisa e a equipe de desenvolvimento.

A documentação funcionou como elemento intermediário entre o modelo conceitual fundamentado no ECD e a implementação tecnológica. Enquanto a arquitetura inferencial especificou o que se pretendia inferir e quais comportamentos poderiam funcionar como evidência, o GDD traduziu essas decisões em cenários, tarefas, regras, telas e eventos passíveis de programação.

### ***Protocolo de avaliação por juízes***

Foi utilizado o instrumento *Neurotasks – Avaliação Técnica, Psicométrica e de Design*, composto por 21 itens distribuídos em cinco domínios: usabilidade e design da interface, jogabilidade e mecânicas, acessibilidade, aspectos psicométricos e avaliativos e instruções. A avaliação dos juízes especialistas foi realizada por meio de um instrumento estruturado, composto por itens avaliados em escala tipo Likert de cinco pontos, distribuídos nos critérios de relevância, clareza e representatividade. Adicionalmente, foram disponibilizados campos abertos para registro de comentários qualitativos e sugestões de aprimoramento. O instrumento foi adaptado com base no protocolo AvaliaJS, cuja versão utilizada neste estudo é apresentada no Apêndice B. O protótipo Neurotasks está disponível em [www.neurotasks.com.br](http://www.neurotasks.com.br).

Além da avaliação quantitativa, os campos abertos permitiram registrar dificuldades observadas, ambiguidades, problemas de interface e sugestões de aprimoramento. O protótipo constituiu o objeto de avaliação, enquanto o protocolo correspondeu ao instrumento de coleta das apreciações dos juízes.

### ***Protocolo de avaliação por usuários***

A experiência dos usuários foi avaliada por meio de protocolo estruturado composto por 30 itens em escala tipo Likert de cinco pontos, organizados nas dimensões de usabilidade e interface, jogabilidade e fluxo da tarefa, acessibilidade percebida, experiência de interação,

aspectos avaliativos e clareza das instruções. O protocolo teve como finalidade examinar a aparência, a aceitabilidade e a experiência de uso do protótipo, além de identificar possíveis elementos da interface ou da tarefa capazes de interferir nos processos de resposta.

O protocolo foi apresentado após a execução do protótipo. Dessa forma, as respostas dos usuários foram formuladas com base na experiência concreta de interação com as tarefas, as instruções, os controles, os feedbacks e o fluxo de navegação. O protocolo não teve como finalidade avaliar tecnicamente a representatividade dos processos cognitivos, mas caracterizar a percepção dos participantes sobre a experiência de uso e identificar fatores potencialmente capazes de interferir nos processos de resposta.

### ***Procedimentos***

#### **Etapa 1 — Desenvolvimento do protótipo**

O desenvolvimento iniciou-se pela delimitação da finalidade avaliativa e dos processos cognitivos de interesse. Em seguida, foram definidas situações cotidianas potencialmente capazes de mobilizar atenção, memória operacional, planejamento, organização, monitoramento da ação, controle inibitório e resolução de problemas. A partir dessa definição, foi construída a narrativa de uma rotina matinal, na qual o participante deveria realizar tarefas nos ambientes do quarto, banheiro, cozinha e escritório.

Os cenários foram selecionados por representarem ambientes familiares à população adulta e permitirem a integração de tarefas de busca, recordação, sequenciamento, organização e resolução de problemas. Para cada tarefa, foram definidos objetivo, estímulos, objetos interativos, regras, ações esperadas, erros possíveis, condições de conclusão, feedbacks e métricas comportamentais. A

linguagem das instruções foi planejada em português, com formulações breves e diretas e busca de neutralidade de gênero.

Após a definição conceitual, foram elaborados os mockups das telas e o fluxo de navegação. A equipe de desenvolvimento implementou os ambientes, objetos clicáveis, transições, instruções, feedbacks visuais e sonoros e mecanismos de registro. A aplicação foi configurada para registrar automaticamente eventos de interação e associá-los a sessões individuais. Foram realizados testes funcionais para verificar carregamento das telas, acionamento dos objetos, progressão entre tarefas, registro das respostas e funcionamento dos formulários.

## Etapa 2 — Documentação do jogo

Paralelamente à implementação, as decisões foram registradas no Game Design Document. O documento foi atualizado ao longo do processo para incorporar alterações nas tarefas, nos cenários, nas mecânicas, nos estímulos e no fluxo de aplicação. Também foi elaborada a matriz inferencial que relacionou processos cognitivos, comportamentos observáveis, tarefas e possíveis ameaças à interpretação.

## Etapa 3 — Avaliação por juízes

Após a disponibilização da versão funcional, os juízes receberam acesso individual à plataforma. Cada participante pôde executar as tarefas e, posteriormente, responder ao protocolo de avaliação técnica, psicométrica e de design no próprio ambiente digital. A participação ocorreu em ambiente remoto, mediante acesso por computador conectado à internet. A aplicação foi assíncrona e realizada de maneira autônoma, sem acompanhamento contínuo do pesquisador.

#### Etapa 4 — Avaliação por usuários

Os usuários foram cadastrados na plataforma e receberam acesso ao protótipo. Inicialmente, visualizaram as orientações de participação e as instruções de uso. Em seguida, executaram as tarefas na ordem definida pelo sistema e, após a interação, responderam ao protocolo de avaliação dos usuários. A aplicação ocorreu remotamente, por meio de navegador de internet, utilizando dispositivo próprio. A execução ocorreu de maneira autônoma e assíncrona.

Durante a aplicação, o sistema registrou automaticamente os acessos, as execuções e a conclusão dos formulários. O pesquisador não interveio durante a execução, permanecendo disponível apenas para dúvidas técnicas.

A versão utilizada não incluiu uma etapa formal de calibração ou treinamento padronizado antes das tarefas avaliativas. Para versões futuras, recomenda-se incluir um pré-treino não pontuado, destinado à familiarização com a mecânica point-and-click, identificação do cursor, seleção de objetos, uso de mouse ou touchpad, controle do som e compreensão dos comandos básicos. Também poderá ser realizada uma verificação inicial de brilho, contraste, escala de tela, resolução, velocidade de carregamento e funcionamento do áudio. Essa etapa deverá utilizar estímulos distintos dos itens avaliativos, de modo a reduzir interferências decorrentes da interface sem treinar diretamente as respostas esperadas.

#### *Análise de dados*

Os dados dos juízes especialistas foram analisados por meio do Índice de Validade de Conteúdo (IVC), calculado pela proporção de respostas com pontuação igual ou superior a 4

em escala de cinco pontos. Foram calculados o Índice de Validade de Conteúdo no nível do item (*Item-Level Content Validity Index* — I-CVI) e o Índice de Validade de Conteúdo da escala pela média (*Scale-Level Content Validity Index/Average* — S-CVI/Ave) para os critérios de relevância, clareza e representatividade. Esses resultados foram interpretados como evidências iniciais de validade baseadas no conteúdo.

Equação 1

$$I-CVI_i = \frac{n_{i,\geq 4}}{N}$$

Em que:

- $n_{i,\geq 4}$  = número de juízes que atribuíram pontuação 4 ou 5 ao item  $i$ ;
- $N$  = número total de juízes.

**Fonte da fórmula:** Polit e Beck (2006) e Polit, Beck e Owen (2007).

Equação 2

$$S-CVI/Ave = \frac{\sum_{i=1}^k I-CVI_i}{k}$$

Em que:

- $I-CVI_i$  = índice de validade de conteúdo do item  $i$ ;
- $k$  = número total de itens avaliados.

**Fonte da fórmula:** Polit e Beck (2006) e Polit et al. (2007).

Equação 3

$$IVC_{\text{geral}} = \frac{\text{número total de avaliações favoráveis dos juízes}}{\text{número total de avaliações realizadas pelos juízes}}$$

Ou, de forma simbólica:

$$IVC_{\text{geral}} = \frac{R_{\geq 4}}{R_T}$$

Em que:

- $R_{\geq 4}$  = total de respostas com pontuação 4 ou 5;
- $R_T$  = total de respostas válidas.

**Fonte metodológica:** Polit e Beck (2006); Polit et al. (2007).

Para os usuários, foi calculada a proporção de respostas favoráveis por item, igualmente considerando como favoráveis as pontuações iguais ou superiores a 4. Esse indicador não foi interpretado como Índice de Validade de Conteúdo, uma vez que os usuários não avaliaram tecnicamente a representatividade dos construtos. As respostas dos usuários foram utilizadas para examinar aspectos de aparência, aceitabilidade, usabilidade, experiência de interação e possíveis interferências nos processos de resposta.

Para cada item avaliado pelos usuários, a proporção de respostas favoráveis foi calculada pela expressão:

$$P_i = \frac{n_{i,\geq 4}}{N_u}$$

Em que:

- $n_{i,\geq 4}$  = número de usuários que atribuíram pontuação 4 ou 5 ao item  $i$ ;
- $N_u$  = número total de usuários com resposta válida ao item.

Para a proporção geral:

$$P_{\text{geral}} = \frac{\sum_{i=1}^k n_{i,\geq 4}}{\sum_{i=1}^k N_{u,i}}$$

Em que:

- $k$  = número total de itens;
- $N_{u,i}$  = número de respostas válidas no item  $i$ .

**Fonte:** elaboração da autora, com base no cálculo de proporções descritivas. Não denominar esse cálculo como IVC.

Inicialmente, foram calculadas estatísticas descritivas para a caracterização dos participantes e do fluxo de participação. Para variáveis contínuas, foram previstas média, desvio-padrão, mediana, valores mínimo e máximo. Para variáveis categóricas, foram calculadas frequências absolutas e relativas. Foram examinadas idade, escolaridade, experiência prévia com jogos, familiaridade tecnológica e autorrelato de condições neurodivergentes, conforme a disponibilidade e a completude dos registros.

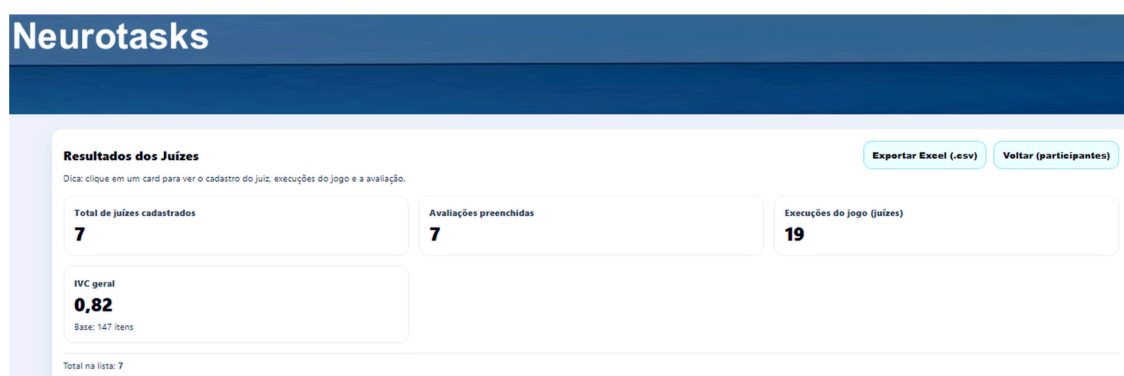
Também foram calculadas frequências referentes às etapas do fluxo amostral: número de pessoas cadastradas, número de participantes com execução registrada, número de avaliações iniciadas, número de avaliações completas e quantidade de respostas válidas.

As respostas abertas foram examinadas qualitativamente, com foco na identificação de padrões recorrentes de dificuldade, sugestões de melhoria, aspectos percebidos como positivos e elementos operacionais potencialmente capazes de introduzir variância irrelevante

ao construto. Em razão do estágio inicial de desenvolvimento do instrumento, todos os resultados foram interpretados como evidências ou indicadores preliminares de adequação.

### Figura 3

*Avaliação de juízes via web.*



### Considerações éticas

Nesta etapa exploratória, os dados analisados concentraram-se nas avaliações fornecidas por juízes e usuários sobre o protótipo, especialmente quanto à validade de conteúdo, aparência, aceitabilidade, usabilidade e experiência de interação. Embora o sistema tenha sido planejado para registrar métricas processuais de desempenho, tais logs não foram utilizados para fins diagnósticos, classificatórios ou inferenciais neste estudo.

### Resultados

Foram registrados 48 cadastros de usuários na plataforma. Desses, 44 apresentaram ao menos uma execução do protótipo, correspondendo a 91,7% dos cadastrados. Quatro participantes cadastrados não apresentaram execução registrada, representando perda inicial de 8,3%. Entre os 44 participantes com execução, 23 concluíram integralmente o protocolo de avaliação dos usuários, correspondendo a 52,3% das execuções e a 47,9% dos cadastros

iniciais. Assim, 21 participantes executaram ou acessaram o protótipo, mas não concluíram o protocolo de avaliação, representando perda de 47,7% após o início da interação.

As análises das respostas dos usuários foram realizadas exclusivamente com as 23 avaliações completas, que produziram 690 respostas válidas. O sistema disponível não permitiu determinar com precisão em qual tarefa ou tela ocorreram todas as interrupções, nem distinguir, em cada caso, abandono voluntário, dificuldade técnica, interrupção externa ou não preenchimento do formulário final. Por essa razão, a perda amostral foi descrita com base nos marcos de cadastro, execução e conclusão da avaliação.

A perda entre a execução registrada e a conclusão do protocolo pode ter introduzido viés de seleção, pois os participantes que completaram todas as etapas podem diferir daqueles que interromperam a participação quanto à familiaridade tecnológica, tolerância à frustração, compreensão das instruções, disponibilidade de tempo ou qualidade do dispositivo e da conexão. Desse modo, a proporção geral de respostas favoráveis pode refletir predominantemente a experiência dos usuários que conseguiram concluir o percurso.

Como o sistema não permitiu identificar de forma uniforme a tarefa, a tela ou o motivo exato da interrupção, não foi possível realizar análise detalhada dos pontos de abandono. Estudos futuros deverão registrar o último evento concluído, a última tela acessada, o tempo de inatividade, mensagens de erro, falhas de carregamento e saída voluntária, permitindo distinguir abandono, interrupção externa, erro técnico e dificuldade de interação.

### **Arquitetura geral do protótipo**

O Neurotasks foi desenvolvido como uma avaliação baseada em jogo centrada em uma rotina funcional. A narrativa apresenta um adulto que acorda atrasado e precisa realizar tarefas antes de sair para o trabalho. Essa estrutura funciona como eixo integrador dos ambientes e das ações, conferindo coerência contextual ao percurso do jogador.

O protótipo foi organizado em quatro ambientes: quarto, banheiro, cozinha e escritório. Em cada ambiente, o participante interage com objetos por meio do mouse, explora o cenário, executa etapas e recebe sinais visuais ou sonoros associados ao progresso. A escolha por interação *point-and-click* buscou padronizar a forma de resposta e reduzir a interferência de demandas motoras mais amplas.

### **Tabela 1**

*Definições operacionais dos processos cognitivos e sua participação nas tarefas do Neurotasks*

| <b>Processo cognitivo</b> | <b>Definição operacional</b>                                                 | <b>Tarefas relacionadas</b>                            | <b>Papel predominante</b>                           |
|---------------------------|------------------------------------------------------------------------------|--------------------------------------------------------|-----------------------------------------------------|
| Atenção seletiva          | Selecionar estímulos relevantes e reduzir interferência de distratores.      | Localizar celular; preparar café; organizar documentos | Primário na busca do celular; secundário nas demais |
| Atenção sustentada        | Manter o foco durante a execução e acompanhar as demandas ao longo do tempo. | Desbloquear computador; organizar documentos           | Secundário                                          |

| <b>Processo cognitivo</b> | <b>Definição operacional</b>                                                 | <b>Tarefas relacionadas</b>                            | <b>Papel predominante</b>                |
|---------------------------|------------------------------------------------------------------------------|--------------------------------------------------------|------------------------------------------|
| Busca visual              | Explorar o ambiente de forma dirigida para localizar um estímulo-alvo.       | Localizar celular                                      | Primário                                 |
| Memória operacional       | Manter e manipular temporariamente informações necessárias à ação.           | Consultar lista; preparar café; desbloquear computador | Primário ou secundário conforme a tarefa |
| Planejamento              | Antecipar etapas e selecionar uma sequência de ações voltada a uma meta.     | Preparar café; rotina geral; organização de tarefas    | Primário                                 |
| Organização               | Estruturar ações ou informações segundo critérios funcionais ou conceituais. | Escovar os dentes; organizar documentos                | Primário                                 |
| Monitoramento da ação     | Acompanhar o desempenho, detectar erros e ajustar a estratégia.              | Todas as tarefas                                       | Secundário transversal                   |

| <b>Processo cognitivo</b>       | <b>Definição operacional</b>                                       | <b>Tarefas relacionadas</b>                                     | <b>Papel predominante</b> |
|---------------------------------|--------------------------------------------------------------------|-----------------------------------------------------------------|---------------------------|
| Controle inibitório             | Evitar respostas impulsivas ou irrelevantes diante de distratores. | Localizar celular; desbloquear computador; organizar documentos | Secundário                |
| Categorização                   | Agrupar informações segundo critérios compartilhados.              | Organizar documentos                                            | Primário                  |
| Resolução de problemas          | Selecionar ações para superar obstáculos e atingir uma meta.       | Preparar café; desbloquear computador                           | Primário                  |
| Raciocínio quantitativo simples | Operar relações numéricas básicas para produzir uma resposta.      | Desbloquear computador                                          | Primário                  |
| Inspeção visual                 | Examinar características dos estímulos relevantes à decisão.       | Organizar documentos; preparar café                             | Secundário                |

*Nota. A atribuição de processos primários e secundários representa uma hipótese teórica de desenvolvimento e não demonstra que cada tarefa mensure isoladamente os processos*

*indicados. A participação relativa dos processos deverá ser examinada em estudos empíricos posteriores.*

### **Tarefas, processos-alvo e métricas observáveis**

A interpretação das métricas foi formulada segundo uma lógica relativa de eficiência e precisão. Em termos preliminares, desempenhos com maior acurácia, menor número de ações irrelevantes, menor repetição de erros e sequência mais coerente com o objetivo foram considerados potencialmente mais eficientes. Entretanto, menor tempo ou menor número de cliques não representa necessariamente melhor funcionamento cognitivo. Respostas muito rápidas podem decorrer de impulsividade, tentativa e erro ou familiaridade prévia com o ambiente digital, enquanto tempos mais longos podem refletir planejamento cauteloso, problemas de interface ou dificuldades de compreensão.

Por essa razão, os indicadores devem ser analisados de forma combinada. O número total de cliques, isoladamente, informa apenas a quantidade de interações. Sua interpretação depende da proporção de cliques relevantes, da presença de cliques repetidos, do percurso realizado, do tempo e da conclusão da tarefa. Uma execução com poucos cliques relevantes e conclusão correta pode representar exploração mais eficiente; muitos cliques irrelevantes, retornos repetidos e ausência de progressão podem sugerir exploração menos direcionada. Essas interpretações permanecem como hipóteses inferenciais preliminares, sem significado diagnóstico nesta etapa.

#### ***Tarefa 1. Minigame “Encontre o celular”***

A primeira tarefa consiste em localizar o celular no quarto. O ambiente foi planejado com múltiplos estímulos visuais e elementos de distração, exigindo busca visual, atenção seletiva e monitoramento da ação. Dessa forma, espera-se que a presente tarefa elicie indicadores comportamentais associados à atenção seletiva, à discriminação visual, ao controle inibitório frente a estímulos distratores e ao monitoramento da ação durante a busca pelo objeto-alvo.

#### **Figura 4**

*Mockup de tela do cenário do quarto da tarefa 1*



**Figura 5**  
*Mockup de tela do cenário cama da tarefa 1*



Os registros incluem tempo até localização do celular, número de cliques, cliques irrelevantes e sequência de exploração.

### ***Tarefa 2. Minigame “Priorize as tarefas escritas no celular”***

A segunda tarefa envolve consultar a lista de atividades no celular. Essa etapa foi planejada para eliciar indicadores relacionados à memória operacional, à manutenção de objetivos e à organização sequencial.

#### **Figura 6**

*Mockup de tela da lista de tarefas no celular da tarefa 2*



Os registros incluem número de consultas à lista, ordem de execução das tarefas e aderência à sequência esperada. A sequência esperada foi estabelecida a partir da análise lógica e funcional de cada tarefa, considerando as ações mínimas necessárias para alcançar a meta, a dependência entre etapas e o critério de eficiência operacional. Quando existiam diferentes percursos capazes de produzir uma resposta correta, não foi pressuposta uma única sequência obrigatória, mas um conjunto de trajetórias funcionalmente aceitáveis. A versão

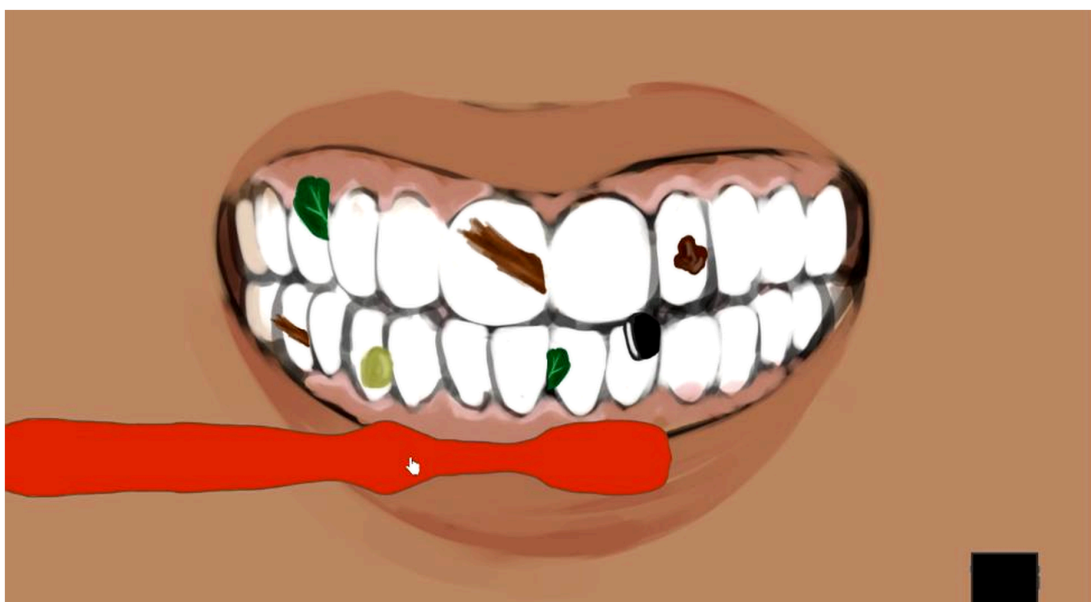
atual não contou com calibração normativa ou estudo-piloto específico para confirmar empiricamente a sequência ótima. Assim, a sequência definida representa uma referência preliminar de desenvolvimento, que deverá ser revisada por especialistas e testada com dados de participantes em estudos posteriores. A definição foi examinada conjuntamente pela equipe de pesquisa e pela equipe de desenvolvimento durante a elaboração do GDD e dos testes funcionais do protótipo.

### ***Tarefa 3. Minigame “Escovar os dentes”***

A terceira tarefa consiste em escovar os dentes. Trata-se de uma rotina procedimental, composta por etapas sucessivas, planejada para observar organização da ação, seguimento de sequência e prontidão para início.

#### **Figura 7**

*Mockup de tela da boca precisando de escovação da tarefa 3*



Os registros incluem latência de início, proporção de passos executados corretamente, tempo total e erros de sequência.

### ***Tarefa 4. Minigame “Fazer o café”***

A quarta tarefa envolve preparar café na cozinha. O participante precisa localizar e manipular componentes da cafeteira, compreender a sequência necessária e concluir a preparação. Essa tarefa foi planejada para eliciar indicadores comportamentais potencialmente relacionados à resolução de problemas práticos, ao planejamento sequencial, à memória operacional e ao monitoramento da ação.

Os registros incluem tempo até conclusão, número de tentativas, erros de sequência, cliques irrelevantes e acertos na execução.

### **Figura 8**

*Mockup de tela da cozinha na tarefa 4*



### **Tarefa 5. Minigame “Desbloquear o computador”**

A quinta tarefa consiste em desbloquear o computador por meio da resolução de operações matemáticas simples. Essa tarefa foi planejada para eliciar indicadores

comportamentais potencialmente relacionados à acurácia, ao controle de resposta, à atenção sustentada e ao monitoramento de erro.

### **Figura 9**

*Mockup de tela do computador da tarefa 5*

**(a) Ambiente de acesso à tarefa**



**(b) Detalhe do computador**



Os registros incluem tempo de resposta, número de tentativas, acertos, erros e correções realizadas durante a tarefa.

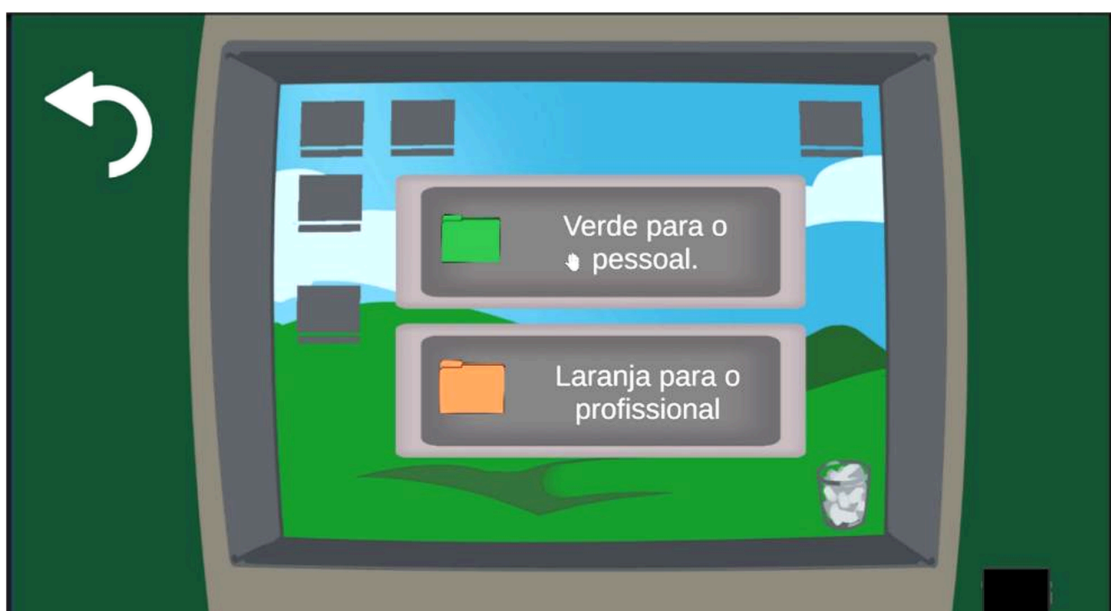
### ***Tarefa 6. Minigame “Organizar as pastas”***

A sexta tarefa envolve organizar documentos em categorias no escritório. O participante deve classificar documentos em grupos temáticos, o que mobiliza categorização, inspeção visual, organização, controle de interferência e monitoramento de erros.

Os registros incluem tempo total, acurácia de classificação, número de tentativas, erros por categoria e desistência. Uma síntese geral das tarefas, os construtos que se pretende medir, os indicadores que se pretende registrar, as hipóteses sobre o desempenho esperado e a identificação das ameaças à interpretação dos resultados é apresentada na Tabela 1.

**Figura 10**

*Mockup de tela da organização de documentos na tarefa 6.*



Todas as tarefas desenvolvidas no presente estudo foram registradas e organizadas no sistema Neurotasks, conforme apresentado na Figura 11. Essa sistematização permitiu centralizar, em ambiente digital, as informações relativas às tarefas, seus objetivos avaliativos, instruções, critérios de execução e parâmetros de análise, favorecendo maior rastreabilidade do processo de construção do instrumento. Além disso, o registro das tarefas no sistema contribui para a padronização da avaliação, para a automatização de etapas da pesquisa e para o controle metodológico das informações utilizadas na análise por juízes e usuários. A aplicação das tarefas ocorre de forma remota e online, o que possibilita maior

flexibilidade na participação dos avaliadores e dos usuários, bem como otimiza o registro automático das respostas e dos indicadores de desempenho. Com isso, o sistema fortalece a organização do banco de dados e a defensabilidade das etapas de desenvolvimento do protótipo.

**Figura 11**  
*Tela inicial do site Neurotasks*

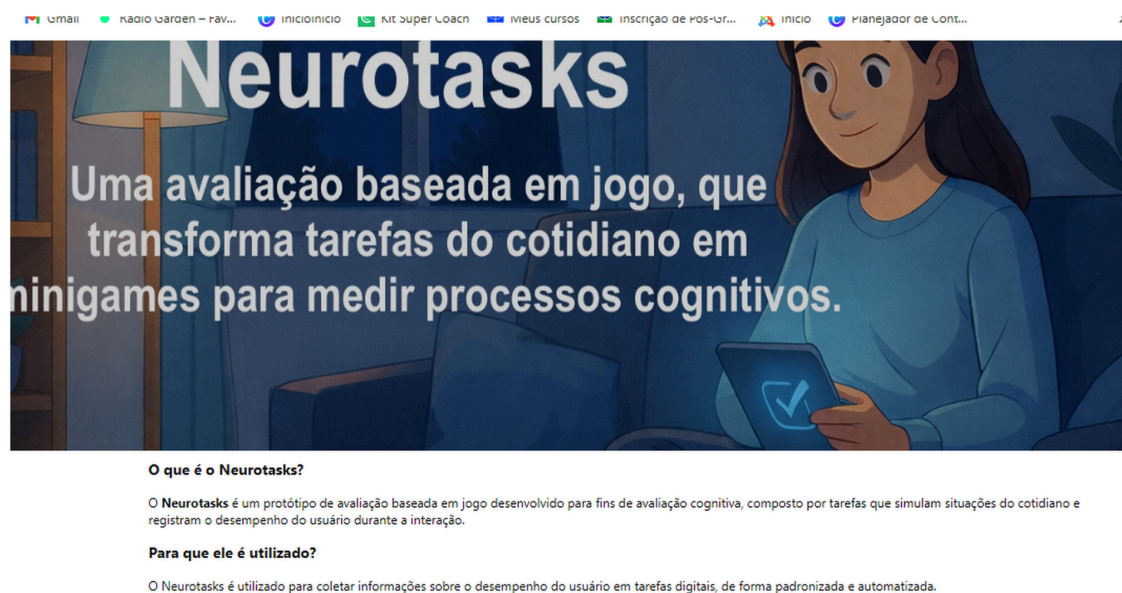


**Tabela 2**

*Arquitetura inferencial preliminar do Neurotasks: tarefas, indicadores observáveis, processos examinados, decisões interpretativas e ameaças à interpretação*

| <b>Tarefa</b>                                  | <b>Processos cognitivos</b>                                                                                                        | <b>Indicadores observáveis</b>                                                                                               | <b>Expectativa de desempenho e hipótese preliminar</b>                                                                                                                                                                                 | <b>Possíveis ameaças à interpretação</b>                                                                                                                                 |
|------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Localizar o celular</b>                     | <b>Primários:</b> atenção seletiva e busca visual.<br><b>Secundários:</b> controle inibitório, manutenção da meta e monitoramento. | Latência até o primeiro clique relevante; tempo total; cliques relevantes e irrelevantes; percurso de exploração; conclusão. | Menor quantidade de interações irrelevantes e busca mais direcionada podem indicar maior eficiência de exploração visual. Latência, cliques e tempo devem ser interpretados conjuntamente, sem significado diagnóstico isolado.        | Tamanho e resolução da tela; contraste; cursor; ambiguidade dos objetos clicáveis; uso de mouse ou touchpad; familiaridade com jogos point-and-click.                    |
| <b>Consultar a lista de tarefas no celular</b> | <b>Primário:</b> memória operacional.<br><b>Secundários:</b> atenção, planejamento, monitoramento e uso de apoio externo.          | Número e duração das consultas; ordem das ações; omissões; retornos à lista; execução das tarefas previstas.                 | Menor necessidade de consultas, associada à execução correta, pode indicar manutenção mais eficiente das informações. Consultas repetidas também podem representar estratégia adaptativa, e não necessariamente dificuldade cognitiva. | Legibilidade; compreensão das instruções; memória prévia da narrativa; tamanho da fonte; dificuldade de navegação; interrupções durante a aplicação.                     |
| <b>Escovar os dentes</b>                       | <b>Primários:</b> organização e sequenciamento.<br><b>Secundários:</b> memória operacional e monitoramento da ação.                | Ordem das etapas; omissões; erros de sequência; repetições; correções espontâneas; tempo de execução.                        | A execução das etapas necessárias em ordem funcional e com poucas repetições pode indicar maior eficiência de organização. Correções podem representar monitoramento e autorregulação.                                                 | Familiaridade com a rotina representada; clareza dos objetos; comandos pouco intuitivos; falhas de feedback; diferenças culturais ou individuais nos hábitos cotidianos. |

|                                 |                                                                                                                                                                                     |                                                                                                                                  |                                                                                                                                                                                                                                                                                |                                                                                                                                                                              |
|---------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Preparar o café</b>          | <p><b>Primários:</b> planejamento e resolução de problemas.</p> <p><b>Secundários:</b> memória operacional, inspeção visual e monitoramento.</p>                                    | <p>Sequência de ações; tentativas; escolha de objetos; erros; correções; tempo; conclusão da tarefa.</p>                         | <p>Uma sequência funcional, com seleção adequada dos objetos e menor número de tentativas desnecessárias, pode representar maior eficiência de planejamento. Não se pressupõe uma única trajetória correta quando diferentes soluções são funcionalmente adequadas.</p>        | <p>Familiaridade com cafeteiras, cápsulas ou utensílios; conhecimento da rotina; ambiguidade dos objetos; feedback insuficiente; diferenças socioeconômicas e culturais.</p> |
| <b>Desbloquear o computador</b> | <p><b>Primário:</b> raciocínio quantitativo e resolução de problemas.</p> <p><b>Secundários:</b> atenção sustentada, memória operacional, controle de resposta e monitoramento.</p> | <p>Tempo para compreender o problema; tentativas; respostas incorretas; revisões; uso de pistas; resposta final.</p>             | <p>Menor número de tentativas acompanhado de resposta correta pode indicar resolução mais eficiente. Respostas rápidas e incorretas podem sugerir tentativa e erro ou baixa verificação, mas essa hipótese necessita de validação empírica.</p>                                | <p>Escolaridade; ansiedade diante de cálculos; compreensão da instrução; dificuldade de leitura; formato de entrada da resposta; familiaridade com teclado e computador.</p> |
| <b>Organizar documentos</b>     | <p><b>Primários:</b> categorização e organização.</p> <p><b>Secundários:</b> atenção seletiva, inspeção visual, memória operacional, controle de interferência e monitoramento.</p> | <p>Acertos de classificação; trocas de categoria; erros; tempo por documento; revisões; sequência de organização; conclusão.</p> | <p>Maior proporção de classificações corretas e menor necessidade de reorganização podem indicar categorização mais consistente. Revisões adequadas podem representar monitoramento, enquanto mudanças repetidas sem progressão podem indicar instabilidade da estratégia.</p> |                                                                                                                                                                              |

---

*Nota. As decisões interpretativas apresentadas são preliminares e correspondem a hipóteses a serem testadas empiricamente. Os indicadores não possuem, nesta etapa, pontos de corte, normas ou capacidade diagnóstica. Resultados desfavoráveis devem ser analisados em*

*conjunto com possíveis fontes de variância irrelevante ao construto, como familiaridade tecnológica, compreensão das instruções, demandas motoras, usabilidade e acessibilidade.*

### **Evidências iniciais de validade de conteúdo, obtidas por juízes especialistas**

Sete juízes especialistas preencheram a avaliação do protótipo. O IVC geral dos juízes foi de 0,82, considerando 147 respostas. O S-CVI/Ave foi de 0,816 para relevância, 0,810 para clareza e 0,830 para representatividade. Esses resultados constituem evidências iniciais de validade baseadas no conteúdo e indicam adequação global preliminar do protótipo quanto à relevância, à clareza e à representatividade dos elementos avaliados.

Os itens com resultados mais robustos concentraram-se nos aspectos psicométricos e avaliativos. O item J16, referente à possibilidade de as tarefas permitirem observar diferenças individuais de desempenho, obteve 100,0% em relevância e clareza e 85,7% em representatividade. O item J15, referente à adequação do jogo para avaliação psicológica, obteve 85,7% em relevância e 100,0% em clareza e representatividade. O item J3, relativo ao aproveitamento do espaço de tela, obteve 85,7% em relevância e 100,0% em clareza e representatividade. Também se destacaram os itens J2 e J5, ambos com 100,0% em representatividade.

Os itens com menores índices indicaram pontos prioritários de refinamento. O item J6, sobre intuitividade dos controles para usuários leigos, obteve 57,1% em relevância, 57,1% em clareza e 71,4% em representatividade. O item J20, referente à adequação das instruções antes do jogo, obteve 71,4% em relevância, 57,1% em clareza e 85,7% em representatividade. Também apresentaram resultados que merecem atenção os itens J13, sobre recursos suficientes para diferentes perfis de usuários, com 71,4% nos três critérios; J14, sobre feedbacks visuais e sonoros, com 85,7% em relevância e 71,4% em clareza e representatividade; J17, sobre aproximação funcional com situações cotidianas, com 85,7%

em relevância e 71,4% em clareza e representatividade; e J19, sobre a tela ou relatório de resultados, com 85,7% em relevância e 71,4% em clareza e representatividade.

A análise individual dos juízes indicou variação importante entre os avaliadores. Os IVCs individuais foram: 0,98; 1,00; 0,21; 0,83; 0,89; 0,89; e 0,94. A presença de um julgamento substancialmente discrepante reduziu o índice global, mas também produziu contribuição qualitativa relevante, ao apontar fragilidades relacionadas à acessibilidade, instruções e clareza dos componentes de interação. Por esse motivo, a avaliação discrepante foi mantida na análise principal, em vez de ser tratada como dado inválido. Em termos metodológicos, essa decisão preserva uma leitura mais conservadora e crítica do protótipo.

**Tabela 2**  
*I-CVI dos juízes por item e critério de avaliação*

| <b>Código</b> | <b>Domínio</b>                    | <b>Item avaliado</b>                                               | <b>Relevância</b> | <b>Clareza</b> |
|---------------|-----------------------------------|--------------------------------------------------------------------|-------------------|----------------|
| J1            | Usabilidade e Design da Interface | O layout das telas é organizado e funcional.                       | 85,7%             | 85,7%          |
| J2            | Usabilidade e Design da Interface | As cores utilizadas são adequadas ao propósito avaliativo do jogo. | 85,7%             | 85,7%          |
| J3            | Usabilidade e Design da Interface | Há bom aproveitamento do espaço da tela.                           | 85,7%             | 100,0%         |
| J4            | Usabilidade e Design da Interface | Os botões e menus funcionaram corretamente.                        | 85,7%             | 71,4%          |
| J5            | Usabilidade e Design da Interface | A navegação é consistente e previsível.                            | 85,7%             | 85,7%          |
| J6            | Usabilidade e Design da Interface | Os controles são intuitivos para usuários leigos.                  | 57,1%             | 57,1%          |
| J7            | Jogabilidade e Mecânicas          | As regras das tarefas são claras e bem definidas.                  | 85,7%             | 85,7%          |
| J8            | Jogabilidade e Mecânicas          | O fluxo das tarefas é coerente com os objetivos do jogo.           | 85,7%             | 85,7%          |

|     |                                      |                                                                                    |        |        |
|-----|--------------------------------------|------------------------------------------------------------------------------------|--------|--------|
| J9  | Jogabilidade e Mecânicas             | O tempo de resposta exigido é adequado às demandas cognitivas.                     | 85,7%  | 71,4%  |
| J10 | Jogabilidade e Mecânicas             | As mecânicas do jogo estão bem integradas às tarefas avaliativas.                  | 85,7%  | 85,7%  |
| J11 | Acessibilidade                       | A linguagem utilizada é clara e apropriada ao público-alvo.                        | 85,7%  | 85,7%  |
| J12 | Acessibilidade                       | O tamanho das fontes é adequado.                                                   | 71,4%  | 85,7%  |
| J13 | Acessibilidade                       | O jogo apresenta recursos suficientes para diferentes perfis de usuários.          | 71,4%  | 71,4%  |
| J14 | Acessibilidade                       | Os feedbacks visuais e sonoros são adequados e não excessivos.                     | 85,7%  | 71,4%  |
| J15 | Aspectos Psicométricos e Avaliativos | O jogo é adequado para avaliação psicológica.                                      | 85,7%  | 100,0% |
| J16 | Aspectos Psicométricos e Avaliativos | As tarefas permitem observar diferenças individuais de desempenho.                 | 100,0% | 100,0% |
| J17 | Aspectos Psicométricos e Avaliativos | As situações do jogo apresentam validade ecológica.                                | 85,7%  | 71,4%  |
| J18 | Aspectos Psicométricos e Avaliativos | O feedback não interfere na mensuração do desempenho cognitivo.                    | 71,4%  | 85,7%  |
| J19 | Aspectos Psicométricos e Avaliativos | A tela/relatório de resultados representa adequadamente o desempenho do avaliado.  | 85,7%  | 71,4%  |
| J20 | Instruções                           | As instruções antes do jogo foram adequadas para o desempenho no jogo.             | 71,4%  | 57,1%  |
| J21 | Instruções                           | A linguagem utilizada nas instruções do jogo é neutra do ponto de vista de gênero. | 71,4%  | 85,7%  |

Nota. I-CVI = Índice de Validade de Conteúdo no nível do item; S-CVI/Ave = Índice de Validade de Conteúdo da escala calculado pela média dos itens. Foram consideradas respostas favoráveis aquelas com pontuação igual ou superior a 4 em escala de cinco pontos. O

S-CVI/Ave foi de 81,6% para relevância, 81,0% para clareza e 83,0% para representatividade, com 147 respostas consideradas.

### **Resultados relativos aos processos de resposta, à aparência e à aceitabilidade entre usuários**

Considerando o fluxo amostral anteriormente descrito, as análises da avaliação dos usuários foram realizadas com as 23 avaliações completas, totalizando 690 respostas válidas. A proporção geral de respostas favoráveis dos usuários foi de 0,76, calculada com base em 690 respostas. Esse resultado indica aceitabilidade inicial e percepção global favorável sobre determinados aspectos da aparência e da experiência de interação. Entretanto, não deve ser interpretado como Índice de Validade de Conteúdo, pois os usuários avaliaram a experiência percebida, a clareza, a usabilidade e a plausibilidade da proposta, e não a representatividade técnica dos construtos cognitivos. Os resultados também fornecem informações preliminares sobre os processos de resposta, especialmente quanto a possíveis interferências da interface, das instruções, dos feedbacks e da navegação.

Os melhores índices ocorreram nos itens U12, referente à adequação do tempo disponível para responder às tarefas, com 95,7%; U14, referente ao tamanho das letras e fonte, com 91,3%; e U28, referente ao sentido das tarefas em relação a situações do cotidiano, também com 91,3%. Também apresentaram índices elevados os itens U11, sobre informações suficientes para começar a jogar, e U30, sobre neutralidade de gênero na linguagem das instruções, ambos com 87,0%. Os itens U1, U17, U18, U21, U22 e U24 obtiveram 82,6%, indicando percepção favorável sobre organização visual, clareza da linguagem, manutenção da atenção, esforço para desempenho, conforto e adequação do jogo para avaliar habilidades cognitivas.

Esses resultados indicam que o protótipo do Neurotasks foi percebido como ecologicamente plausível, legível, compreensível em sua proposta geral e potencialmente adequado à avaliação de habilidades cognitivas. A força desse achado está na experiência concreta de uso: os participantes compreenderam que as tarefas faziam sentido no cotidiano e perceberam o jogo como uma proposta avaliativa plausível.

Entretanto, alguns itens apresentaram índices abaixo do desejável. O menor resultado ocorreu no item U3, referente ao quanto os elementos da interface ajudaram a realizar as tarefas com pouco esforço, com 43,5%. O item U8, sobre clareza das mensagens e sinais quando algo não deu certo, obteve 47,8%. O item U16, relativo ao fato de os sons não atrapalharem a concentração, obteve 60,9%. O item U6, sobre clareza e lógica da navegação entre telas, obteve 65,2%. Os itens U4, U26 e U27 obtiveram 69,6%, indicando fragilidades na facilidade de uso dos controles, compreensão do feedback e representação do desempenho na tela de resultados.

As respostas abertas convergiram com os resultados quantitativos. Entre os pontos positivos, os usuários destacaram inovação, familiaridade com jogos digitais, estética amigável, tarefas do cotidiano, engajamento e possibilidade de observar aspectos cognitivos durante a execução. Entre os pontos críticos, foram recorrentes comentários sobre dificuldade para identificar locais de interação, tamanho do ponteiro, ausência de pistas em momentos de erro, necessidade de sinais de conclusão de tarefa, dificuldade na tarefa do café, pouca legibilidade dos documentos, ambiguidade na organização das pastas, problemas em tela cheia e necessidade de recursos de acessibilidade sonora e visual.

**Tabela 3**

*Proporção de respostas favoráveis dos usuários sobre aparência, aceitabilidade, usabilidade e experiência de interação com o Neurotasks*

| Item | Conteúdo avaliado                                                     | Proporção favorável |
|------|-----------------------------------------------------------------------|---------------------|
| U1   | Visual organizado e consistente ao longo das telas.                   | 82,6%               |
| U2   | Jogo visualmente agradável.                                           | 78,3%               |
| U3   | Elementos da interface ajudaram a realizar tarefas com pouco esforço. | 43,5%               |
| U4   | Controles fáceis de entender e usar.                                  | 69,6%               |
| U5   | Informações em tela suficientes para jogar sem ajuda externa.         | 73,9%               |
| U6   | Navegação entre telas clara e lógica.                                 | 65,2%               |
| U7   | Compreensão sobre atingir ou não os objetivos das tarefas.            | 73,9%               |
| U8   | Mensagens e sinais claros quando algo não deu certo.                  | 47,8%               |
| U9   | Compreensão rápida do que precisava fazer no jogo.                    | 78,3%               |
| U10  | Regras das tarefas claras.                                            | 78,3%               |
| U11  | Informações suficientes para começar a jogar.                         | 87,0%               |
| U12  | Tempo disponível adequado para responder às tarefas.                  | 95,7%               |
| U13  | Ritmo do jogo favoreceu concentração.                                 | 73,9%               |
| U14  | Tamanho das letras e fonte facilitaram a leitura.                     | 91,3%               |
| U15  | Alertas visuais adequados.                                            | 78,3%               |
| U16  | Sons não atrapalharam concentração.                                   | 60,9%               |
| U17  | Linguagem clara e fácil de entender.                                  | 82,6%               |
| U18  | Jogo manteve atenção durante a interação.                             | 82,6%               |

|     |                                                           |       |
|-----|-----------------------------------------------------------|-------|
| U19 | Jogo motivador para concluir tarefas.                     | 73,9% |
| U20 | Nível de desafio adequado.                                | 73,9% |
| U21 | Esforço para bom desempenho nas tarefas.                  | 82,6% |
| U22 | Conforto durante a experiência.                           | 82,6% |
| U23 | Jogaria novamente.                                        | 73,9% |
| U24 | Jogo parece adequado para avaliar habilidades cognitivas. | 82,6% |
| U25 | Objetivos das tarefas claros.                             | 73,9% |
| U26 | Feedback ajudou a compreender o desempenho.               | 69,6% |
| U27 | Tela de resultados representa bem o desempenho.           | 69,6% |
| U28 | Conteúdo das tarefas fez sentido em relação ao cotidiano. | 91,3% |
| U29 | Instruções antes do jogo adequadas.                       | 78,3% |
| U30 | Linguagem das instruções neutra quanto a gênero.          | 87,0% |

---

Nota. Foram consideradas respostas favoráveis aquelas com pontuação igual ou superior a 4 em escala de cinco pontos. A proporção geral de respostas favoráveis dos usuários foi de 0,76. Esse indicador expressa a frequência de avaliações positivas sobre aparência, aceitabilidade, usabilidade e experiência de interação, não constituindo Índice de Validade de Conteúdo.

Os resultados dos usuários não foram diretamente comparados aos índices dos juízes por meio de testes estatísticos, pois os dois grupos responderam a instrumentos com finalidades e critérios distintos. Os juízes examinaram relevância, clareza e representatividade técnica dos elementos do protótipo, produzindo evidências iniciais baseadas no conteúdo. Os

usuários avaliaram a experiência concreta de interação, incluindo aparência, aceitabilidade, usabilidade, instruções, navegação e feedback. Assim, os indicadores são complementares, mas não equivalentes.

A aproximação entre os grupos foi realizada de forma interpretativa, observando convergências e discrepâncias. Avaliações técnicas favoráveis, acompanhadas de percepção menos favorável dos usuários, podem indicar que uma tarefa é conceitualmente coerente, mas apresenta barreiras práticas de interação. Em sentido inverso, uma experiência agradável não garante que os elementos sejam representativos dos processos cognitivos pretendidos. Portanto, coerência técnica e experiência prática devem ser examinadas conjuntamente, sem transformar a avaliação dos usuários em evidência de validade de conteúdo em sentido estrito.

No presente estudo, a avaliação geral dos juízes foi favorável, mas os usuários atribuíram proporções menos elevadas a aspectos relacionados ao esforço exigido pela interface, à clareza das mensagens de erro, à navegação, ao som, ao feedback e à representação dos resultados. Os itens referentes à facilidade proporcionada pelos elementos da interface e à clareza dos sinais de erro apresentaram, respectivamente, 43,5% e 47,8% de respostas favoráveis. Esses resultados sugerem que a coerência conceitual inicial do protótipo não eliminou dificuldades operacionais durante a interação. A análise dos usuários, portanto, constitui uma etapa necessária de refinamento antes que os logs sejam utilizados como evidências psicométricas do desempenho cognitivo.

### **Síntese integrada dos achados**

A análise integrada indica que o Neurotasks apresenta adequação conceitual inicial, mas ainda demanda refinamento na camada de interação. Os juízes avaliaram positivamente a

proposta avaliativa, a possibilidade de observar diferenças individuais e a coerência geral do instrumento. Os usuários, por sua vez, confirmaram a plausibilidade ecológica e o engajamento, mas apontaram dificuldades concretas na interface, nos *feedbacks*, na navegação e na interpretação de resultados. Portanto, os achados sustentam melhor a afirmação de que o protótipo apresenta uma arquitetura avaliativa promissora do que a afirmação de que já mede, de forma isolada e validada, os processos cognitivos propostos.

Esse padrão é metodologicamente relevante. Ele sugere que o instrumento está mais maduro em sua arquitetura inferencial do que em sua experiência operacional. Em avaliações digitais, essa distinção é central: um protótipo pode ter boa fundamentação teórica, mas se a interface cria ruído, a evidência comportamental coletada pode ser contaminada por variância irrelevante ao construto. Assim, os resultados não invalidam a proposta; eles indicam uma agenda de refinamento para as próximas versões do instrumento. A principal implicação é que as próximas versões do Neurotasks devem reduzir a ambiguidade interacional antes que os logs comportamentais sejam utilizados em modelos psicométricos mais complexos.

### **Discussão**

O presente estudo teve como objetivo descrever o desenvolvimento do Neurotasks e apresentar evidências iniciais de validade baseadas no conteúdo, obtidas por meio da avaliação de juízes especialistas, e resultados preliminares relativos aos processos de resposta, à aparência, à aceitabilidade e à experiência de interação dos usuários. Os resultados indicaram que o protótipo apresenta adequação conceitual inicial e aceitabilidade geral, mas também revelaram pontos críticos relacionados à experiência de uso, especialmente em interface, instruções, *feedbacks*, acessibilidade e tela de resultados.

As tarefas atuais constituem uma primeira versão de complexidade fixa. Futuras versões poderão incorporar níveis progressivos de dificuldade, com aumento controlado do

número de estímulos, extensão das sequências, necessidade de alternância de regras, redução de pistas e maior interferência. Também poderá ser investigada uma lógica adaptativa, na qual a sequência ou a dificuldade das tarefas seja ajustada segundo desempenho, idade, perfil funcional, familiaridade digital e necessidades de acessibilidade.

A adaptação, contudo, deverá ser psicometricamente controlada, preservando comparabilidade, registro das rotas administradas e critérios claros de progressão. A personalização não pode transformar a experiência em tarefas incomparáveis ou obscurecer o significado dos escores.

### **Equidade e fairness**

A fairness deve ser investigada como eixo próprio da qualidade da medida, e não apenas como consequência geral da validade. Isso envolve verificar se diferentes grupos têm oportunidades equivalentes para compreender, acessar e executar as tarefas; se a interface produz efeitos diferenciais; se os indicadores apresentam precisão comparável; e se as interpretações mantêm o mesmo significado entre participantes com diferentes idades, escolaridades, níveis socioeconômicos, condições de deficiência, neurodivergência e familiaridade digital.

Para isso, estudos futuros deverão examinar invariância, funcionamento diferencial, precisão por subgrupo, padrões distintos de abandono e possíveis interações entre características pessoais e tecnológicas. A disponibilização de recursos de acessibilidade não deve alterar o construto-alvo, mas reduzir barreiras irrelevantes à interação.

## **Níveis progressivos e tarefas adaptativa**

As tarefas atuais constituem uma primeira versão de complexidade fixa. Futuras versões poderão incorporar níveis progressivos de dificuldade, com aumento controlado do número de estímulos, extensão das sequências, necessidade de alternância de regras, redução de pistas e maior interferência. Também poderá ser investigada uma lógica adaptativa, na qual a sequência ou a dificuldade das tarefas seja ajustada segundo desempenho, idade, perfil funcional, familiaridade digital e necessidades de acessibilidade.

A adaptação, contudo, deverá ser psicometricamente controlada, preservando comparabilidade, registro das rotas administradas e critérios claros de progressão. A personalização não pode transformar a experiência em tarefas incomparáveis ou obscurecer o significado dos escores.

A principal contribuição do estudo consiste na explicitação da arquitetura inferencial do instrumento. Em avaliações baseadas em jogos, a inovação tecnológica pode produzir uma falsa sensação de validade. O fato de o participante interagir com um ambiente digital, executar tarefas e gerar logs não significa, automaticamente, que os dados coletados sustentam inferências psicológicas. Ao utilizar o ECD como lógica de desenvolvimento, o Neurotasks foi planejado a partir da articulação entre modelo de proficiência, modelo de evidência e modelo de tarefa. Essa estrutura fortalece o argumento de validade ao explicitar a articulação entre construtos-alvo, comportamentos observáveis e tarefas planejadas para eliciar evidências relevantes, conforme proposto pelo *Evidence-Centered Design* (Mislevy et al., 2003, 2006; Mislevy, 2018). Tal perspectiva é coerente com a concepção de validade como sustentação das interpretações e usos da medida (American Educational Research Association et al., 2014; Kane, 2013; Messick, 1994). Contudo, o argumento permanece em

fase inicial, pois o estudo sustenta a coerência conceitual e a plausibilidade avaliativa do protótipo, mas ainda não demonstra empiricamente que os indicadores processuais discriminam processos cognitivos específicos. Essa etapa exigirá estudos posteriores de validade, precisão e relação com medidas externas (Kim & Shute, 2015; Shute & Ventura, 2013; von Davier & Mislevy, 2013).

Os resultados dos juízes reforçam essa leitura. O IVC geral de 0,82 e os valores de S-CVI/Ave superiores a 0,80 para relevância, clareza e representatividade constituem evidências iniciais de validade baseadas no conteúdo. Esses resultados sugerem adequação preliminar do protótipo quanto à correspondência entre os elementos avaliados, os objetivos das tarefas e os processos cognitivos propostos. Particularmente, os itens relativos à adequação do jogo para avaliação psicológica e à possibilidade de observar diferenças individuais de desempenho obtiveram índices elevados. Esse achado sustenta a plausibilidade do Neurotasks como proposta avaliativa em desenvolvimento, ao indicar coerência inicial entre os construtos-alvo, as tarefas digitais e os indicadores comportamentais planejados. Tal interpretação é compatível com o *Evidence-Centered Design*, que enfatiza a necessidade de explicitar a relação entre modelo de proficiência, modelo de evidência e modelo de tarefa para sustentar inferências avaliativas (Mislevy et al., 2003, 2006; Mislevy, 2018), bem como com a concepção contemporânea de validade como sustentação das interpretações e usos da medida (American Educational Research Association et al., 2014; Kane, 2013).

A análise dos juízes também indicou fragilidades importantes. Os menores índices em controles intuitivos, instruções, acessibilidade, feedbacks e tela de resultados mostram que a camada operacional ainda precisa ser aprimorada. Esse ponto é mais do que uma questão estética ou de experiência do usuário. Em avaliações psicológicas digitais, instruções, botões, cursores, sinais, sons e telas de resultado integram o sistema de elicitación da evidência, pois

organizam as condições de interação entre participante e tarefa. Conforme o *Evidence-Centered Design*, esses elementos compõem o modelo de tarefa e devem favorecer a produção de comportamentos observáveis coerentes com o modelo de evidência (Mislevy et al., 2003, 2006; Mislevy, 2018). Quando dificuldades de interface interferem na compreensão da tarefa ou na emissão da resposta, os indicadores registrados podem incorporar variância irrelevante ao construto, comprometendo a interpretação dos processos cognitivos avaliados e a equidade da medida (American Educational Research Association et al., 2014; Huff & Sireci, 2001; Kim & Shute, 2015). Assim, a qualidade psicométrica futura do Neurotasks dependerá não apenas da consistência das tarefas em relação aos processos cognitivos-alvo, mas também da redução sistemática de barreiras de interface, familiaridade tecnológica e acessibilidade que possam introduzir variância irrelevante ao construto.

A avaliação dos usuários reforça essa preocupação. A proporção geral de respostas favoráveis de 0,76 indica aceitabilidade inicial e percepção global favorável sobre parte da experiência de uso. Esses dados não constituem evidência de validade baseada no conteúdo, mas oferecem informações preliminares sobre aparência, usabilidade, experiência de interação e processos de resposta. O item U3, com 43,5%, indica que muitos participantes não perceberam a interface como facilitadora da execução com pouco esforço. O item U8, com 47,8% de concordância, indicou fragilidade nas mensagens e nos sinais de erro do sistema. Esse achado é relevante para o argumento inferencial do instrumento, pois, em avaliações digitais, o feedback e a sinalização da tarefa compõem as condições pelas quais a evidência é produzida. Conforme o *Evidence-Centered Design*, a interpretação dos dados observáveis depende da articulação entre tarefa, evidência e construto-alvo (Mislevy et al., 2003, 2006; Mislevy, 2018). Portanto, quando o usuário não recebe sinais claros sobre o que ocorreu durante a interação, cliques, tentativas e tempos de resposta podem refletir não apenas

processos cognitivos, mas também ruído de interface, configurando possível fonte de variância irrelevante ao construto (American Educational Research Association et al., 2014; Huff & Sireci, 2001; Kim & Shute, 2015).

Já os itens de melhor desempenho entre os usuários indicam pontos fortes relevantes. O tempo disponível para responder foi considerado adequado por 95,7% dos participantes. A legibilidade das fontes e o sentido cotidiano das tarefas obtiveram 91,3%. A neutralidade da linguagem e a suficiência de informações para iniciar o jogo também apresentaram resultados elevados. Esses achados sugerem que as fragilidades identificadas não se concentram na proposta avaliativa geral do Neurotasks, mas em aspectos específicos da interação usuário-sistema. Esse padrão é coerente com a literatura sobre *game-based assessment*, que indica que elementos de design, usabilidade, feedback e experiência do usuário podem influenciar a qualidade dos dados produzidos e, conseqüentemente, a validade, a fidedignidade e a equidade das interpretações derivadas do desempenho (Kim & Shute, 2015; Mantell et al., 2025). Em termos clínico-metodológicos, os resultados sugerem que o Neurotasks apresenta uma hipótese avaliativa inicial plausível, mas ainda requer refinamentos operacionais para reduzir interferências de interface e favorecer que os indicadores registrados expressem, de modo mais específico, os processos cognitivos de interesse.

As evidências obtidas neste estudo não incluem, ainda, relações com variáveis externas, estrutura interna ou precisão das métricas processuais. Portanto, não é possível afirmar que os indicadores do Neurotasks convergem com medidas consolidadas dos processos cognitivos propostos ou que discriminam níveis de funcionamento. Estudos futuros deverão comparar as métricas do protótipo com instrumentos de atenção, memória operacional, planejamento, controle inibitório, categorização e funcionamento cotidiano, examinando a magnitude e a especificidade das associações.

A ausência de correlação com medidas externas não deverá ser interpretada automaticamente como prova de inexistência do processo cognitivo na tarefa. Poderá indicar inadequação do indicador digital, baixa precisão, restrição de variância, dificuldade insuficiente, influência excessiva da interface, diferença entre o construto medido pelas tarefas ou fragilidade da medida externa utilizada. Do mesmo modo, uma correlação positiva isolada não será suficiente para demonstrar validade. O significado das relações deverá ser avaliado em conjunto com conteúdo, processos de resposta, estrutura interna, precisão e possíveis explicações alternativas.

A discrepância entre juízes e usuários constitui um achado metodologicamente relevante. Os especialistas contribuíram para examinar a coerência conceitual e a representatividade das tarefas, aspectos centrais à validade de conteúdo, enquanto os usuários evidenciaram dificuldades práticas relacionadas à interação com o protótipo, como usabilidade, compreensão das instruções e experiência de navegação (Haynes et al., 1995; Rubio et al., 2003). Essa complementaridade é compatível com abordagens contemporâneas de desenvolvimento de instrumentos digitais, nas quais a avaliação técnica e psicométrica deve ser articulada à experiência concreta dos usuários e a ciclos iterativos de refinamento da plataforma (Ginige et al., 2025; Mantell et al., 2025). Em avaliações digitais, essa distinção é especialmente importante, pois elementos de interface e feedback podem interferir nos indicadores registrados, introduzindo variância irrelevante ao construto e afetando a validade e a equidade das interpretações (American Educational Research Association et al., 2014; Kim & Shute, 2015).

As respostas qualitativas trouxeram insumos importantes para o refinamento. A tarefa do café foi mencionada como ponto de dificuldade por diferentes participantes, especialmente em relação ao tamanho dos alvos, identificação do reservatório de água, compreensão da

sequência e necessidade de pistas visuais. A tarefa de organização de documentos também apareceu como crítica, principalmente por legibilidade insuficiente e ambiguidade na classificação. Esses achados sugerem que algumas tarefas do Neurotasks podem estar gerando carga cognitiva acessória, relacionada a demandas operacionais da interface, compreensão das instruções, feedbacks ou navegação no ambiente digital. Nesse caso, parte do desempenho registrado pode refletir não apenas os processos cognitivos-alvo, mas também exigências adicionais impostas pelo formato da tarefa. Essa interpretação é coerente com a Teoria da Carga Cognitiva, que diferencia demandas inerentes ao conteúdo da tarefa de cargas extrínsecas produzidas pelo modo como a informação é apresentada (Sweller, 1994; de Jong, 2010). Em GBAs, elementos do design, da interface e da experiência de interação podem influenciar validade, fidedignidade, equidade e qualidade dos dados produzidos. Revisões e estudos recentes mostram que dificuldades de usabilidade, baixa familiaridade com jogos e diferenças na percepção de domínio podem alterar a experiência de execução e introduzir demandas não pretendidas pela tarefa, reforçando a necessidade de controlar interferências operacionais na interpretação dos indicadores processuais (Kim & Shute, 2015; Mantell et al., 2025; Markovitch et al., 2025).

À luz dos Standards for Educational and Psychological Testing (American Educational Research Association et al., 2014), os resultados deste estudo concentram-se predominantemente em evidências de validade baseadas no conteúdo e em informações preliminares relacionadas aos processos de resposta. A avaliação dos juízes especialistas contribuiu para examinar a relevância, a clareza e a representatividade dos elementos do protótipo, constituindo a principal fonte de evidência baseada no conteúdo. A avaliação dos usuários, por sua vez, forneceu informações sobre compreensão das tarefas, usabilidade, para a percepção superficial de plausibilidade, clareza e adequação, aceitabilidade, experiência de

interação e possíveis interferências operacionais na produção das respostas. Esses resultados ainda não constituem evidências conclusivas baseadas nos processos de resposta, uma vez que não foram realizadas análises sistemáticas dos logs comportamentais, entrevistas cognitivas ou protocolos verbais. Permanecem necessárias evidências baseadas na estrutura interna, nas relações com variáveis externas, na precisão, na estabilidade e nas consequências do uso.

Evidências recentes também reforçam que a validade de avaliações cognitivas baseadas em jogos deve ser construída de forma cumulativa. Meta-análises indicam que os indicadores derivados de jogos apresentam relações relevantes, porém não necessariamente equivalentes, com medidas tradicionais de habilidade cognitiva (Bipp et al., 2024). Dessa forma, os estudos futuros do Neurotasks deverão examinar não apenas correlações com instrumentos de referência, mas também a estabilidade das métricas processuais, a estrutura interna dos indicadores, as diferenças associadas à familiaridade digital, a invariância entre grupos e a contribuição incremental dos dados de processo para a interpretação do funcionamento cognitivo.

A familiaridade com as tarefas não depende apenas de idade ou experiência com jogos. Escolaridade, nível socioeconômico, ocupação, acesso a computadores, frequência de uso da internet, tipo de moradia e experiências domésticas podem influenciar o reconhecimento dos objetos e a compreensão dos cenários. Participantes com maior acesso a tecnologias ou familiaridade com ambientes semelhantes aos representados podem iniciar a tarefa com vantagem operacional. Por outro lado, uma pessoa pode apresentar funcionamento cognitivo preservado e, ainda assim, desconhecer uma cafeteira específica, um tipo de documento ou determinada convenção de interface.

Esse aspecto reforça a necessidade de diversificar cenários e selecionar atividades amplamente reconhecíveis em diferentes contextos de vida adulta. Futuras versões deverão incluir análise cultural e socioeconômica dos estímulos, consultas a participantes de diferentes grupos e possibilidade de tarefas alternativas com demandas cognitivas equivalentes.

A variância irrelevante ao construto corresponde à parcela do desempenho produzida por fatores que não pertencem ao processo cognitivo que se pretende avaliar. No Neurotasks, por exemplo, maior tempo para localizar um objeto pode decorrer de atenção seletiva menos eficiente, mas também de brilho reduzido, tela pequena, cursor pouco visível, uso pouco familiar do touchpad ou ambiguidade sobre quais elementos são clicáveis. Da mesma forma, dificuldades na preparação do café podem refletir planejamento, mas também desconhecimento de cafeteiras de cápsula; erros na organização de documentos podem refletir categorização, mas também baixa legibilidade ou pouca familiaridade com os documentos representados.

A redução dessa variância exige padronização das condições de aplicação, treinamento básico de interface, estímulos mais claros, acessibilidade, registro das características do dispositivo, revisão dos objetos e instruções, testes de usabilidade e análise diferenciada dos tipos de erro. O sistema deverá separar, tanto quanto possível, erro cognitivo hipotetizado, erro induzido pela interface, erro motor e erro relacionado à baixa familiaridade digital.

Como o sistema apresenta cálculo automatizado dos indicadores, recomenda-se que versões futuras disponibilizem, em área técnica ou relatório metodológico, as fórmulas utilizadas, os critérios de recodificação das respostas, o número de observações válidas e as

regras aplicadas aos dados ausentes. Essa transparência permitirá auditar os cálculos e assegurar correspondência entre o procedimento descrito no manuscrito e o resultado apresentado pela plataforma.

Do ponto de vista aplicado, os achados apontam prioridades mais diretivas para a próxima versão do Neurotasks. Primeiro, é necessário aprimorar instruções iniciais e instruções contextuais, reduzindo a dependência de memória de instruções longas. Segundo, devem ser ampliados sinais de erro, conclusão e progresso. Terceiro, recomenda-se aumentar a visibilidade de alvos clicáveis, cursores e objetos pequenos. Quarto, é necessário revisar os elementos sonoros, com possibilidade de controle de volume, e equivalentes visuais para participantes com limitações auditivas. Quinto, a tela de resultados deve ser redesenhada para comunicar desempenho de forma mais clara, distinguindo informações úteis ao participante e informações técnicas destinadas ao pesquisador. Sexto, recomenda-se inserir fase breve de treino, especialmente para familiarização com a lógica *point-and-click*.

Em síntese, o estudo mostra que o Neurotasks é uma proposta importante, mas ainda em desenvolvimento. Sua principal força é não depender apenas do apelo tecnológico. O instrumento foi construído a partir de uma lógica inferencial explícita, alinhada ao ECD, e apresentou índices iniciais favoráveis entre juízes e usuários. Sua principal fragilidade está na necessidade de reduzir ruídos de interface para que os logs comportamentais possam, em estudos futuros, ser interpretados com maior segurança.

Este estudo apresenta limitações. A primeira limitação refere-se ao estágio inicial do instrumento. O Neurotasks é um protótipo em desenvolvimento; portanto, os resultados não autorizam uso diagnóstico, classificatório ou decisório. As evidências apresentadas são preliminares e devem ser compreendidas como parte do processo de construção do argumento de validade. Desse modo, os achados não autorizam inferências diagnósticas ou

classificatórias sobre funcionamento cognitivo, mas sustentam a continuidade do programa de pesquisa voltado ao refinamento do protótipo e à produção progressiva de evidências psicométricas.

A segunda limitação refere-se à amostragem por conveniência e ao tamanho da amostra. Embora o número de juízes e usuários seja compatível com uma etapa inicial de desenvolvimento, ainda é insuficiente para análises psicométricas robustas. Estudos posteriores deverão ampliar a amostra e diversificar perfis de participantes quanto à idade, escolaridade, familiaridade digital e condições cognitivas.

A terceira limitação refere-se à ausência de análise empírica dos logs comportamentais. Embora o instrumento tenha sido planejado para registrar métricas de processo, este manuscrito ainda não testa a relação entre essas métricas e os processos cognitivos propostos. Portanto, a ligação entre indicadores observáveis e construtos permanece, nesta etapa, em nível teórico e inferencial.

A quarta limitação refere-se ao instrumento de avaliação dos juízes. Como apontado por uma das especialistas, parte dos itens do protocolo avaliou características do jogo, da interface e da experiência, e não diretamente a representatividade dos construtos cognitivos. Essa observação é metodologicamente importante. Em versões futuras, recomenda-se separar dois protocolos: um voltado à avaliação técnica, de usabilidade e design; e outro especificamente destinado à análise de conteúdo cognitivo, com itens que perguntem diretamente se cada tarefa e cada métrica representam adequadamente os processos-alvo. Essa separação permitirá distinguir validade de conteúdo do construto, qualidade da interface e aceitabilidade do usuário, evitando que dimensões distintas sejam colapsadas em um único índice global.

O estudo não examinou sistematicamente diferenças relacionadas à idade, escolaridade, experiência prévia com jogos, familiaridade digital ou neurodivergência. Essas características podem influenciar a compreensão das tarefas, a navegação e a experiência de interação, devendo ser registradas e controladas em estudos posteriores.

O nível de dificuldade das tarefas ainda não foi calibrado empiricamente. Como as atividades representam rotinas relativamente simples, é possível que algumas apresentem efeito de teto entre adultos cognitivamente preservados, reduzindo sua capacidade de discriminar diferenças individuais. Por outro lado, dificuldades de interface podem elevar artificialmente a complexidade e produzir desempenho inferior sem que haja aumento real da demanda cognitiva. Estudos posteriores deverão examinar distribuições de tempo, erros, tentativas e acurácia, identificar efeitos de piso e teto e verificar se as tarefas oferecem variação suficiente para diferenciar níveis de desempenho.

Estudos futuros deverão contemplar novas rodadas de refinamento, reaplicação com juízes e usuários, análise dos logs processuais, investigação da estrutura interna, estimativas de precisão, análise de estabilidade temporal e estudos de relação com medidas externas de processos cognitivos. Recomenda-se também examinar o impacto da familiaridade com jogos digitais, do tipo de dispositivo, do tamanho da tela e da escolaridade sobre o desempenho. Adicionalmente, será necessário definir critérios para tratamento de logs inválidos, identificação de padrões aleatórios de resposta, controle de desistência, modelagem de tempo extremo e diferenciação entre erro cognitivo e erro induzido pela interface. Essa agenda é necessária para que o Neurotasks avance de protótipo promissor para instrumento psicometricamente defensável.

A finalidade imediata do Neurotasks é apoiar pesquisa e desenvolvimento metodológico em avaliação cognitiva digital. Em etapas posteriores, o protótipo poderá ser

investigado como ferramenta complementar de triagem, monitoramento ou apoio clínico, sempre integrado a outras fontes de informação. Qualquer utilização diagnóstica dependerá de normas, estudos de precisão, evidências de validade, definição de pontos de corte e investigação de sensibilidade e especificidade. Na etapa atual, o Neurotasks não deve ser utilizado como medida diagnóstica independente.

Caso o Neurotasks venha a ser proposto para finalidade de triagem ou apoio diagnóstico, será necessário examinar sensibilidade, especificidade, valores preditivos e desempenho classificatório em grupos clínicos e não clínicos. Essas análises deverão ser realizadas somente após a definição de escores, estudo da estrutura interna, estimação da precisão e estabelecimento de critérios externos adequados. Antes disso, a classificação de indivíduos produziria risco elevado de falso positivo ou falso negativo.

O presente estudo descreveu o desenvolvimento do Neurotasks, um protótipo de *game-based assessment* fundamentado no *Evidence-Centered Design* para avaliação de processos cognitivos em adultos, e apresentou evidências iniciais de validade de conteúdo e a percepção superficial de plausibilidade, clareza e adequação. O instrumento foi estruturado em tarefas contextualizadas em ambientes cotidianos e planejado para registrar métricas processuais de desempenho, como tempo, latência, cliques, tentativas, erros, omissões, sequência de ações e acurácia.

Os resultados indicaram IVC geral de 0,82 entre juízes e índice geral de validade de percepção superficial de plausibilidade, clareza e adequação de 0,76 entre usuários, sugerindo adequação inicial do protótipo. Os juízes reconheceram a pertinência da proposta avaliativa e o potencial das tarefas para observar diferenças individuais de desempenho. Os usuários perceberam positivamente o sentido cotidiano das tarefas, a legibilidade, o tempo de resposta e a adequação geral do jogo para avaliar habilidades cognitivas. Entretanto, os

resultados também apontaram fragilidades em interface, feedback, navegação, sons, acessibilidade e tela de resultados.

A principal contribuição do Neurotasks está na tentativa de articular inovação digital e psicometria. O instrumento não deve ser compreendido apenas como um jogo aplicado à Psicologia, mas como uma arquitetura avaliativa em desenvolvimento, na qual construtos, tarefas, evidências e interpretações são organizados de forma explícita. As próximas etapas deverão refinar a experiência de uso e testar empiricamente os logs comportamentais, fortalecendo progressivamente o argumento de validade do instrumento.

### **Considerações finais**

O presente estudo contribuiu para o campo da avaliação psicológica digital ao sistematizar uma revisão de escopo e traçar um percurso metodológico aplicável à construção e análise inicial de *game-based assessments* (GBAs) voltado à avaliação de processos cognitivos em adultos. Ao articular fundamentos teóricos, critérios psicométricos, princípios do *Evidence-Centered Design* e indicadores da experiência do usuário, o estudo buscou demonstrar que o desenvolvimento de instrumentos digitais com finalidade avaliativa exige mais do que inovação tecnológica ou atratividade visual. Exige, sobretudo, clareza sobre os construtos avaliados, coerência entre tarefas e evidências esperadas, controle de fontes de erro e análise sistemática das condições que sustentam as inferências produzidas a partir do desempenho.

Nesse percurso, a descrição das etapas de concepção, estruturação, avaliação e refinamento das tarefas digitais oferece uma contribuição metodológica para futuras pesquisas interessadas no desenvolvimento de instrumentos gamificados. A explicitação desse processo permite compreender como elementos narrativos, mecânicas de interação,

instruções, regras, feedbacks, demandas motoras e registros de desempenho podem ser organizados de modo a favorecer maior alinhamento entre o ambiente digital e os processos cognitivos de interesse. Assim, o estudo reforça que, em GBAs, o jogo não deve ser entendido apenas como uma interface lúdica, mas como um sistema avaliativo no qual cada decisão de design pode impactar a qualidade dos dados coletados e a defensabilidade das interpretações.

O protocolo AvaliaJS mostrou-se um recurso relevante para a análise crítica do processo de construção do jogo, ao permitir a avaliação de dimensões como usabilidade, jogabilidade, acessibilidade, clareza das instruções, mecânicas de interação e adequação dos elementos psicométricos envolvidos nas tarefas. Essa avaliação é particularmente importante porque, em instrumentos digitais, a interação do usuário com o ambiente não constitui apenas uma condição operacional de aplicação. Ela pode interferir diretamente nos indicadores de desempenho, especialmente quando dificuldades de navegação, ambiguidades instrucionais, baixa familiaridade com jogos, sobrecarga visual ou demandas acessórias passam a competir com o construto que se pretende avaliar.

Dessa forma, os achados reforçam a necessidade de atenção às possíveis fontes de variância irrelevante ao construto em avaliações digitais. Em tarefas gamificadas, uma resposta pode refletir não apenas planejamento, atenção, tomada de decisão, flexibilidade ou monitoramento, mas também fatores externos ao processo cognitivo-alvo, como compreensão da interface, habilidade motora, experiência prévia com ambientes digitais e adaptação às regras do jogo. Reconhecer e controlar essas fontes de variação é condição central para fortalecer o argumento de validade e para ampliar a equidade no uso de instrumentos digitais em diferentes perfis de usuários.

Assim, o estudo oferece uma contribuição inicial, mas estratégica, para o avanço dos GBAs no contexto da avaliação psicológica. Ao integrar análise teórica, estruturação metodológica, avaliação psicométrica preliminar e experiência do usuário, a pesquisa aponta caminhos para o desenvolvimento de instrumentos digitais mais rigorosos, acessíveis e cientificamente defensáveis. Além disso, o site do Neurotasks constitui uma contribuição operacional relevante, ao automatizar etapas do processo de pesquisa, especialmente no que se refere à avaliação por juízes especialistas em ambiente digital. Por meio do sistema, torna-se possível organizar o acesso às tarefas, registrar as avaliações, sistematizar respostas e otimizar a coleta de dados de forma mais padronizada, reduzindo etapas manuais e favorecendo maior controle metodológico. Recomenda-se que estudos futuros ampliem a amostra, investiguem novas evidências de validade e precisão, analisem os dados processuais de forma mais robusta e examinem o funcionamento das tarefas em diferentes perfis de usuários. Esse percurso será fundamental para fortalecer a qualidade inferencial do Neurotasks e sua potencial aplicabilidade clínica, educacional e científica no campo da avaliação psicológica digital.

A agenda futura deverá contemplar: (a) refinamento das tarefas, instruções, interface e feedbacks; (b) implementação de pré-treino e registro das condições tecnológicas; (c) análise dos logs processuais, incluindo tempo de reação, latência, cliques, erros, correções, consultas, trajetórias e abandono; (d) estudo da dificuldade e do desempenho esperado; (e) investigação da estrutura interna, precisão e estabilidade temporal; (f) estudos de relações com variáveis externas e validade convergente; (g) análise de invariância, funcionamento diferencial e fairness; (h) construção de referências normativas; (i) investigação de sensibilidade e especificidade em grupos clínicos; e (j) avaliação da utilidade incremental do Neurotasks em relação a instrumentos consolidados.

Também deverão ser realizadas comparações entre faixas etárias, níveis de escolaridade, familiaridade tecnológica, experiência com jogos, condições socioeconômicas, perfis neurodivergentes e grupos clínicos. O desenvolvimento de normas exigirá amostras amplas e representativas, condições de aplicação padronizadas e definição clara da população e das finalidades interpretativas.

Considerando a rápida evolução das tecnologias de avaliação, recomenda-se que as próximas etapas de desenvolvimento também incorporem procedimentos iterativos de experiência do usuário, testes com participantes de diferentes níveis de familiaridade digital e análises específicas de equidade. Estudos recentes indicam que a qualidade de avaliações cognitivas baseadas em jogos depende não apenas de sua arquitetura psicométrica, mas também da adequação da interface, do feedback, da acessibilidade e do modo como diferentes perfis de usuários interagem com o ambiente digital (Mantell et al., 2025; Markovitch et al., 2025).

A contribuição original deste trabalho está na articulação, em contexto brasileiro, entre game-based assessment, Evidence-Centered Design, processos cognitivos, registros comportamentais e avaliação do protótipo por juízes especialistas e usuários. Essa articulação representa uma proposta metodológica inicial e não implica superioridade ou exclusividade em relação a outras iniciativas nacionais ou internacionais.

Conclui-se que o Neurotasks constitui um protótipo de game-based assessment em estágio inicial de desenvolvimento. As evidências obtidas sustentam a continuidade do processo de refinamento, mas ainda não autorizam interpretações individuais, classificação diagnóstica ou uso independente em decisões clínicas. Seu potencial mais próximo está na pesquisa e no desenvolvimento de indicadores complementares do funcionamento cognitivo,

cuja utilidade dependerá do acúmulo progressivo de evidências psicométricas e do controle das interferências produzidas pela interface e pelas condições de aplicação.

## Referências

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Andrade, B. G. da S., Carvalho, F. S. de, & Carvalho, C. A. P. de. (2025). Desenvolvimento e validação de um serious game para educação em saúde bucal de adolescentes. *Revista Brasileira de Informática na Educação*, 33, 972–989.  
<https://doi.org/10.5753/rbie.2025.5128>
- Arieli-Attali, M., Wylie, E. C., & Bauer, M. I. (2019). The expanded evidence-centered design (e-ECD) for learning and assessment systems: A conceptual framework for incorporating learning goals and processes within assessment design. *Frontiers in Psychology*, 10, Article 853. <https://doi.org/10.3389/fpsyg.2019.00853>
- Bennett, R. E., LaMar, M., & Mazzeo, J. (2025). Technology-based assessment: Validity, modeling, and analysis issues. In R. L. Brennan (Ed.), *Educational measurement* (5th ed.). Oxford University Press.
- Bipp, T., Wee, S., Walczok, L., & Hansal, L. (2024). The relationship between game-related assessment and traditional measures of cognitive ability—A meta-analysis. *Journal of Intelligence*, 12(12), Article 129. <https://doi.org/10.3390/jintelligence12120129>
- Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A., & Hogan, R. (2016). New talent signals: Shiny new objects or a brave new world? *Industrial and Organizational Psychology*, 9(3), 621–640. <https://doi.org/10.1017/iop.2016.6>

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design:

Some food for thought. *Instructional Science*, 38, 105–134.

<https://doi.org/10.1007/s11251-009-9110-0>

DiCerbo, K. E., & Mislevy, R. J. (2017). Evidence-centered design and the future of

assessment. In M. von Davier, C. Zhu, & P. C. Kyllonen (Eds.), *Innovative assessment of collaboration* (pp. 19–37). Springer.

Frensch, P. A., & Funke, J. (Eds.). (1995). *Complex problem solving: The European perspective*. Lawrence Erlbaum Associates.

Funke, J. (1993). Microworlds based on linear equation systems: A new approach to complex problem solving and experimental results. In G. Strube & K. F. Wender (Eds.), *The cognitive psychology of knowledge* (pp. 313–330). Elsevier.

[https://doi.org/10.1016/S0166-4115\(08\)62663-1](https://doi.org/10.1016/S0166-4115(08)62663-1)

Georgiou, K., Gouras, A., & Nikolaou, I. (2019). Gamification in employee selection: The development of a gamified assessment. *International Journal of Selection and Assessment*, 27(2), 91–103. <https://doi.org/10.1111/ijsa.12240>

Ginige, J. A., Rana, K., Belcastro, M., Gunaratne, P., Lillis, S., Numbers, K. T., Lin, P.-H., Nguyen, K., Tanvir, M., & Reppermund, S. (2025). Technical development and user experience evaluation of the computerised version of the Sydney Test of Activities of Daily Living in Memory Disorders (C-STAM). *Frontiers in Digital Health*, 7, Article 1637916. <https://doi.org/10.3389/fdgth.2025.1637916>

- Gomez, M. J., Ruipérez-Valiente, J. A., & García-Clemente, F. J. (2023). A systematic literature review of game-based assessment studies: Trends and challenges. *IEEE Transactions on Learning Technologies*, 16(4), 500–515.  
<https://doi.org/10.1109/TLT.2022.3226661>
- Greiff, S., Holt, D. V., & Funke, J. (2013). Perspectives on problem solving in educational assessment: Analytical, interactive, and collaborative problem solving. *The Journal of Problem Solving*, 5(2), 71–91. <https://doi.org/10.7771/1932-6246.1153>
- Griffin, P., McGaw, B., & Care, E. (Eds.). (2012). *Assessment and teaching of 21st century skills*. Springer. <https://doi.org/10.1007/978-94-007-2324-5>
- Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238–247. <https://doi.org/10.1037/1040-3590.7.3.238>
- Huff, K. L., & Sireci, S. G. (2001). Validity issues in computer-based testing. *Educational Measurement: Issues and Practice*, 20(3), 16–25.  
<https://doi.org/10.1111/j.1745-3992.2001.tb00066.x>
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. Pfeiffer.

- Kim, Y. J., & Shute, V. J. (2015). The interplay of game elements with psychometric qualities, learning, and enjoyment in game-based assessment. *Computers & Education*, 87, 340–356. <https://doi.org/10.1016/j.compedu.2015.07.009>
- Kleitman, S., Huang, P.-S., & Adjerid, I. (2022). Applying evidence-centered design to game-based assessment of psychological resilience. *Frontiers in Psychology*, 12, Article 717568. <https://doi.org/10.3389/fpsyg.2021.717568>
- Leutner, F., Codreanu, S.-C., Brink, S., & Bitsakis, T. (2023). Game-based assessments of cognitive ability in recruitment: Validity, fairness, and test-taking experience. *Frontiers in Psychology*, 13, Article 942662. <https://doi.org/10.3389/fpsyg.2022.942662>
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–385. <https://doi.org/10.1097/00006199-198611000-00017>
- Mantell, R., Hwang, Y. I. J., Dark, M., Radford, K., Kasumovic, M., Monds, L., Schofield, P. W., Butler, T., & Withall, A. (2025). Evaluating the user experience and usability of game-based cognitive assessments for older people: Systematic review. *JMIR Aging*, 8, Article e65252. <https://doi.org/10.2196/65252>
- Markovitch, B., Kamps, J. C. C., Markopoulos, P., & Birk, M. V. (2025). Do cognitive assessment games leave infrequent video game players behind? Evaluating frequent and infrequent players' gaming experience and data quality. *Computers in Human Behavior*, 172, Article 108720. <https://doi.org/10.1016/j.chb.2025.108720>

- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design* (ETS Research Report No. RR-03-16). Educational Testing Service. <https://doi.org/10.1002/j.2333-8504.2003.tb01908.x>
- Mislevy, R. J., Oranje, A., Bauer, M. I., von Davier, A. A., Hao, J., Corrigan, S., & John, M. (2015). Psychometrics and game-based assessment. In J. C. Impara & B. S. Plake (Eds.), *The next generation of testing: Common core standards, smarter-balanced, PARCC, and the nationwide testing movement* (pp. 207–233). Information Age Publishing.
- Peters, H., Kyngdon, A., & Stillwell, D. (2021). Construction and validation of a game-based intelligence assessment in Minecraft. *Computers in Human Behavior, 119*, Article 106701. <https://doi.org/10.1016/j.chb.2021.106701>
- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489–497. <https://doi.org/10.1002/nur.20147>
- Quiroga, M. A., Escorial, S., Román, F. J., Morillo, D., Jarabo, A., Privado, J., Hernández, M., Gallego, B., & Colom, R. (2015). Can we reliably measure the general factor of intelligence (g) through commercial video games? Yes, we can! *Intelligence, 53*, 1–7. <https://doi.org/10.1016/j.intell.2015.08.004>
- Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research, 27*(2), 94–104. <https://doi.org/10.1093/swr/27.2.94>

- Salen, K., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. MIT Press.
- Santana, A. N., Melo, M. R. A., & Minervino, C. A. S. M. (2019). Instrumentos de avaliação das funções executivas: Revisão sistemática dos últimos cinco anos. *Avaliação Psicológica*, 18(1), 96–107. <https://doi.org/10.15689/ap.2019.1801.14668.11>
- Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503–524). Information Age Publishing.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312. [https://doi.org/10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5)
- von Davier, M., Khorramdel, L., He, Q., Shin, H. J., & Chen, H. (2019). Developments in psychometric population models for technology-based large-scale assessments: An overview of challenges and opportunities. *Journal of Educational and Behavioral Statistics*, 44(6), 671–705. <https://doi.org/10.3102/1076998619881789>
- Weiner, E. J., & Sanchez, D. R. (2020). Cognitive ability in virtual reality: Validity evidence for VR game-based assessments. *International Journal of Selection and Assessment*, 28(3), 215–235. <https://doi.org/10.1111/ijsa.12295>
- Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving—More than reasoning? *Intelligence*, 40(1), 1–14. <https://doi.org/10.1016/j.intell.2011.11.003>

## **Apêndices**

### **Apêndice A**

#### **Termo de Consentimento Livre e Esclarecido (TCLE)**

##### **“Protótipo de um game-based assessment para avaliação de processos cognitivos em adultos, fundamentado na abordagem Evidence-Centered Design”**

Universidade de Brasília– UnB

Pesquisadoras responsáveis: Michelle Ferreira de Andrade.

Orientadora: Cristiane Faiad; coorientador: Alexandre Serpa.

Você está sendo convidado(a) a participar do projeto de pesquisa acima citado. O texto abaixo apresenta todas as informações necessárias sobre o que estamos fazendo.

O nome deste documento que você está lendo é Termo de Consentimento Livre e Esclarecido (TCLE).

Antes de decidir se deseja participar (de livre e espontânea vontade), você deverá ler e compreender todo o conteúdo. Ao final, caso decida participar, você será solicitado a assiná-lo e receberá uma via dele.

Antes de assinar, faça perguntas sobre tudo o que não tiver entendido bem. A equipe deste estudo responderá às suas perguntas a qualquer momento (antes, durante e após o estudo).

#### **Natureza e objetivos do estudo**

- Você está sendo convidado a participar desta pesquisa que desenvolveu um jogo eletrônico para a avaliação psicológica de processos cognitivos em adultos por meio de

tarefas digitais de desempenho e investigar suas evidências de validade baseadas no Evidence-Centered Design.

### **Procedimentos do estudo**

- Sua participação consiste em 15 min de aplicação de 1 jogo.
- Após o jogo, será aplicado 1 questionário de avaliação do jogo.
- Não será necessária a filmagem do procedimento, porém toda a sua performance será armazenada em nosso banco de dados.

- ( ) Estou de acordo com a utilização dos meus dados neste estudo.
- ( ) Não estou de acordo com a utilização dos meus dados neste estudo.
- A pesquisa será realizada em um local reservado, garantindo o sigilo das informações do participante.

### **Riscos e benefícios**

Os riscos apresentados pela presente pesquisa são mínimos, uma vez que a participação

É voluntária e o/a participante poderá se recusar a responder qualquer questão que lhe traga constrangimento ou mal-estar, podendo desistir em qualquer momento sem nenhum prejuízo.

Os dados serão coletados pela tecnologia do jogo, e a pesquisa será divulgada em mídias sociais, universidade e em grupos da comunidade gamer. Os participantes não se identificarão por meio de nome ou números identitários, como CPF ou RG.

Ao participar desta pesquisa, o/a participante não terá nenhum benefício direto e imediato. Entretanto, após analisar os dados, poderão ser realizados estudos futuros com amostras ampliadas, com vistas ao desenvolvimento de parâmetros interpretativos e novas

evidências psicométricas. Esse instrumento, em estudos futuros, poderá contribuir como fonte complementar de informação sobre processos cognitivos, desde que novas evidências psicométricas sustentem sua interpretação e uso.

### **Confidencialidade**

- Seus dados serão manuseados somente pelos pesquisadores e não será permitido o acesso a outras pessoas.
- Os dados e instrumentos utilizados ficarão guardados sob a responsabilidade de Michelle Ferreira de Andrade com a garantia de manutenção do sigilo e confidencialidade, e arquivados por um período de 5 anos; após esse tempo serão destruídos.
- Os resultados deste trabalho poderão ser apresentados em encontros ou revistas científicas. Entretanto, ele mostrará apenas os resultados obtidos como um todo, sem revelar seu nome, instituição à qual pertence ou qualquer informação que esteja relacionada com sua privacidade.

Se houver alguma consideração ou dúvida referente aos aspectos éticos da pesquisa, entre em contato com o Comitê de Ética em Pesquisa da Universidade de Brasília – CEP/UnB, que aprovou esta pesquisa, pelo telefone 3107 2294 ou pelo e-mail [eticaunb@unb.br](mailto:eticaunb@unb.br). O horário de atendimento do CEP-UNB é de segunda a sexta: das 8h às 12h e das 13h às 17h. Também entre em contato para informar ocorrências irregulares ou danosas durante a sua participação no estudo.

O CEP é um grupo de profissionais de várias áreas do conhecimento e da comunidade, autônomo, de relevância pública, que tem o propósito de defender os interesses dos participantes da pesquisa em sua integridade e dignidade e de contribuir para o desenvolvimento da pesquisa dentro de padrões éticos.

Ao assinar abaixo, você confirma que leu as afirmações contidas neste termo de consentimento, que foram explicados os procedimentos do estudo, que teve a oportunidade de fazer perguntas, que está satisfeito com as explicações fornecidas e que decidiu participar voluntariamente deste estudo. Uma via será entregue a você e a outra será arquivada pelo pesquisador responsável.

Caso tenha qualquer dúvida sobre a pesquisa, incluindo os danos possíveis, entre em contato com a pesquisadora responsável, Michelle Ferreira de Andrade, no telefone (61) 983877007 ou pelo e-mail michellefandrade@gmail.com.

Eu \_\_\_\_\_, após receber a explicação completa dos objetivos do estudo e dos procedimentos envolvidos nesta pesquisa, concordo voluntariamente em fazer parte deste estudo.

Brasília, \_\_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_.

---

PARTICIPANTE

---

—  
Michelle Ferreira de Andrade

CRP 01/20131

(61) 993877007 – michellefandrade@gmail.com

**Endereço do responsável pela pesquisa:**

Instituição: Universidade de Brasília

Endereço: *Campus* Universitário Darcy Ribeiro, Brasília-DF | CEP 70910-900.

Telefones p/contato: +55 (61) 3107-2294

## Apêndice B

### Protocolo de Avaliação Remota baseado no AvaliaJS

| Protocolo de Avaliação Remota                                          |                               |                                                                                        |                                                               |                                                                |
|------------------------------------------------------------------------|-------------------------------|----------------------------------------------------------------------------------------|---------------------------------------------------------------|----------------------------------------------------------------|
| 1. DADOS GERAIS DO NEUROTASKS                                          |                               |                                                                                        |                                                               |                                                                |
| <b>Data:</b>                                                           | <b>Versão:</b> Protótipo<br>1 | <b>Autores:</b> Michelle Ferreira de Andrade, Cristiane Faiad e equipe de programação. |                                                               |                                                                |
| <b>Título do Projeto:</b> Neurotasks                                   |                               |                                                                                        | <b>Título da Avaliação:</b> Avaliação de Processos Cognitivos |                                                                |
| <b>Objetivo Geral do Jogo:</b> Avaliar processos cognitivos em um GBA. |                               |                                                                                        | <b>Foco (temática):</b> Processos Cognitivos.                 |                                                                |
| <b>Tipo e formato:</b> Point and Click.                                |                               |                                                                                        | <b>Parte do Neurotasks:</b> Protótipo, Versão 1               |                                                                |
| <b>Disciplinas:</b>                                                    |                               | <b>Público-alvo:</b> adultos.                                                          |                                                               |                                                                |
| <b>Idioma:</b><br>português                                            | <b>Custo:</b> R\$ 7.500.      | <b>Licença:</b>                                                                        | <b>Ano e Versão:</b><br>2025/2026.                            | <b>Plataforma:</b> Web.                                        |
| <b>Disponibilidade:</b> restrita à pesquisa enquanto fase de teste.    |                               |                                                                                        |                                                               | <b>Status:</b> <input type="checkbox"/> Final   x<br>Protótipo |

| II. CONTEXTUALIZAÇÃO                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                     |                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------|-------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| II.1. OBJETIVO DO PROTOCOLO.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                     |                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| <p>Desenvolver e analisar um protótipo digital de <i>game-based assessment</i>, denominado Neurotasks, com o propósito de investigar indicadores potenciais do funcionamento cognitivo de adultos a partir de tarefas digitais estruturadas. O estudo examina, em perspectiva psicométrica, clínica e processual, a relação entre o desempenho nas tarefas e os dados de processo gerados pela interação do participante, incluindo tempo de resposta, precisão, erros, tentativas, latência, sequência de ações e estratégias de resolução. No contexto da avaliação psicológica informatizada, busca-se discutir a aplicabilidade inicial do protótipo, suas evidências preliminares de validade e os caminhos necessários para estudos futuros de precisão, validade convergente, aproximação funcional com situações cotidianas e utilidade clínica.</p> |                                                                     |                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| II. PLANEJAMENTO DA AVALIAÇÃO                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                     |                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| <b>II.2.1. Foco da avaliação (O QUÊ?)</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | Avaliar a jogabilidade do jogo.<br><br>Avaliar as métricas do jogo. | <b>II.2.2. Objetivo da avaliação (POR QUÊ?)</b> | <input checked="" type="checkbox"/> <b>Conhecimento</b> – Identificar indicadores de compreensão, raciocínio e aplicação de informações em tarefas cognitivas estruturadas.<br><br><input checked="" type="checkbox"/> <b>Habilidade</b> – Avaliar habilidades cognitivas a partir do desempenho em tarefas digitais.<br><br><input checked="" type="checkbox"/> <b>Engajamento</b> – Verificar o nível de envolvimento do avaliando com as tarefas e com a dinâmica do jogo. |

|                                     |  |  |                                                                                                                                                                                                                                                                                                                                                                                                   |
|-------------------------------------|--|--|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                     |  |  | <p>☑ <b>Autoavaliação</b> – Favorecer a reflexão do avaliando sobre seu próprio desempenho cognitivo, quando aplicável.</p> <p>☑ <b>Progresso do jogo</b> – Monitorar a evolução do desempenho ao longo das etapas do Neurotasks.</p> <p>☑ <b>Interação com sistema</b> – Analisar padrões de interação com a interface, incluindo tempo de resposta, precisão, erros e estratégias adotadas.</p> |
| <b>II.3. ASPECTOS DE AVALIAÇÃO.</b> |  |  |                                                                                                                                                                                                                                                                                                                                                                                                   |

☑ **Usabilidade:** Avalia a facilidade de uso do Neurotasks, considerando a clareza da interface, a intuitividade dos controles, a fluidez da navegação, a qualidade da interatividade e a efetividade dos feedbacks fornecidos ao usuário durante a execução das tarefas.

☑ **Jogabilidade:**

Analisa a coerência das regras, o sistema de recompensas e punições, bem como a interação do usuário com o ambiente virtual, objetos, conteúdos e atividades propostas, verificando se a dinâmica do jogo sustenta o engajamento e a execução adequada das tarefas cognitivas.

☑ **Acessibilidade:** examina a adequação dos controles, opções de configuração e linguagem utilizada, considerando diferentes perfis de usuários, com foco na redução de barreiras de acesso e na promoção de uma experiência funcional e inclusiva.

☑ **Experiência do Usuário (UX):** avalia indicadores subjetivos e objetivos relacionados à satisfação, nível de desafio, atenção sustentada, imersão, motivação e respostas emocionais durante a interação com o Neurotasks, integrando dados autorrelatados e métricas comportamentais.

☑ **Aspectos Avaliativos:** Analisa a percepção do efeito global do instrumento digital em desenvolvimento sobre o processo de compreensão da Tarefa, considerando a clareza dos objetivos, a progressão das tarefas e a coerência entre desafio proposto e desempenho apresentado.

☑ **Conteúdo:** avalia a adequação dos objetivos avaliativos às tarefas propostas e a efetividade do conteúdo em elicitar os processos cognitivos pretendidos, considerando a consistência teórica e a mensurabilidade dos resultados obtidos.

☑ **Compreensão da tarefa:** mede indicadores de avalia indicadores de compreensão, adaptação à tarefa e percepção de clareza durante a interação com o Neurotasks, quando aplicável. e na compreensão das tarefas, por meio da comparação de dados obtidos antes, durante e após a interação com o Neurotasks, quando aplicável.

☑ **Outros Aspectos – Dados de Processo (Process Data):** Avalia métricas objetivas de interação, como tempo de resposta, número de tentativas, erros, acertos, padrões de escolha e estratégias utilizadas, com finalidade psicométrica e clínica.

## II.4. TÉCNICAS DE AVALIAÇÃO (COMO?)

*As técnicas de avaliação adotadas neste protocolo são definidas com base no Game Design Document (GDD) do Neurotasks, assegurando alinhamento entre objetivos teóricos, mecânicas de jogo, tarefas cognitivas e métricas coletadas. A avaliação fundamenta-se nas teorias dos processos cognitivos, priorizando dados de desempenho e dados de processo (process data).*

### II.4.1. Teorias:

#### ☑ **Teorias sobre Processos Cognitivos**

A avaliação baseia-se em **teorias sobre processos cognitivos**, compreendendo a inteligência como um sistema hierárquico e interdependente de habilidades amplas e específicas. As tarefas do Neurotasks, conforme descritas no GDD, são projetadas para elicitare processos cognitivos, permitindo análise funcional do desempenho ao longo da interação.

☑ **Teorias sobre treinamento:** adota-se o princípio de ativação e mobilização de habilidades cognitivas, compreendendo o desempenho como expressão funcional de habilidades cognitivas previamente consolidadas, e não como resultado de processos instrucionais ou de treino. O desempenho observado é interpretado a partir da resolução de situações-problema ecologicamente contextualizadas, concebidas como tarefas únicas e não repetitivas, conforme descrito no Game Design Document (GDD) do Neurotasks, preservando a validade da inferência psicométrica ao evitar efeitos de aprendizagem ou prática (Ackerman, 1996; Ackerman & Kanfer, 2009).

### ☑ Teorias sobre motivação, engajamento ou satisfação

O engajamento é compreendido como uma variável contextual e moderadora do desempenho cognitivo, inferida indiretamente por meio de métricas comportamentais e dados de processo, tais como tempo de resposta, persistência na tarefa, abandono e padrões de tentativa, conforme operacionalizado no Game Design Document (GDD) do Neurotasks. Nessa abordagem, o engajamento não constitui um construto avaliativo primário, mas um fator que influencia a mobilização de recursos cognitivos durante a execução das tarefas, sem substituir a mensuração direta do desempenho cognitivo (Mayer, 2014).

### ☑ Teorias sobre feedback

No Neurotasks, o feedback é definido exclusivamente como feedback informativo de desempenho, destinado a possibilitar que o jogador tome ciência de seu próprio resultado e acompanhe sua execução ao longo da tarefa. Esse feedback não possui função instrucional, corretiva ou de reforço de aprendizagem, não orienta estratégias de resposta nem altera a progressão da atividade, sendo apresentado de modo a não interferir nos processos cognitivos mobilizados, preservando, assim, a validade da mensuração cognitiva conforme os pressupostos psicométricos (Hattie & Timperley, 2007).

☑ **Psicometria em Rede (Network Psychometrics)** — A modelagem dos dados do Neurotasks fundamenta-se na Psicometria em Rede, abordagem que concebe construtos psicológicos como sistemas de componentes interdependentes, cujas interações produzem o desempenho observado. Nessa perspectiva, habilidades cognitivas, indicadores comportamentais e métricas de processo são representados como nós de uma rede dinâmica, permitindo analisar padrões de conectividade, centralidade e interação funcional entre componentes cognitivos. Essa abordagem está alinhada a modelos mutualistas da inteligência e à literatura contemporânea em psicometria computacional (van der Maas et al., 2006; Borsboom, 2017; Epskamp et al., 2018).

#### II.4.2. Técnicas, instrumentos e artefatos:

- ☑ **Observação Eletrônica (automática):** Registro contínuo e automatizado das interações do usuário com o Neurotasks, incluindo tempo de resposta, acertos, erros, padrões de escolha, sequência de ações e estratégias implícitas, conforme definido no GDD.
- ☑ **Teste Baseado em Jogo Digital (*game-based assessment*):** aplicação do Neurotasks como instrumento de avaliação cognitiva, no qual cada tarefa representa um cenário funcional projetado para elicitare habilidades cognitivas, conforme mapeamento teórico-operacional descrito no GDD.
- ☑ **Pré-teste:** coleta de informações iniciais para contextualização do desempenho cognitivo, sem finalidade instrucional, respeitando os critérios psicométricos do protocolo.
- ☑ **Pós-teste (quando aplicável):** Análise comparativa de indicadores de desempenho e processo, exclusivamente para fins exploratórios e de validação do instrumento, conforme delineado no GDD.
- ☑ **Questionário de Perfil:** Coleta de dados sociodemográficos, educacionais e tecnológicos relevantes para controle de variáveis intervenientes na análise psicométrica.
- ☑ **Cadastro de Perfil Cognitivo:** Organização estruturada dos indicadores extraídos do jogo, permitindo análise individual e grupal do funcionamento cognitivo.
- ☑ **Análise de Dados de Processo (Process Data Analytics):** Extração e modelagem de métricas temporais, comportamentais e estratégicas, utilizadas como indicadores observáveis para inferência de habilidades cognitivas e para construção de redes psicométricas.
- ☑ **Modelagem em Rede Psicométrica:** Análise das inter-relações entre habilidades cognitivas, tarefas do jogo e métricas de processo, permitindo identificar padrões de centralidade, conectividade e dependência funcional entre componentes cognitivos.

☑ **Matriz ECD**, definida no GDD, que organiza a correspondência entre habilidades cognitivas, mecânicas do jogo e indicadores observáveis, garantindo rastreabilidade teórica e metodológica do protocolo.

## **II.5. PARTICIPANTES.**

A avaliação do instrumento digital em desenvolvimento Neurotasks envolve a participação de usuários-avaliadores e usuários-avaliados, caracterizados como indivíduos que interagem diretamente com o sistema com a finalidade de fornecer dados sobre o funcionamento do jogo e outro tipo de usuário sobre o desempenho cognitivo, métricas de processo e informações complementares para análise psicométrica, conforme previsto no Game Design Document (GDD).

|                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|---------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Usuário-Avaliador</b></p> | <p>Usuário-Avaliador</p> <p><b>5.1. Fase Piloto</b></p> <p><b>Perfil dos participantes:</b> adultos, de ambos os sexos, com escolaridade mínima compatível com o uso de tecnologias digitais, selecionados para testar o funcionamento geral do Neurotasks. Nesta fase, os participantes atuam como usuários-avaliadores com foco na verificação de usabilidade, estabilidade do sistema, clareza das tarefas e adequação das métricas de processo. Nesta fase, os avaliadores foram os orientadores do projeto.</p> <p><b>Total de participantes:</b> 49.</p> <p><b>Total de Pessoas com Deficiência (PcD):</b></p> <p><input type="checkbox"/> Não questionado</p> <p><input checked="" type="checkbox"/> Não há</p> <p><b>II.5.2. Fase de Validação</b></p> <p><b>Perfil dos participantes/juízes:</b> adultos, de ambos os sexos, recrutados para a coleta sistemática de dados de desempenho cognitivo e métricas de processo, com finalidade psicométrica. Nesta fase, os participantes/juízes atuam como usuários-avaliadores do Neurotasks, permitindo análises de validade, precisão e modelagem em redes psicométricas, conforme os pressupostos teóricos do modelo de processos cognitivos.</p> <p><b>Total de participantes:</b></p> <p><b>Total de menores de idade:</b></p> <p><input checked="" type="checkbox"/> Não há <input type="checkbox"/> Há, quantidade:</p> |
|---------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

**Total de Pessoas com Deficiência (PcD):**

☐ Não questionado

☐ Não há

☐ Há, quantidade:

**Em caso afirmativo, especificar a(s) condição(ões):**

☐ PcDV – Deficiência Visual/Baixa Visão

☐ PcDF – Deficiência Física/Mobilidade Reduzida

☐ PcDI – Deficiência Intelectual

☐ PcTEA – Transtorno do Espectro do Autismo

☐ PcSurdoCegueira – Surdocegueira

☐ PcDA – Deficiência Auditiva

☐ PcAHSD – Altas Habilidades/Superdotação

☐ Outro(s):

### **Papéis dos Participantes**

Os participantes da avaliação do Neurotasks podem assumir diferentes papéis, conforme os objetivos específicos de cada etapa do estudo. Os usuários-avaliadores podem ser classificados como usuários-piloto, atuando como avaliados-juizes que podem ser especialistas no domínio da psicologia, neurociências e avaliação psicológica, bem como especialistas de apoio, incluindo profissionais das áreas de jogos digitais, tecnologia, usabilidade e acessibilidade, entre outros, quando pertinente.

Para fins de caracterização amostral, um formulário de perfil poderá ser aplicado previamente ao início das sessões de teste, com o objetivo de coletar informações relevantes sobre o histórico, a experiência prévia e o papel desempenhado pelo participante no processo avaliativo.

☐ **Usuário-piloto (orientadora e coorientador)**

Quantidade: 2

☐ **Usuário-Juiz (Especialistas)**

Quantidade:

☐ **Especialista em Jogos Digitais**

Quantidade: 1

☐ **Especialista em Tecnologia**

Quantidade: 1

☐ **Especialista em Usabilidade e/ou Acessibilidade**

Quantidade: 1

|                                        |                                                                                                                                             |                                                                                                                                 |
|----------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|
|                                        | <input type="checkbox"/> <b>Usuário-Final (avaliado)</b><br>Quantidade:<br><br><input type="checkbox"/> <b>Outro(s):</b><br>Quantidade:     |                                                                                                                                 |
| <b>Equipe de Avaliação<sup>3</sup></b> | <input type="checkbox"/> <b><u>Pesquisador</u></b> . Quantidade: 1<br><br><input type="checkbox"/> <b><u>Observador</u></b> . Quantidade: 1 | <i>O <u>pesquisador</u> é responsável por conduzir a avaliação. O <u>observador</u> é responsável por observar a interação.</i> |
| <b>II.6. PERÍODO E LOCAL.</b>          |                                                                                                                                             |                                                                                                                                 |

As sessões de avaliação do **Neurotasks** ocorrerão em ambiente **remoto**, conforme o planejamento metodológico definido no Game Design Document (GDD), podendo ser realizadas de forma **síncrona** e/ou **assíncrona**, a depender da etapa do estudo e do papel desempenhado pelo participante. As atividades serão conduzidas em plataforma digital segura, assegurando registro adequado dos dados de desempenho e das métricas de processo.

#### Síncrono (comunicação em tempo real)

|                             |                                                                                                                                                                                                                                             |
|-----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>1</p> <p>Ex</p> <p>.</p> | <p>Dia: 15/01/2026, hora: 14h00.</p> <p>Modalidade: remota, via plataforma de videoconferência (ex.: Google Meet).</p> <p>Duração aproximada: 60 minutos.</p> <p>Participantes: usuários-avaliadores (especialistas e usuários-finais).</p> |
| <p>2</p>                    | <p>Dia:            Hora:</p> <p>Modalidade: remota, via site.</p> <p>Duração aproximada:            minutos.</p> <p>Participantes:</p>                                                                                                      |
| <p>3</p>                    | <p>Dia:            Hora:</p> <p>Modalidade: remota, via site.</p> <p>Duração aproximada:            minutos.</p> <p>Participantes:</p>                                                                                                      |
| <p>4</p>                    | <p>Dia:            Hora:</p> <p>Modalidade: remota, via site.</p> <p>Duração aproximada:            minutos.</p>                                                                                                                            |

|                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|
|                                                                                                                                                                                                                                                                                                                                                     | Participantes:                                                                                                      |
| 5                                                                                                                                                                                                                                                                                                                                                   | Dia:            Hora:<br>Modalidade: remota, via site.<br>Duração aproximada:            minutos.<br>Participantes: |
| Anotações:                                                                                                                                                                                                                                                                                                                                          |                                                                                                                     |
| <b>Assíncrono (comunicação não simultânea).</b><br>Quando aplicável, a avaliação poderá incluir etapas assíncronas, nas quais os recursos necessários à interação com o Neurotasks serão disponibilizados aos participantes para acesso em período previamente definido.                                                                            |                                                                                                                     |
| <b>Descrição:</b><br>Dia(s):<br>Período/horário estimado:<br>Meio de envio/acesso:                                                                                                                                                                                                                                                                  |                                                                                                                     |
| <b>II.7. MATERIAIS.</b>                                                                                                                                                                                                                                                                                                                             |                                                                                                                     |
| Os materiais utilizados consistem no <b>protótipo funcional do Neurotasks</b> , recurso digital interativo desenvolvido para avaliação cognitiva, conforme descrito no Game Design Document (GDD). O protótipo é composto por módulos e tarefas digitais contextualizadas, com registro automatizado de dados de desempenho e métricas de processo. |                                                                                                                     |

Como materiais de apoio, serão utilizados **logs de interação, dashboards de monitoramento e formulários eletrônicos** para caracterização dos participantes. A interação com o sistema ocorrerá por meio de **computador ou tablet**, com acesso à internet. Todos os materiais destinam-se exclusivamente às atividades previstas no protocolo.

### III. PROCEDIMENTOS

A avaliação do Neurotasks será realizada de forma remota, por meio de sessões síncronas e/ou assíncronas, conforme a etapa do estudo, utilizando protótipo funcional do sistema, com registro automatizado de dados de desempenho e métricas de processo.

#### **Procedimentos Síncronos**

**Resumo:** Sessão remota mediada por pesquisador para interação do usuário-avaliador com o protótipo do Neurotasks.

**Modelo de objeto:** protótipo funcional do Neurotasks.

**Composição da sessão:** 01 usuário-avaliador, 01 pesquisador e 01 observador.

#### **Procedimentos:**

- ☒ Aplicação do TCLE e questionário de perfil.
- ☒ Breve apresentação do sistema pelo pesquisador.
- ☒ Aplicação do cadastro.
- ☒ Interação do usuário-avaliador com o protótipo, com compartilhamento de tela e registro de áudio (webcam opcional).
- ☒ Registro automático dos dados de desempenho e métricas de processo.
- ☒ Aplicação de questionário de reação/autoavaliação e pós-teste, quando previsto (juízes).

☒ Realização de entrevista breve ou *debriefing*, quando aplicável.

☒ Encerramento da sessão.

**Resultados:**

☒ Registros de interação (logs e gravações).

☒ Planilhas de respostas aos questionários.

☒ Relatório sintético das avaliações. (Quando aplicável)

**Procedimentos Assíncronos**

**Resumo:** Interação autônoma do usuário-avaliador com o Neurotasks, sem mediação em tempo real.

**Modelo de objeto:** protótipo funcional do Neurotasks.

**Composição:** usuários-avaliadores, de forma individual.

**Procedimentos:**

☒ Envio de instruções e acesso ao sistema por meio eletrônico.

☒ Aceite do TCLE e preenchimento de questionários previstos.

☒ Realização das tarefas no Neurotasks de forma autônoma.

☒ Registro automático dos dados de desempenho e métricas de processo.

☒ Entrevista ou *debriefing* síncrono, quando necessário.

**Resultados:**

☒ Logs de interação.

☒ Planilhas de questionários.

☒ Relatório sintético das avaliações.

#### IV. MEDIDAS E INSTRUMENTOS

##### **SÍNCRONO (comunicação em tempo real)**

##### **Repositório de modelos e artefatos:**

<https://bit.ly/Pro-AvaliaJS-Artefatos>

<http://www.neurotasks.com.br/>

<https://acesse.one/gamedesigndocument>

##### **Planejamento e Execução**

##### **☒ Termo de Consentimento Livre e Esclarecido (TCLE) e Questionário de Perfil do Usuário:**

Instrumento eletrônico aplicado previamente ao início da sessão.

##### **☒ Roteiro de Observação:**

Instrumento padronizado para registro de comportamento, tempo de execução e verbalizações do usuário-avaliador durante a interação.

##### **[ROTEIRO DE OBSERVAÇÃO – NEUROTASKS](#)**

##### **☒ Termo de Livre consentimento - TCLE:**

Aplicado para os usuários-finais

##### **[Termo de Consentimento Livre e Esclarecido – TCLE](#)**

##### **☒ Questionário de Pré-Teste:**

Aplicado antes da interação, para o cadastro do usuário.

##### **[QUESTIONÁRIO DO PARTICIPANTE](#)**

##### **☒ Questionário de Pós-Teste:**

Aplicado após a interação, para coleta de percepções dos juízes.

## INSTRUMENTO B — AVALIAÇÃO POR JUÍZES / ESPECIALISTAS

### ☒ **Entrevista Semiestruturada / Debriefing:**

Realizada após a interação, com base nas seguintes questões norteadoras:

- Impressões gerais sobre o objeto avaliado;
- Destaques da experiência de avaliação;
- Aspectos positivos identificados;
- Aspectos negativos identificados.

[Questionário – Avaliação do Neurotasks - Usuário -final](#)

[Instrumento A – Avaliação da Experiência do usuário, Usabilidade e Função Avaliativa](#)

## **ASSÍNCRONO (Comunicação não simultânea)**

### **Repositório de modelos e artefatos:**

<https://bit.ly/Pro-AvaliaJS-Artefatos>

<http://www.neurotasks.com.br/>

### **Planejamento e Execução**

### ☒ **Termo de Consentimento Livre e Esclarecido (TCLE) e Questionário de Perfil do Usuário:**

Aplicados eletronicamente antes do início da interação.

[Termo de Consentimento Livre e Esclarecido – TCLE](#)

### ☒ **Roteiro do Usuário:**

Documento com instruções padronizadas para acesso e realização das tarefas no Neurotasks.

[Roteiro para o usuário do Neurotasks](#)

**☑ Documento de Descrição do Objeto Avaliado:**

Material informativo de apoio.

[Documento de Descrição do Objeto Avaliado – Neurotasks](#)

**☑ Questionário de Avaliação (Reação):**

Aplicado após a conclusão da interação.

[Questionário – Avaliação do Neurotasks - Usuário -final](#)

**☑ Questionário de Pré-Teste:**

Aplicado previamente à interação, quando previsto no delineamento do estudo.

[QUESTIONÁRIO DO PARTICIPANTE](#)

**☑ Questionário de Pós-Teste:**

Aplicado após a interação, quando previsto no delineamento do estudo.

[Documento de Descrição do Objeto Avaliado – Neurotasks](#)

**Resultados**

**☑ Diário de Bordo:** registros eletrônicos da interação e anotações dos pesquisadores, quando aplicável.

**☑ Relatório de Resultados da Avaliação:** Documento sintético com a sistematização dos dados coletados.

**V. TAREFAS**

***Nota.* Embora o protocolo original contemple procedimentos síncronos, no presente estudo a aplicação foi realizada de forma assíncrona e autônoma.**

**SÍNCRONO (comunicação em tempo real)**

Durante a sessão síncrona, o usuário-avaliador interage com o protótipo funcional do Neurotasks, seguindo um roteiro estruturado de tarefas digitais contextualizadas, apresentado pelo pesquisador. As tarefas consistem na resolução de situações-problema simuladas, que exigem tomada de decisão, organização de informações e execução de ações no ambiente digital, conforme instruções exibidas na interface.

O usuário-avaliador realiza as tarefas uma única vez, de forma contínua, podendo verbalizar seus pensamentos durante a execução, quando aplicável. O pesquisador acompanha a interação em tempo real, sem fornecer orientações instrucionais ou corretivas, intervindo apenas para garantir o cumprimento do roteiro. Durante a execução, o sistema registra automaticamente dados de desempenho e métricas de processo.

### **ASSÍNCRONO (comunicação não simultânea)**

Na modalidade assíncrona, o usuário-avaliador acessa o protótipo funcional do Neurotasks de forma autônoma, seguindo um roteiro padronizado de tarefas previamente disponibilizado pelo pesquisador. As tarefas envolvem a interação com cenários digitais e a resolução de situações-problema apresentadas pelo sistema, conforme instruções exibidas na tela.

O usuário-avaliador realiza as tarefas individualmente, sem mediação em tempo real, respeitando a sequência proposta e sem repetição das atividades. Todo o desempenho e as interações são registrados automaticamente pelo sistema, para posterior análise.

#### **Nota técnica**

Em ambas as modalidades, as tarefas não possuem caráter instrucional ou de treinamento e são destinadas exclusivamente à avaliação do desempenho cognitivo, conforme o delineamento metodológico do Neurotasks.

## Referências

### Referências

- Köhler, B., Haladjian, J., Simeonova, B., & Ismailović, D. (2012). Feedback in low- vs. high-fidelity visuals for game prototypes. In *Proceedings of the Second International Workshop on Games and Software Engineering: Realizing User Engagement with Game Engineering Techniques (GAS '12)* (pp. 42–47). IEEE Press.
- Melo, A. M., Crespo, I. M., Medeiros, G. C., & Oliveira, A. B. (2020). Estratégias remotas à avaliação de interfaces de usuário. In *Anais da IV Escola Regional de Engenharia de Software*. Universidade Federal do Pampa.
- Oliveira, R. N. R.(2019). Avaliações em jogos educacionais: instrumentos de avaliação da reação, aprendizagem e comparação de jogos. In *Anais do XXX Simpósio Brasileiro de Informática na Educação (SBIE 2019)* (pp. 972–981).

## Apêndice C

### Ficha de validade de conteúdo: avaliação por juízes especialistas

**Instrumento avaliado:** Neurotasks – Avaliação técnica, psicométrica e de design

**Objetivo do instrumento:** avaliar a adequação técnica, psicométrica e conceitual de um protótipo de *game-based assessment* destinado à avaliação de processos cognitivos em adultos, fundamentado na abordagem *evidence-centered design*.

Identificação do juiz

- Formação: \_\_\_\_\_
- Área(s) de atuação: \_\_\_\_\_
- Tempo de experiência profissional: \_\_\_\_\_
- Data: \_\_\_\_ / \_\_\_\_ / \_\_\_\_

Instruções

Avalie cada item quanto à sua adequação ao construto, clareza semântica e pertinência para avaliação psicológica, considerando sua experiência profissional.

Relevância refere-se à importância do item para o construto avaliado; clareza diz respeito à compreensibilidade semântica do item; e representatividade indica o quanto o item expressa adequadamente o domínio teórico ao qual pertence.

**Utilize a escala abaixo:**

**Escala de Concordância (Likert – 5 pontos)**

- 1 = Discordo totalmente  
 2 = Discordo  
 3 = Nem concordo nem discordo  
 4 = Concordo  
 5 = Concordo totalmente

**Parte I - Avaliação dos itens**

**Domínio 1 – Usabilidade e Design da Interface**

| <b>Item</b> | <b>Descrição do Item</b>                                           | <b>Relevância e importância do item para o construto avaliado.</b>                                                                        | <b>Clareza</b><br>Compreensibilidade semântica do item.                                                                                   | <b>Representatividade</b><br>O item expressa adequadamente o domínio teórico ao qual pertence.                                         |
|-------------|--------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| <b>J1</b>   | O layout das telas é organizado e funcional.                       | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 |
| <b>J2</b>   | As cores utilizadas são adequadas ao propósito avaliativo do jogo. | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 |
| <b>J3</b>   | Há bom aproveitamento do espaço da tela.                           | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 |
| <b>J4</b>   | Os botões e menus funcionaram corretamente.                        | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 |
| <b>J5</b>   | A navegação é consistente e previsível.                            | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 |
| <b>J6</b>   | Os controles são intuitivos para usuários leigos.                  | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 |



## Domínio 2 – Jogabilidade e Mecânicas

| Item | Descrição do Item                                                 | Relevância                                                                                                                                | Clareza                                                                                                                                   | Representatividade                                                                                                                        |
|------|-------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| J7   | As regras das tarefas são claras e bem definidas.                 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 |
| J8   | O fluxo das tarefas é coerente com os objetivos do jogo.          | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 |
| J9   | O tempo de resposta exigido é adequado às demandas cognitivas.    | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 |
| J10  | As mecânicas do jogo estão bem integradas às tarefas avaliativas. | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 |

## Domínio 3 – Acessibilidade

| Item | Descrição do Item                                                         | Relevância                                                                                                                                | Clareza                                                                                                                                   | Representatividade                                                                                                                     |
|------|---------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| J11  | A linguagem utilizada é clara e apropriada ao público-alvo.               | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 |
| J12  | O tamanho das fontes é adequado.                                          | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 |
| J13  | O jogo apresenta recursos suficientes para diferentes perfis de usuários. | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 |
| J14  | Os feedbacks visuais e sonoros são adequados e não excessivos.            | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 |

#### Domínio 4 – Aspectos Psicométricos e Avaliativos

| Item | Descrição do Item                                                                  | Relevância                                                                                                                                | Clareza                                                                                                                                   | Representatividade                                                                                                                     |
|------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| J15  | O jogo é adequado para avaliação psicológica.                                      | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 |
| J16  | As tarefas permitem observar diferenças individuais de desempenho.                 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 |
| J17  | As situações do jogo apresentam aproximação funcional com atividades cotidianas.   | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 |
| J18  | O feedback não interfere na mensuração do desempenho cognitivo.                    | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 |
| J19  | A tela/relatório de resultados representa adequadamente o desempenho do avaliado.  | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 |
| J20  | As instruções antes do jogo foram adequadas para o desempenho do jogo.             | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 |
| J21  | A linguagem utilizada nas instruções do jogo é neutra do ponto de vista de gênero. | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4<br><input type="checkbox"/> 5 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 |

## **Parte II – Avaliação semântica**

Caso algum item apresente problemas de clareza, descreva abaixo:

Item(s): \_\_\_\_\_

Sugestão de reformulação:

---

---

## **Parte III – Avaliação global**

O conjunto de itens representa adequadamente a proposta do Neurotasks?

☐ Sim ☐ Parcialmente ☐ Não

Sugestões gerais:

---

---

Assinatura do Juiz: \_\_\_\_\_

**Data:** \_\_\_\_ / \_\_\_\_ / \_\_\_\_

## **Apêndice D**

### Instruções ao jogador

#### **Fase do Café**

Objetivo: preparar café utilizando a cafeteira elétrica, seguindo corretamente todas as etapas do processo.

Antes de começar:

- Leia atentamente todas as instruções antes de iniciar.
- Observe os itens disponíveis e certifique-se de que estão limpos e prontos para uso.
- Itens necessários
  1. Cafeteira elétrica
  2. Água da torneira
  3. Cápsula de café
  4. Jarra de vidro da cafeteira
  5. Ligar o botão da cafeteira.

Passo a passo

1. Coloque a água no reservatório atrás da cafeteira.

Despeje a água da torneira no compartimento da cafeteira. Leve o compartimento até a torneira. Ligue a torneira.

2. Adicione a cápsula de café na cafeteira.
3. Ligue a cafeteira.
4. Aguarde o preparo.

5. Terminou a tarefa, vá até a próxima.

Quando o preparo estiver concluído, dirija-se a outro compartimento da casa.

Fase da Higiene Bucal – Escovar os Dentes

Objetivo:

Realizar corretamente a higiene bucal, escovando os dentes.

Itens necessários

- Escova de dentes
- Creme dental
- Copo
- Pia

Passo a passo

- Pegue a escova de dentes e o creme dental
- Coloque o creme na escova
  - Objetivo: escove os dentes até a sujeira desaparecer
- Guarde a escova
- Terminou a tarefa, vá até a próxima.

Quando o preparo estiver concluído, se dirija até outro compartimento da casa.

Fase do Computador – Desbloqueio por Cálculos

Como funciona

- O computador está bloqueado por um sistema de segurança baseado em cálculos.
- Cada resposta correta libera uma etapa do desbloqueio.
- Respostas incorretas exigem nova tentativa.

Passo a passo

- Aperte no símbolo de dois passinhos para começar.

Um cálculo matemático aparecerá na tela do computador.

- Identifique a operação.

Verifique se o cálculo envolve: adição. (+)

- Realize o cálculo mentalmente

Resolva o cálculo com calma, evitando respostas impulsivas.

- Digite a resposta

Insira o resultado do cálculo no campo indicado.

- Confirme a resposta: pressione o botão de confirmação para validar sua resposta.

- Acompanhe o feedback

- Resposta correta: o sistema avança para o próximo cálculo ou libera parte do acesso.

- Resposta incorreta: o sistema solicita uma nova tentativa.

- Complete a sequência

Continue resolvendo os cálculos até concluir todas as etapas do desbloqueio.

- Terminou a tarefa, separe as pastas.

Fase dos Documentos – Separação e Organização

Como funciona:

1. Os documentos estão misturados e precisam ser organizados.
2. Cada documento pertence a uma categoria específica.
3. A organização correta permite avançar na tarefa.

Passo a passo

I. Observe todos os documentos.

Analise os documentos disponíveis na mesa ou na tela.

II. Leia as informações principais.

Preste atenção em títulos, datas, nomes, símbolos ou cores que ajudem na identificação.

III. Identifique os critérios de separação.

Os documentos podem ser organizados por:

a. Tipo de documento

b. Categoria

c. Data

d. Importância

e. Destino correto

IV. Selecione um documento por vez

V. Classifique o documento: arraste ou coloque o documento no local correspondente à sua categoria.

VI. Confirme a organização.

Verifique se o documento foi colocado no local correto antes de seguir para o próximo.

VII. Continue até finalizar.

Repita o processo até que todos os documentos estejam organizados.