

Pensando a Análise de Protocolo, seus contextos e variações.

Analyzing the Think-aloud protocol method, its contexts and variations.

KELLER, Luciana; Doutora em Design; UERJ/ESDI

keller.luciana@gmail.com.br

Atualmente, a aplicação da Análise de Protocolo (*Think-aloud Protocol*) se concentra nos testes de usabilidade e no desenvolvimento de interfaces. No entanto, o método tem sua origem, sob denominações diversas, nos estudos de Neurolinguística e de Psicologia Cognitiva no início do século passado, que exploravam o funcionamento da mente humana por meio da fala. O presente artigo discute este método, apresentando sua trajetória até Design, abordando suas potencialidades e críticas. Acredita-se que as questões levantadas nesta discussão são enriquecedoras para pesquisadores que desejem reproduzir ou adaptar a análise de protocolo. Dentre estas questões, destacam-se: as diferenças entre sua aplicação original e no Design; e as variações desenvolvidas para contornar limitações metodológicas. Desse modo, busca-se auxiliar na identificação de parâmetros relevantes para inclusão do usuário no projeto, além de servir como um referencial teórico para uma melhor compreensão do desenvolvimento de testes de usabilidade no geral.

Palavras-chave: Análise de Protocolo; Ferramentas de pesquisa em Design; Usabilidade.

Currently, the application of Think-aloud Protocol (or Protocol Analysis) focuses on usability tests and interface development. However, the method has its origin (under different names) in the studies of Neurolinguistics and Cognitive Psychology at the beginning of the last century, exploring the functioning of the human mind through speech. This article discusses this method, presenting its trajectory to Design, its potentialities and criticisms. We believe that the issues raised in this discussion are enriching for researchers who wish to reproduce or adapt the protocol analysis method. Among these issues, the following are highlighted: the differences between its original application and in Design; variations developed to bypass methodological limitations. Therefore, we seek to assist in the identification of relevant parameters for the inclusion of the user in the project, in addition to provide theoretical references for a better understanding of the development of usability tests in general.

Keywords: Think-aloud Protocol; Design research tools; Usability.

1 Introdução

O presente artigo tem por objetivo apresentar o método *Think-aloud Protocol*, ou Análise de Protocolo, discutindo sua aplicação ao campo do Design. Acredita-se que as questões levantadas ao longo desta discussão possam ser enriquecedoras para os pesquisadores que desejem se apropriar desta técnica de pesquisa, seja para reproduzi-la ou adaptá-la. Dentre estas questões, encontram-se dois pontos centrais de interesse: (1) as diferenças entre a aplicação atual do método ao campo da Usabilidade em relação às suas origens, direcionadas às necessidades de outras áreas, como a psicologia cognitiva; e (2) as variações desenvolvidas ao longo do tempo para contornar limitações metodológicas, influenciadas pela importância de cada variável em seu contexto de aplicação. Através de um breve apanhado histórico baseado em uma revisão de literatura, pretende-se elencar os elementos-chave que caracterizam o método, bem como as consequentes alterações sofridas por ele ao longo do tempo. Acredita-se que, perpassando os elementos que motivaram as adaptações e desdobramentos do Think-aloud Protocol, este artigo pode auxiliar o pesquisador/designer na identificação de parâmetros relevantes para a definição da melhor técnica a ser empregada para inclusão do usuário no projeto, além de servir como um referencial teórico para uma melhor compreensão do desenvolvimento de testes de usabilidade no geral.

Atualmente, a principal aplicação da análise de protocolo se concentra nos testes de usabilidade, em especial no desenvolvimento de interfaces, apresentando como principais vantagens o baixo custo e a confiabilidade dos dados coletados (visto que derivam da interação do produto com usuários reais). Além de facilitarem sua execução, estes fatores possibilitam sua aplicação à avaliação de produtos de naturezas diversas, como artefatos físicos e digitais, customização e sincronização de aparelhos eletrônicos, manuais de instruções e até mesmo questionários.

A técnica é aplicada nas fases de Concepção/Prototipagem e de Avaliação/Refinamento do projeto de produtos ou sistemas, tendo tanto o intuito de identificar falhas e corrigi-las, quanto de gerar novas ideias a partir das necessidades demonstradas pelos usuários. Desse modo, possibilita-se que seja testado apenas um elemento do protótipo, e não apenas o todo, podendo também ser utilizado para análise de produtos concorrentes ou gerar ideias a partir de problemas apresentados por uma categoria de produtos de modo geral.

Em resumo, tais testes consistem em propor a um potencial usuário de um produto ou sistema (geralmente um protótipo) que se pretende analisar, para que verbalize constantemente seus pensamentos (pense em voz alta – em inglês, think-aloud) durante a execução de uma dada tarefa (ou conjunto de tarefas), expressando dificuldades e impressões ao longo da interação. O teste é aplicado por um profissional, em um ambiente controlado, com registros em vídeo, áudio e (quando aplicável) gravações de tela. Em alguns casos, utiliza-se um vidro de sentido único para que outras pessoas possam acompanhar o andamento do teste, sem que o participante se sinta intimidado pela plateia. Segundo Lewis e Rieman (1993) :

A ideia básica do thinking aloud é bem simples. Pede-se aos usuários para que executem uma tarefa-teste, mas também pede-se para que falem com você enquanto trabalham. Peça para que lhes digam o que estão pensando: o que estão tentando fazer, perguntas que surgem enquanto trabalham, coisas que eles lêem. Você pode fazer uma gravação dos comentários deles ou pode apenas tomar notas. Isso deve ser feito de tal modo que seja possível distinguir o que estão fazendo e onde seus comentários se inserem na sequência. (p.83)

Há um cuidado especial em minimizar a interferência por parte do pesquisador/moderador que acompanha a sessão, dirigindo-se ao participante apenas quando há pausas na fala, para lembrá-lo de manter suas verbalizações constantes. Isso ocorre tanto com o objetivo de não influenciar a fala dos participantes ou sua percepção em relação ao produto, quanto para não influenciar a carga cognitiva exigida e/ou o tempo de resolução da tarefa, nos casos em que estes fatores são relevantes. Apesar desse risco de enviesamento das respostas, a presença de um mediador é essencial para garantir que os participantes não se sintam perdidos, inseguros, dispersos, ou avaliados.

No geral, as variações do teste são simples e baratas de serem executadas (se comparadas a outros testes de usabilidade), pois não são necessários equipamentos caros, nem um grande número de participantes. Além disso, a técnica possui como principal diferencial a confiabilidade das informações obtidas, visto que partem diretamente da interação com o usuário. Por meio da Análise de Protocolo, é possível identificar falhas no produto/protótipo através da interação direta com o usuário, em oposição às técnicas que envolvem critérios definidos a priori e/ou participação de especialistas de determinada área, como a avaliação heurística.

2 A trajetória o método

Com denominações diversas, a análise de protocolo (*Think-aloud Protocol*, *Thinking Aloud Method*, *Protocol Analysis*, dentre variações similares) tem sua origem nos estudos de Neurolinguística e de Psicologia Cognitiva no início do século passado, que buscavam explorar o funcionamento da mente humana, partindo de informações obtidas através da fala. Neste primeiro momento, o método consistia em pedir a um voluntário que expusesse seus pensamentos em voz alta enquanto executasse uma determinada tarefa, como resolver um problema matemático/lógico, em um ambiente controlado. A intenção da aplicação da técnica era compreender os mecanismos através dos quais o cérebro processa e armazena informações. Ou seja, diferente de sua aplicação usual ao Design, o foco estava nos processos cognitivos do indivíduo, e não nos artefatos (ou produtos) envolvidos na tarefa.

Com o crescente interesse em diversas áreas pelo funcionamento da mente humana, pesquisadores passaram a importar o método para seus respectivos campos de conhecimento, como no estudo de processos de tomada de decisão, aprendizagem (dos mais variados assuntos e recortes), leitura e escrita, memorização, compreensão de mapas e imagens, preenchimento de questionários, etc. Naturalmente, essa expansão fez com que tais pesquisadores buscassem adequar o emprego do *Think-aloud* às necessidades específicas de suas áreas. Nesse contexto transitório, o trabalho de Ericsson e Simon (1993) representa um marco, pois a obra se propunha a servir como uma base metodológica com diretrizes para a utilização dessas verbalizações (ou *verbal reports*) como fonte primária e confiável de dados para diversas áreas, empregando o termo análise de protocolo (Protocol Analysis).

Ao ser trazida da Psicologia Cognitiva para o campo da Usabilidade e do Design, a análise de protocolo sofreu diversas alterações, como a incorporação de dispositivos de eye-tracking e outros instrumento de mapeamento, além de variações na dinâmica de aplicação como a divisão em etapas (concorrente e retrospectivo) ou a participação de mais de um voluntário (avaliação cooperativa), detalhados mais à frente. A análise de protocolo passou a ser empregada no desenvolvimento de produtos de diversas naturezas, sendo aplicável tanto a avaliação de artefatos físicos, quanto a sistemas operacionais, interfaces, customização e sincronização de aparelhos eletrônicos, manuais de instruções e até mesmo questionários. De modo mais significativo que em outras áreas, o método sofreu uma drástica mudança: o objetivo que antes era analisar um processo cognitivo passa a ser a avaliação do produto. Essa

transição de objeto resulta em um problema central (de especial interesse para esse trabalho) que são as inconsistências metodológicas.

3 O Think-aloud na Psicologia cognitiva

Dentro da Psicologia Cognitiva a técnica se propunha a explorar o funcionamento da mente humana, especialmente o processamento de informações pelo cérebro. Como diferencial, as informações eram obtidas através de um comportamento observável, no caso, através da fala (*verbal reports* ou *talk-aloud reports*). Contudo, se questionava a validade de dados dessa natureza em pesquisas científicas, principalmente na vertente behaviorista. O principal foco de crítica ao uso de verbalizações como fonte primária de dados em pesquisas científicas dessa natureza era a confiabilidade e a veracidade dessas informações, visto a complexidade dos processos mentais (introspectivos) frente às limitações impostas pela verbalização (um comportamento externo observável).

Antes dessa crítica, o uso de dados dessa natureza era tido como um conhecimento privilegiado, como um insight da experiência do indivíduo e do funcionamento da mente humana. Porém, com a ascensão do behaviorismo, a fala passou a ser associada à subjetividade e, conseqüentemente, o uso de verbalizações (para além de respostas objetivas) passou a ser desincentivado, pois seria pouco confiável cientificamente. Segundo Ericsson e Simon (1993) :

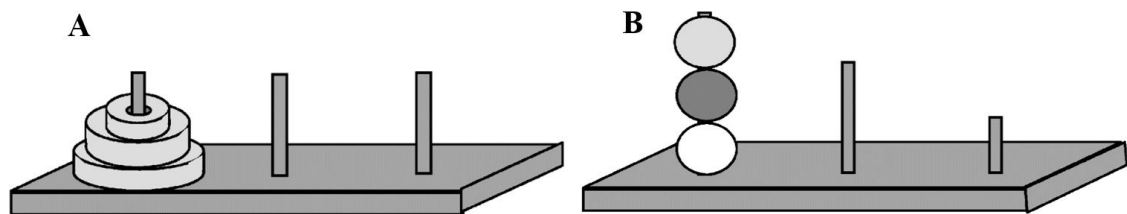
Desde o início da psicologia enquanto uma ciência, pesquisadores impelidos pelas dificuldades de dependerem unicamente de observações externas no estudo dos processos mentais, têm questionado os sujeitos de suas pesquisas sobre suas experiências, processos de pensamento e estratégias. Alegações a favor da validade desse tipo de registros verbais foram baseados principalmente na noção de que indivíduos têm um acesso privilegiado às suas próprias experiências; na medida em que sejam verdadeiros, suas respostas são confiáveis. Com a emergência do behaviorismo as instáveis suposições de confiança e acesso privilegiado foram rejeitadas como não científicas, e verbalizações passaram a ser tratadas como qualquer outro tipo de comportamento observável – como um comportamento verbal – mas, para além de simples respostas de “sim” ou “não”, foram relativamente pouco usadas em pesquisa experimental. (p.xii)

Neste primeiro momento, o método visava atender a essa preocupação, buscando refletir, da maneira mais fiel o possível, os mecanismos através dos quais o cérebro processa e armazena informações, assunto de especial interesse para a psicologia comportamental. De modo similar ao empregado hoje na Usabilidade, o teste também consistia em pedir ao voluntário para que expusesse seus pensamentos em voz alta enquanto executa uma determinada tarefa em um ambiente controlado, na presença de um mediador. Contudo, essa tarefa, ou conjunto de tarefas, era definida de acordo com o tipo de processo de interesse do estudo, tendo como exemplo resolver um problema matemático/lógico, traduções de textos curtos, quebra-cabeças físicos (como a Torre de Hanoi e a Torre de Londres¹, exemplificadas na figura 1) e jogos de memorização.

¹ O objetivo da Torre de Hanói é mover a torre de discos para um pino diferente, em uma configuração específica. Um disco maior nunca deve ser colocado sobre um menor. O objetivo da Torre de Londres é combinar o padrão e a cor das bolas com o arranjo específico. Não mais do que duas bolas podem ser colocadas no pino do meio, e não mais do que uma bola pode ser colocada no pino menor.

Nas décadas de 80 e 90, pesquisadores como Ericsson e Simon (1980 e 1993), Lewis e Rieman (1993), Fonteyn, Kuipers e Grobe (1993), Nielson (1993), buscavam estabelecer diretrizes para o emprego desta ferramenta de modo padronizado, pois sua aplicação se estendia crescentemente a uma gama de ciências aplicadas, como Psicologia, Educação e Ciência cognitiva. Nessa época, o uso de verbalizações como fonte de dados já era empregado ao estudo dos mais diversos temas: tomada de decisões, aprendizado de idiomas, interpretação de texto, ergonomia, processos cognitivos na escrita, testes de QI, memorização, percepção social, contabilidade, desenvolvimento de questionários, deficiências que impactam na aprendizagem, validação de questões de múltipla escolha e testes de usabilidade em produtos digitais (ERICSSON e SIMON, 1993).

Figura 1 - Esquemas da Torre de Hanoi (A) e da Torre de Londres (B).



Fonte: Berg e Byrd (2002, p. 587)

Para diversos autores, em especial Ericsson e Simon (1980 e 1993), um importante elemento para a validade do uso das verbalizações no estudo dos processos cognitivos é sua relação com a memória de curto-prazo (Short-Term Memory - STM) e a memória de longo-prazo (Long-Term Memory - LTM):

A mais genérica e fraca hipótese que nós preconizamos é a de que a cognição humana é processamento de informação: que um processo cognitivo pode ser visto como uma sequência de declarações internas sucessivamente transformadas por uma série de processos de informação. Uma suposição mais importante e específica é a de que informação é armazenada em várias memórias, que possuem diversas capacidades e características de acesso: diversos armazenamentos sensoriais de breve duração, uma memória de curto (STM) prazo com espaço de armazenamento limitada e/ou duração intermediária; e a memória de longo prazo (LTM) com uma extensa capacidade de armazenamento e relativamente permanente, mas com baixa capacidade de fixação e momentos de acesso, se comparado aos demais tipos de memória. (ERICSSON e SIMON, 1993 p.11)

A obra de Ericsson e Simon (1993) aborda pontualmente diversas variáveis, dificuldades e questões metodológicas envolvidas no uso de verbalização em estudos científicos, sendo uma das fontes mais citadas de diretrizes para análise de protocolo. Dentre as dificuldades encontradas por pesquisadores na obtenção de verbalizações que refletissem os processos cognitivos apropriadamente, dois tipos de processos são particularmente problemáticos: os processos de percepção-motora (perceptual-motor processes) e os processos de decodificação visual (visual encoding processes). Os problemas da verbalização de processos da percepção motora são mais claramente visíveis em situações nas quais o problema é representado fisicamente (como na Torre de Hanoi), e/ou a tarefa envolve a manipulação de objetos físicos. Quando engajados em tarefas deste tipo, os participantes não verbalizam (ou mesmo indicam ciência) dos processos mentais mais básicos envolvidos em suas ações (ERICSSON E SIMON 1993, p.91-2).

Em relação aos processos de decodificação visual, os autores apontam que, nos processos cognitivos, há uma distinção entre uma representação/código visual e uma representação/código oral-simbólica. Portanto, quando os participantes interagem com imagens, há um extensivo processo de descrição e codificação oral da imagem, que afeta a natureza de suas verbalizações. No modelo proposto pelos autores, essas explicações verbais (de um comportamento) não alteram o percurso natural da tarefa (ERICSSON E SIMON 1993, p.95-102).

Para lidar com questões como estas, Ericsson e Simon (1993) dividem as verbalizações coletadas na análise de protocolo em três níveis de acordo com a necessidade de processamento cognitivo e, portanto, de confiabilidade. A Saber: (nível 1) em que a informação exigida na tarefa é verbalizada do exato modo em que é processada pela memória de curto prazo (STM), sendo portanto o nível que reflete os processos mentais da mais confiavelmente; (nível 2) em que há algum tipo de descrição, explicação ou interpretação do conteúdo do pensamento, contanto que não haja a adição de informações ao foco do participante, apenas a especificação de informações que não se encontram codificadas isomorficamente a linguagem (como as informações sensoriais, imagens e símbolos que precisam ser traduzidos oralmente); (nível 3) em que o participante explica seus pensamentos, motivações, preferências ou acrescenta algo para além das informações relativas a tarefa puramente, o que inclui a verbalização de tarefas motoras às quais não estaria atento normalmente (ERICSSON e SIMON, 1993 p.79).

Essa classificação visa não apenas a interpretação dos dados coletados, mas o estabelecimento de diretrizes mais claras para os participantes, visto sua influência nas verbalizações. Como exemplo, os autores apresentam as instruções utilizadas em dois estudos aos quais o método é frequentemente creditado: os psicólogos Dunker e Claparède. As instruções são divididas em duas partes, uma principal e a outra complementar. No trecho principal das instruções temos: “Tente pensar em voz alta. Creio que você faça isso frequentemente quando está sozinho, trabalhando em um problema” (DUNKER, 1926 apud ERICSSON e SIMON); e “Pense, reflita em voz alta, me diga tudo o que passar por sua cabeça durante o seu trabalho em busca da solução do problema” (CLAPARÈDE, 1934 apud ERICSSON e SIMON). O primeiro exemplo favoreceria verbalizações de nível 1, ao contrário do segundo que, por ser mais genérico, exigiria a verbalização de informações diversas, sejam elas codificadas ou não (ERICSSON e SIMON, 1993 p.80).

Já as instruções propostas por Ericsson e Simon (1993) para o CTA são:

Nesse experimento, estamos interessados no que você diz para si mesmo enquanto realiza a tarefa que nós lhe passamos. Para isso, iremos solicitar para que você fale em voz alta, enquanto trabalha nos problemas. O que quero dizer com falar alto é que eu necessito que você diga em voz alta tudo o que você diz para si mesmo silenciosamente. Apenas aja como se você estivesse sozinho em um quarto falando consigo mesmo. Se você ficar em silêncio por qualquer quantidade de tempo, eu irei lembrá-lo para continuar falando em voz alta. Você entende o que preciso que você faça? (p.377)

As instruções seguem com pequenos exercícios, para que primeiro o participante possa praticar a ação simultânea de falar e trabalhar no problema em uma situação mais simples (como cálculos matemáticos básicos e construção de anagramas), e para que o mediador tenha certeza de que o participante entendeu as regras com clareza e se sente seguro. Além dessas instruções, os autores sugerem diversos cuidados por parte do mediador ao instigar o participante a retomar as verbalizações quando ocorrem pausas. Esses lembretes devem ser feitos após pausas de 15-60 segundos, preferencialmente com palavras como “Continue

falando.” ou “O que você está pensando?”, porém não mais específico do que isso e com o mínimo de interferências o possível.

Outra sugestão neste sentido é a ausência completa do mediador na sala do teste, tendo o lembrete substituído por um sinal sonoro. No entanto, nos casos em que se deseja uma forma específica de verbalização, essa estratégia pode levar o participante a não verbalizar exatamente o que se espera dele. No mesmo sentido, Lewis e Rieman (1993 p.84) ressaltam que “não dizer nada aos participantes após as instruções iniciais geralmente não funciona. A maioria das pessoas não manterá um bom fluxo de comentários sem um pouco de incentivo”. Os autores sugerem o uso de expressões como “me diga o que está pensando” ou “continue falando”, enquanto uma má escolhas de palavras seriam “porque você fez isso?” ou “o que você acha que isso significa?”.

4 Variações do *Think-Aloud*

Devido ao amplo leque de possibilidades de aplicação do teste pode, com o tempo, surgiram variações que visam atender as especificidades de cada caso. Cada uma dessas variações busca solucionar uma questão metodológica específica, adequada às variáveis mais relevantes a determinado problema. Em suas versões clássicas, as verbalizações (o “pensar em voz alta”) podem se dar simultaneamente à execução da tarefa, de modo Concorrente (*Concurrent Think-Aloud* - CTA) ou, no caso de tarefas muito curtas ou em que o tempo de execução seja relevante, a verbalização pode se dar posteriormente a tarefa, de modo Retrospectivo (*Retrospective Think-Aloud* - RTA), com o auxílio de uma gravação. Se comparadas a outros testes de usabilidade, tanto estas quanto outras variações mais recentes do método são simples e baratas de serem executadas, pois não exigem equipamentos caros, nem um grande número de participantes (replicações).

Em ambos os modelos, o método possui como diferencial a confiabilidade das informações obtidas, visto que partem diretamente da interação com o usuário. Sendo assim, o objetivo do método é observar o desempenho das características do produto durante a interação, ajudando os pesquisadores a identificarem suas potencialidades e limitações. Por meio do teste é possível identificar falhas no produto/protótipo através da interação direta com o usuário, em oposição às técnicas que envolvem critérios definidos a priori e/ou participação de especialistas de determinada área, como a avaliação heurística. A técnica é aplicada nas fases de Concepção/Prototipagem e de Avaliação/Refinamento do projeto de produtos ou sistemas, tanto com o intuito de identificar falhas e corrigi-las, quanto de gerar novas ideias a partir das necessidades demonstradas pelos usuários. Isso permite também que seja testado apenas um elemento do protótipo, além de possibilitar sua aplicação à análise de produtos concorrentes ou por uma categoria de produtos.

Em geral, a análise de protocolo é empregada com a intenção de possibilitar um vislumbre do ponto de vista do usuário, permitindo ao pesquisador insights construtivos a respeito do produto que não lhe ocorreriam de outro modo. No entanto, como qualquer outra técnica, a análise de protocolo possui limitações. Primeiramente, a perspectiva do usuário representa uma amostra da realidade, não correspondendo necessariamente a de todos os usuários. Além disso, podem ocorrer o (já mencionado) enviesamento das respostas, comprometendo a qualidade dos dados obtidos. Estes vieses podem ser causados por fatores de naturezas diversas, especialmente os relacionados às instruções e à interação entre o mediador e o participante. Por esta razão, é importante que os parâmetros para realização do teste sejam definidos com base no aspecto da interação mais relevante que se deseja avaliar, visto que estas múltiplas variações resultam na verbalização de diferentes tipos de informação.

Elementos como a dinâmica de aplicação do teste (HAAK e JONG, 2015; BOREN e RAMEY, 2000), a natureza e tempo de execução da tarefa (HERTZUM e HOLMEGAARD, 2015), o ambiente, o modo como são dadas as instruções ou mesmo a presença do pesquisador na sala (HAAK e JONG, 2005; RIIHIAHO, 2015) podem alterar os resultados do teste. Essas variações podem levar, portanto, à identificação de problemas de diferentes naturezas e graus de profundidade.

A carga cognitiva imposta pela ação simultânea de agir, pensar e falar, por exemplo, é um elemento amplamente explorado e debatido, que justifica variações no modo de aplicação do teste (ALSHAMMARI, ALHADRETI e MAYHEW, 2015). Essa sobrecarga cognitiva é a principal variável entre os dois modelos mais difundidos da análise de protocolo: o modelo concorrente (CTA) e o modelo retrospectivo (RTA).

No modelo concorrente, os participantes verbalizam seus pensamentos simultaneamente à execução da tarefa proposta, enquanto que no modelo retrospectivo as verbalizações se dão apenas após a conclusão da tarefa, enquanto o participante assiste ao registro em vídeo, acompanhado pelo(s) pesquisador(es). A justificativa para essa divisão é a possibilidade de permitir ao participante que se concentre nas duas tarefas separadamente, dando mais ênfase em suas opiniões e insights. Por outro lado, essa divisão também cria um lapso temporal que possibilita que informações sejam perdidas, visto que o participante pode não se lembrar com detalhes do porquê de suas ações. Além disso, em ambos os modelos, “expressar pensamentos em voz alta”, somado a presença de um avaliador atento à fala, pode ser percebido como um modo pouco natural de falar, inibindo comentários tanto pela dificuldade quanto pelo constrangimento do participante.

Para lidar com desvantagens como essas, diversas variações subsequentes da técnica foram criadas. Dentre elas, destacam-se: Question Asking Protocol, Cooperative Evaluation, Cooperative Usability Testing, Backtracking Analysis e Experience Clip (RIIHIAHO, 2015). Todas estas variações partem da inclusão de usuários reais dos produtos analisados para identificação de problemas de usabilidade e questões projetuais. No entanto, em suas particularidades, cada variação aborda essa inclusão de modo distinto, apresentando vantagens e limitações frente a outras técnicas.

No Cognitive Interviewing, por exemplo, questionários são testados previamente por voluntários, que respondem as perguntas verbalizando também suas dúvidas, impressões e sentimentos. Assim, é possível identificar e corrigir problemas no questionário antes de sua aplicação em campo, ajudando a evitar que sejam abandonados pela metade. Portanto, a técnica é particularmente útil na área da Saúde, garantindo uma abordagem adequada a grupos específicos (como pacientes pediátricos e geriátricos), bem como para a coleta de dados que perpassam questões delicadas e/ou invasivas (DRENNAN, 2002).

5 Críticas ao método

Como mencionado anteriormente, a expansão do uso da análise de protocolo da Psicologia cognitiva para outras áreas até chegar ao Design, implicou em inconsistências metodológicas, resultantes da mudança de foco da análise (do processo cognitivo para o artefato sendo utilizado pelo usuário). Ainda que o trabalho de Ericsson e Simon (1993) se proponha a organizar e fundamentar teoricamente essa transição, Boren e Ramey (2000) apontam que, apesar da obra ser frequentemente citada em testes de usabilidade, raramente a metodologia citada pelos autores é seguida à risca. Por vezes, os trabalhos nem mesmo descrevem as diretrizes metodológicas utilizadas ou as instruções dadas aos participantes, dificultando a comparação posterior de resultados entre estudos. Além disso, os autores ressaltam que as

diretrizes oriundas da psicologia cognitiva não possuem recomendações para elementos importantes para a Usabilidade, como a presença de inconsistências por parte do protótipo (*bugs*, mal funcionamento do sistema, travamentos, erros, etc.) ou da parte do participante (que pode apresentar comportamentos ou dificuldades inesperadas pelo pesquisador), tornando os testes de usabilidade mais complexos e com mais variáveis de difícil controle (BOREN e RAMEY, 2000).

No levantamento realizado pelos autores foram observadas diversas incoerências entre o modelo de Ericsson e Simon e sua aplicação em campo. Das 125 interferências feitas por mediadores ao se dirigirem aos participantes, por exemplo, apenas 16 eram lembretes para a retomada das verbalizações. O intervalo observado no tempo de espera para a realização dessas interferências também variou consideravelmente, com intervalos de menos de 5 segundos em alguns casos e em outros sem interferência alguma por parte do mediador. A escolha das palavras também era por vezes inconsistente, com lembretes como “João, você poderia nos dizer porque você apertou a tecla enter?” que, além de longo, sugere uma conexão pessoal entre “João” e “nós”, exige uma pausa no fluxo de pensamento, incita um enviesamento de foco e ainda exige uma explicação para o comportamento. Ou seja, uma intervenção precisamente contrária às diretrizes de Ericsson e Simon (BOREN e RAMEY, 2000).

Boren e Ramey (2000) argumentam que, apesar do amplo uso da análise de protocolo, as inconsistências entre a teoria e a prática são tão significativas que, dentro do campo da Usabilidade, não há rigor suficiente para que a análise de protocolo seja designada como um método. Para os autores, além das discrepâncias observadas por eles em campo, essa mudança de foco implica em uma diferença fundamental na comunicação que ocorre ao longo do teste.

Na perspectiva adotada por Ericsson e Simon (1993) e outros autores preocupados com a compreensão do processo cognitivo, há o pressuposto de que, durante o teste, o receptor das mensagens (a pessoa a qual as verbalizações se destinam) é puramente passivo, neutro ou mesmo inexistente. Portanto, os participantes são instruídos a “pensar em voz alta” (think-aloud) como se estivessem sozinhos na sala, falando consigo mesmos, simulando a ausência de um ouvinte. Contudo, Boren e Ramey (2000) criticam fortemente essa ideia de uma “ausência de comunicação” no caso dos testes de usabilidade. Para os autores, na prática, a abstração do mediador é inalcançável e, conseqüentemente, as verbalizações constituem sim uma forma de comunicação, pois possuem necessariamente um emissor (o participante) e um receptor (o mediador). Nesse mesmo sentido, os autores também ressaltam que grande parte da cognição que é objeto de interesse da Usabilidade “se manifesta não apenas através das verbalizações, mas também das interações com o sistema; isso fornece um contexto útil para a interpretação e a determinação da validade das verbalizações” (BOREN e RAMEY, 2000 p.263).

Ainda segundo Boren e Ramey (2000), diversos pesquisadores parecem valorizar verbalizações de formas completamente distintas do modelo proposto por Ericsson e Simon, como por exemplo a preferência por explicações e reflexões, em detrimento de informações estritamente procedimentais em relação à tarefa. Isso sugere que, para a Usabilidade, verbalizações de nível 3 são de maior importância do que as de nível 1 e 2.

Sendo assim, teste de usabilidade, frequentemente é perguntado aos participantes sobre suas preferências e aversões, o que também é desincentivado pelo modelo de Ericsson e Simon. Além disso, as verbalizações raramente são analisadas minuciosamente, o que ocorre apenas em casos mais problemáticos. Contudo, os autores identificaram que, para os profissionais que trabalham com usabilidade, “a experiência de observar testes de usabilidade, ou ao menos ler citações dos participantes, foi frequentemente apontada como a mais poderosa ferramenta

para convencer membros de uma equipe de que um problema existe” (BOREN E RAMEY, 2000 p.266) .

Riihiaho (2015) aponta em sua tese que, apesar destas inconsistências metodológicas, o CTA ainda é essencial aos testes de usabilidade por auxiliar na “identificação de problemas de usabilidade e suas causas, além disso, ajuda os clientes a entenderem os problemas [...]”. Desse modo, o impacto que método pode ter na equipe de desenvolvimento geralmente supera os potenciais efeitos no desempenho dos usuários” (p.143) . Já Haak e Jong apontam para o fato de que, em testes de usabilidade com CTA, ao verbalizarem suas ações (e não apenas seus pensamentos), os participantes não deixam de contribuir para detecção de problemas reais: “o sucesso [do teste] não depende tanto da veracidade da expressão dos pensamentos dos participantes, como é o caso em muitos experimentos psicológicos, mas da utilidade desses pensamentos na detecção de problemas do usuário” (p.326) .

6 Testes em dupla

Uma solução apontada por diversos trabalhos para o problema da comunicação na análise de protocolo é a inclusão de dois participantes (usuários) que trabalham conjuntamente na tarefa e/ou solução do problema. No campo da Usabilidade, esse tipo de teste apresenta nomenclaturas variadas: *constructive interaction*; *co-discovery learning*; *team usability testing*; *paired-user testing*; e *co-participation* (RIIHIAHO, 2015, p.95). Segundo a autora, apesar dos efeitos da presença de outras pessoas sobre a performance dos participantes ser um assunto tratado em diversos estudos experimentais em outras áreas, este ainda é um assunto pouco abordado no campo da Usabilidade, apresentando resultados conflitantes (RIIHIAHO, 2015, p.70).

Geralmente estes estudos se detêm na comparação das variações entre as modalidades Concorrente (CTA) e Retrospectiva (RTA). Nesse sentido, um exemplo interessante no campo da Usabilidade, é o trabalho de Haak e Jong (2015), que compara o CTA com uma destas variações do teste em dupla, o Constructive Interaction Test (CI). Neste levantamento, os autores identificaram que a presença de um segundo usuário não representa necessariamente uma mudança significativa na performance dos participantes, quantidade de problemas detectados ou mesmo na natureza desses problemas. Portanto, das duas opções, a solução financeiramente mais eficiente é o CTA. Todavia, essa modalidade possui algumas desvantagens em relação ao CI. A principal delas se refere ao que os autores denominam de validade ecológica (*ecological validity*) do método: para que os problemas com o produto venham à tona durante o teste, é essencial que o ambiente onde é conduzido reflita as condições de uso do modo mais natural possível. (HAAK e JONG, 2015).

Nesse sentido, a presença do mediador (muitas vezes percebido como um “avaliador”), somada aos equipamentos de gravação, pode ser intimidadora para o participante, se refletindo negativamente tanto em sua performance quanto em suas verbalizações. A participação de um segundo usuário tem por objetivo desviar o foco do mediador e direcioná-lo para a tarefa. Ao trabalharem como um time e dialogarem entre si, as verbalizações “pouco naturais” do CTA (falar consigo mesmo como se estivesse sozinho na sala) dão lugar a uma modalidade de comunicação mais usual (BOREN E RAMEY, 2000; HAAK e JONG, 2015; RIIHIAHO, 2015).

Haak e Jong (2015) apontam que em ambos os métodos (CTA e IC) os participantes demonstram em suas verbalizações que estão cientes da presença do mediador, ainda que lhes seja instruído ao início do teste para ignorá-lo. Para os autores, a percepção do mediador enquanto um avaliador das ações dos participantes torna improvável que as ações e

pensamentos expressados nas verbalizações reflitam situações reais de uso. Os participantes, ao se sentirem avaliados e/ou culpados por não serem capazes de operar adequadamente o produto, atribuirão a culpa das falhas a si próprios e não ao produto, deixando assim de verbalizar os problemas que interessam aos pesquisadores. Isso dificulta a identificação correta da origem de um problema específico, especialmente no caso de artefatos com poucas pistas visuais (como sites que consistem em blocos de texto). Consequentemente, essa pressão seria a ameaça mais grave à validade dos dados coletados em ambos os casos (CTA e IC), pois ela pode levar a identificação de diferentes números e/ou tipos de problemas (HAAK e JONG, 2015).

Apesar da baixa efetividade em simplesmente dizer aos participantes que o objeto de estudo é o produto (e não a pessoa), Haak e Jong (2015) sugerem que um modo de minimizar este problema em ambos os casos é, no início da sessão, enfatizar as imperfeições do produto, explicando-lhes que ainda está em fase de testes ou que ainda não está de acordo com o padrão desejado. Desse modo, os participantes são encorajados a uma postura menos competitiva, contribuindo para a validade das verbalizações.

Outra comparação interessante realizada pelos autores é a identificação de duas diferenças principais entre as verbalizações obtidas com o CTA e com o IC:

Em primeiro lugar, ao contrário dos participantes do CTA, os participantes do IC dificilmente descreveram qualquer uma de suas ações. Como eles estavam trabalhando juntos lado a lado, eles obviamente não tinham nenhum motivo para descrever suas ações para o parceiro. Parece que o trabalho conjunto os tornava alheios ao monitor de teste, no sentido de que não o viam como um ouvinte que deveria ser mantido informado do que estava acontecendo o tempo todo. Uma segunda diferença [...] é que [os participantes do CI] se comportaram de forma menos insegura. [...] Seu comportamento mais confiante pode ser explicado pelo fato de trabalharem em equipe: eles nunca se sentiam sozinhos ao serem incapazes de realizar uma determinada ação, o que lhes dava mais motivos para acreditar que a culpa não era deles, mas do site. (HAAK e JONG, 2015 p. 325)

Essa primeira característica (a ausência de descrições de ações) representa uma diferença significativa em relação ao CTA, impactando tanto na natureza das verbalizações registradas, quanto no meio de registro da sessão (exigindo um registro em vídeo que não é tão necessário no CTA). Se por um lado esses elementos descritivos são suprimidos das verbalizações, por outro novos elementos são trazidos à tona, em especial aqueles ligados às frustrações. Essa mudança exemplifica com clareza como as variáveis do teste influenciam na natureza das verbalizações e, portanto, devem ser consideradas de acordo como o elemento da interação que se deseja avaliar com o teste.

7 Considerações finais

Ao compreendermos o método da análise de protocolo como fruto de uma trajetória que perpassa diferentes áreas, épocas e funções, podemos analisar com maior clareza os elementos que a compõem. Por meio da compreensão e articulação de suas variáveis é então possível adaptar o método para atender melhor às necessidades e contextos de cada pesquisa. Considerando-se que em suas origens o foco do método não estava na análise de um produto, mas nos processos cognitivos do sujeito que executava a tarefa, é esperado que suas instruções, parâmetros e lentes de análise possuam preocupações distintas das empregadas pelo Design. No entanto, isso não impede que essas perspectivas, oriundas da psicologia cognitiva, impactem na apropriação deste método pelos designers, positiva ou negativamente.

Ao resgatarmos essas origens, podemos observar que, ainda que o trabalho de Ericsson e Simon (1980 e 1993) seja frequentemente citado como fonte, as preocupações metodológicas apresentadas pelos autores são incoerentes com a aplicação da análise de protocolo no campo do Design. Isso se reflete especialmente nas instruções dadas aos participantes, visto que as verbalizações de níveis 2 e 3, que para Ericsson e Simon devem ser evitadas, são justamente aquelas que mais interessam aos designers (BOREN e RAMEY, 2000).

A questão das instruções exemplifica o modo como as variáveis metodológicas influenciam diretamente nos resultados do teste, direcionando a natureza das verbalizações. Nesse sentido, uma variável que merece destaque é a participação de dois usuários no teste que, apesar de elevar custos e exigir alguns ajustes, ajuda a solucionar algumas limitações. No formato clássico, frequentemente busca-se contornar o problema da insegurança sentida pelos participantes, causada pelo desejo de performar bem diante da figura de autoridade que é o mediador (por vezes entendido negativamente como um avaliador). Ao adicionarmos um segundo usuário, o foco da atenção dos participantes se desloca do mediador para o uso do produto, direcionando as verbalizações para elementos mais úteis ao pesquisador. Além disso, o ganho na sensação de segurança proporcionado por essa mudança permite que sejam verbalizados sentimentos e frustrações que possivelmente seriam suprimidos em outras modalidades.

Referências

- ALSHAMMARI, T.; ALHADRETI, O.; MAYHEW, P. When to ask participants to think aloud: a comparative study of concurrent and retrospective think-aloud methods. *International Journal of Human Computer Interaction*, vol.6, n3, 48-64, 2015.
- BERG, W; BYRD, Dana. The Tower of London Spatial Problem-Solving Task: Enhancing Clinical and Research Implementation. *Journal of clinical and experimental neuropsychology*. 24. 586-604. 10.1076/jcen.24.5.586.1006. (2002).
- BOREN; RAMEY. Thinking aloud: Reconciling theory and practice. *IEEE transactions on professional communication*, vol. 43, no. 3, september 2000.
- DRENNAN, J. An evaluation of the role of the clinical placement coordinator in student nurse support in the clinical area. *Journal of Advanced Nursing*, vol.40, 475-483, 2002.
- ERICSSON, K.A.; SIMON, H.A. *Protocol Analysis: verbal reports as data*. Revised ed. Cambridge: MIT Press, 1993.
- ERICSSON, K.A.; SIMON, H.A. Verbal reports as data. In *Psychological Review*, Vol. 87, no. 3, 1980.
- FORTEYN, Marsha; KUIPERS, Benjamin; GROBE, Susan. A description of think aloud method and protocol analysis. *Qualitative Health Research - QUAL HEALTH RES*. Vol3, 1993.
- HERTZUM, M.; HOLMEGAARD, K. D. Thinking aloud influences perceived time. *Human Factors*, Vol. 57, n 1, 2015.
- LEWIS, C; RIEMAN, J. *Task-centered user interface design: a practical introduction*. Boulder: University of Colorado, 1993
- RIIHIAHO, Sirpa. *Experiences with usability testing: effects of thinking aloud and moderator presence*. Tese de doutorado. Helsinki: Aalto University publication series, 2015.

VAN DEN HAAK, M.J.; DE JONG, M.D.T. Analyzing the interaction between facilitator and participants in two variants of the think-aloud method. In Proceedings of the International Professional Communication Conference, 2005.